February 16, 2018

Re: Resubmission of manuscript "Analysis of the risk associated to coastal flooding hazards: A new historical extreme storm surges dataset for Dunkirk, France", nhess-2017-417

Copernicus Publications
Editorial Support

Dear Editor,

Thank you for the opportunity to revise our manuscript, "Analysis of the risk associated to coastal flooding hazards: A new historical extreme storm surges dataset for Dunkirk, France". We appreciate the careful review and constructive suggestions. It is our belief that the manuscript is substantially improved after making the suggested edits.

Following this letter are the reviewers comments with our responses, including how and where the text was modified. As suggested by the reviewers, our manuscript had been checked by a professional English editing service. The revision has been developed in consultation with all coauthors, and each author has given approval to the final form of this revision.

Thank you for your consideration.

Sincerely,

Yasser Hamdi

Point-by-Point response / reviewer # 1

Yasser Hamdi

General comments

Comments	Responses to comments
It is well known that deriving water levels for large return periods using limited duration of observation can result in very significant errors. In this paper, the authors attempted to address this issue for the city of Dunkirk through the collection of long-term historical data. Although this approach is not really a first of its kind, it is still extremely interesting and valuable, and as such, it could be considered favorably for publication, provided that the authors address the two main concerns (and minor issues, see below) that I have regarding this paper: 1. The authors state in the abstract that the aim is to demonstrate the technical feasibility of including long-term historical information to improve the statistical assessment of extreme water level return periods. But what is really a technically challenging when doing this? It is indeed challenging to find old HI, and the authors really deserve to be thanked for this valuable effort. But what is at least as challenging (in my opinion) is to answer this kind of questions: - to what extent are the storm surge dynamics that occurred hundreds of years ago representative of the actual level of risk? - has the bathymetry, topography, or land cover of the studied area evolved since then, and what could be the impact on storm surge dynamics? - how accurate are the historical water level data, considering, for example, potential sea level rise, land subsidence, uncertainties relative to the distinction between overflowing and overtopping when assessing maximum water levels, etc? Answering these questions could be extremely difficult, but I think that the authors should at least discuss them in the paper.	What is really a technically challenging is : - Find the right source of information, cross with other sources (to find the same information elsewhere and if the event is described in the same way or not). It is then necessary to quantify the information (estimate the value of the storm surge from qualitative information and quantitative information about other physical quantities that can lead to the estimation of storm surges - There are several types of historical data (range, exact value, lower bound, threshold of perception). Transform the historical information to these different types is not always easy… - In the statistics, one must calculate the empirical probabilities of the historical data (which is not the same as the systematic ones), calculate a likelihood taking into account the heterogeneous data (the historical data) and assign an effective duration to the collected data which an important setting in the frequency model that will be used to estimate the high return levels ... - Ensuring the completeness of the information is a task that requires a remarkable effort especially by the historian
2-This study (as others) relies on one extremely strong hypothesis: the maximum water level is supposed to be spatially homogeneous, not only in Dunkirk, but also for the nuclear plant in Gravelines, 20km away. To what extent can this assumption be considered realistic? For example, tide gauges generally do not capture the whole wave setup component of the surge, which can vary by a few dozens of centimeters between a harbor, and nearby beaches. They also generally poorly capture infragravity waves, which have been observed in many places along the shoreline of France and can have huge impacts on coastal flooding. The bathymetry and topography can be also quite different in Dunkirk, and 20km away from the city. Considering the very high stakes, it is hence important I think to address these issues (or at least to discuss them) in the paper.	Indeed, the assumption is strong and the state of current practice is to neglect some local effects… Nevertheless, the spatial homogeneity assumption was not used arbitrarily in the present study (and in other studies dealing with the use of information in a regional context). Indeed, It was concluded in two regional frequency models (which were developed to estimate extreme storm surges in Dunkirk and other sites) that the Gravelines NPP is located in a physically (based the calculation of the extremal dependence coefficient) and statistically homogeneous region centered on the Dunkirk harbor.

Specific Comments

Section	Comment	Response to reviewer	Response in the paper
Abstract & Introduction (§2)	Line 23: "Dunkirk site, representative of the Gravelines NPP": this statement should be tempered or discussed in greater details, as mentioned above.	All the studies about the Gravelines NPP use data from the Dunkirk harbor. Indeed, in the nuclear safety field, the representativeness of stations (for rainfall, discharges, sea levels, etc.) is being taken quite seriously. An in-depth study comparing sea levels (and storm surges) in Dunkerque and Gravelines (a short series) has been shown that the impact of the local effects is not significant.	A sentence (justifying the use of the Dunkirk harbor to analyze storm surges in Gravelines) was added in the 2nd § of the introduction.
1. the § before the last one	Lines 117-119: it is unclear whether or not you have already built a complete historical database for the entire French coasts (at least Atlantic and English Channel). Please reformulate.	This database is completed and is currently the subject of a working group involving several French organizations to share, complete and maintain it.	The sentence was reformulated.
2.2.2 Table 1 4.2 & 4.3 Tables 4-5	Table 1: please specify (when possible) where exactly the water and surge levels were obtained (at tide gauges? Dikes? in the streets? In houses? in areas exposed to waves or not?). It would be also interesting to know for each case to what extent the values could be affected or not by wave setup, wave run-up, or overtopping. If an area at a given altitude is flooded because of overtopping, it should not be treated the same way as if it was flooded by overflowing for example). Please do the same whenever possible for Tables 4-5.	Unfortunately, we do not have these details in the archives.	
3.1	Lines 267-268: "A POT threshold equal to 0,75m […] is an adequate choice". Please give at least a few indications or a reference to explain how you came to choose this value.	OK	The 1st § in section 3.1 was modified.
4.3	Lines 445-446: this seems to suggest that some historical data (for the 18th century for example) have been collected but were not used in this paper. Why? If the purpose of this study is really to demonstrate the technical feasibility of a long-term historical study, then it is more important to describe this kind of information than computing new extreme water levels. On the other hand, if the objective is to do these computations, then all the available data have to be taken into account.	It is currently the subject of a second article that will be written by the historian and colleagues.	

2. Typographical-technical corrections

All minor corrections proposed by the reviewer were accepted and performed directly in the paper. The line by line review was considered as well.

Point-by-Point response / reviewer # 2
Yasser Hamdi

General comments

Comments	Responses to comments
The manuscript nhess-2017-417 presents the reconstruction of the storm surge level in Dunkirk, utilizing data from different sources and dating back to the 16th century. It is a remarkable effort towards reconstructing the storm surge climate in Dunkirk and the detailed literature review provided is of invaluable significance.	Indeed, the effort to build the database is important (thanks to the historian with whom we worked).
The current form of the manuscript requires major revision since the syntax of the language is often problematic. The incoherent structure throughout the text and especially the description of the results, along with the poor quality of the presented results, makes it difficult for the reader to follow. The publication has the potential to be useful for future studies related with the impact of coastal floods, as soon as a proper justification of some technical approaches is provided.	- A major revision of English for errors (syntax, grammar, spelling and vocabulary) was made; - The structure and the quality of the presented results have been improved.

Specific Comments

Comment	Response to reviewer	Response in the paper
The description of the data needs to be improved – see also some recommendations at the next section. In section 4.3 the historical surge dataset is presented, but it is not clear whether the hydrodynamic component under study is the storm surge level or the total water level (including the contribution of other hydrodynamic components).	As it can be seen in the list of historical data obtained (Tables 1, 2, 6, 7), we seek to estimate the storm surges in the end. Total seal levels are also collected when available. Storm surges are then deducted from total levels by subtracting the predicted levels.	Sentence added to text in the 1st § of section 3.2.
Additionally, it is not clear how the surge level is estimated when only the meteorological conditions are available from the historical records.	Only when the tidal coefficient is given, using the approach described in §2 - section 3.3, one can estimate a value of the surge.	
While it is a fair assumption that during a storm event, the water level along neighbouring areas may exhibit a similar level, the local bathymetric features and the man-made structures may alter the local water level. Therefore these data should be considered only as qualitatively accurate and not quantitatively. Should these data be used, a comparison with numerical simulations would decrease the level of uncertainty.	Indeed, these data are considered qualitatively accurate and quantitatively uncertain. Although the distance between the places where information were collected and our point of interest (the target site: the Graveline NPP) is very small, the impact of certain local effects can creep in the inference. A relatively short series on the target site is used to compare the extreme levels at this site and Dunkirk. The comparison shows that in most cases the impact is not significant. Otherwise, the comparison with the simulations will make it possible to appreciate this uncertainty.	

