Response point-by-point to Referee #1

Dear Referee #1,

We would like to thank for your interesting review. We are glad to see that you think that this study is a relevant approach to reduce uncertainty associated with the estimation of return periods. When we deal with coastal hazards, we cannot neglect any extreme data available. Historical data may store very important information. This methodology enables to use historical data with gauged data in order to define events of different severity for coastal management purposes.

In addition, thank you to have described our manuscript as clearly structured.

Overall, we took into account your suggestions to improve the quality and to improve even more the clarity and the form of our study.
In the following, we will give to the Referee #1 point to point detailed answer to every specific comments:

1) The title is descriptive, but very long. If possible, could this be shortened?

   We agree with this helpful consideration did by Referee #1. A paper title must to provide to readers synthetically the main topic of the study. For this reason, it has been changed as follows: "The use of historical information on the regional frequency analysis of extreme skew surge". We think that this helps to understand that our work provides a suitable framework for the use of historical information; the credible duration is the innovative tool that we develop for achieving this framework. The reader will discover it reading the paper.

2) The term FAB is not given in full within the manuscript what does it stand for? Why is FAB not a keyword if this is the approach used? It must be a key method to be in the title and initiate the start of the conclusion.
3) P8, L11, if the RFA and FAB method are the same why is FAB used in the title when RFA is introduced/discussed in more detail. What does FAB stand for? I suggest RFA is used throughout with 1 introduction to FAB as an alternative name, giving the name in full and as the acronym.
4) P11, L25, The conclusion starts with the FAB method, when the RFA terminology is used more frequently in the paper. Try to consistently use 1 terminology, only indicating it could also have another name once.

   All these helpful considerations are taken into account in the new version of the manuscript. In particular, note that the RFA (Regional Frequency Analysis) is a general term for a statistical approach using regional data. Several declinations exist in the literature, such as the Hosking and Wallis approach or the pooling approach. FAB is thus a new declination of the RFA, allowing the use of historical information.
   The term FAB is thus introduced as a keyword of this study and it is defined in the **abstract (line 19).** FAB is a regional method that allows to take into account not only systematic extremes but also historical extremes based on the new concept of credible duration. In order to better clarify what is FAB method, some parts are added in the new paper version as follows:

   Abstract, L20-22: *"The proposed RFA method (hereinafter referenced as FAB, from the name of the authors) allows the use of historical data together with systematic data thanks to the use of the credible duration concept."*

P4, L14-15: *"Thanks to this new concept of duration, in this study, the FAB methodology is developed in order to use together historical data and systematic data in a Regional Frequency Analysis of extreme events."*

5) In the abstract return periods up to 1000 years are mentioned, is there a reference for this level being used for rare events that can be added. The need for cost effective defence to withstand this level of event is required in the introduction.

This is an important point that allows readers to understand why we are estimating extreme hazards for big return period. In order to provide a clear explanation, we add the reference of the French Nuclear Safety Authority to the introduction. In addition, the need of coast effective defence is outlined as follows:

P1, L26-32: *"Flood risk is a noteworthy threat for anthropic and industrial settlement along the coasts. Coast defence constructions have to withstand extreme events. Therefore, in this context, a statistical estimation of extreme marine variables (such as sea levels or skew surges, the difference between the maximum observed sea level and the maximum predicted astronomical tide level during a tidal cycle (Simon, 2007; Weiss, 2014c)) occurrence probability is necessary. These variables associated to high return periods (up to 1000 years (ASN, 2013)) are estimated by statistical methods linked to the Extreme Value Theory (Fréchet, 1928; Gnedenko, 1943; Gumbel, 1958; Coles, 2001)."*

6) Skew storm surge can simply be referred to as skew surge throughout. However, the term needs to be defined with reference in the introduction.

The skew storm surge term has been replaced by skew surge throughout the article. The term has been defined with reference as follows:

P1, L28-29: *"the difference between the maximum observed sea level and the maximum predicted astronomical tide level during a tidal cycle (Simon, 2007; Weiss, 2014c)."*

7) P1, L28, delete 'The' and start the sentence 'Flood risk is'.

The error has been corrected (**P1, L26** of the new paper version).

8) P2, L1, suggested updates to text: In the past, return level combinations have...... recorded at a single......

The suggestion has been accepted and performed. Anyway, we are preserving the word "estimations" instead of the suggested word "combinations" that seems more suitable to this framework (**P1, L33** of the new paper version).

9) P2, L6, move 'together' from before 'all' to after 'data'.

The error has been corrected (**P2, L3** of the new paper version).

10) P2, L18, 'compared with a'.

The sentence has been removed in the new paper version.

11) P2, L29, 'In particular.....' This sentence is unclear please reword.

The sentence has been completely reworded and better explained as follows:

"Specifically, historical data are all the skew surge events that were not recorded by tide gauge and, consequently, systematic data are all the measurements of skew surge recorded by tide gauge." (**P2 L25-27** of the new paper version)

12) P2, L30-31, suggested updates to text: ... and look for information during....

The suggestion has been accepted and performed (**P2, L33** of the new paper version).

13) P2, L32, 'sea level measurements'

The error has been corrected (**P2, L34** of the new paper version).

14) P3, L2, replace 'whichever' with 'applying a'

The error has been corrected (**P3, L7** of the new paper version).

15) P3, L6, delete 'the support of'

The error has been corrected (**P3, L11-12** of the new paper version).

16) P3, L9, suggested update to text: is a key step....

The suggestion has been accepted and performed (**P3, L13** of the new paper version).

17) P3, L10, this paragraph describes how data gaps in historical data suggest there were no extremes. Previously gaps in gauge data were associated to instrument failure during extremes. Ensure there is always clarity about which data is being discussed throughout the paper, historical data or gauge data.

This point are better discussed and clarified in the new Introduction. We defined as "historical data" are all the isolated data points derived from non-systematic measurements of skew surges. Note that this definition includes also isolated data point reconstructed during gaps of the systematic measurements. This is because statistically speaking the FAB methodology allows to treat them similar to the isolated data reconstructed from pre-gaugement period. We decided however to keep the historical data wording because this has been widely used in the literature and help focusing on the need for an effort (and the potential) in data archaeology.

P2-3, L25-2: *"Specifically, historical data are all the skew surge events that were not recorded by tide gauge and, consequently, systematic data are all the measurements of skew surge recorded by tide gauge. One obvious and very promising way for collecting information not recorded by a tide gauge is to investigate and reconstruct historical events occurred before the starting date of gauge observations. This activity is also known as data archaeology and has been demonstrated very promising and useful for several study in the literature. Note however, that this definition also include isolated data point reconstructed during gaps of the systematic measurements. This is because statistically speaking they could be treated similarly to the isolated data reconstructed from pre-gaugement period. To find historical events it is thus necessary to go back in time and to look for information before the gauging period or during each data gap in a time series data. In fact, during important storms, tide gauge may be affected by a partial or total failure in the sea level measurements and so this kind of events cannot be detected. Thanks to non-systematic records, we can*

hold a numerical value to these ungauged extreme events. This allows to use a bigger number of storms in the statistical assessment of the extreme skew surges".

P3, L11-14: *"As historical data are isolated data points thus the duration of the observation is not defined. What happened during the time period between two or more past extreme events remains unknown. However, the coverage period estimation of historical data is a key step in the extreme value statistical models (Prosdocimi, 2017)."*

18) P3, L14, add 'the' estimate 'of' the coverage

The sentence has been removed in the new paper version.

19) P3, L14, reword: the exhaustively of the.

The sentence has been reworded as follows:

P3, L19-21: *"This allows the estimation of the coverage period and the indirect assumption that the collected historical data are exhaustive throughout the coverage period assessed."*

20) P3, L19, 'assess' might be more appropriate than 'guess'.

The error has been corrected (**P4, L9** of the new paper version).

21) P3, L24, suggest updating text: gauges positioned along the UK, French and Spanish coasts....

The suggestion has been accepted and we switch directly "placed" with *"located"* which seems more appropriate (**P4, L16** of the new paper version).

22) P3, L29, update 'called systematic'.

The sentence has been removed in the new paper version.

23) P3, L30, add This allows 'us' to....

The sentence has been removed in the new paper version.

24) P3, L30, Is the duration of the biggest event only considered as the duration after theat event until the next, or is if the window over that event that falls midway between this event and an earlier/later event? Please clarify.

All the concept of duration is better explained in the new paper version. This comment can be replied through this paragraph:

P5, L9-15: *"When historical skew surge are available, the computation of λ is a tricky topic. In fact, although the number of historical data n_{hist} over the sampling threshold is known, we do not have any information on which happens during non-gauged period. It is thus impossible to define a duration over which the n_{hist} data are observed. In other words, fixing whatever sampling threshold, we need this information for defining the frequency of occurrence of the events over the threshold (see the Eq.2). We called the missing information "historical duration" d_{hist} or, in other words, the time period for which historical skew surge above sampling threshold are estimated to be exhaustively recorded at the best of our knowledge."*

25) P3, 30, 'been the biggest events'

The sentence has been removed in the new paper version.

26) P3, L33, 'exceedance'.

The sentence has been removed in the new paper version.

27) P4, L8, suggested update to text: alters the duration of the systematic events

The sentence has been removed in the new paper version.

28) P4, L9, 'that occur in'

The sentence has been removed in the new paper version.

29) P4, L11, 'tide gauge data with'

The sentence has been removed in the new paper version.

30) P4, L25, reword 'trend lack on skew surge frequency'. It is unclear what is meant at present.

The sentence has been reformulated as follows:

P5, L23-24: *"Eq.3 is in accordance with the lack of a significant trend on the extreme skew surges frequency during the second part of 19th century, 20th century and the first part of 21st century in the regions considered in this study (Fig.5)."*

31) P4, L26, introduce Fig 7 at this point so the reader knows the region that is being considered.

The Fig.5 (the old Fig.7) has been introduced at the end of the sentence in brackets (**P5 L24** of the new paper version).

32) P4, L27, suggested updates to text: The lack of trend in skew surge frequency also means there is a lack of trend in storm frequency.

The sentence has been removed in the new paper version.

33) P4, L28, can you provide a reference as an example in the literature.

The paragraph has been rewritten with some references as follows:

P5, L25-27: *"According to current scientific literature, based on past observations, frequency of storms has no defined trend (Hanna et al., 2008; Barring and Fortuniak, 2009; MICORE Project, 2009). Therefore, the hypothesis that storm frequency is constant has been assumed."*

34) P4, L33, Update the brackets so the authors' names are within the sentence.

The update has been performed (**P5 L31** of the new paper version).

35) P5, L5, this sentence 'If frequency...' is unclear. Please reword.

The sentence has been reformulated as follows:

P6, L3-6: *"Therefore, the hypothesis that storm frequency is constant may be a reasonable working hypothesis for demonstrating the approach. Obviously, it is crucial to test this hypothesis before the application of the approach in the present state to any new dataset. Note that the methodology could be adapted quickly in the future for non-stationary datasets, where the introduced concept can be adjusted. For skew surges, this hypothesis is ensured".*

36) P5, L5-10, UKCP09 should be referenced as a source of climate information for the European shelf seas

The UKCP09 is an excellent reference that has to be added in this paper. Thanks to Referee #1 for the suggestion.

37) P5, L29 & P6, L9, u is not used in any equations so does not need to be defined.

