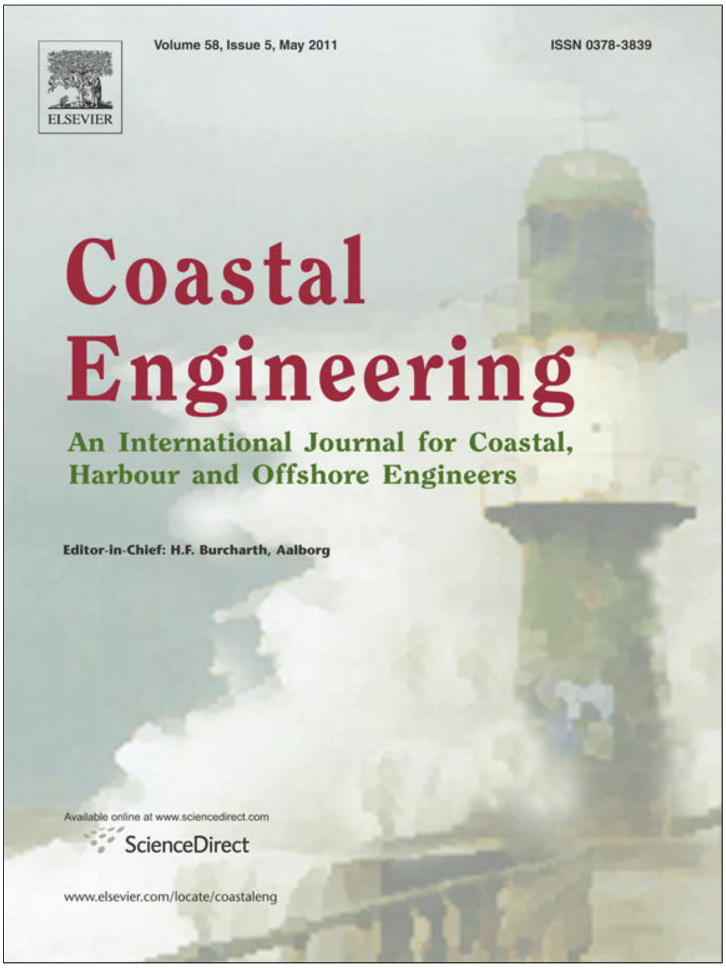
Volume 58, Issue 5, May 2011
ISSN 0378-3839
ELSEVIER
Coastal Engineering
An International Journal for Coastal, Harbour and Offshore Engineers
Editor-in-Chief: H.F. Burcharth, Aalborg
Available online at www.sciencedirect.com
ScienceDirect
www.elsevier.com/locate/coastaleng

Coastal Engineering 58 (2011) 385–394

Contents lists available at ScienceDirect

Coastal Engineering

journal homepage: www.elsevier.com/locate/coastaleng

A multi-distribution approach to POT methods for determining extreme wave heights

Franck Mazas *, Luc Hamm

SOGREAH Consultants, 6 rue de Lorraine 38130 Echirolles, France

ARTICLE INFO

Article history:
Received 2 June 2010
Received in revised form 17 December 2010
Accepted 20 December 2010
Available online 14 January 2011

Keywords:
Extreme wave heights
POT
GPD
Multi-distribution
BIC/AIC
Climate variability

ABSTRACT

Determination of extreme wave heights using a Peaks-Over-Threshold (POT) approach is revisited. Firstly, the GPD-Poisson model is recalled. A double threshold is presented and justified, with objective tools for determining the high threshold. This model is then extended to other statistical distributions, namely the Weibull and Gamma distributions. Objective criteria (BIC and AIC) based upon likelihood are used to select the best-fitting distribution. This method is tested on two locations: the historical IAHR Haltenbanken dataset and a location at the entry of the Strait of Gibraltar. Finally, sensitivity analyses are carried out with respect to the high threshold and to the duration of the dataset to estimate the robustness of the approach presented.