Technical correction

Comment	Response to reviewer	Response in the paper
The MS should be proofread by a native English speaker for errors in syntax, grammar, spelling and vocabulary.	Ok.	
Informal expressions and language (e.g. "an important surge", "horrible storm") are used for the context of a scientific journal, while the terminology is not the most appropriate (e.g. "marine flooding", "marine submersion").	"marine flooding" and "marine submersion is replaced by "coastal flooding".	"marine flooding" and "marine submersion is replaced by "coastal flooding".
The manuscript lacks structure and is very difficult for the reader to follow, as the presentation of the data takes place together with the analysis. It is recommended first to describe the data that will be analyzed; this section should be followed by a short description of the methods and finally a section that presents the results after incorporating all the types of available data.	Structure reviewed.	The paper is restructured as proposed by the reviewer.
Although the POTH method has been described in previous publications, it is recommended to provide a short summary at the Methods section. This would give a better overview to the reader, regarding the analysis of the data and would enhance the clarity of the paper.	A general description of the POTH model and settings was provided in sections 4 and 4.3. Ok for further description.	A section 4.2 "The POTH frequency model" is added.
There is an abundance of information (the damage and the fatalities triggered by the storm, the weather description, etc) scattered around the essay that is loosely connected to the main argument. It would be helpful to move this to a supplementary material section; this would tidy up the main points and would make the argument read in a clearer way. For the same reason, measurements obtained from other sites may be omitted too from the main body of the manuscript, since they are not considered at the analysis (e.g. the section from line 206 to 213).	- The §s on the description of HI are moved to an appendix. - Description of the measurements which are not considered in the analysis (like the 1808 event) is removed from the main body of the manuscript to the appendix.	Please see Appendix 1.
Please provide a map displaying all the places mentioned at the MS.	OK	Fig.1 updated.
Section 2 lacks structure, coherence and paragraph unity. The main title of the section as well as the ones of the following paragraphs are misleading and do not correspond to the topic of the paragraphs. Additionally, section 2.2.1 should be renumbered to 2.1.1 as it refers to the tide gauge record and not to the short-term HI.	Structure reviewed.	The paper is restructured as proposed by the reviewer.
The quality of the figures, the tables and their captions is poor and should be improved. Fig.6 does not provide any extra information to the reader.	Quality of figures and tables reviewed. Fig. 6 removed; Fig. 4 & 5 improved and merged; Fig. 2 and 7 improved and merged.	
Consider merging Tables 1, 2, 4 preferably presenting only the information related with the storm surge level and the data included in the analysis. All the information with respect to the sources, meteorological conditions etc should be provided in a tabular form at the supplementary material for future reference and for reproducing the analysis of this study.	Table 2 removed (merged with Table 1) Tables 5 and 6 merged	

Minor corrections Line by line review

All minor corrections proposed by the reviewer were accepted and performed directly in the paper. The line by line review was considered as well.

Analysis of the risk associated ~~to~~ with coastal flooding hazards: A new historical extreme storm surges dataset for Dunkirk, France

Abstract

This paper aims to demonstrate the technical feasibility of a historical study devoted to French Nuclear Power Plants (NPPs) which can be prone to extreme coastal flooding events. It has been shown in the literature that the use of Historical Information (HI) can significantly improve the probabilistic and statistical modelling of extreme events. There is a significant lack of historical data ~~about~~ on marine flooding (storms and storm surges) compared to river flooding events. To address this data scarcity and to improve the estimation of the risk associated with ~~to the~~ coastal flooding hazards, a dataset of historical storms and storm surges that hit the Nord-Pas-de-Calais region during the ~~five~~ past five centuries was recovered from archival sources, examined and used in a frequency analysis (FA) in order to assess its impact on ~~the~~ frequency estimations. This work on the Dunkirk site (representative of the Gravelines NPP) is a continuation of previous work performed on the La Rochelle site in France. Indeed, the frequency model (FM) used in the present paper had some success in the field of coastal hazards and it has been applied in previous studies to surge datasets to prevent coastal flooding in the La Rochelle region in France.

In a first step, only information collected from the literature (published reports, journal papers and PhD theses) is considered. Although this first historical dataset has extended the gauged record back in time to 1897, serious questions related to the exhaustiveness of the information and about the validity of the developed FM have remained unanswered. Additional qualitative and quantitative HI ~~were~~ was extracted in a second step from many older archival sources. This work has led to the construction of storms and coastal flooding sheets summarizing key data on each identified event. The quality control and the cross-validation of the collected information, which have been carried out systematically, indicate that it is valid and complete as regards extreme storms and storm surges. Most of the HI collected ~~displays a~~ is in good agreement with other archival sources and documentary climate reconstructions. The probabilistic and statistical analysis of a dataset containing an exceptional observation considered as an outlier (i.e. the 1953 storm surge) ~~has been~~ is significantly improved when the additional HI collected in both literature and archives ~~are~~ is used. As the historical data tend to be extreme, the right tail of the distribution has been reinforced and the 1953 "exceptional" event does not appear as an outlier any more. This new dataset provides a valuable source of information on storm surges for future characterization of coastal hazards.

1 Introduction

As the coastal zone of the Nord-Pas-de-Calais region in Northern France is densely populated, coastal flooding represents a natural hazard threatening the costal populations and facilities in several areas along the shore. The Gravelines Nuclear Power Plant (NPP) is one of those coastal facilities. It is located near the community of Gravelines in North~~ern~~ France, approximately 20 km from Dunkirk and Calais. The Gravelines NPP is the sixth largest nuclear power station in the world, the second largest in Europe and the largest in Western Europe.

Extreme weather conditions could induce strong surges that could cause coastal flooding. The 1953 North Sea flood was a major flood caused by a heavy storm that occurred on the night of Saturday, 31 January and morning of Sunday, 1 February. The floods struck many European countries and France had not been the exception. This was particularly the case along the northern coast of France, from Dunkirk to the Belgium border. Indeed, it has been shown in an unpublished study that Dunkirk is fairly representative of the Gravelines NPP in terms of extreme sea levels. In addition, the harbor of Dunkirk is an important military base containing a lot of archives. The site of Dunkirk ~~is then~~ has therefore been selected as ~~the~~ site of interest in the present paper (**Fig. 1**). An old ~~plan~~ map of ~~the~~ Dunkirk city is presented in the right panel of **Fig. 1** (we shall return to this ~~plan~~ map at a later stage in this paper). It ~~'~~is a common belief today that the Dunkirk region is vulnerable and subject to several climate risks (e.g.~~,~~ Maspataud et al. 2013). More severe coastal flooding events such as the November 2007 North Sea and the March 2008 Atlantic storms~~,~~ could have had much more severe consequences especially if they had occurred at ~~high tide~~hightide (Maspataud et al. 2013; Idier et al. 2012). It is ~~then~~ important for us to take into account the return periods of such events (especially in the current context of global change and projected sea-level rise) in order to manage and reduce coastal hazards, implement risk ~~policies~~ prevention policies and ~~to~~ enhance and strengthen coastal defence against coastal flooding.