We define u as sampling threshold for each case shown in Section 3 and displayed in Fig.2.

38) P6, L1, the random locations should be presented as example locations and the place names given in the figure captions.

We are in according with your comment and the paragraph is rewritten as follows:

P8, L16-21: *"Fig. 1 shows the functional example of the whole skew surge series recorded at La Rochelle tide gauge (France). Systematic skew surges are represented by the continuous series of data plotted in black from about 1940 to 2017. They are observed during a fixed period equal to the recording time period, in this case, 32.88 years.*
The historical skew surges are artificial data generated by ourselves for this descriptive illustration and they are represented as green cross points. As they are not supposed to be obtained by continuous systematic records of skew surge measurements, the duration of their support is unknown."

In addition the titles of the Figures 1,2 have been updated.

39) P6, L12, The four cases need to be linked to a table and the figures.

A link of the figures 1 and 2 of the table 1 has been added in the sentence (**P8 L29-30** of the new paper version).

40) P6, L15 &L16, the word 'more' needs replacing. Moreover might be more appropriate.

The sentence has been reworded as follows:

P9, L11-13: *"In particular, when the threshold is extremely high, it is almost sure that no extreme skew surges occurred during the coverage period of historical data. If the threshold is lower, then it is less certain that all extreme skew surges have been observed in the historical part."*

41) P6, L24-32, the cases need to be introduced before L29 and Fig. 5. This information could be moved to follow the paragraph that end L5 so the four cases are introduced.

The paragraph with the four cases has been moved after the paragraph that end L8. We corrected also some grammatical errors in the first sentence as follows:

P8 L32-33: *"Case 1 uses a sampling threshold u_{C1} corresponding to 0.5 extremes per year (λ=0.5) and lead to a credible duration of 42.88 years (Tab.1)."*

42) P9, L9, 'by Ward's'

The error has been corrected (**P10, L18** of the new paper version).

43) P9, L29, suggest updates to text: A regional sample is formed of normalised... are divided by....

The suggestion has been accepted and corrected (**P11, L5** of the new paper version).

44) P9, L30 & P10, L28, 'enables us to'.

The errors have been corrected (**P11, L6 & P12, L2** of the new paper version).

45) P10, L1, 'a storm's frequency'

The sentence has been removed in the new paper version.

46) P10, L1, 'equal at all'

The error has been corrected (**P11, L9** of the new paper version).

47) P10, L10, 'visual look of the regional'

The error has been corrected (**P11, L17** of the new paper version).

48) P10, L16, 'years at each'

The error has been corrected (**P11, L23** of the new paper version).

49) P11, L25, 'trend in storm frequency'.

The sentence has been removed in the new paper version.

50) P12, L7, 'of storm frequency'.

The error has been corrected as *"of storm magnitude"* (instead of "of storm frequency") (**P13, L18** of the new paper version).

51) Figures 1-4, where there is no data there should be no line joining the periods of data collection. This space should be left empty to indicate no data.

The Figures 1-4 have been updated and the spaces you suggest have been left empty. In addition, all the figures are resumed in only one with different thresholds.

Response point-by-point to Referee #2

Dear Referee #2,
We would like to thank you for the time you dedicated to our paper. We agree with a lot of your comments and we are sure that your help is adding a lot of value to our work, in particular, in the way in which the new approach is now presented.
Thank you for acknowledge that the idea to improve the RFA including historical data point and non-systematic data have some interest.

Moreover, we agree on the fact that at this stage the manuscript needs some rewritings and clarifications. The fact that the idea of credible duration is acknowledged to be rather simple, but that it took to the Referee #2 quite a long of time to grasp it from explanation given in section 2 (As for the "systematic skew storm surges" "historical data" meaning and definition) is a brilliant example of the fact that the presentation and definitions needs to be clarified. With this in mind, we largely rewrote the manuscript. We thanks the Referee #2 for pointing this out and this is what we did in the new version of the manuscript.

In the following, we will show to the Referee #2 how we took into account these general remarks and we gave point to point detailed answer to every specific comments

1) The authors don't really discuss any reason why their technique should work better than using a perception threshold.

This is a very important point, pointing directly to the first motivation of the work. We were indeed quite frustrated with the fact that the perception threshold hypothesis were constantly challenged and basically impossible to verify. In the new version of the manuscript we point it out in a new paragraph as follows:

P3-4, L21-5: *"So far, the perception threshold has allowed to use historical information in an extreme frequency analysis (Payrastre et al., 2011; Payrastre et al., 2013; Bulteau et al., 2014) but the hypothesis for using this approach seems sometimes difficult to prove.*
Indeed, the perception threshold is usually defined equal to an historical event value, based on the principles that, if any other historical event would have occurred it would have been recorded (Payrastre et al., 2011; Bulteau et al., 2014). This is a quite strong hypothesis, which is difficult to be verified and validated. Moreover, the duration associate to the perception threshold is usually defined as the duration between the oldest historical data and the start of the systematic series (Payrastre et al., 2011; Hamdi et al., 2015).
The main point of novelty of the paper is to propose a new approach in order to deal with these questions in a more satisfactory way, the equivalent duration of historical data.
Note that the new approach is required for the application of the RFA, at least in the POT approaches (Van Gelder and Neykov, 1998; Bernardara et al., 2011; Bardet et al., 2011; Weiss et al., 2014b). Indeed, the definition of the perception threshold equal to an historical value precludes the possibility to use RFA methodology proposed by Weiss (2014c). Specifically, in Weiss regional methodology, the number of extreme events per year λ must be equal in each site of a region in order to estimate a regional distribution. This is impossible to achieve if the threshold is linked to the characteristic of the local observations.
In addition, when historical information (i.e. non-systematic data) is collected during a total failure of tide gauge, this influences the frequency of the event. One could only re-estimate λ if the credible duration is re-estimated (for details, see the section 2)."

2) They explain how to estimate this credible duration, but in my opinion they don't provide persuasive explanation of why this should be considered a reasonable estimation of the time horizon.

We agree that a direct explanation on why credible duration should be considered a reasonable estimation of the time period is needed. We added in into the text. To resume we consider this new approach needed because it takes into account physical consideration (storm frequency) for the estimation of the credible duration instead of frustrating strong hypothesis on the exhaustiveness of data (such as in the perception threshold approach). For this reason, we present this in our paper as follows:

P4, L7-13: *"This evaluation is based on the hypothesis that the frequency of the extreme events is not changing significantly in time. In other words, the average number of extreme events should be the same (obviously this has to be carefully verified and demonstrated case by case). In this way, we can assess the coverage period of historical data. Based on the frequency value of systematic storms occurrence, our hypothesis introduces a physical and tangible element in the estimation of the historical duration. For this reason, we state that this novelty allows to overtake the concept of perception threshold providing a credible estimation of the unknown historical duration used to estimate the extreme quantiles."*

P13, L5-9: *"In the past, the estimation of the period covered by the historical data collection has usually limited the use of historical data. The concept of perception threshold was employed so far to use historical data even though its hypothesis are quite strong and difficult to verify. For this reason, the new concept of credible duration has been formulated that it is based on physical considerations supported by several scientific studies. This leads to use historical isolated data points on the extreme value analysis."*

3) They don't explain how exactly they make use of this quantity in the extreme value analysis.

In the previous of the paper, we guessed wrongly that the use of this quantity in the extreme value analysis was implicit. However, as the Referee #2 suggests, where we make use of the credible duration quantity, needs an explicit explanation to understand the real purpose of this study. In particular, the duration of the observations are used in the EVA approach for transforming a probability of occurrence in a frequency of occurrence, introducing a time unit in the estimation.

This is exposed in the new paper version as follows:

P4-5, L21-8: *"In the Extreme Value Analysis (EVA) of natural hazards, the estimations of extreme quantiles are performed applying statistical tools to a dataset of extreme variables. Usually, these datasets of extremes can be generated, from continuous time series data, or taking as extremes the maximum value in several predefined time blocks (Block Maxima approach) or fixing a sampling threshold (Peak-Over-Threshold approach).*
Based on a POT approach, our study examines datasets of skew surges. From now on, we use skew surge as extreme variable but all concepts that we will describe can be extended also to other natural hazards.
In the POT framework, an extreme sampling threshold has to be considered in order to extract the dataset of extreme skew surges from a continuous time series data of systematic skew surges recorded by tide gauge. In this way, the extreme sample is

created and the extreme quantiles computation of skew surge can be performed in according to the statistical distribution used. The relation between return periods T and quantiles x_T is expressed as:

$$P(X > x_T) = \frac{1}{\lambda T} \qquad (1)$$

where $P(X>x_T)$ is the exceedance probability of x_T or, thoroughly, the probability that in every λT years an event X, strongest than x_T, occur. Thanks to Eq.1, the results can be illustrated through a return level plot.
The variable λ represents the number of skew surge events per year that exceed the sampling threshold. In other words, λ is the frequency of storms occurrence and it can be expressed as the ratio between the number of skew surges n over the sampling threshold and the recording time of systematic skew surge measurements in years d (hereinafter called systematic duration):

$$\lambda = \frac{n}{d} \qquad (2)$$

The knowledge of n and d is necessary to compute the frequency of extreme skew surges occurrence λ and to estimate extreme quantiles of skew surges. In other words, a duration estimation is needed to transform the probability [no units] in frequency [time-1 dimension]."

P7, L12-24: *"The knowledge of the regional sample duration, as exposed for a generic frequency analysis in the previously paragraph of this section, is a needed factor. In fact, the evaluation of the regional return period T_r of a particular regional quantile x_{Tr} is expressed as in Eq.8:*

$$T_r = \frac{1}{\lambda_r P(X_s > x_{T_r})} \qquad (8)$$

where $P(X_s>x_{Tr})$ is the probability that a storm s impacts at least one site in the region with a normalized intensity greater than x_{Tr}. Therefore, thanks to the regional distribution Fr, the local return period of the storm s is calculated as:

$$T = \frac{1}{\lambda \left(1 - F_r(x_{T_r})\right)} \qquad (9)$$

In addition, Weiss (2014b) use for the computation of the empirical local return periods a Weibull plotting position. This element is correlated with the regional duration Dr:

$$\tilde{T}_{s,loc} = \frac{D_r + 1}{n_r + 1 - rank(s)} \qquad (10)$$

Then, the biggest storm s_{max} of the regional sample is linked to a local return period thanks to D_r."

4) Last but not least, the quality of the exposition is poor.

A good quality of exposition is a key point for whatever scientific paper. An exhaustive understanding of paper concept has to be realised in the simplest way by the reader.

For this reason, after thanking the Referee #2 for this useful comment, we have largely rewritten the paper and we have renovated the methodological part **(Section 2)**.

5) The title is too long, the term "historical data" is puzzling, and I would suggest of not putting the achronym FAB here

We agree the suggestion of the Referee #2. In fact, a long and elaborate title does not tempt a reading. The title has been reduced and simplified in this way: "The use of historical information on the regional frequency analysis of extreme skew surge". The FAB acronym has been defined and replaced and the historical data has been redefined as all the isolated data points derived from non-systematic measurements of skew surges. Note that this definition includes also isolated data point reconstructed during gaps of the systematic measurements. This is because statistically speaking the FAB methodology allows to treat them similar to the isolated data reconstructed from pre-gaugement period. We decided however to keep the historical data wording because this has been widely used in the literature and help focusing on the need for an effort (and the potential) in data archaeology.