1. Introduction

Statistical methods to determine extreme wave heights using the Peaks-Over-Threshold approach (POT) have been significantly improved for several years. The IAHR Working Group on Extreme Wave Analysis issued recommendations about the most appropriate way to proceed when determining extreme wave heights (Mathiesen et al., 1994). They recommended the use of the POT method along with a Weibull distribution estimated by maximum likelihood. A little later, several authors introduced the GPD-Poisson model (e.g. Coles, 2001), which is the most natural way to proceed when using the POT approach. While respecting the general guidelines of the IAHR Maritime Hydraulics-Working Group on Extreme Wave Analysis (Mathiesen et al., 1994), this model notably improves several key steps of the analysis, particularly by fitting a Generalized Pareto Distribution (GPD) to storm peaks while assuming that the number of storms in one year follows a Poisson distribution. It is now recommended (Hawkes et al., 2008) and widely used (e.g. Méndez et al., 2006; Thompson et al., 2009), although many authors still prefer other distributions, mainly the classical extreme distributions: GEV, Weibull, and Gumbel.

However, it should be recalled that the GPD-Poisson model is an asymptotic model. For this reason, other distributions might give better results.

We therefore propose to extend this model to a multi-distribution approach, using the Weibull and Gamma distributions in addition to the GPD. Objective criteria for choosing the most appropriate threshold and determining the best-fitting distribution are also presented.

This method is illustrated by case studies in the Northern Atlantic and in the Strait of Gibraltar.

2. POT method revisited

2.1. Brief justification of the GPD-Poisson model

Let us consider a sample of wave height data $(X_1,\ldots,X_n)$. These data follow an unknown continuous distribution, say F. Let u be a threshold and $y = x_{|x>u} - u$ the exceedance by x of the threshold u. So $Y = (Y_1,\ldots,Y_N)$ is the sample of the N threshold exceedances. The law of threshold exceedance is given by:

$$\mathbb{P}[Y<y] = \mathbb{P}[X<u+y\,|\,X>u] = \frac{F(u+y)-F(u)}{1-F(u)} \tag{1}$$

According to Pickands, 1975 (see also Embrechts et al., 1997), when u is large, this law is very nearly in the form of the Generalized Pareto Distribution defined as:

$$\begin{cases} G_{Y;k,\sigma}(y) = 1-\left(1+k\dfrac{y}{\sigma}\right)^{-\frac{1}{k}} \text{ if } k\neq 0 \\ G_{Y;\sigma}(y) = 1-\exp\left(-\dfrac{y}{\sigma}\right) \text{ if } k=0 \end{cases} \tag{2}$$

where k is the shape parameter and σ is a scale parameter. When $k>0$, the distribution has a heavy and unbounded tail and belongs to the

* Corresponding author. Tel.: +33 4 56 38 46 91; fax: +33 4 76 33 43 33.
E-mail addresses: franck.mazas@sogreah.fr (F. Mazas), luc.hamm@sogreah.fr (L. Hamm).

0378-3839/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.coastaleng.2010.12.003

Fréchet domain of attraction (a heavy tail is not exponentially bounded, and extreme values are more likely to occur than in distributions with exponential or lighter tails). When $k<0$, the distribution is bounded by $x_{max}=u+\sigma/k$ and belongs to the Weibull domain of attraction. Finally, when this parameter is zero, the GPD is the exponential distribution with scale parameter σ.

Still, it must be kept in mind that the GPD is an asymptotic law. This means we must be in its range of validity, i.e. u must be high enough. However, the higher the threshold, the greater the uncertainties because of the very small number of data left. It is the well known dilemma between bias and variance.

If we consider that the number of events (i.e. storms) in one year follows a Poisson distribution with parameter λ, we obtain the so-called GPD-Poisson model: the law of the exceedances is a Generalized Pareto Distribution and the storm occurrence is a Poisson process.

A Poisson distribution should thus be fitted to the data. However, the most common estimator for its unique parameter (e.g. the maximum likelihood estimator) is the empirical mean. We are thus able to link the number of storm occurrences with the return period T.

2.2. The multi-distribution POT model

2.2.1. Data homogenization

The first step in the analysis is to extract homogenous time series from the main continuous sea states time series (buoy measurement, hindcast data, etc.). If this step is omitted, storms from very different meteorological phenomena will be treated together, although it is most likely they are not identically distributed. Such homogenization can be carried out by separation in carefully chosen directional sectors, seasonal analysis (e.g. summer/winter monsoon) and separation of sea states into independent wave systems. Rare but very strong events such as hurricanes should also be checked if necessary. Actually, homogenization may be the most important step in the analysis (this point was stressed by Mathiesen et al., 1994), although it is often the least considered: the best statistical analysis cannot extract the "truth" out of wrongly prepared data ("Garbage In, Garbage Out").