The storm surge frequency analysis (FA) represents a key step in the evaluation of the risk associated ~~to~~ with coastal hazards. The frequency estimation of extreme events (induced by natural hazards) using probability functions has been extensively studied for more than a century (e.g.~~,~~ Gumbel, 1935; Chow, 1953; Dalrymple, 1960; Hosking and Wallis, 1986, 1993, 1997, Hamdi et al. 2014, 2015). We generally need to

estimate the risk associated with an extreme event in a given return period. Most extreme value models are based on available at-site recorded observations only. A common problem in FA and estimation of the risk associated with extreme events is the estimation from a relatively short gauged record of the flood corresponding to 100-1000 year return periods. The problem is even more complicated when this short record contains an outlier (an observation much higher than any other in the dataset). This is the case with several sea-level time series in France and characterizes the Dunkirk surge time series as well.

The 1953 storm surge was considered as an outlier in our previous work (Hamdi et al, 2014) and in previous research (e.g. Bardet et al, 2011). Indeed, although the Gravelines NPP is designed to sustain very low probabilities of failure and despite the fact that no damage was reported at the French NPPs, the 1953 coastal flooding had shown that the extreme sea levels estimated with the current statistical approaches could be underestimated. It seems that the local FA is not really suitable for a relatively short dataset containing an outlier.

Indeed, a poor estimation of the distribution parameters may be related to the presence of an outlier in the sample (Hamdi et al, 2015), and must be properly addressed in the FA. One would expect that one or more additional extreme events in a long period (500 years for instance) would, if properly included in the frequency model (FM), improve the estimation of a quantile at the given high-return period. The use of other sources of information with more appropriate FMs is required in the frequency estimation of extremes. Worth noting is that this recommendation is not new and dates back several years. The value of using other sources of data in the FA of extreme events has been recognized by several authors (e.g. Hosking and Wallis, 1986 and Stedinger and Cohn, 1986). By other sources of information we refer here to events that occurred not only before the systematic period (gauging period) but also during gaps of the recorded time series. Water marks left by extreme floods, damage reports and newspapers are reliable sources of Historical information (HI). It can also be found in the literature, archives, unpublished written records, etc. It may also arise from verbal communications from the general public. Paleoflood and dendrohydrology records (the analysis and application of tree-ring records) can be useful as well. A literature review on the use of HI in flood FAs with an inventory of methods for its modeling has been published by Ouarda et al, (1998). Attempts to evaluate the usefulness of HI for the frequency estimation of extreme events are numerous in the literature (e.g. Guo and Cunnane, 1991; Ouarda et al, 1998; Gaal et al, 2010; Payrastre et al, 2011; Hamdi, 2011; Hamdi et al, 2015). Hosking and Wallis (1986) have assessed the value of HI using simulated flood series and historical events generated from an extreme value distribution and quantiles are estimated by the maximum likelihood method with and without the historical event. The accuracy of the quantile estimates was then assessed and it was concluded that HI is of great value provided either that the flood frequency distribution has at least three unknown parameters or that gauged records are short. It was also stated that the inclusion of HI is unlikely to be useful in practice when a large number of sites are used in a regional context. Because HI is often imprecise, its inaccuracy should be considered in the analysis. Nevertheless, the influence of an outlier can be decreased by increasing its representativity in the sample when using the HI, knowing that its uncertainty is sometimes considerable (e.g. Payrastre et al, 2011; Hamdi et al, 2015). A frequency estimation of extreme storm surges based on the use of HI has rarely been studied explicitly in the literature (Bulteau et al, 2014, Hamdi et al, 2015, 2016) despite its significant impact on social and economic activities and on NPPs' safety. Bulteau et al. (2014) have estimated extreme sea-levels by applying a Bayesian model to the La Rochelle site in France. This same site was used as a case study by Hamdi et al, (2015) to characterize the coastal flooding hazard. The use of a skew surge series containing an outlier in local frequency estimation is limited in the literature as well.

It is often possible to augment the storm surges record with those that occurred before and after gauging began. Before embarking on a thorough and exhaustive research of any HI related to coastal flooding that hit the area of interest, potential sources of historical coastal flooding data for the French coast (Atlantic and English Channel) and more specifically for the Charente-Maritime region were identified in the literature (e.g. Garnier and Surville, 2010). The HI collected has been very helpful in the estimation of extreme surges at La Rochelle, which was heavily affected by the storm Xynthia in 2010 that generated a water level considered so far as an outlier (Hamdi et al, 2015). Indeed, these results for the La Rochelle site have encouraged us to build a more complete historical database covering all the extreme coastal flooding that occurred over the past five centuries on the entire French coast (Atlantic and English Channel). This database has been completed and is currently the subject of a working group involving several French organizations for maintenance. However, only the historical storm surges that hit the Nord-Pas-de-Calais region during this period are presented herein.

The main objective of the present work is the collection of HI on storms and storm surges that occurred in the last five centuries and to examine its impact on the frequency estimation of extreme storm surges. The paper is organized as follows: HI collected in the literature and its impact on the FA results is presented in sections 2 and 3. The fourth section presents the HI recovered from archival sources, the quality control thereof, and validation. In section 5, the FM is applied using both literature and archival

120 sources. The results are discussed in the same section before concluding and presenting some perspectives
121 in section 6.

2 Use of HI to improve the frequency estimation of extreme storm surges

123 The systematic storm surge series is obtained from the corrected observations and predicted tide levels. The
124 tide gauge data is managed by the French Oceanographic Service (SHOM - Service Hydrographique et
125 Océanographique de la Marine) and measurements are available since 1956. The R package
126 TideHarmonics (Stephenson, 2015) is used to calculate the tidal predictions. In order to remove the effect of
127 sea level rise, the initial mean sea level (obtained by tidal analysis) is corrected for each year by using an
128 annual linear regression, before calculating the predictions. The regression is obtained by calculating daily
129 means using a Demerliac Filter (Simon 2007). Monthly and annual means are calculated with respect to the
130 Permanent Service for Mean Sea Level (PSMSL) criteria (Holgate, ~~et al.,~~et al, 2013). This method is inspired
131 by the method used by SHOM for its analysis o~~f~~in high water levels during extreme events (SHOM, 2015).
132 The available systematic surge dataset was obtained for the period from 1956 to 2015.
133   The effective design of ~~the~~ coastal defense is dependent on how high a design quantile (1000-year~~s~~
134 storm surge for instance) will be. But this is always estimated with uncertainty and not precisely known.
135 Indeed, any frequency estimation is given with a confidence interval (CI) ~~whose~~ of which the width depends
136 mainly on the size of the sample used in the estimation. Some other sources of uncertainties (such as the
137 use of trends related to ~~the climatic evolution~~climate change) can be considered in the frequency estimation
138 (Katz ~~et al.~~et al, 2002). As mentioned in the introductory section, samples are often short and characterized
139 by the presence of outliers. The CIs are rather large and in some cases ~~exceed~~ more than 2 or 3 times (and
140 even more) the value of the quantile. Using the upper limit of this CI would likely lead to a more expensive
141 design of the defense structure. One could just use the most likely estimate and neglect the CI but it is more
142 interesting to consider the uncertainty as often estimated in frequency analyses. The width of the CI (i.e.
143 inversely related to the sample size) can be reduced by increasing the sample size. In the present work, we
144 focus on increasing the number of observations by adding information about storm surges induced by
145 historical events. Additional storm surges can be subdivided into two groups:
146   1. HI during gaps ~~of~~ in systematic records;
147   2. HI before the gauging period (can be found in the literature and/or collected by historians in archives).

3 HI during and before the gauging period

149 A historical research devoted to the French NPPs located ~~at~~on the Atlantic and English Channel coast was a
150 genuine scientific challenge due to the ~~timely implementation~~time factor and the geographic dispersion of the
151 nuclear sites. To be considered in the FA, a historical storm surge must be well documented; its date must
152 be known and some information on its magnitude must be available. Mostly, available information concerns
153 the impact and the societal disruption caused at the time of the event (Baart, 2011).