P2-3, L25-2: "Specifically, historical data are all the skew surge events that were not recorded by tide gauge and, consequently, systematic data are all the measurements of skew surge recorded by tide gauge. One obvious and very promising way for collecting information not recorded by a tide gauge is to investigate and reconstruct historical events occurred before the starting date of gauge observations. This activity is also known as data archaeology and has been demonstrated very promising and useful for several study in the literature. Note however, that this definition also include isolated data point reconstructed during gaps of the systematic measurements. This is because statistically speaking they could be treated similarly to the isolated data reconstructed from pre-gaugement period. To find historical events it is thus necessary to go back in time and to look for information before the gauging period or during each data gap in a time series data. In fact, during important storms, tide gauge may be affected by a partial or total failure in the sea level measurements and so this kind of events cannot be detected. Thanks to non-systematic records, we can hold a numerical value to these ungauged extreme events. This allows to use a bigger number of storms in the statistical assessment of the extreme skew surges."

P3, L11-14: "As historical data are isolated data points thus the duration of the observation is not defined. What happened during the time period between two or more past extreme events remains unknown. However, the coverage period estimation of historical data is a key step in the extreme value statistical models (Prosdocimi, 2017)."

6) From the abstract one does not understand what's really the point in this research

In the new version, the **abstract** has been almost completely rewritten in order to point out straight on the central point of this study:

"The design of effective coastal protections requires an adequate estimation of the annual occurrence probability of rare events associated to a return period up to 10^3 years. Regional Frequency Analysis has been proved to be an applicant way to estimate extreme events using region into large and spatially distributed datasets. Nowadays, historical data are available providing new insight on past events estimations. The utilisation of historical information would increase the precision and the reliability of RFA low probability extrapolation. However, the information from past events usually looks like isolated data and they are different from continuous data

from systematic measurements of tide gauges. This make the definition of the duration of our observations period. The duration of the observation period is however crucial the frequency estimation of the extreme occurrence. For this reason, we introduced here the concept of "credible duration". The proposed RFA method (hereinafter referenced as FAB, from the name of the authors) allows the use of historical data together with systematic data thanks to the use of the credible duration concept."

7) Abstract, line 20, ..events data provided by single event "analysis" .. do you mean .. sparse records or estimations of extreme events in absence of systematic measurements ?

We agree that it was a very unclear sentence. Having rewritten the abstract, this sentence is not more present.

8) Abstract, line 23. To answer these questions a new method is introduced here (hereinafter referenced as FAB, from the name of the authors)

This suggested part has been added in the abstract (**Abstract, P1 L21**).

9) pag 1, line 30. "stochastic segment of sea level", what do you mean? Any sea level variable can be treated as stochastic, from tides to waves.

The incorrect definition of skew surge has been replaced by a proper skew surge definition as follows:

P1, L28-29: *"the difference between the maximum observed sea level and the maximum predicted astronomical tide level during a tidal cycle"*

10) pag 4, line 2, "duration for systematic skew storm surges". This is really unclear. Say rather "time horizon of systematic observations/measurements of ...". In general, I would not speak about "duration", but rather "time horizon"

This concepts has been clarified thanks to the reviewer comments. The notion of "duration for systematic skew storm surges" has been provide in the introduction part as follows:

P3, L8-11: *"The extreme statistical analysis requires a knowledge of the time period for which all the events above an extreme threshold are known (Leese, 1973; Payrastre et al., 2011). For the gauged or, better, systematic data, this time is represented by the recording period (i.e. systematic duration)."*

We are, however a bit more reluctant in introducing the "time horizon" wording. The point is that the terms duration is really widespread in the scientific publications, published among the other by NHESS, for the definition of the frequency of occurrence of a probabilistic variable. We worried that the introduction of a brand new wording using horizon may be confusing. Horizon is often used for the long terms climate change scenarios. It looks like more linked to the future than the pasty. Some of these studies have been referenced as follows:

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

- Hamdi, Y., Bardet, L., Duluc, C.-M., and Rebour, V.: Use of historical information in extreme surge frequency estimation: case of the marine flooding on the La Rochelle site in France, Natural Hazards and Earth System Sciences, 15, 1515-1531, doi:10.5194/nhess-15-1515-2015, 2015.

- Sabourin, A., Renard, B.: Combining regional estimation and historical floods: A multivariate semiparametric peaks-over-threshold model with censored data, Water Resources Reseach, 51(12), doi:10.1002/2015WR017320, 2015.

- Weiss, J., Bernardara, P., and Benoit, M.: Modeling intersite dependence for regional frequency analysis of extreme marine events, Water Resources Research, 50, 5926-5940, doi:10.1002/2014WR015391, 2014b.

11) pag 4, line 23: "the lack of a trend on skew storm surge frequency means the storm frequency has not a trend" this is a tautology

During the paper rewriting, this real unclear sentence has been removed.

12) pag 4-5, lines 25-3: this is also a consequence of the requirements for a stationary EVA. I believe this long explanation can be simplified and shortened.

For a stationary EVA, the frequency and intensity stationarity is required. Intensity stationary is verified by a Student's T-test. In order to verify the frequency stationarity, we need to know the frequency of the whole sample. As Eq.2 expresses, the frequency of extremes depends on the duration value that is an unknown value (at the beginning of the study) for extreme samples with historical data. For this reason, the frequency is considered stationary as several scientific studies state. Thanks to this hypothesis, the credible duration can be estimated.

13) section 2: this whole explanation is not very clear. I would start explaining the concept of "credible time horizon", and then I would define it in the simplest way possible.

We agree with this comment. The methodological part has been re-examined, restructured and, largely rewritten in order to define the credible duration as simple as possible. Moreover, the **section 2** of the new paper version begins with the explanation of the concept of credible duration, as suggested by Referee #2.

14) figures 1-4: I would summerize these figures (that are almost identical) in a unique figure, trying also to explain better why they are interesting.

The figures are summerized in a unique graph (Figure 2) where all the different thresholds are represented by different colours. The motivation of why these thresholds are interesting is expressed during all the **new Section 3** and, in particular:

P8, L29-31: *"Four cases of different values of high sampling threshold are examined (Fig.2 and Tab.1). These particular cases are chosen in order to let us know the effects of different sampling thresholds on the computation of the credible duration."*

The use of historical information on the regional frequency analysis of extreme skew surge

Roberto Frau[1,2], Marc Andreewsky[1], and Pietro Bernardara[3]

[1]EDF R&D Laboratoire National d'Hydraulique et Environnement (LNHE), Chatou, 78401, France
[2]Université Paris-Est, Saint-Venant Hydraulics Laboratory (ENPC, EDF R&D, CEREMA), Chatou, 78401, France
[3]CEREA, Joint Laboratory École des Ponts ParisTech – EDF R&D, Université Paris-Est, 77455 Marne la Vallée, France

Correspondence to: Roberto Frau (*roberto.frau@edf.fr*)

Abstract. The design of effective coastal protections requires an adequate estimation of the annual occurrence probability of rare events associated to a return period up to 10^3 years. Regional Frequency Analysis has been proved to be an applicant way to estimate extreme events using region into large and spatially distributed datasets. Nowadays, historical data are available providing new insight on past events estimations. The utilisation of historical information would increase the precision and the reliability of RFA low probability extrapolation. However, the information from past events usually looks like isolated data and they are different from continuous data from systematic measurements of tide gauges. This makes the definition of the duration of our observations period tricky. The duration of the observation period is however crucial for the frequency estimation of the extreme occurrence. For this reason, we introduced here the concept of "credible duration". The proposed RFA method (hereinafter referenced as FAB, from the name of the authors) allows the use of historical data together with systematic data thanks to the use of the credible duration concept.

Key words: Regional Frequency Analysis; Historical data; Data archaeology; Skew surge; Credible duration; Regional credible duration; Historical duration; FAB method.

1 Introduction

Flood risk is a noteworthy threat for anthropic and industrial settlement along the coasts. Coast defence constructions have to withstand extreme events. Therefore, in this context, a statistical estimation of extreme marine variables (such as sea levels or skew surges, the difference between the maximum observed sea level and the maximum predicted astronomical tide level during a tidal cycle (Simon, 2007; Weiss, 2014c)) occurrence probability is necessary. These variables associated to high return periods (up to 1000 years (ASN, 2013)) are estimated by statistical methods linked to the Extreme Value Theory (Fréchet, 1928; Gnedenko, 1943; Gumbel, 1958; Coles, 2001).

In the past, time series recorded at a single site have been used to estimate return levels for marine hazards. However, these extreme quantiles are linked to huge estimation uncertainties due to too short local recordings (typically 30 years to 50 years long). A greater amount of data is therefore necessary to get reliable estimations of extreme marine hazards.

1

For this reason and thanks to the availability of several local datasets, the Regional Frequency Analysis (RFA) is employed to reduce statistical uncertainties allowing the use of bigger extreme samples. The primary purpose of RFA is to cluster different locations in a region and to use all the data together (Cunnane, 1988; Hosking and Wallis, 1997; Van Gelder and Neykov, 1998; Bernardara et al., 2011; Weiss, 2014c). RFA is based on the index

5 flood principle (Dalrymple, 1960) that introduces a local index for each site in order to preserve their peculiarities into a region. A regional distribution can be defined only after checking its statistical homogeneity (Hosking and Wallis, 1997). In particular, the probability distribution of extreme values must be the same everywhere in the region to allow the fitting of a probability distribution with a big data sample. Several applications exist for marine variables and, mostly, for skew surge (Bernardara et al., 2011; Bardet et al., 2011; Weiss et al., 2013).

10 Specifically, Weiss (2014c) proposed a RFA approach to treat extreme waves and skew surges. According to this approach, homogeneous regions are first built relying on typical storm footprints and their statistical homogeneity is subsequently tested. The addition of physical elements permits to get regions that are both physically and statistically homogeneous. Furthermore, this methodology, and in particular the pooling method, yields an equivalent regional duration in years that enables the evaluation of extreme events' return periods.

15 As proved in hydrology during the last 30 years, extending datasets with historical data or paleofloods is an alternative way to reduce uncertainties in the estimations of extremes (Hosking and Wallis, 1986; Stedinger and Baker, 1987; Ouarda et al., 1998; Benito et al., 2004; Payrastre et al., 2011) as well as representing better potential outliers. Recent works on local sea levels (Bulteau et al., 2014) and on local skew surges (Hamdi et al., 2015) pointed out that the incorporation of historical data leads to a positive benefit in an extreme frequency analysis.

20 The combination of these two approaches increases the amount of available data allowing to get better extreme return levels. The use of extraordinary floods events in RFA has already been explored by Nguyen et al. (2014) for peak discharges and, only Hamdi et al. (2016) have explored it for oceanic and meteorological variables.