2.2.2. Peak selection and double threshold

Once we have time series of homogenous sea states, we have to extract storm peaks. If we keep in mind that a rigorous statistical analysis requires independent and identically distributed (i.i.d.) data, we will pay special attention to obtaining independent storm peaks. Firstly, we should be careful concerning possible fluctuations in storms around the threshold. If the wave height falls below the threshold for a short period, say 3 h in a 24-hour storm, we should not cut the storm in two. Secondly, we should set a minimum period between two storms to ensure their peaks are independent. Finally, once the storm peaks are identified, outliers (i.e. values significantly larger than the other ones) must be checked carefully in order to be sure they really belong to the population and are not the result of some measurement error. If so, they could have a return period T much larger than the duration of the time series K. Thus they provide valuable information and we recommend keeping them in the sample.

The interest of a threshold is to consider that storm peaks above it have a statistically extreme behavior, i.e. they follow the same extreme distribution. However, we do not know the threshold value *a priori*. A simple way to proceed is therefore to use a double threshold (u_1,u_2). A low value u_1 is set to select both weak and strong storms. There is no need for precise criteria in the choice of u_1 because the procedure relies more heavily on u_2 (see below). Its aim is only to extract the storm peaks from the time series, reducing the sample size from 10,000 to 100,000 values to a few hundreds of peaks. u_1 shall be high enough to discriminate two consecutive storms and low enough to be below the "extreme area", i.e. the strong storms showing genuine statistically extreme behavior.

We obtain N_T peaks over a period of K years. Hence, the mean number of storms per year above u_1 is:

$$\lambda_T = \frac{N_T}{K} \tag{3}$$

Our experience in extra-tropical areas led us to set u_1 so as to have λ_T approximately between 5 and 10, although it is not an absolute constraint.

We have now to determine the high threshold u_2 above which storms have a statistically extreme behavior. As the GPD is the asymptotic law, it seems quite reasonable to use its properties to determine u_2. In particular, if a sample follows a GPD, the shape parameter k and the modified scale parameter $\sigma^*=\sigma-ku_2$ remain constant when u_2 increases. So if we fit a GPD to the exceedances of a threshold varying between u_1 and, for instance, a threshold corresponding to one storm per year, we can draw graphs of shape and modified scale parameters with respect to u_2 and search for "domains of stability" where they will remain roughly constant. As we want to be in the asymptotic domain, we are interested in the highest domain of stability. And as we want to have as much information as possible, we will choose the lowest threshold of this highest domain.

Thus we select N storm peaks over K years, namely $\lambda=N/K$ storms per year (as we have seen, this empirical mean is also the estimator of the Poisson parameter). After many tests, we believe it is appropriate for λ to stand approximately between 2 and 5. If K is low, a value around 5 is more advisable in order to ensure that N is large enough (with a minimum of 20–30). In contrast, if K is quite large (around 40–50), a value of around 2 is more appropriate.

2.2.3. Fit to multiple distributions

Storm peaks above u_2 are now to be fitted to a statistical distribution. As we have seen, the GPD is the asymptotic law, and thus a natural candidate. However, we do not know whether we are within the asymptotic domain. Thus, other distributions might fit the data better. We can try many of them and then determine the best-fitting one.

When looking for suitable distributions, it is useful to know their domain of attraction for maxima (Castillo and Sarabia, 1992). If they belong to the Fréchet domain (e.g. Pareto or beta laws), their tails are heavy and unbounded, which means they give too much weight to extreme events. Practice shows they are not appropriate for coastal engineering applications where the wave heights are physically bounded. If they belong to the Gumbel domain, their tails decrease exponentially. If they belong to the Weibull domain (e.g. GPD with negative shape parameter), their tails are bounded. We can thus limit our study to distributions belonging to Weibull or Gumbel domains of attraction for maxima.