3.1 HI collected in the literature

155 As mentioned above, a common issue in frequency estimations is the presence of gaps within the datasets.
156 Failure of the measuring devices and damage~~s~~, ~~whose main cause is especially the~~ mainly caused by
157 natural hazard~~s~~ (storms, for instance), are often the origin of these gaps. Human errors, strikes, wars, etc.,
158 can also give rise to these gaps. Nevertheless, these gaps are themselves considered as dependent events.
159 It is therefore necessary to ensure that the occurrence of the gaps and the observed variable are
160 independent. Whatever the origin and characteristics of the missing period, the use of the full set of ~~the~~
161 extreme storm surges that occurred during the gaps is strongly recommended to ensure the exhaustiveness
162 of the information. This will make the estimates more robust and reduce associated uncertainties. Indeed, by
163 delving into the literature and the web, one can obtain more information about this kind of events. Maspataud
164 (2011) was able to collect sea-level measurements that were taken by regional maritime services during a
165 storm event in the beginning of 1995, a time where the Dunkirk tide gauge was not working. This allowed the
166 calculation of the skew surge, which was estimated by the author at 1.15m on January 2nd, 1995. This storm
167 surge is high enough to be considered as an extreme event. In fact, it was exceeded only twice during the
168 systematic period (January 5th, 2012 and December 6th, 2013). For convenience, we would like to recall here
169 the definition of a skew surge: It is the difference between the maximum observed water level and the
170 maximum predicted tidal ~~one~~ level regardless of their timing during the tidal cycle (a tidal cycle contains one
171 skew surge).
172   For the relatively short-term pre-gauging period, a literature review was conducted in order to get an
173 overview of the storm events and associated surges that hit the Nord-Pas-de-Calais region in France during
174 the last two centuries. Some documents and storm databases on local, regional or national scales are
175 available:

- the "-Plan de Prévention de Risques Littoraux (PPRL)-": ~~: are~~ refers to documents made by the French state on a communal scale, describing the risks a coastal zone is subject to, e.g. coastal flooding and erosion, and preventive measures in case of a hazard happening. To highlight the vulnerability of a zone, an inventory of storms and marine inundation within the considered area is attached to this document~~:~~.
- Deboudt (1997) and Maspataud (2011) describe the impact of storms on coastal areas for the study region;
- the VIMERS Project: gives information on the evolutions of the Atlantic depressions that hit Brittany (DREAL Bretagne 2015);
- NIVEXT Project: presents historical tide gauge data and the corresponding extreme water and surge levels for storm events (SHOM, 2015)~~-~~;
- Lamb 1991~~-~~: provides synoptic reconstructions of the major storms that hit the British ~~l~~Isles ~~for~~ from the 16th century up ~~to~~ till today.

According to the literature, the storm of ~~the 31st~~ January 31st to ~~4st~~ February 1st, 1953 caused the greatest surge and was the most damaging within the study area. This event ~~is~~ has been well analyzed and documented (Sneyers, 1953, Rossiter 1954, Gerritsen, 2005, Wolf and Flather 2005): A depression formed over the Northern Atlantic Ocean close to Iceland moving eastward over Scotland and then changing its direction to south-eastwards over the North Sea, ~~was~~ accompanied by strong northerly winds. An important surge was generated by this storm that, in conjunction with a high ~~spring tide~~springtide, resulted in particularly high sea levels. Around the southern parts of the Northern Sea the maximum surges exceeded 2.25m, reaching 3.90m at Harlingen, Netherlands. Large areas were flooded in ~~the~~ Great Britain, ~~n~~Northern ~~P~~parts of France, Belgium, the Netherlands and the German Bight, causing the death of more than 2,000 people. Le Gorgeu and Guitonneau (1954) indicate that during this event, the water level exceeded ~~over 2,40m~~ the predicted water level at the Eastern Dyke of Dunkirk by more than 2.40m (Table 1). Bardet et al. (2011) included a storm surge equal to 2.13m in their regional frequency analysis. Both authors indicate the same observed water level, i.e. 7.90m, but the predicted water level differs: While in 1954 the predicted water level was estimated at 5.50m, the predictions were reevaluated to 5.77m by the SHOM using the harmonic method. A storm surge of 2.13m is therefore used in the present study. Nevertheless, as also shown in Table 1, some other storms (1897, 1949 and 1995) inducing important storm surges and coastal floods ~~were~~ occurred within the area of interest. Appendix 1 presents a description of these events which are quite well documented in the literature. In the appendix, the description of some other historical events (~~whose~~ of which the information provided did not allow ~~estimating~~ the estimation of a storm surge value) ~~are~~ is included as well.

3.2 HI collected in the archives

For the longer term, the HI collection process involves the exploration and consultation, in a context of a permanent multi-scalar approach, of HI which can be seen as a real documentary puzzle with a large number of historical sources and archives. Indeed, NPPs are generally ~~implemented~~located, for obvious safety reasons, in sparsely populated and isolated areas ~~and that's~~ which is why these sites ~~know~~ were subject to little anthropogenic influence in the past. However, this difficulty does not ~~mortgage~~ forfeit a historical perspective due to the rich documentary resources for studying an extreme event ~~to~~ on different scales ranging from the site itself to that of the Region (Garnier, 2015 and 2017 ~~bis~~a). In addition, this may be an opportunity for researchers and a part of the solution because it also allows a risk assessment at ungauged sites.

First, it is important to distinguish between "direct data" (also referred to as "direct evidence") and "indirect data" (also referred to as "proxy data"). The first refers to all information from the archives that describes an extreme event (a storm surge event for instance) that occurred at a known date. If their content is mostly instrumental, such as meteorological records presented in ~~some commonplace~~ certain ordinary books or by the Paris Observatory (since the 17th century), sometimes accurate descriptions of extreme climatic events are likewise found. The "proxy data" rather ~~inform~~ indicate the influence of ~~some~~ certain storm initiators and triggers such as wind and pressure. Concretely, they provide information indirectly on coastal flooding for example.

Private documents or "ego-documents" (accounts and ~~commonplace~~ ordinary books, private diaries, etc.) are used in many ways during 16th to 19th ~~Century~~centuries. Authors recorded local facts, short news and ~~last~~ latest events, and amongst them, weather incidents. These misidentified historical objects may contain many valuable meteorological data. These private documents most often take the form of a register or a journal in which the authors record various events (economic, social and political) ~~and~~ as well as weather information ~~as well~~. Other authors use~~d~~ a more integrated approach to describe a weather event by combining observations of extreme events, instrumental information, phenology (impact on harvests), prices in local markets and possibly its social expression (scarcity, emotions, riots, etc.). All these misidentified sources are another opportunity for risk and climate historians to better understand the natural and coastal

hazards (coastal flooding, earthquakes, tsunamis, landslides, etc.) of the past. Some of these private documents may be limited to weather tables completely disconnected from their socio-economic and climatic contexts. Most of the consulted documents and archives describe the history of coastal flooding in the area of interest. Indeed, the historical inventory identifies and describes ~~the~~ damaging coastal flooding that occurred on the northern coast of France (Nord-Pas-de-Calais and Dunkirk) over the past five ~~past~~ centuries. It presents a selection of remarkable marine floods that occurred in this area and ~~it~~ integrates not only ~~the~~ old events but also those occurr*ing*~~ed~~ after the gauging period ~~has~~ beg*a*~~u~~n. The information is structured around storms and coastal flooding summary sheets. Accompanied and supported by a historian, several research and field missions were carried out and a large number of archival sources ~~have been then~~ explored and, whenever possible, exploited. The historical analysis began with the consultation of the documentary information stored in the rich library of the communal archive of Dunkirk, Gravelines, Calais and Saint-Omer. The most consulted documents were obtained directly from the Municipal archives because the Municipal Acts guarantee a chronological continuity at least ~~since~~ *from* the end of the 16th century *up* to the French Revolution (1789). Very useful for spotting extreme events, they unfortunately provide poor instrumental information. ~~Therefore w~~*W*e *therefore* also considered data ~~out of~~ *from* local chronicles of annals of the city of Dunkirk, as well as reports written by scientists or naturalists to describe tides at Calais, Gravelines, Dunkirk, Nieuport and Oostende. Most of them contain old maps, technical reports, sketches or plans of dykes, sluices and docks designed by engineers of the 18th *to* ~~-~~20th centuries and from which it may be possible to estimate water levels reached during extreme events. ~~Mostly, the b~~*B*ibliographical documents are *mostly* chronicles, annals and memoirs written after the disaster. Finally, for the more recent period*,* available local newspapers ~~have been~~*were* consulted.