In this study, we investigate the use of historical data in RFA of systematic measurements of skew surge levels. The proposed FAB method allows to estimate reliable extreme skew surge events taking into account all the

25 available non-systematic records of skew surge levels. Specifically, historical data are all the skew surge events that were not recorded by tide gauge and, consequently, systematic data are all the measurements of skew surge recorded by tide gauge. One obvious and very promising way for collecting information not recorded by a tide gauge is to investigate and reconstruct historical events occurred before the starting date of gauge observations. This activity is also known as data archaeology and has been demonstrated very promising and useful for several

30 study in the literature. Note however, that this definition also include isolated data point reconstructed during gaps of the systematic measurements. This is because statistically speaking they could be treated similarly to the isolated data reconstructed from pre-gaugement period. To find historical events it is thus necessary to go back in time and look for information before the gauging period or during each data gap in a time series data. In fact, during important storms, tide gauge may be affected by a partial or total failure in the sea level measurements and so this

35 kind of events cannot be detected. Thanks to non-systematic records, we can hold a numerical value to these

2

ungauged extreme events. This allows to use a bigger number of storms in the statistical assessment of the extreme skew surges.

Anyway, Baart et al. (2011) pointed out the hitch in recovering accurate historical information. An historical event is the result of a wide investigation of several qualitative and quantitative sources. Historical data are collected from different sources as ancient newspapers, archives, water marks, palaeographic studies and others. For this reason, each historical data has its own origin and its quality has to be checked. Therefore, a critical analysis of historical data must be performed before applying a statistical approach.

Using historical data with time series data recorded by a tide gauge is not trivial. The extreme statistical analysis requires a knowledge of the time period for which all the events above an extreme threshold are known (Leese, 1973; Payrastre et al., 2011). For the gauged or, better, systematic data, this time is represented by the recording period (i.e. systematic duration). As historical data are isolated data points thus the duration of the observation is not defined. What happened during the time period between two or more past extreme events remains unknown. However, the coverage period estimation of historical data is a key step in the extreme value statistical models (Prosdocimi, 2017). Reasonable computation of historical duration is the base of reliable statistical estimation of rare events.

In the past, to overcome this issues, the perception threshold concept has been introduced. The perception threshold represents the minimum value over which all the historical data are reported and documented (Payrastre et al., 2011; Payrastre et al., 2013). The hypothesis is thus set that all the events over the perception threshold must have been recorded and the time period without records corresponds to time period without extreme events. This allows the estimation of the coverage period and the indirect assumption that the collected historical data are exhaustive throughout the coverage period assessed. So far, the perception threshold has allowed to use historical information in an extreme frequency analysis (Payrastre et al., 2011; Payrastre et al., 2013; Bulteau et al., 2014) but the hypothesis for using this approach seems sometimes difficult to prove.

Indeed, the perception threshold is usually defined equal to an historical event value, based on the principles that, if any other historical event would have occurred it would have been recorded (Payrastre et al., 2011; Bulteau et al., 2014). This is a quite strong hypothesis, which is difficult to be verified and validated. Moreover, the duration associated to the perception threshold is usually defined as the duration between the oldest historical data and the start of the systematic series (Payrastre et al., 2011; Hamdi et al., 2015).

The main point of novelty of the paper is to propose a new approach in order to deal with these questions in a more satisfactory way, the equivalent duration of historical data.

Note that the new approach is required for the application of the RFA, at least in the POT approaches (Van Gelder and Neykov, 1998; Bernardara et al., 2011; Bardet et al., 2011; Weiss et al., 2014b). Indeed, the definition of the perception threshold equal to an historical value precludes the possibility to use RFA methodology proposed by Weiss (2014c). Specifically, in Weiss regional methodology, the number of extreme events per year λ must be

equal in each site of a region in order to estimate a regional distribution. This is impossible to achieve if the threshold is linked to the characteristic of the local observations.

In addition, when historical information (i.e. non-systematic data) is collected during a total failure of tide gauge, this influences the frequency of the event. One could only re-estimate λ if the credible duration is re-estimated (for details, see the section 2).

In order to deal with these questions, in this paper, a different approach is proposed to estimate the equivalent duration of historical data. This evaluation is based on the hypothesis that the frequency of the extreme events is not changing significantly in time. In other words, the average number of extreme events should be the same (obviously this has to be carefully verified and demonstrated case by case). In this way, we can assess the coverage period of historical data. Based on the frequency value of systematic storms occurrence, our hypothesis introduces a physical and tangible element in the estimation of the historical duration. For this reason, we state that this novelty allows to overtake the concept of perception threshold providing a credible estimation of the unknown historical duration used to estimate the extreme quantiles.

Thanks to this new concept of duration, in this study, the FAB methodology is developed in order to use together historical data and systematic data in a Regional Frequency Analysis of extreme events. This methodology, described in details in section 2, is applied to the extreme skew surges given by 71 tide gauges located in the UK's, French's and Spanish's coasts and 14 historical collected skew surges is shown in section 4. In addition, a functional example on the computation of credible duration is illustrated in section 3.

2 FAB Methodology

2.1 The concept of credible duration

In the Extreme Value Analysis (EVA) of natural hazards, the estimations of extreme quantiles are performed applying statistical tools to a dataset of extreme variables. Usually, these datasets of extremes can be generated, from continuous time series data, or taking as extremes the maximum value in several predefined time blocks (Block Maxima approach) or fixing a sampling threshold (Peak-Over-Threshold approach).

Based on a POT approach, our study examines datasets of skew surges. From now on, we use skew surge as extreme variable but all concepts that we will describe can be extended also to other natural hazards.

In the POT framework, an extreme sampling threshold has to be considered in order to extract the dataset of extreme skew surges from a continuous time series data of systematic skew surges recorded by tide gauge. In this way, the extreme sample is created and the extreme quantiles computation of skew surge can be performed in according to the statistical distribution used. The relation between return periods T and quantiles x_T is expressed as:

$$P(X > x_T) = \frac{1}{\lambda T} \qquad (1)$$

where $P(X > x_T)$ is the exceedance probability of x_T or, thoroughly, the probability that in every λT years an event X, strongest than x_T, occur. Thanks to Eq.1, the results can be illustrated through a return level plot.

4

The variable λ represents the number of skew surge events per year that exceed the sampling threshold. In other words, λ is the frequency of storms occurrence and it can be expressed as the ratio between the number of skew surges n over the sampling threshold and the recording time of systematic skew surge measurements in years d (hereinafter called systematic duration):

$$\lambda = \frac{n}{d} \quad (2)$$

The knowledge of n and d is necessary to compute the frequency of extreme skew surges occurrence λ and to estimate extreme quantiles of skew surges. In other words, a duration estimation is needed to transform the probability [no units] in frequency [time^{-1} dimension].

When historical skew surge are available, the computation of λ is a tricky topic. In fact, although the number of historical data n_{hist} over the sampling threshold is known, we do not have any information on which happens during non-gauged period. It is thus impossible to define a duration over which the n_{hist} data are observed. In other words, fixing whatever sampling threshold, we need this information for defining the frequency of occurrence of the events over the threshold (see the Eq.2). We called the missing information "historical duration" d_{hist} or, in other words, the time period for which historical skew surge above sampling threshold are estimated to be exhaustively recorded at the best of our knowledge. The inclusion of historical data in a systematic dataset of extreme skew surge is shown as an example in Fig.1. It needs the formulation of a credible hypothesis in order to estimate the duration of the whole dataset composed by systematic and historical records.

Taking advantage to the perfect knowledge of systematic skew surges, the basic hypothesis proposed here for the computation of historical duration is that the frequency of storms λ remains unchanged in time. More precisely, λ_{syst}, namely the number of systematic skew surges per year over the sampling threshold, is supposed to be equal to λ_{hist}, namely the number of historical skew surges per year over the same sampling threshold.

$$\lambda = \lambda_{syst} = \lambda_{hist} \quad (3)$$

Eq.3 is in accordance with the lack of a significant trend on the extreme skew surges frequency during the second part of 19th century, 20th century and the first part of 21st century in the regions considered in this study (Fig.5).

According to current scientific literature, based on past observations, frequency of storms has no defined trend (Hanna et al., 2008; Barring and Fortuniak, 2009; MICORE Project, 2009). Therefore, the hypothesis that storm frequency is constant has been assumed. No significant trends are detected in terms of magnitude or frequency of storms between 1780 and 2005 in the North Atlantic and European regions since the Dalton minimum (Barring and Fortuniak, 2009). Over the past 100 years in the North Atlantic basin, no robust trends in annual numbers of tropical storms, hurricanes and major hurricanes counts have been identified by the last IPCC report of 2013. Hanna et al. (2008), Matulla et al. (2008) and Allan et al. (2009) state that observations have no clear trends over the past century or longer with substantial decadal and longer fluctuations but with the exceptions of some regional and seasonal trends (Wang et al., 2009; Wang et al., 2011). Moreover, MICORE (2009) in the "Review of climate

change impacts on storm occurrence" affirms that results from coastal areas do not detect significant trend in frequency of storms for the existing and available datasets.

The assumption that the annual frequency of skew surges is constant in Europe during the last two centuries is widely warrantied by the several references mentioned above. A solid check must be done for others marine variables to know if they can fit this hypothesis as well.

Anyway, a new duration of observation d_{cr} is defined as "credible" (i.e. based on credible hypothesis) considering a steady λ. This novelty is expressed as:

$$d_{cr} = \frac{n_{tot}}{\lambda} = \frac{n_{syst}+n_{hist}}{\lambda} = \frac{n_{syst}}{\lambda} + \frac{n_{hist}}{\lambda} \qquad (4)$$

Eq.4 shows how the number of extreme data is divided in two sides according to the nature of data: n_{syst} is the number of extreme systematic data above the sampling threshold u and n_{hist} is the number of extreme historical data over the same threshold u. Therefore, the credible duration depends also on the sampling threshold value. Moreover, this splitting provides indirectly the coverage period of historical skew surges in addition to the duration of recorded data. The historical side of the duration d_{cr}^{hist} can be formulated as follows:

$$d_{cr}^{hist} = \frac{n_{hist}}{\lambda} \qquad (5)$$

The coverage period on average that every historical skew surge hold is called associated duration $d_{cr,ass}$

$$d_{cr,ass} = \frac{d_{cr}^{hist}}{n_{hist}} = \frac{1}{\lambda} \qquad (6)$$

Eq.6 displays how the associated duration can be evaluated just knowing λ. Therefore, every extreme data not originated from a recorded time series has an associated duration. This is also the case for historical events occurred during a time gap in gauged skew surges series.

Anyway, a functional example is illustrated in Section 3 order to grasp all the notions explained up to here and, in particular, their relevance in the sampling threshold choice.

2.2 The use of credible duration in RFA framework

Weiss (2014c) defined a methodology to fit a regional probability distribution for sea extreme variables. This approach enables to put together data from different locations and to form regions. The probability distribution of extreme skew surges must be the same inside a region. Thus, the extreme data of a same region are gathered together by the use of a normalisation parameter that preserves the peculiarities of each site and the regional probability distribution is so defined. The formation of regions based on typical storm footprints (Weiss et al.,

2013) is used to create physically and statistically homogeneous regions and it leads to know a master component of this regional process: the regional effective duration D_{eff}.

The regional effective duration D_{eff} represents the number of years in which the regional sample N_r is observed. D_{eff} is defined as the product between the degree of regional dependence φ $[1; N]$ and the mean of local durations

$$D_{eff} = \varphi \times \bar{d} = \frac{\lambda r}{\lambda} \bar{d} = \frac{Nr}{\lambda} \quad (7)$$

in Eq.7, λ_r is the number of regional storms per year and $\bar{d}$ is the mean of local observations' duration of the N regional sites $\bar{d} = \sum_{i=1}^{N} d_i / N$.