Our tests have shown that along with the GPD, the Gamma distribution and 2-parameter Weibull distribution for minima often behave quite well. Although other distributions may be studied, we will work here with these two laws, whose cumulative distribution functions are respectively:

$$\text{Gamma}: F_{Y;k,\sigma}(y) = \mathbb{P}\{Y\leq y\} = \frac{\gamma\left(k,\frac{y}{\sigma}\right)}{\Gamma(k)} \tag{4}$$

$$\text{Weibull}: F_{Y;k,\sigma}(y) = \mathbb{P}\{Y\leq y\} = 1-\exp\left[-\left(\frac{y}{\sigma}\right)^k\right] \tag{5}$$

As for the GPD, we work with y, that is the threshold exceedance $x-u_2$, provided $x>u_2$. k and σ are respectively the shape and the scale parameters. Γ is the Gamma function, and γ is the lower incomplete gamma function.

The choice of the estimator is also quite important. Mathiesen et al. (1994) mention three estimators: least squares methods, the method of moments and the Maximum Likelihood Estimator (MLE). The statistical theory says that an estimator must be *robust*, i.e. it is not disturbed by an outlier and it must be *consistent*, i.e. the bias and the variance tend to zero when the sample size increases. Least squares methods, though easy to implement, are neither robust nor consistent. In particular, they are found to be sensitive to outliers (Mathiesen et al., 1994). They are therefore rejected. It is nevertheless noteworthy that Goda (2000) recommends this method with modified plotting position formulae. The method of moments may be used as first approximations but the small sample sizes hinder it. In particular, the method of moments gives too much bias for the typical sample sizes we are handling (Goda, 2000). To handle this difficulty, Hosking and Wallis (1987) have proposed an estimator based upon the Probability Weighted Moments. But it is known to be less efficient than the MLE. Finally, the most handy and appropriate method is to use the Maximum Likelihood Estimator (MLE). This estimator maximizes the likelihood function of the fit, which is defined by:

$$L(X_1,\dots,X_N|\theta) = \Pi_{i=1}^{N} f_\theta(X_i;\theta) \tag{6}$$

where f_θ is the joint density function (with parameter vector θ) at the sample observations X_i. The log-likelihood function is usually used, since it is much easier to derive:

$$l(X_1,\dots,X_N|\theta) = \Sigma_{i=1}^{N} \ln(f_\theta(X_i;\theta)) \tag{7}$$

Thus we have an optimization problem, as the likelihood function has as many variables as the distribution has parameters. In some cases, optimization algorithms may fail to maximize the likelihood.

However, the use of the MLE for two-parameter distributions such as the Weibull and Gamma distributions has a very disturbing drawback. These distributions are very sensitive to the distance between u_2 and the first peak. In other words, the estimated parameters will be quite different if the smallest value of the ordered sample of the threshold exceedances Y_1 is 0.1, 0.01 or 0.001. When we look at the 100-year wave height, the result varies between 14 and 16 m! The GPD is much less sensitive to this phenomenon. We think the explanation could be related to the shape of the density functions just above 0. A comparison with the method of moments estimator was carried out. From these tests, it appeared that the two-parameter distributions could be used with MLE, but only when u_2 meets a storm peak. As this peak is excluded, the first value of the exceedance sample is as far from u_2 as possible.

A solution would be to use the three-parameter Weibull and Gamma distributions (the latter being known as Pearson-III distribution) by adding a location parameter μ ($\mu < Y_1$, the first and smallest exceedance):

$$\text{Pearson}-\text{III}\,(3-\text{parameter Gamma})\text{:}\ F_{Y;k,\sigma}(y) = \mathbb{P}\{Y \le y\} = \frac{\gamma\left(k, \frac{y-\mu}{\sigma}\right)}{\Gamma(k)} \tag{8}$$

$$\text{Weibull}: F_{Y;k,\sigma,\mu}(y) = P\{Y \le y\} = 1-\exp\left[-\left(\frac{y-\mu}{\sigma}\right)^k\right] \tag{9}$$

However, ML estimation of such distributions is very difficult, and the algorithms usually fit two-parameter distributions inside a discrete range of location parameters (Panchang and Gupta, 1989). Actually, it appears that quite often the maximum likelihood with respect to this location parameter μ is obtained for $\mu \to Y_1$ (with $\mu < Y_1$). Now, the Maximum Likelihood Estimator is known to provide poor results when the maximum is at the limit of the interval of validity of one of the parameters. In our applications, this is a major drawback of this estimator. We are currently carrying out further investigation on this subject and shall soon submit our results.