Multiplying the sources and trying to crosscheck events allowed us to constitute a database of 73 events. We focused the research on the period between 1500 and 1950, ~~as for~~ *since* most of the time tide gauge observations are available after 1950. The first event took place in 1507 and the last in 1995. Depending on how it is mentioned in the archive and as shown in the left panel of Fig. 2, the collated events were split~~ed~~ in two groups. Storm surge events are events~~,~~ where there is a clear mention of flooding within the sources. Are considered as storms, events where only information about strong wind and gales are available. Except for *the* 19th century, we have much more storm surge events~~,~~ than storms events. All the collected events are summarized in Table 2.

3.3 Data quality control

First of all, it is appropriate to remember that the storm surge is the variable of interest in our historical research. It should~~,~~ however be stressed here that the total sea level, ~~as it~~ *which* is a more operational *piece of* information, is likely to be available most often. The conversion to ~~the~~ storm surge is easily performed afterwards by subtracting the predicted levels. Nevertheless, all types of data require quality control and need to be corrected and homogenized if necessary to ensure that the data ~~are reflecting~~*reflect* real and natural variations of the studied phenomena rather than the influence of other factors. This is particularly the case ~~for~~ *with* historical data ~~that have been~~ taken in different site conditions ~~and have not been taken using~~ *without the use of* modern standards and techniques (Brázdil ~~et al.,~~*et al,* 2010). As mentioned earlier, archival documents are of different nature and qualities. We therefore decided to classify them by their degree of reliability according to a scale ranging between 1 and 4:

- ~~The d~~*D*egree 1: not very reliable historical source (it is impossible to indicate the exact documentary origin). It is particularly the case for HI found *o*~~i~~n the web.
- ~~The d~~*D*egree 2: information found in scientific books talking about storms without clearly mentioning the sources.
- ~~The d~~*D*egree 3: books, newspapers, reports and eyewitness statements citing historical events and clearly specifying its archival sources.
- ~~The d~~*D*egree 4: is the highest level of reliability. Information is taken ~~in~~ *at* a primary source (e.g.~~,~~ an original archival report talking about a storm written by an engineer in the days following the event).

Although ~~the~~ information classified as a category 1 document is not very reliable, it still gives the information that something happened at a date and is therefore not definitely ignored. Typically this type of document needs to be crosschecked with other documents. As shown in Fig. 2 (to the right), the classification of the data reveals a good reliability of collected information as there are no sources classified in category 1 and less than 10% of the sources are in category 2. It is worth noting that paradoxically, the older the information, the more reliable the archival document is.

3.4 The historical surge dataset

The concern is that it is not always possible to estimate a storm surge or a sea level from the information collected for each event. We focus herein on the reconstruction of some events of the 18th century (1720-

1767) where certain HI makes it possible to estimate water levels. As depicted in Fig. 2 (to the right), out of the 73 events, 40 are identified as events causing coastal floods, but not all the sources contain qualitative data or at least some information about water level reached. We selected herein the events with the most information about some characteristics of the event (the water level reached, wind speed and direction and in some cases measured information). The tide coefficient is a ratio of the semi-diurnal amplitude by the mean spring neap tide amplitude introduced by Laplace in the 19th century and commonly used in France since then. Today, the coefficient 100 is attributed by definition to the semi-diurnal amplitude of equinox springtides of Brest. Therefore the range of the coefficient lies between 20 and 120, i.e. the lowest and highest astronomical tides. Calculated for each tide at Brest harbor, it is applied to the complete French metropolitan Atlantic and Channel coastal zone (Simon, 2007). Table 3 shows a synthesis of the six events which we will analyze in more detail, showing the tide coefficient (obtained from the SHOM website), some wind characteristics and water levels reached in Dunkirk and other cities. As with the short-term HI, a description of these events which are quite well documented in the literature is presented in Appendix 2 with a description of some other historical events (of which the available information did not allow an estimate of a storm surge value). Some other HI about other extreme storms, occurring in the period 1767-1897, were collected in the archives and identified as events causing coastal floods, but the sources are not informative enough to lead to the estimation of storm surge values. A description of these events is also presented in Appendix 2. To be able to reduce the CI of the high RLs (the 1000-year one for instance), it is insufficient to have the time-window (the historical period), as the observations or estimates of high surges are unknown. A fixed time-window and magnitudes of the available high storm surges are required to improve the estimates of probabilities of failure. The exhaustiveness assumption of the HI on this time-window will therefore be too crude and will make no sense. The historical period 1770-1897 was therefore eliminated from inference. Fortunately, these discontinuities in the historical period can be managed in the FM (Hamdi et al, 2015). Two non-successive time-windows, 1720-1770 and 1897-2015, will therefore be used as historical periods in the inference.

The extreme storm surges that occurred during the 1720–1767 time-window are then analyzed and the development of a methodology to estimate the surges induced by the events from the last part of the 18th and the 19th century is undergoing. Table 3 shows estimated water levels (for Dunkirk, Gravelines, Calais, Oostende and Nieuport) compared to the associated Mean High-Water Springs (MHWS) which is the highest level reached by springtides (on the average over a period of time often equal to 19 years). De Fourcroy D-R. (1780) presented the water levels in royal foot of Paris, where 1 foot corresponds to 0.325 m and is divided into 12 inches (1 inch = 0.027m) except for the Oostende levels that are given in Flemish Austrian Foot (corresponding to 0.272m and divided in 11 inches). As a first approach the height of the surge above the MHWS level was estimated, which has the advantage that the local reference level does not need to be transposed into the French leveling system and as the historic sea level is considered, there is no need to assess sea level rise which due to climate change can be discarded. De Fourcroy D-R. (1780) gave water levels for the five cities in their respective leveling system: In Calais, zero corresponds to a fixed point on the Citadelle sluice, in Gravelines, zero corresponds to a fixed point on the sluice of the river Aa. For Dunkirk, the "likely low tide of mean springtides" is considered as a zero point and marked on the docks of the Bergues Sluice; we will subsequently refer to this zero as Bergues Zero. The location of the measure point of the Bergues Sluice is presented in Fig. 1 (to the right) on an old map of Dunkirk city. The difference between the observed water levels and the MHWS is the surge above MHWS. The three levels are about the same height, ranging from 1.46m to 1.62m. We calculated the surge above MHWS for Calais, Gravelines, Nieuport and Oostende; they are shown in the second-to-last column of Table 3. It is interesting to note that, for the 1763 and 1767 events, the highest levels were reconstructed in Oostende and the lowest levels in Calais.