The local frequency of storms λ must be the same in all the region. More details about the D_{eff} are clarified in Weiss (2014b). The whole RFA approach is set for systematic data.

The knowledge of the regional sample duration, as exposed for a generic frequency analysis in the previously paragraph of this section, is a needed factor. In fact, the evaluation of the regional return period T_r of a particular regional quantile x_{T_r} is expressed as in Eq.8:

$$T_r = \frac{1}{\lambda_r P(X_s > x_{T_r})} \quad (8)$$

where $P(X_s > x_{T_r})$ is the probability that a storm s impacts at least one site in the region with a normalized intensity greater than x_{T_r}. Therefore, thanks to the regional distribution F_r, the local return period of the storm s is calculated as:

$$T = \frac{1}{\lambda(1 - F_r(x_{T_r}))} \quad (9)$$

In addition, Weiss (2014b) use for the computation of the empirical local return periods a Weibull plotting position. This element is correlated with the regional duration D_r:

$$\tilde{T}_{s,loc} = \frac{D_r + 1}{n_r + 1 - rank(s)} \quad (10)$$

Then, the biggest storm s_{max} of the regional sample is linked to a local return period thanks to D_r.

However, the need to employ historical data in the extreme analysis has been considered as highly relevant by scientific community to get reliable quantiles and to not miss significant information. In the RFA, the use of historical records is allowed by the new notion of credible duration. A new way to process historical skew surges at regional scale is introduced. The addition of one or more historical skew surges in one or more sites of a region requires the knowledge of the duration of each site. Thanks to the credible duration concept applied previously to a single site, a coverage period for both types of data can be evaluated. It allows to switch from the notion of the regional effective duration $D_{r,eff}$ to the regional credible duration $D_{r,cr}$:

$$D_{r,cr} = \varphi \times \bar{d} = \varphi \sum_{i=1}^{N} \frac{d_{i,cr}}{N} \quad (11)$$

Regional credible duration depends also on the degree of regional dependence φ that increases when the data are independent. Moreover, Eq.11 shows that a local duration surplus brings consequently an excess of the coverage period of regional sample. With some simplifications, the Eq.11 can thus be so reformulated:

$$D_{r,cr} = \frac{\lambda r}{\lambda} \bar{d} = \frac{Nr}{\lambda} \quad (12)$$

The lack of a trend on storms' frequency is a necessary condition to assume a steady λ. If this hypothesis is satisfied, the regional credible duration can be computed.

The regional frequency analysis of skew surges including historical data or, FAB method, can be performed and these two combined approaches ensure better estimations of extreme values.

3 Illustration of credible duration

In order to better illustrate the concept of credible duration and how it varies according to different thresholds, a specific illustration on connection between these variables is described in this section by the example shown in Figure 1.

Fig. 1 shows the functional example of the whole skew surge series recorded at La Rochelle tide gauge (France). Systematic skew surges are represented by the continuous series of data plotted in black from about 1940 to 2017. They are observed during a fixed period equal to the recording time period, in this case, 32.88 years.

The historical skew surges are artificial data generated by ourselves for this descriptive illustration and they are represented as green cross points. As they are not supposed to be obtained by continuous systematic records of skew surge measurements, the duration of their support is unknown.

As expressed above, d_{cr} and in particular d_{cr}^{hist} depends on the sampling threshold value u. Eq.4 and Eq.5 show this dependence through the λ value.

The application of the new proposed concept allows the computation of the historical duration of skew surges that represents a basic way to take into account the historical data on the estimation of extreme values.

Although the period of systematic observation is known and fixed by tide gauge's records, the coverage period of historical observations is associated to the value of sampling threshold u. The hypothesis that the frequency of storms λ has no trend leads to different values of historical duration for each λ selected value and consequently, for each likely sampling threshold. Four cases of different values of high sampling threshold are examined (Fig.2 and Tab.1). These particular cases are chosen in order to let us know the effects of different sampling thresholds on the computation of the credible duration.

Case 1 uses a sampling threshold u_{C1} corresponding to 0.5 extremes per year ($\lambda=0.5$) and lead to a credible duration of 42.88 years (Tab.1). A higher threshold u_{C2} is used in Fig.2 for the Case 2 ($\lambda=0.23$) and how you can see the credible duration grows until 54.62 years because the number of historical data over the threshold does not change.

In Case 3, the threshold (u_{C3}) increases (λ=0.2) but the credible duration decreases (52.88 years) because the smallest value of historical skew surge does not exceed the threshold. On average, the coverage period of each historical event still rises. Case 4 displays the use of a very high threshold u_{C4} (λ=0.07) to define the sample of extremes. In this case, only one historical event is bigger than threshold and its coverage period or, associated duration, corresponds to the historical duration. More the historical data value is big and more high is the probability that the historical event covers a longer period.

As stated above, the four cases (resume in Fig.2 and Tab.1) show the behaviour of credible duration for different values of threshold. As threshold value increases, associated duration increases as well. Eq.4 expresses the credible duration as the sum of systematic duration and historical duration. Systematic duration is steady and it is independent from threshold value. For this reason, the variation of credible duration depends only on the variation of historical duration. In particular, when the threshold is extremely high, it is almost sure that no extreme skew surges occurred during the coverage period of historical data. If the threshold is lower, then it is less certain that all extreme skew surges have been observed in the historical part.

However, it is not always true that the credible duration is biggest when threshold grows (Fig.3). In fact, the credible duration and historical duration depends not only of the threshold value but also of the number of historical events exceeding it. If the threshold is extremely high, not all historical data are taken into account in the extremes' sample. On the contrary, the associated duration of each historical data increases when threshold grows for the reasons explained in the previous paragraph (shown by the blue line in Fig.3).

In conclusion, the sampling threshold choice is an important step in the extreme quantiles estimation allowing the computation of the credible duration. Finally, with all of these elements, we can deal with historical data in a regional framework.

4 Application of FAB method on a real dataset

4.1 Systematic skew surge database and historical data

The systematic skew surge database is originated from time series of sea levels recorded by 71 tide gauges located along French, British and Spanish coasts of the Atlantic Ocean, the English Channel, the North Sea and the Irish Sea (Fig.4).

French data are provided by SHOM (Service Hydrographique et Océanographique de la Marine, France), English data by BODC (British Oceanographic Data Center, UK) and Spanish data by IEO (Instituto Espanol de Oceanografia, Spain). All data, except Spanish data, are updated until January 2017. Every site has different systematic durations depending on the tide gauge records. Starting from 1846, the longest time series is recorded at Brest for a systematic period of 156.57 years. Sea levels are hourly recorded except for British ports. BODC supplies sea levels every 15 minutes from 1993.

9

Skew surge data are generated in each site from the difference between the maximum sea levels recorded around the time of the predicted astronomical high tide and the self-same predicted astronomical high tide. Before computing skew surge, sea levels must be corrected by a likely long-term alteration of mean sea levels, or eustatism. Only sea level data with significant linear trends are edited. More details on the computation of skew surges are described by Simon (2007), Bernardara et al. (2011) and Weiss (2014c). In this way, the systematic database of 71 skew surge series is created.

Finding accurate historical data is hard and tricky and, at the moment, only 14 historical skew surge are available. These historical records are located in 3 of the 71 ports: 9 at La Rochelle (Gouriou, 2012; Garnier and Surville, 2010; Breilh, 2014), 4 at Dunkerque (DREAL, 2009; Parent et al., 2007; Maspataud, 2011) and 1 at Dieppe (provided by the Service hydrographique et océanographique de la Marine SHOM, France). The historical data are introduced in corresponding skew surge series of the concerned sites.

This skew surge database of systematic and historical data is exploited in this study for the estimation of regional return periods.

4.2 Formation of regions

The method to form homogeneous regions proposed by Weiss et al. (2014a) is used to get physically and statistically homogeneous regions for the database shown above. This methodology is based on a storm propagation criterion identifying the most typical storms footprints. The configuration of parameters to detect storms are set using the same criteria of Weiss (2014c). The most typical storm footprints are revealed by Ward's hierarchical classification (Ward, 1963) and correspond to four physical regions. The homogeneity test (Hosking and Wallis, 1997) is applied to check if all regions are also statistically homogeneous. Only one physical region is not statistically homogeneous and, after an inner partition and a further check, five physically and statistically homogeneous regions are obtained.

Moreover, a double threshold approach (Bernardara et al., 2014) is employed to separate the physical part from the statistical part of the used extreme variables and two different thresholds are used: the first to define physical storms (physical threshold) and the second one to define regional samples (statistical sampling threshold or, more simply, sampling threshold).

Fig.5 displays the five regions founded out through the application of Weiss' methodology. Although a different input database is used, the number and the position of regions are similar to Weiss' study.

The regionalisation processing shows that only Region 1 and Region 2 hold historical data. The three sites with historical data belong actually to these two regions: La Rochelle to Region 1 and, Dieppe and Dunkerque to Region 2.

The region with more historical information is the Region 1 (nine historical skew surges) and the results of FAB method proposed are focused on this region.

10

4.3 Focus on Region 1

4.3.1 Statistical threshold and credible duration

The sample of regional extremes is analysed with statistical tools in order to estimate regional return levels associated to very extreme return periods. A Regional Pooling Method is used to select independent maximum data. A regional sample is formed through normalised local data. Maximum local data are divided by a local index, equal to the local sampling threshold value (Weiss, 2014c). Moreover, this methodology enables us to estimate the regional distribution F_r considered as the distribution of the normalised maximum data above the statistical sampling threshold.

This threshold corresponds to a storm's frequency λ that is assumed equal at all sites of a region. The sampling threshold value selection or, better, λ value selection is a tricky topic (Bernardara et al., 2014). There is not an univocal approach to define the exact sampling threshold. Twelve sensitivity indicators depending on all possible λ values are proposed to assign the best λ value for the regional analysis of extreme skew surge using historical data: the statistical homogeneity test (Hosking and Wallis, 1997), the stationarity test applied to storm intensities of regional sample (Hosking and Wallis, 1993), the chi-squared test for the regional distribution (Cochran, 1952), the statistical test to detect outliers in regional sample (Hubert and Van der Veeken, 2008; Weiss, 2014c), the value of regional credible duration, the number of regional sample's data, the value of the degree of regional dependence φ, the visual look of the regional return plot and the stability of regional credible duration, local return levels, shape parameter and scale parameter of the GPD distribution (Coles, 2001). The λ value selection can not be carried out through one of these methods and only the better combination of all indicators can enable us to appoint the optimal λ. All sensitivity indicators have the same relevance in the proposed sensitivity analysis.

A skew surge's frequency of 0.36 is chosen for Region 1 thanks the sensitivity analysis of the statistical factors mentioned above. Taking λ=0.36 means to have on average one extreme skew surge over the sampling threshold every 2.78 years at each site. This threshold selection permits the computation of the credible duration for every site belonging to the Region 1. In the Region 1, all ports, except for La Rochelle, have only systematic skew surge and their credible duration, equal to systematic duration, is determined and independent from the sampling threshold value.

On the contrary, the site of La Rochelle has nine historical skew surges and its credible duration depends both on systematic duration and historical duration. The systematic duration does not change and it is equal to the recording period of La Rochelle tide gauge. As shown in the fictive illustration of section 3, the historical duration differs while threshold value changes.