2.2.4. Best fit selection

Once several distributions are fitted to the data, we have to determine the best fit. For this purpose, we use objective Bayesian criteria. The first one is the Bayesian Information Criterion (BIC), also known as the Schwarz Criterion (Schwarz, 1978). It minimizes the bias between the fitted model and the unknown "true" model. Assuming asymptotic conditions (N large enough), BIC is given by:

$$\text{BIC} = -2\ln L + k_p \ln N \tag{10}$$

where L is the likelihood of the fit, N is the sample size (number of storm peaks above u_2) and k_p is the number of parameters of the distribution.

We can also use the closely related Akaike Information Criterion (AIC), which gives the model providing the best compromise between bias and variance (Akaike, 1973). It can be interpreted as the sum of two terms, the first one measuring bias and the second one measuring variance. Under the same assumptions as BIC, AIC is given by:

$$\text{AIC} = -2\ln L + 2k_p \tag{11}$$

For BIC as for AIC, the lower the criterion, the better the fit, so we will select the distribution providing the lowest criteria. Most of the time, both criteria give the same result. If they do not, we recommend keeping the distribution giving the most conservative return values.

2.2.5. Return values and confidence intervals

We now have only one distribution left, with MLE estimated parameters. We are interested in wave heights of return period T. It is actually a quantile of the estimated distribution, whose non-exceedance probability is $1-1/\lambda T$. These quantiles for GPD, Gamma and Weibull distributions are given by:

$$\text{GPD}: Hs_T = u_2 + \frac{\hat{\sigma}}{\hat{k}}\left[(\lambda T)^{\hat{k}} - 1\right] \tag{12}$$

$$\text{Gamma}: Hs_T = u_2 + \Gamma^{-1}_{\hat{k},\hat{\sigma}}\left(1-\frac{1}{\lambda T}\right) \tag{13}$$

$$\text{Weibull}: Hs_T = u_2 + \hat{\sigma}\,[\ln(\lambda T)]^{\frac{1}{\hat{k}}} \tag{14}$$

Finally, confidence intervals are to be computed. Many authors (Coles, 2001) use the classical asymptotic method. Mathiesen et al. (1994) advocate the use of Monte-Carlo simulation techniques. A robust way is to use parametric bootstrap methods (Thompson et al., 2009). The principle is quite simple (see for instance Efron and Tibshirani, 1993a,b). From the estimated distribution with estimated parameter vector $\hat{\theta}_0$, a random sample of size N is generated and the same distribution is then fitted to this sample, leading to a slightly different estimated parameter vector $\hat{\theta}_1$, which will give a slightly different quantile $Hs_{T,1}$. After 100,000 iterations, a sample of 100,000 $Hs_{T,i}$ is obtained. The 90% confidence interval will be given by the percentiles $[Hs_{T,5\%}; Hs_{T,95\%}]$. It is advisable to correct the bias of the bootstrap. Bias is given by the difference between the empirical mean of $Hs_{T,i}$ and $Hs_{T,0}$, and it simply has to be removed from the percentiles previously obtained.

3. Case studies

3.1. Datasets

We shall study two different locations. The first one is the historical Haltenbanken dataset, provided by the IAHR Working Group on Extreme Wave Analysis (van Vledder et al., 1994). The Haltenbanken