For the sake of convenience and for more precision, we needed to refine the surges above MHWS presented in the second-to-last column of Table 3. This refinement required the development of a tide coefficient-based methodology. Indeed, the tide coefficient for each storm event indicates whether surge above MHWS is over- or underrated or approximately right. As this coefficient is calculated for the Brest site and applied to the whole coastal zone, a table showing expected mean levels in Dunkirk for each tide coefficient was established. One tide coefficient estimated at Brest can have different high-water levels at Dunkirk. For this study, it was assumed that the historic MHWS corresponds to the tide coefficient 95. In the developed methodology, all the 2016 hightides for each tide coefficient are used and the water levels for each tide coefficient are averaged. The difference Δ_{WL} between this averaged level and the water level corresponding to the tide coefficient 95 (the actual MHWS) is then calculated and added (or subtracted) to the historic surge above MHWS. Where we have two surges, the mean of the two values is considered. Results for the Dunkirk surges are shown in the last column of Table 4.

In addition to the water levels reached during events and in specific years, other types of HI (lower bounds and ranges) can be collected. For instance, De Fourcroy D-R. (1780) stated that the highest water level measured during the period 1720-1767 was the one induced by the 1767 extraordinary storm. Paradoxical though it may seem at first sight, the skew surge caused by the 1763 storm is greater than the

355 1767 one. A plausible explanation is that the 1767 event ~~was~~ occurred when the tide was higher than that of
356 1763. Fig. 3 shows two examples of HI collected in the archives.

357 For the Dunkirk series, it is interesting to see that it is easier to estimate storm surges induced by events
358 from the 18th century, as the water levels were either measured or reconstructed only a few years after the
359 events took place. During research for his thesis, N. Pouvreau (2008) started an inventory of existing tide
360 gauge data available in different archive services in France. According to him, the first observations of the
361 sea level in Dunkirk were made in the years 1701 and 1702, where time and height were reported.
362 Observations were also made in 1802 and another observation campaign was held during 1835. The first
363 longer series ~~is~~ date~~s~~d from 1865 to ~~—~~1875. For the 20th century, only sparse data is available for the first
364 half of the century. Pouvreau (2008) only listed the data found in the archives of the National Geographic
365 Institute (Institut Géographique National IGN), the Marine Hydrographic and Oceanographic Service (Service
366 Hydrographique et Océanographique de la Marine SHOM) and the Historical Service of Defense (Service
367 Historique de la Défense SHD). During the present study we found evidence that sea levels were measured
368 at the Bergues sluice during the 18th century and that ~~diverse~~ various hydrographic campaigns were ~~made~~
369 carried out during the 19th century (De Fourcroy D-R., 1780). This research and first analysis of historical
370 data shows the potential of the data collected, as we were able to quantify some historical skew surges, but it
371 also shows how difficult and time-~~-~~consuming the transformation of descriptive information into skew surge
372 values is, and that more detailed analysis will be necessary to estimate the other historical surges.

4 Frequency estimation of extreme storm surges using HI

374 In this work, we suggest a method of incorporating the HI developed by Hamdi et al. (2015). The proposed
375 FM (POTH) is based on the Peaks-Over-Threshold with HI. The POTH method uses two types of HI: Over-
376 Threshold Supplementary (OTS) and Historical Maxima (HMax) data which are structured in historical
377 periods. Both kinds of historical data can only be complementary to the main systematic sample. The POTH
378 FM was applied to the Dunkirk site to assess the value of historical data in characterizing the coastal flooding
379 hazard and more particularly in improving the frequency estimation of extreme storm surges.

4.1 Settings of the POT frequency model

381 To prepare the systematic POT sample and in order to exploit all available data separated by gaps, the
382 surges recorded since 1956 were concatenated to form one systematic series. However, it makes for
383 subjectivity in what should be taken as a reasonable threshold for the POT frequency model. Indeed, the use
384 of a too-low threshold can introduce a bias in the estimation by using observations which may not be
385 extreme data, ~~and this~~ which violates the principle of the extreme value theory. On the other hand, the use of
386 a too-high threshold will reduce the sample of extreme data. Coles (2001) has shown that stability plots
387 constitute a graphical tool for selecting the optimal value of the threshold. The stability plots are the
388 estimates of the GPD parameters and the mean residual life-plot as a function of the threshold when using
389 the POT approach~~.~~. It was concluded that a POT threshold equal to 0.75m (corresponding to a~~n~~ rate of
390 events equal to 1.4 events/year) is an adequate choice. The POT sample with an effective duration w_s of
391 46.5 years (from 1956 to 2015) is represented by the grey bars in the left panel of Fig. 4 (a, b and c). As
392 homogeneity, stationarity and randomness of time series are prerequisites in a FA (Rao & Hamed, 2001),
393 non-parametric tests such as the Wilcoxon test for homogeneity (Wilcoxon, 1945), the Kendall test for
394 stationarity (Mann, 1945), and the Wald-Wolfowitz test for randomness (Wald & Wolfowitz, 1943) are
395 applied. These tests were passed by the Dunkirk station at the 5% level of significance.

4.2 The POTH frequency model

4.3 Settings of the frequency model with HI (POTH)

398 An important question arises ~~is related~~ with regard to the exhaustiveness of the HI collected in a well-~~-~~
399 defined time -window (called herein the historical period). In order to properly perform the FA, this criterion
400 must be fulfilled. Indeed, we have good evidence to believe that other than the 1995 storm surge, the surges
401 induced by the 1897, 1949 and 1953 storms are the biggest ~~on~~ for the period 1897-2015. The POTH FM was
402 first applied with a single historical dat~~um~~a which is that of 1953 represented by the red bar in Fig. 4-a. It ~~has~~
403 not ~~been~~ complicated to demonstrate that this event is undoubtedly an outlier. Indeed, in order to detect
404 outliers, the Grubbs-Beck test was used (Grubbs and Beck, 1972). As mentioned in the previous section,
405 some historical extreme events experienced by ~~the~~ Dunkirk city are available in the literature. Only this
406 information (including the 1953 ~~once~~event) is considered in this first part of the case study.

407 Otherwise, HI is most often considered in the FA models for pre-gauging data. Less or no attention has
408 been given to ~~the~~ non-recorded extreme events that occurred during the systematic missing periods. As
409 mentioned earlier in this paper, the sea level measurement induced by the 1995 storm was missed and a

value of the skew surge (1.,15m) was reconstructed from information found in the literature (Maspataud, 2011). As this event is of ordinary intensity and has taken place very recently, it is considered as ~~a~~ systematic data even if this type of data can be managed by the POTH FM by considering ~~them~~ it as HI (Hamdi ~~et al.~~et al, 2015). The HI collected from both literature and archives with some model settings are summarized in Table 5 and the POTH sample with a historical period of 72.,51 years is presented in Fig. 4-b. Parameters characterizing datasets including both systematic and HI were introduced in Hamdi ~~et al.,~~et al, (2015). The HI is used herein as HMax data that complements the systematic record (with an effective duration D_{eff} equal to w_s) on one historical period (1897-2015) with a known duration $w_h = w_{HMax} = 2015 - 1897 + 1 - D_{eff}$ ($w_h = 72,51\ years$) and three historical data ($n_k = 3$). Other features of the POTH FM have been used. A parametric method (based on the Maximum Likelihood) for estimating the Generalized Pareto Distribution (GPD) parameters considering both systematic and historical data have been developed and used. The maximum likelihood method was selected for its statistical features especially for large series and for the ease with which any additional information (i.e. the HI) is incorporated in it. On the other hand, the plotting positions exceedance formula based on both systematic observations and HI (Hirsch, 1987; Hirsch and Stedinger, 1987; Guo, 1990) is proposed to calculate the observed probabilities and ~~it~~ has been incorporated into the POTH FM considered herein. The reader is referred to Hamdi et al. (2015) for more theoretical details on the POTH model and on the Renext package used to perform all the estimations and fits.