The credible duration at La Rochelle is 57.88 years and the contribution of historical data is almost the 43% of the total duration. Obviously, without these nine historical data, the coverage period of extreme data at La Rochelle would have been less than 25 years, equivalent to the historical duration. All the historical data are over the chosen threshold and, consequently, the associated duration of each historical data is equal to 2.77 years.

11

4.3.2 Regional credible duration

Eq.11 enables us to compute the regional credible duration for Region 1. The introduction of nine historical data in the local time series and, specifically, only eight in the regional sample after the use of Regional Pooling method, grants an earning of 11.11 years of regional duration. As Tab.2 shows, historical data increase the credible durations for sites where there are historical records and, consequently, the regional credible duration increases likewise. In Region 1, only the site of La Rochelle increases its duration. This contribution allows to increase the regional credible duration.

Regional credible duration depends not only on the mean of local durations but also on the degree of regional dependence φ (Eq.11). Obviously, this degree does not change with only nine more added data. In conclusion, the rise of the regional credible duration of Region 1 is produced mostly by the increase of credible duration at La Rochelle.

Tab.2 shows the difference with the FAB method that allows to take into account historical data and a traditional method of RFA (Bernardara et al., 2011; Weiss, 2014c) that does not enable the use of historical data. Fig.6 displays how more λ decreases in value, the more the difference between regional credible duration with the nine historical data (blue line) and the regional effective duration without historical data (green line) for Region 1 increases (red line). This is due to the greater certainty that historical data above the sampling threshold are the biggest during a longer time.

4.3.3 Regional return levels

Regional return levels for Region 1 are obtained through the fit of a Generalized Pareto Distribution (Fig.7). Penalized maximum likelihood estimation (PMLE) (Coles and Dixon, 1999) is used to estimate the regional GPD parameters (γ, k). As shows Fig.7, the regional distribution is unbounded due to a positive shape parameter k. The return level plot of Region 1 is produced with the 95% confidence intervals generated by a bootstrap of observed storms. A Weibull plotting position is employed for the regional empirical return period. Return levels for each site are achieved through multiplication between regional return levels and local indices, equivalent to the local sampling threshold.

Red dotted points represent the eight historical data taken into account in Region 1. As appears, the benefit of using historical data is not only an increase of the coverage period of extreme data but also able to consider in a frequency analysis non-systematic storms occurred. The historical skew surges used in this study represent very big storms and, for this reason, it is newsworthy to take into account their large magnitude. Indeed, the biggest normalised skew surge considered in Region 1 is 3.01 compared with 2.13 of the regional sample generated without historical data (Tab.2).

5 Conclusions

The application of FAB method, based on the new concept of credible duration, allows to get more reliable extreme estimations of return levels thanks to the inclusion of historical data to a systematic skew surge database.

The introduction of historical data on a frequency analysis of extremes is an innovative way to increase the reliability on estimations. In the past, the estimation of the period covered by the historical data collection has usually limited the use of historical data. The concept of perception threshold was employed so far to use historical data even though its hypothesis are quite strong and difficult to verify. For this reason, the new concept of credible duration has been formulated that it is based on physical considerations supported by several scientific studies. This leads to use historical isolated data points on the extreme value analysis.

The hypothesis, on which credible duration concept is based, is satisfactory for skew surge. For other variables, a hypothesis check has to be done. Following the hypothesis check, all the considerations that we raised in this study can be applied also for other natural hazards.

When historical data are available, the credible duration depends on sampling threshold. This novelty allows to expand the credible duration's concept on regional scale. The Regional Frequency Analysis is an another efficient way to reduce uncertainties on the extreme estimations. These two ways can be used together by the proposed FAB method.

The results describe the relevance of using historical records on a regional analysis both in terms of regional duration and of storm magnitude. Sometimes, these data can represent very large events and so we have to consider them during the estimation of extreme variables. Moreover, the existence of historical records enables to raise the coverage period of the extreme sample.

No extreme data must be missed in an extreme frequency analysis. It is very important to use all of the available data in order to get extreme estimations increasingly more reliable and, regional analysis and historical data are two tools that can be merge with satisfactory results.

New perspectives are opened to improve extreme quantiles estimations using increasingly new historical data. An historical data is the result of different investigations on several qualitative and quantitative sources. The aggregation of all of these data in an historical database is a need to exploit the maximum amount of extremes. Furthermore, every historical data value is associated to different uncertainties and it is challenging to be able to take into account these uncertainties in the regional analysis.

References

Allan, R., Tett, S. and Alexander, L.: Fluctuations in autumn-winter severe storms over the British Isles: 1920 to present, International Journal of Climatology, 29(3), 357-371, doi:10.1002/joc.1765, 2009.

ASN: Guide de l'ASN n°13 : Protection des installations nucléaires de base contre les inondations externes, 2013.

Baart, F., Bakker, M. A. J., van Dongeren, A., den Heijer, C., van Heteren, S., Smit, M. W. J., van Koningsveld, M., and Pool, A.: Using the 18th century storm-surge data from the Dutch Coast to improve the confidence in flood-risk estimates, Natural Hazards and Hearth System Sciences, 11, 2791-2801, doi:10.5194/nhess-11-2791-2011, 2011.

Bardet, L., Duluc, C.-M., Rebour, V., and L'Her, J.: Regional frequency analysis of extreme storm surges along the French coast, Natural Hazards and Earth System Sciences, 11(6), 1627-1639, doi:10.5194/nhess-11-1627-2011, 2011.

Barnett, V., and Lewis, T.: Outliers in Statistical Data, 3rd Edition. John Wiley, Chichester, 1994.

Barriendos, M., Coeur, D., Lang, M., Llasat, M.C., Naulet, R., Lemaitre, D., and Barrera, A.: Stationarity analysis of historical flood series in France and Spain (14th–20th centuries), Natural Hazards and Earth System Sciences, 3(6), 583–592, doi:10.5194/nhess-3-583-2003, 2003.

Barring, L. and Fortuniak, K.: Multi-indices analysis of southern Scandinavian storminess 1780-2005 and links to interdecadal variations in the NW Europe-North sea region, International Journal of Climatology, 29(3), 373-384. doi:10.1002/joc.1842, 2009.

Benito, G., Lang, M., Barriendos, M., Llasat, M., Frances, F., Ouarda, T., Thorndycraft, V., Enzel, Y., Bardossy, A., Cœur, D., and Bobée, B.: Use of Systematic, Paleoflood and Historical Data for the Improvement of Flood Risk Estimation, Review of Scientific Methods, Natural Hazards, 31(3), 623-643, 2004.

Benson, M. A.: Use of historical data in flood‐frequency analysis. Eos, Transactions American Geophysical Union, 31(3), 419-424, doi:10.1029/TR031i003p00419, 1950.

Bernardara, P., Andreewsky, M., and Benoit, M.: Application of regional frequency analysis to the estimation of extreme storm surges, Journal of Geophysical Research, 116, C02008, doi:10.1029/2010JC006229, 2011.

Bernardara, P., Mazas, F., Kergadallan, X., and Hamm, L.: A two-step framework for over-threshold modelling of environmental extremes, Natural Hazards and Earth System Sciences, 14, 635-647, doi:10.5194/nhess-14-635-2014, 2014.

Breilh, J.F., Bertin, X., Chaumillon, E., Giloy, N. and Sauzeau, T,: How frequent is storm-induced flooding in the central part of the Bay of Biscay?, Global and Planetary Change, 122, 161-175. doi:10.1016/j.gloplacha.2014.08.013, 2014a.

Breilh, J.F.: Les surcotes et les submersions marines dans la partie centrale du Golfe de Gascogne : les enseignements de la tempete Xynthia, Ph.D. thesis, Sciences de la Terre, Université de La Rochelle, France, available at: https://tel.archives-ouvertes.fr/tel-01174997/document, 2014b.

14

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

Cochran, W. G.: The $\chi 2$ Test of Goodness of Fit, The Annals of Mathematical Statistics, 23(3), 315-345, 1952.

Cohn, T. A.: The incorporation of historical information in flood frequency analysis, M. S. thesis, Dep. of Environ. Eng., Cornell Univ., Ithaca, N.Y., U.S., 1984.

Cohn, T. A., and Stedinger, J. R.: Use of historical information in a maximum-likelihood framework, Journal of Hydrology, 96(1-4), 215-223, doi:10.1016/0022-1694(87)90154-5, 1987.

Coles, S. and Dixon, M.: Likelihood-based inference for extreme value models, Extremes, 2(1), 5-23, doi:10.1023/A:1009905222644, 1999.

Coles, S.: An Introduction to Statistical Modeling of Extreme Values, Springer, London, doi:10.1007/978-1-4471-3675-0, 2001.

Condie, R., and Lee, K. A.: Flood frequency analysis with historic information, Journal of Hydrology, 58(1-2), 47-61, doi:10.1016/0022-1694(82)90068-3, 1982.

Cunnane, C.: Methods and merits of regional floods frequency analysis, Journal of Hydrology, 100(1/3), 269-290, doi:10.1016/0022-1694(88)90188-6, 1988.

Dalrymple, T.: Flood Frequency Analyses, Manual of Hydrology: Part 3, "Flood Flow Techniques" U.S. Geological Survey Water-Supply Paper, 1543-A, US GPO, 1960.

DREAL (Ministère de l'Ecologie, de l'Energie, du Développement durable, et de l'Aménagement du territoire), 59/62 SREI/DRNHM: Détermination de l'aléa de submersion marine intégrant les conséquences du changement climatique en région Nord – Pas-de-Calais – Etape 1: compréhension du fonctionnement du littoral, available at: http://www.hauts-de-france.developpement-durable.gouv.fr/IMG/pdf/Determination-alea-submersion-marine-integrant-consequences-changement-climatique-dreal-npdc.pdf, 2009.

Fréchet, M.: Sur la lois de probabilité de l'écart maximum, Annales de la Société Polonaise de Mathématique, 6, 93–122, 1928.

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

Gnedenko, B.: Sur la distribution limite du terme maximum d'une série aléatoire, Annals of Mathematics, 44(3), 423-453, 1943.

Gouriou, T.: Evolution des composantes du niveau marin à partir d'observations de marégraphe effectuées depuis la fin du 18e siècle en Charente-Maritime, Ph.D. thesis, Océanographie physique, Université de La Rochelle, France, 2012.

Gumbel, E. J.: Statistics of Extremes, Columbia University Press, NewYork, U.S., 1958.

15

Hamdi, Y., Bardet, L., Duluc, C.-M., and Rebour, V.: Use of historical information in extreme surge frequency estimation: case of the marine flooding on the La Rochelle site in France, Natural Hazards and Earth System Sciences, 15, 1515-1531, doi:10.5194/nhess-15-1515-2015, 2015.

Hamdi, Y., Bardet, L., Duluc, C.-M., and Rebour, V.: Development of a target-site-based regional frequency model using historical information, European Geosciences Union General Assembly 2016 (EGU 2016), Vienna, Austria, 17-22 April 2016, 18, EGU2016-8765, 2016.