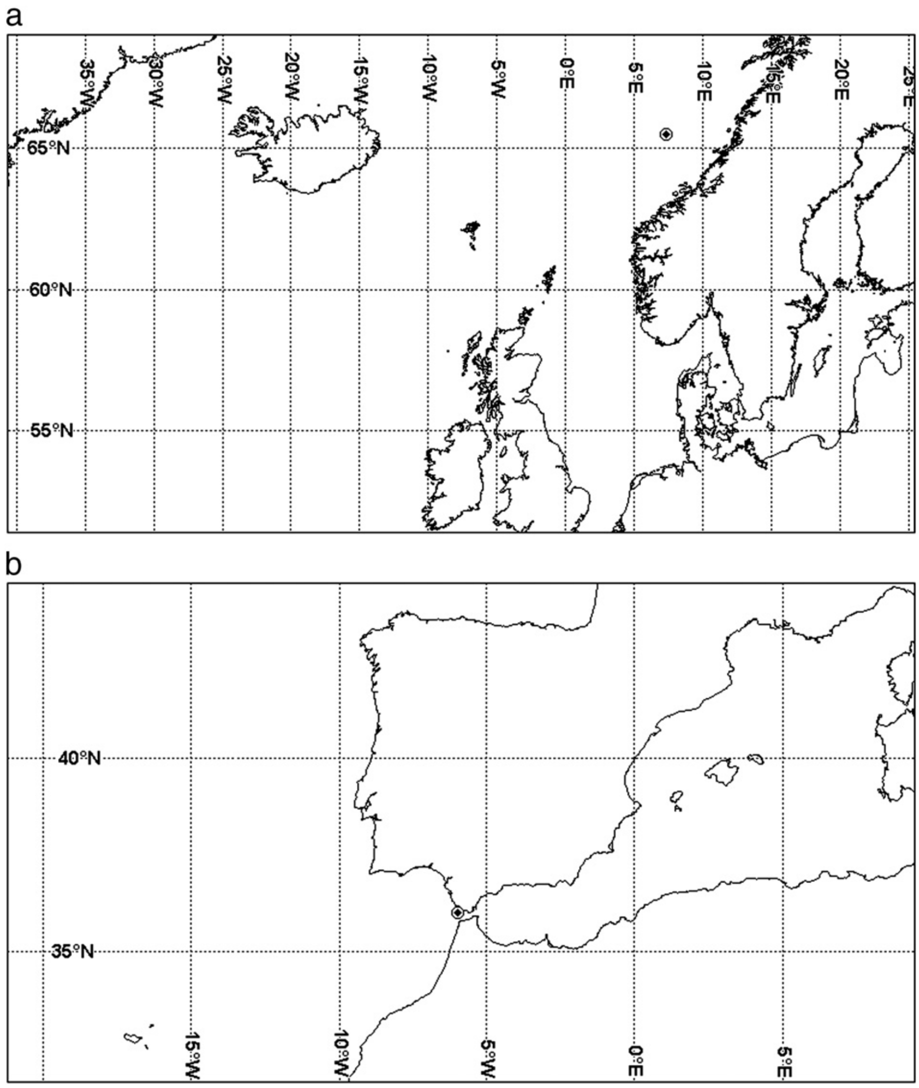

Fig. 1. a) Location of the Haltenbanken dataset. b) Location of the Gibraltar dataset.

buoy is located off the coast of Norway. Its coordinates are 65°5′N; 7°34′ E (Fig. 1a). The original dataset consists of 128 buoy-measured storm peaks above 7 m for a period of 9 years, so no pre-treatment on this sample was done and the peaks were considered to be independent and identically distributed. The shortness of the period must be stressed and will be discussed later.

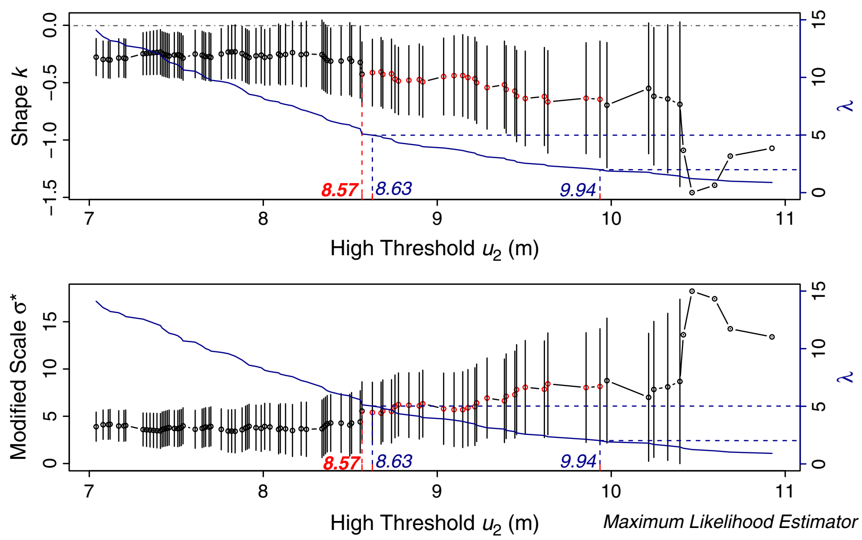

Fig. 2. Haltenbanken dataset: stability of shape and modified scale parameters for Generalized Pareto Distribution.