5 Results and discussion

We report herein the results of the FA applied to the Dunkirk tide gauge. As with any sensitive facility, high Return Levels (RLs) (100, 500 and 1000-year extreme surges, for instance) are needed for the safety of NPPs. The results are presented in the form of probability plots in the right panel of Fig. 4 (d, e, and f). The theoretical distribution function is represented by the solid line in this figure, while the dashed lines represent the limits of the 70% CIs. The HI is depicted by the empty red circles, while the black full ones represent the systematic sample. The results (estimates of the desired RLs and uncertainty parameters) are also summarized in Table 6. Fitting the GPD to the sample of extreme POTH storm surges yields the relative widths $\Delta CI/S_T$ of the 70% CIs (the variance of the RL estimates are calculated with the delta method).

The FA was firstly performed considering systematic surges and the 1953 storm surge as ~~a~~ historical data. It can be seen that the fit of the POTH sample including the 1953 historical event (with w_h equal to 16.,5 years) presented in Fig. 4-d (called hereafter the initial fitting), is poor at the right tail and more specifically, at the largest storm surge (the historical data of 2.,13 m occurred in 1953) which have a much lower observed return period than its estimated one. The estimates of the RLs of interest and uncertainty parameters (the relative width $\Delta CI/S_T$ of the 70% CIs) are presented in columns 2-3 of Table 6. These initial findings are an important benchmark as we follow the evolution of the results to evaluate the impact of additional HI. 100-, 500- and 1000-year quantiles given by the POTH FM with the 1897, 1949 and 1953 historical storm surges included are about 3-6% higher than those obtained by the initial POTH FM. This result was expected as the additional historical surges are higher than all the systematic ones. The relative widths of the CIs are about 20-25% narrower.

Unlike the 1897 historical event, the 1949 and 1953 ones have a lower observed return period than their estimated one. A plausible explanation for this result is that the body of the distribution is better fitted than the right tail one and this is a shortcoming directly related to the exhaustiveness assumption used in the POTH FM. Indeed, as stated in Hamdi et al. (2015) and as mentioned above, a major limitation of the developed FM arises when the assumption related to the exhaustiveness of the information is not satisfied. This is obviously worrying for us because the POTH FM is based on this assumption. Overall, using additional data in the local FM has improved the variances associated ~~to~~ with the estimation of the GPD parameters but did not conduct to robust estimates with a better fitting (particularly at the right tail, the high RLs being very sensitive to the historical values) if the assumption of exhaustiveness is still strong. This first conclusion is likewise graphically backed by the CIs plots shown in Fig. 4-e. Nevertheless, as the impact of historical data becomes more significant, there is an urgent need to carry out a deeper investigation of all the historical events that occurred in the region of interest (Nord-Pas-de-Calais) over ~~the~~a longest historical period. In order to have robust estimates and reduced uncertainties, it is absolutely necessary that the collected information be as complete as possible.

The robustness of the POTH FM is one of the more significant issues we must deal with. The main focus of this discussion is the assessment of the impact of the additional HI (collected from the archives) on the frequency estimates for high RLs. The same FM was performed but with the long-term additional HI (collected in the archives) and different settings (Table 5). The results of the POTH FM using HI from both literature and archives (called hereafter the full FM) are likewise summarized in the last two columns of Table

467 6. The results are also presented in the form of a probability plot (Fig. 4-f). Fig. 7 consists of two subplots
468 related to the FA of the Dunkirk extreme surges. The left side (Fig. 4-c) shows collected data: the systematic
469 surges are represented ~~with~~ by the grey bars, the historical surges extracted from the literature ~~with~~ by red
470 bars and those extracted from the archives (estimated and corrected with regards to the tide coefficients) are
471 represented ~~with~~ by the green ones. We can also see the two time -windows (the blue background areas in
472 the graph) 1720-1770 and 1897-2015 used in the POTH FM as historical periods. The right side shows the
473 results of the full FM. As mentioned earlier in this paper, to consider the full POTH FM, six historical storm
474 surges distributed equally ($n_k = 3$) over two not-successive time -windows: 1720-1770 ($w_{HMax1} = 50$ *years*) and
475 1897-2015 ($w_{HMax2} = 72.5$ years, knowing that $w_s = 46.5$ years) are used as historical data. In the plotting
476 positions, the archival historical surges are represented by green squares, while those found in the literature
477 are depicted by red circles. The fitting presented in Fig. 4-f shows a good adequacy between the plotting
478 positions and theoretical distribution function (calculated probabilities of failure). Indeed, all the points of the
479 observed distribution are not only inside the CI, but even better, they are almost on the theoretical
480 distribution curve. The results of Table 6 show that:
481 - The RLs of interest had increased by only 10 to 20 cm. This is an important element of robustness. Indeed,
482   adding or removing one or more extreme values from the dataset does not significantly affect the desired
483   RLs. In other words, it is important that the developed model is not very sensitive (in terms of RLs used as
484   design bases) to a modification in the data regarding very few events. As a matter of fact, the model owes
485   this robustness to the exhaustiveness of the available information.
486 - The relative widths of CIs with no archival HI included are 1.~~,~~5 times larger than those given by the full
487   model. This means that the user of the developed model is more confident in the estimations when using
488   the additional HI collected in the archives.
489   After collecting HI about the most extreme storm surge events in the 18th and 20th centuries, it was first
490 found that the 1953 event is still the most important one in term~~s~~ of magnitude. The developed POTH FM
491 attributed a 200-year return period to this event. The value of the surge induced by the 1953 storm is
492 between 1.~~,~~75m and 2.~~,~~50m. That said, it is interesting to note that this CI includes the value of 2.~~,~~40m
493 estimated by Le Gorgeu and Guitonneau (1954). This may be a reason to think that the continuation of our
494 work on the quantification of the skew surges that occurred in the 19th century will ~~may be~~perhaps reveal
495 extreme surges similar to that induced by the 1953 storm.

6 Conclusion & perspectives

497 To improve the estimation of ~~the~~ risk associated ~~to~~ with exceptional high surges, HI about storms and coastal
498 flooding events for the Nord-Pas-de-Calais was collected by historians for the 1500-1950 period. Qualitative
499 and quantitative information about all the extreme storms that hit the region of interest were extracted from a
500 large number of archival sources. In this paper, we presented the case study of Dunkirk in which the
501 exceptional surge induced by the 1953 violent storm appears as an outlier. In a second step, the information
502 collected (in both literature and archives) was examined. Quality control and cross validation of the collected
503 information indicate~~s~~ that our list of historic storms is complete as regards extreme storms. Only events that
504 occurred in the periods 1720-1770 and 1897-2015 were estimated and used in the POTH FM as historical
505 data. To illustrate challenges and opportunities for using this additional data and analyzing extremes over a
506 longer period than was previously possible, the results of the FA of extreme surges was presented and
507 analyzed. The assessment of the impact of additional HI is carried out by comparing theoretical quantiles
508 and associated confidence intervals, with and with no archival historical data, and constitutes the main result
509 of this paper.
510   The conclusions drawn in previous studies were examined in greater depth in the present paper. Indeed,
511 on the basis of the results obtained previously (Hamdi et al, 2015) and in the present paper, the following
512 conclusions are reached:
513 - The use of additional HI over longer periods than the gauging one, can significantly improve the
514   probabilistic and statistical treatment of a dataset containing an exceptional observation considered as an
515   outlier (i.e. the 1953 storm surge).
516 - As the HI collected in both literature and archives tend to be extreme, the right-~~-~~tail distribution has been
517   reinforced and the 1953 "exceptional" event does not appear as an outlier any more.
518 - As this additional information is exhaustive (relatively to the corresponding historical periods), the RLs of
519   interest ~~had~~ increased very slightly and the confidence intervals were reduced significantly.
520 An in-depth study could help to thoroughly improve the quantification method of the historical surges and
521 apply the developed model on other sites of interest.