Hanna, E., Cappelen, J., Allan, R., Jonsson, T., Le Blancq, F., Lillington, T. and Hickey, K.: New Insights into North European and North Atlantic Surface Pressure Variability, Storminess, and Related Climatic Change since 1830, Journal of Climate, 21, 6739-6766, doi:10.1175/2008JCLI2296.1, 2008.

Hosking, J. R. M. and Wallis, J. R.: Paleoflood hydrology and flood frequency analysis, Water Resources Research, 22(4), 543-550, doi:10.1029/WR022i004p00543, 1986a.

Hosking, J. R. M. and Wallis, J. R.: The value of historical data in flood frequency analysis, Water Resources Research, 22(11), 1606-1612, doi:10.1029/WR022i011p01606, 1986b.

Hosking, J. R. M. and Wallis, J. R.: Some statistics useful in regional frequency analysis, Water Resources Research, 29(2), 271-281. doi:10.1029/92WR01980, 1993.

Hosking, J. R. M. and Wallis, J. R.: Regional Frequency Analysis. An approach based on L-moments, Cambridge University Press, 1997.

Hubert, M., and Van der Veeken, S.: Outlier detection for skewed data, Journal of chemometrics 22(3-4), 235-246, doi:10.1002/cem.1123, 2008.

IPCC: Climate change 2013: the physical science basis, Contribution to the IPCC fifth assessment report (AR5), Chapter 2: Observations: Atmosphere and Surface, 2013.

Jenkins, G. J., Murphy, J. M., Sexton, D. M. H., Lowe, J. A., Jones, P. and Kilsby, C. G.: UK Climate Projections: Briefing report, Met Office Hadley Centre, Exeter, UK, 2009.

Leese, M. N.: Use of censored data in the estimation of Gumbel distribution parameters for annual maximum flood series, Water Resources Research, 9(6), 1534-1542, doi:10.1029/WR009i006p01534, 1973.

Maspataud, A.: Impacts des tempêtes sur la morphodynamique du profil côtier en milieu macrotidal, Ph.D. thesis, Océanographie, Université du Littoral Côte d'Opale, France, available at: https://tel.archives-ouvertes.fr/tel-00658671/document, 2011.

Matulla, C., Schoner, W., Alexandersson, H., von Storch, H. and Wang, X. L.: European storminess: late nineteenth century to present, Climate Dynamics, 31(2-3), 125-130, doi:10.1007/s00382-007-0333-y, 2008.

MICORE Project 202798: Review of climate change impacts on storm occurrence, edited by O.Ferreira (UALG), M.Vousdoukas (UALG) and P.Ciavola (UniFe), available at: http://www.micore.eu/file.php?id=4, 2009.

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

16

Ouarda, T. B. M. J., Rasmussen, P. F., Bobée, B., and Bernier, J.: Utilisation de l'information historique en analyse hydrologique fréquentielle, Revue des sciences de l'eau, 11, 41-49, doi:10.7202/705328ar, 1998.

Parent, P., Butin, T., Vanhee, S. and Busz, N.: Un territoire soumis au risque de submersion marine. Les tempetes de 1949 et 1953 à Dunkerque, Institution départementale des Wateringues, available at : http://www.ingeo.fr/fichier/file/Wateringues/Affiche4-Dunkerque.pdf, 2007.

Payrastre, O., Gaume, E., and Andrieu, H.: Usefulness of historical information for flood frequency analyses: developments based on a case study, Water Resources Research, 47, W08511, doi:10.1029/2010WR009812, 2011.

Payrastre, O., Gaume, E., and Andrieu, H.: Historical information and flood frequency analyses: which optimal features for historical floods inventories?, La Houille Blanche - Revue internationale de l'eau, 3, 5-11, doi:10.1051/1hb/2013019, 2013.

Picklands, J.: Statistical inference using extreme order statistics, The Annals of Statistics, 3(1), 119-131, 1975.

Pouvreau, N.: Trois cents ans de mesures marégraphiques en France : outils, méthodes et tendances des composantes du niveau de la mer au port de Brest, Ph.D. thesis, Climatologie, Université de La Rochelle, France, available at: https://tel.archives-ouvertes.fr/tel-00353660/document,2008.

Prosdocimi, I.: German tanks and historical records: the estimation of the time coverage of ungauged extreme events, Stochastic Environmental Research and Risk Assessment, 1-16, doi:10.1007/s00477-017-1418-8, 2017.

Reis, D. S. and Stedinger, J. R.: Bayesian MCMC flood frequency analysis with historical information, Journal of Hydrology, 313(1-2), 97-116, doi:10.1016/j.jhydrol.2005.02.028, 2005.

Sabourin, A., Renard, B.: Combining regional estimation and historical floods: A multivariate semiparametric peaks-over-threshold model with censored data, Water Resources Reseach, 51(12), doi:10.1002/2015WR017320, 2015.

SHOM (Service Hydrographique et Océanographique de la Marine): Références Altimétriques Maritimes - Edition 2016, 2016.

Simon, B.: La marée océanique côtière, Editions de l'Institut Océanographique, 2007.

Stedinger, J. R. and Cohn, T. A.: Flood frequency analysis with historical and paleoflood information, Water Resources Research, 22(5), 785-793, doi:10.1029/WR022i005p00785, 1986.

Stedinger, J. R. and Baker, V. R.: Surface water hydrology: historical and paleoflood information, Reviews of Geophysics, 25(2), 119-124, doi:10.1029/RG025i002p00119, 1987.

Van Gelder, P. H. A. J. M.: A new statistical model for extreme water levels along the Dutch coast, edited by: Tickle, K. S., Goulter, I. C., Xu, C. C., Wasimi, S. A., Bouchart, F., Stochastic Hydraulics '96, Proceedings of the 7th IAHR International Symposium, Mackay, Queensland, Australia, 243–249, 1996.

Van Gelder, P. H. A. J. M., and Neykov, N. M.: Regional frequency analysis of extreme water level along the Dutch coast using L moments: A preliminary study, Stochastic models of hydrological processes and their applications to problems of environmental preservation, NATO Advanced Research Workshop, 23-27 November 1998, Moscow, Russia, 14-20, 1998.

5

10

15

20

25

30

35

Wang, X. L., Zwiers F. W., Swail V. R., and Feng Y.: Trends and variability of storminess in the Northeast Atlantic region, 1874-2007. Climate Dynamics, 33(7-8), 1179-1195, doi:10.1007/s00382-008-0504-5, 2009.

Wang, X. L., Wan, H., Zwiers, F. W., Swail, V. R., Compo, G. P., Allan, R. J., Vose, R. S., Jourdain, S., and Yin, X.: Trends and low-frequency variability of storminess over western Europe, 1878–2007, Climate Dynamics, 37(11-12), 2355–2371, doi:10.1007/s00382-011-1107-0, 2011.

Ward, J.H.: Hierarchical grouping to optimize an objective function, Journal of the American Statistical Association, 58(301), 236-244, 1963.

Weiss, J., Bernardara, P., and Benoit, M.: A method to identify and form homogeneous regions for regional frequency analysis of extreme skew storm surges, Proceedings of the 1st International Short Conference on Advances in Extreme Value Analysis and Application to Natural Hazards, 18-20 September 2013, Siegen, Germany, 2013.

Weiss, J., Bernardara, P., and Benoit, M.: Formation of homogeneous regions for regional frequency analysis of extreme significant wave heights, Journal of Geophysical Research: Oceans, 119(5), 2906-2922, doi:10.1002/2013JC009668, 2014a.

Weiss, J., Bernardara, P., and Benoit, M.: Modeling intersite dependence for regional frequency analysis of extreme marine   events, Water Resources Research, 50, 5926-5940, doi:10.1002/2014WR015391, 2014b.

Weiss, J.: Analyse régionale des aléas maritimes extrêmes, Ph.D. thesis, Mécanique des fluides, Université Paris-Est, France, available at: https://pastel.archives-ouvertes.fr/tel-01127291/document, 2014c.

Table 1. Summary table of several variables involved in the computation of credible duration for 4 different cases of study.

	Case 1	Case 2	Case 3	Case 4
Lambda λ	0.5	0.23	0.2	0.07
Sampling threshold (m)	0.7	0.77	0.81	1.02
Systematic duration (years)	32.88	32.88	32.88	32.88
Historical credible duration (years)	10	21.74	20	14.29
Number of historical data OT	5	5	4	1
Associated duration (years)	2	4.35	5	14.29
Credible duration (years)	42.88	54.62	52.88	47.17

5

Table 2. Summary table of the results obtained with the application of FAB method for Region 1.

	RFA without historical data	RFA with 9 historical data (FAB, 2017)
Lambda λ	0.36	0.36
Local duration on average (years)	41.4	42.36
Regional credible duration (years)	558.33	569.44
Intensity of biggest normalized surge	2.13 (SD)	3.01 (HD)
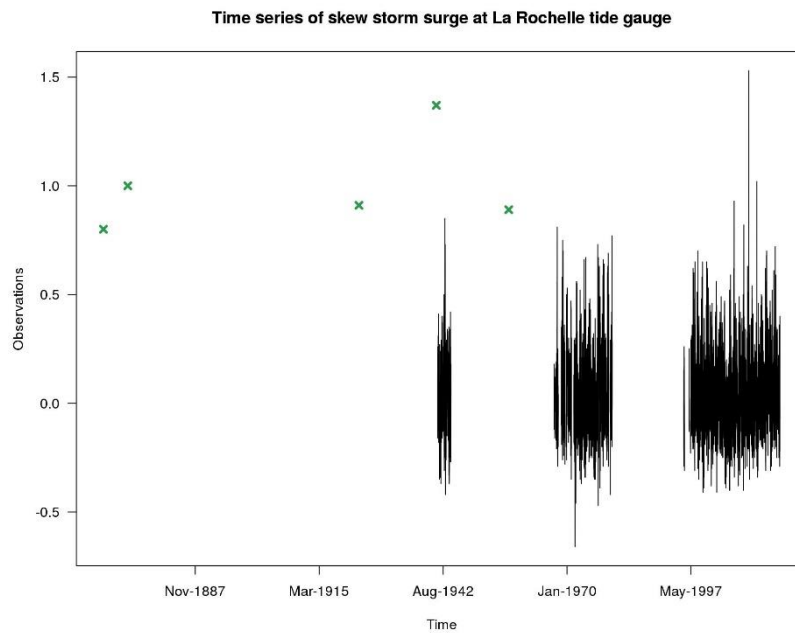
10 **Fig. 1. Illustrative example of skew surge dataset at La Rochelle tide gauge. In black the systematic skew surge and in green the historical isolated data point (non-systematic data).**