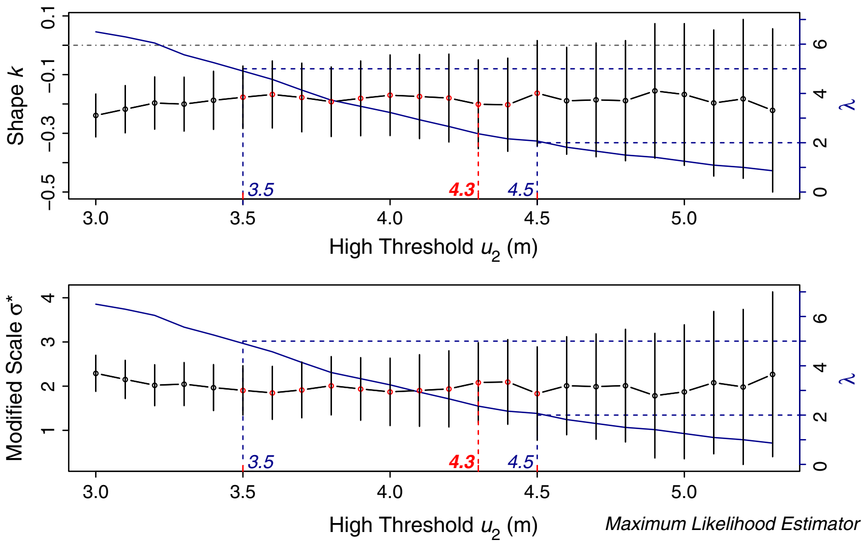

Fig. 3. Gibraltar dataset: stability of shape (above) and modified scale (below) parameters for Generalized Pareto Distribution.

The second dataset comes from the SIMAR-44 hindcast database provided by Puertos del Estado. We chose point 1056044, whose coordinates are 36°N; 6°W (Fig. 1b). It is located at the western entry of the Strait of Gibraltar. Wave and wind data are provided every 3 h from 1958 to 2001 for a total of 44 years. In contrast to Haltenbanken, the storm peaks have to be extracted here. As has been said above, the first and most important step is to homogenize the sample. A simple method, directional analysis, will be used. We will only consider the western sector, facing the Atlantic. From the model point, we can draw lines to Cape St-Vincent, Portugal north-westwards and towards the Moroccan coast near El-Jadida. Thus we obtain the following sector: [220°; 295°]. It is not a very wide sector, but unsurprisingly it is the dominant one, as 86% of the hindcasted waves come from this sector. We consider that all these waves are homogenous and, in particular, that all these western storms are generated by the same kind of Atlantic depressions and thus are identically distributed.

If we set a low threshold u_1 equal to 3 m, we select $N_T = 288$ storm peaks. As $K = 44$ years, we then have a mean number of total storms per year $\lambda_T = 6.55$, which seems quite reasonable.

3.2. Selection of high thresholds

We will now try to determine the best high threshold. For this purpose, as has been explained in Section 2.2.2, we will adjust a GPD to the data over a wide range of thresholds and look at the stability of the shape parameter k and of the modified scale parameter σ^* (Fig. 2). As has been previously discussed, the thresholds tested are those meeting the dataset values. We use tools available in the "ismev" package (Coles and Stephenson, 2006) developed for the R language (R Development Core Team, 2009). We modified these tools in order to take into account the remarks made previously. On the secondary axis, we draw the change in λ, so as to see easily the thresholds corresponding to a value of λ between 2 ($u_2 = 9.94$) and 5 ($u_2 = 8.63$). These limit thresholds are written in italics.

We can see two domains of stability where the parameters remain approximately constant. The lowest threshold of the highest domain, just below the value of 8.63, is 8.57 m. For this threshold, λ is 5.11, which is slightly higher than 5, but as K is very low (9 years), we can allow this small exceedance in order to have N large enough (46).