Appendix 1: HI collected in the literature

523 01/03/1949: A violent storm with mean hourly wind speeds reaching almost 30m.s[-1] and gusts of up to
524 38.~~,~~5m.s[-1] (Volker, 1953) was the cause of a storm surge that reached the coast of ~~the North of~~ Northern
525 France and Belgium in the beginning of March 1949. The tide gauge of Antwerp in the Escaut estuary

526 measured a water level higher than 7m TAW (a reference level used in Belgium for water levels) which classifies
527 this event as a "*buitengewone sStormvloed*", an extraordinary storm surge (Codde and De Keyser 1967). For
528 the Dunkirk area two sources reporting water levels were found: tThe first saying that 7.30m was reached as
529 a maximum water level at the eastern Dike in Dunkirk, exceeding the predicted high tidehightide, i.e. 5.70m,
530 with 1.60m (Le Gorgeu and Guittoneau 1954). A second document relates that the maximum water level
531 reached was about 7.55m at Malo-les-Bains, which would mean a surge of 1.85m (DREAL Nord-Pas-de
532 -Calais). It is worth noting that the use of proxy data (i.e. the descriptions of events in the historical sources
533 summarized in Table 1) to extract sea-level values and to create storm-surges databases is seriously
534 limited. For the 1791 and 1808 storms, there is sufficient evidence that extreme surge events have taken
535 took place (extreme water level on Walcheren Island) but the sources are not informative enough to estimate
536 water levels reached in Dunkirk. A surge of 1.25m is given for the storm of 1921. The problem is that the
537 type of the surge (instantaneous or skew), the exact location at which it was recorded and the hydro-
538 meteorological parameters are not reported. For the skew surge of 1949, two different values at two
539 locations are given. There are predicted and observed water levels for the storms of 1905 and 1953 in
540 Calais, which indicated that the differences is a skew surge, but likewise neither the exact location nor the
541 information about the reference level were are furnished. The need offor tracing back to "direct data"
542 describing a storm and its consequences becomes clear, as well as performing a cross-check of the data on
543 a spatial and factual level, as Brázdil (2000) also suggests.

544 28/11/1897: What was felt as stormy winds were felt in Ireland on the 27th November 27th, 1897 became an
545 eastward-moving storm with gale-force winds over Great Britain, Denmark and Norway (Lamb, 1991). This
546 storm caused interruption of telephone communications between the cities of Calais, Dunkirk and Lille and
547 great damage to the coastal areas (Le Stéphanois, November 30th, 1897). At Malo-les-Bains, a small town
548 close to Dunkirk, the highest water level reached 7.36m although the high tidehightide was predicted with at
549 5.50m, resulting in a skew surge of 1.86m that caused huge damage to the port infrastructures (DREAL
550 Nord-Pas-de-Calais).

551 14/01/1808: During the night from 14th to 15th January 14th to 15th, 1808, "a terrible storm, similar to a storm
552 that hit the region less than one year before on 18 February 18, 1807" hit the coasts of the most northern
553 parts of France up to the Netherlands. This storm caused severe flooding as well in the Dunkirk area as also
554 in the Zeeland area in the south western parts of the Netherlands where the water rose up to 25 feet on the
555 isle of Walcheren (i.e. 7.62m). The journal also reports more than 200 deaths. For the Dunkirk area, the last
556 time the water levels rose as high as in January 1808 was 2nd February 2nd, 1791. Unfortunately, this source
557 did does not provide any information that we can quantify or any information on the meteorological and
558 weather conditions that we can use to reconstruct the storm surge value.

Appendix 2: HI collected in the archives

560 1720-1767: In essays written by a mathematician of the royal academy of science, De Froucroy D-R, who
561 describes the tide phenomenon oin the Flemish coast, some extreme water levels observed within the study
562 area are reported and described. The author refers to five events that occurred during the period 1720 to
563 1767. The same information is confirmed by a Flemish scientist, Dom Mann (1777, 1780). De Froucroy D-R
564 witnessed the water levels induced by the 1763 and 1767 storms and reconstructed the level induced by the
565 1720 event in Dunkirk. Water levels at that time are given for the cities of Dunkirk, Gravelines and Calais in
566 the "pied du roi" unit ("foot of the king" was a French measuring unit, corresponding to 0.325m) above local
567 mean low-water springs. The French water levels are completed by measurements made in ancient Flemish
568 feet above the highest astronomical tides for the cities of Oostende and Nieuport (De Fourcroy D-R., 1780;
569 Mann, 1777, 1780). The upper panel of Fig. 3 shows an example of HI as presented in the archives (De
570 Fourcroy D-R., 1780).
571 The 1720 event is a memorable event for the city of Dunkirk, as the water level during spring
572 tidespringtide was increased by the strong gales blowing from north-western direction which destroyed the
573 cofferdam built by the British in the year 1714, cutting the old harbor off from sea access and prohibiting any
574 maritime trade, and thus slowly causing the ruin of the city. The socio-cultural impact of the natural
575 destruction of the cofferdam was huge, as it restarted the trading of in the city (Chambre de Commerce de
576 Dunkerque 1895, Plocq, 1873, Belidor, 1788). In 1736, the only sea level available is given for Gravelines
577 harbor, but extreme water levels are confirmed in the sources as they mention at least 4 feet of water in a
578 district of Calais, and water levels that overtopped the docks of the harbor in Dunkirk (Municipal Archive of
579 Dunkirk DK291, Demotier, 1856). As mentioned above, communal and municipal archives contain plans of
580 dykes, docks and sluices of in Dunkirk harbor designed by engineers with the means available at that time,
581 and such sketches were recovered. A 1740 sketch showing a profile of the Dunkirk harbor dock is presented
582 in the lower panel of Fig. 3 for illustrative purposes only. The use of these plans and sketches in the
583 estimation of some historical storm surges is ongoing. The lower-lying streets of Gravelines were accidently
584 flooded by the high water levels in March 1750. The fact that an extreme water level was also reported also

in Oostende for the same day confirms the regional aspect of the event. The surge of 1763 occurred in a period with mean tidal range, but water level exceeded the level of mean spring high tidehightide in Dunkirk, Calais and Oostende. Unfortunately no more information about the flooded area is available. Strong west-north-westerly winds caused by a quick drop of the in pressure produced high water levels from Calais up to the Flemish cities. It is, at least for the period from 1720 to 1767, the highest water level ever seen and known. The 1720 and 1767 events show good evidence of the wind direction and wind intensity, while in various sources, except for the water levels reported, the events from 1736, 1750 and 1763 are in different sources always cited together and described as "*extraordinary sea-levels that are accompanied or caused by strong winds blowing from South-West to North*" (De la Lande, 1781, De Fourcroy D-R., 1780, Mann, 1777, 1780). As with the 1897-2015 historical/systematic periods, the same question related to the exhaustiveness of the HI collected in the 1720-1770 historical period arises. As our historical research on extreme storm surges occurred in this time -window was very thorough, we have good reasons to believe that the surges induced by the 1720, 1763 and 1767 storms are the biggest on for that historical period.

1767–1897: For the 1778, 1791, 1808 and 1825 events, the sources report strong that winds were blowing from north-westerly directions and that in Dunkirk the quays and docks of the harbor were overtopped as the highest water levels were reached. We know that, after the event of February 1825, at least 19 storms events occurred and we have good evidence to believe that some of them induced extreme surges, but either the information available is not sufficient to draw an approximate value of the water level, or. tThe quantification of the storm surges induced by these events is complicated and time--consuming.

1936: The 1936 event can be considered as a lower bound, as the document from the archive testifies that the "water level was at least 1m higher than the predicted tide" during the storm that occurred on the night of 4st December 1st, 1936 (Municipal Archives of Dunkirk 4S 881). The 1936 event, which can be designated described as a moderately extreme storm, is the only one collected on the 50-year time -window (1897-1949). As the surge lower bound value induced by this event is too small (i.e. exceeded more than 10 times during the systematic period), it could be exceeded several times during the 1897-1949 period. Its involvement in the statistical inference will have the opposite effect and will not only increase the width of the CI but will also degrade the quality of the fit. The 1936 historical event was therefore eliminated from inference.