19

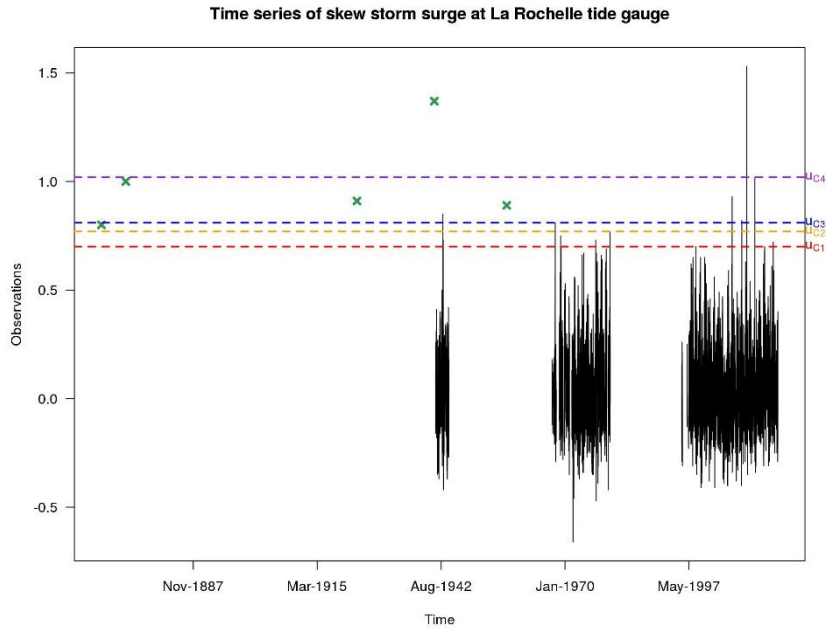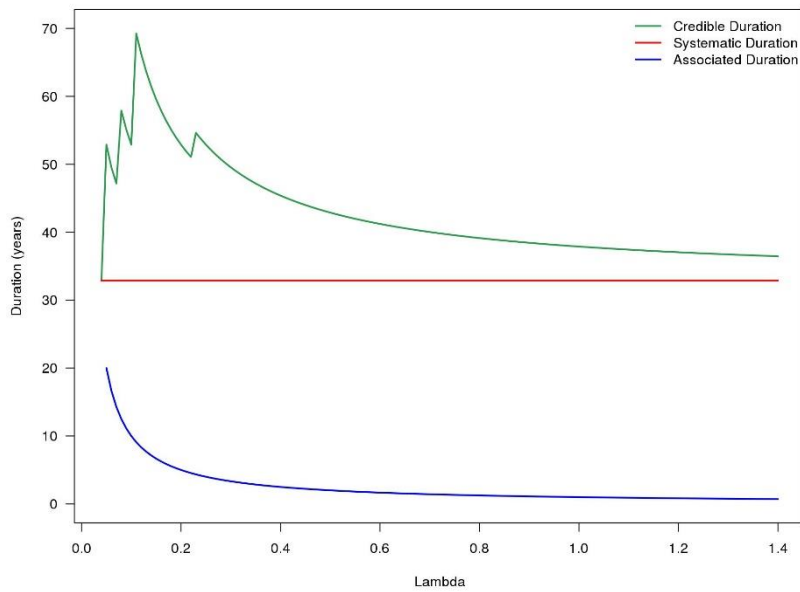**Time series of skew storm surge at La Rochelle tide gauge**

Fig. 2 Skew surge dataset of Section 3 (composed by systematic and historical data) used in the computation of credible duration with 4 different thresholds u_{c1}, u_{c2}, u_{c3}, u_{c4}. The four sampling thresholds u_c are used linked to a rising value o λ ($\lambda_{c1}=0.5$, $\lambda_{c2}=0.23$, $\lambda_{c3}=0.2$, $\lambda_{c4}=0.07$).

5

Fig. 3. Fictive example of skew surge dataset of systematic and historical data - Credible duration (green line), systematic duration (red line) and associated duration (blue line) for several threshold values.
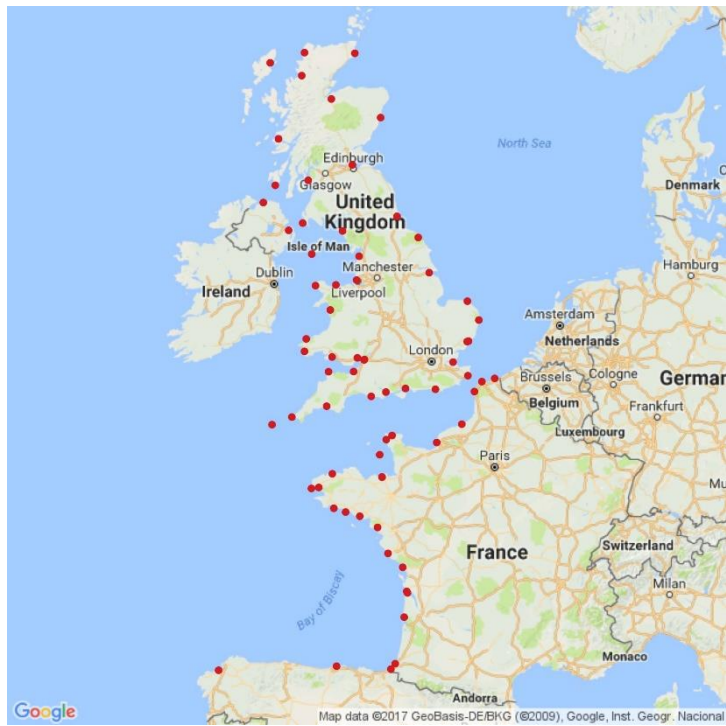
20

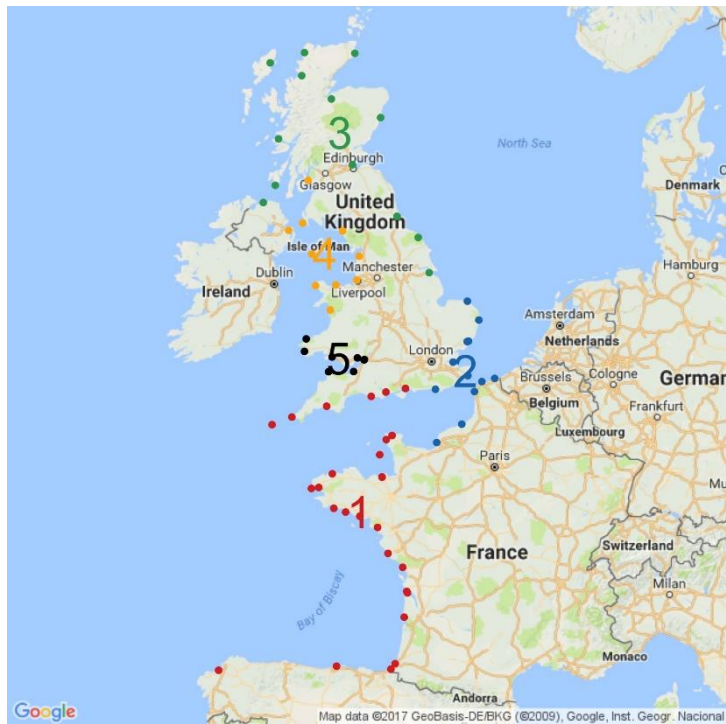**Fig. 4. Location of the 71 tide gauges used in this application.**

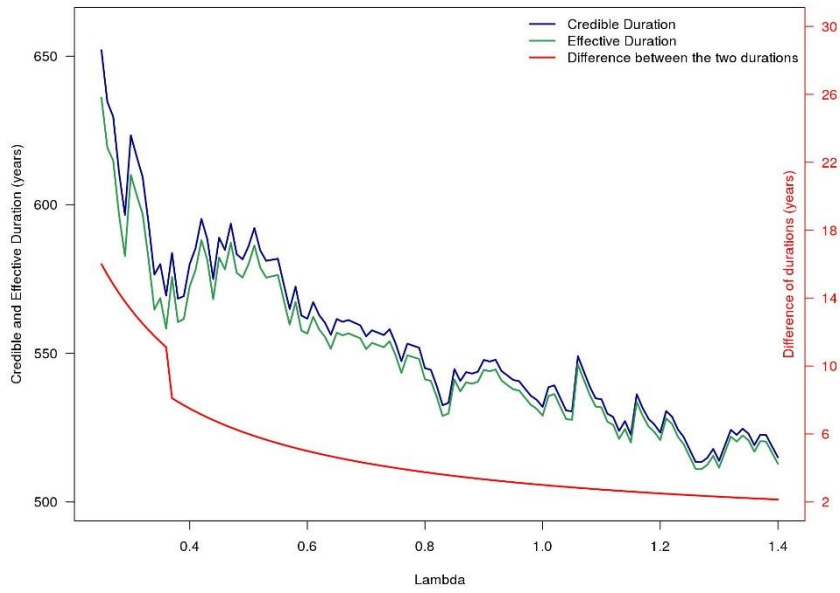**Fig. 5. The five physically and statistically homogeneous regions for the 71 tide gauges.**
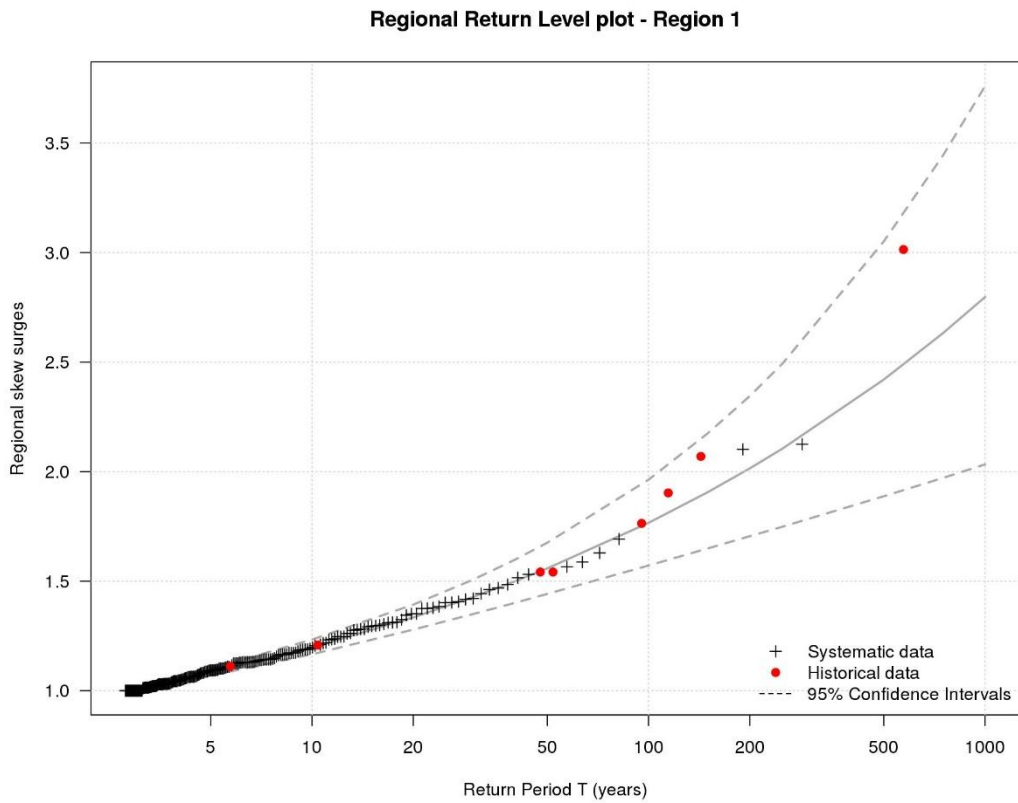
21

Fig. 6. Variation of regional credible duration (taking into account the nine historical data) and regional effective durations (only with systematic data) and their difference (in red) for Region 1.

Regional Return Level plot - Region 1

5   **Fig. 7. Return level plot for the regional distribution of Region 1 with the confidence intervals of 95% estimated by bootstrap. Red points represent the regional historical data and black crosses the regional systematic data.**

22