For Gibraltar, the choice is more difficult (see Fig. 3). The curves are rather flat when u_2 is higher than 3.5. (corresponding to $\lambda = 5$) This value could therefore be adopted. However, it is important to bear in mind that here K is large (44 years). A value closer to $\lambda = 2$ (corresponding to $u_2 = 4.5$ m) may therefore be more appropriate. Since there is small bump for 4.5 m, we will choose $u_2 = 4.3$ m. It is clear that choosing the right threshold is not always a straightforward matter. Thompson et al. (2009) presented methods for automated threshold selection, but these should be used rather when working with too many datasets for visual examination.

Table 1 recapitulates the characteristics of the samples.

Table 1
Characteristics of the samples.

	Haltenbanken	Gibraltar
K (years)	9	44
u_1 (m)	7	3
N_T (−)	128	288
λ_T (yr^{-1})	14.22	6.55
u_2 (m)	8.57	4.3
N (−)	46	104
λ (yr^{-1})	5.11	2.37

3.3. Fit

For both datasets, the three distributions (GPD, Weibull, Gamma) are now fitted to the exceedances of the high threshold with the Maximum Likelihood Estimator. Table 2 provides BIC and AIC criteria for the two locations and the three distributions.

We can see that both criteria give the same result. For Haltenbanken, GPD is clearly selected, with Weibull then Gamma quite far off. In contrast, for Gibraltar, GPD gives poor results with respect to these criteria. The Gamma distribution is selected since it minimizes both BIC and AIC criteria.

Actually, as we use distributions with the same number of parameters (two), we could consider only the (log-)likelihood of the fit: this would give the same results as BIC and AIC. However, it is convenient to have a means of discriminating between fits for distributions with one, two or three parameters together.

Table 2
BIC and AIC criteria for the fits of the three distributions for both datasets.

		GPD	Weibull	Gamma
Haltenbanken	BIC	120.0	122.0	122.4
	AIC	116.3	118.3	118.8
Gibraltar	BIC	216.6	212.4	211.4
	AIC	211.3	207.1	206.1

Table 3
Return values for the best-fitting distribution with 90% confidence intervals.

Return period (years)	Haltenbanken (GPD)	Gibraltar (Gamma)
100	12.7	8.8
	12.0–13.5	8.0–9.6
50	12.6	8.3
	11.9–13.3	7.6–8.9
20	12.4	7.6
	11.8–13.0	7.1–8.1
10	12.2	7.1
	11.7–12.7	6.7–7.5
5	11.9	6.5
	11.5–12.3	6.2–6.9
2	11.4	5.8
	11.0–11.8	5.6–6.0
1	10.8	5.3
	10.4–11.2	5.1–5.4

3.4. Return values and confidence intervals

The last step of the analysis is now to compute the return values for the return periods of interest using the quantile functions defined above for the best-fitting distribution. 90% confidence intervals are also computed using a parametric bootstrap approach with 100,000 iterations (bootstrap bias is corrected). Table 3 gives the results for 1, 2, 5, 10, 20, 50 and 100 years.

It can be seen that the Haltenbanken results are much lower than those given in van Vledder et al. (1994), where Hs_{100} varies between 14.2 and 15.8 m. The methods used in this paper were quite different, and nobody used a GPD at that time. The closest analysis was carried out by member D, who applied a 3-parameter Weibull distribution to the 46 storm peaks above 8.6 m. This member obtained a Hs_{100} of 14.7 m (12.6–16.9 90% CI). In our analysis, the 2-parameter Weibull distribution also gives 14.7 m for Hs_{100} (13.1–16.6 90% CI) but is rejected by the BIC/AIC criteria. It is also noteworthy that bootstrap confidence intervals for this GPD fit are much narrower than in the case of the Working Group analysis (1.5 m versus 2.5 to 5 m). Nonetheless, an explanation of such differences will be given later.

As for Gibraltar, the confidence intervals remain quite narrow; the fit seems reasonable.

4. Sensitivity analysis with respect to the high threshold

4.1. Purpose of the analysis

We have proposed an objective method for determining the high threshold. Nevertheless, the case of Gibraltar shows that a part of subjectivity may remain when choosing it. It follows that studying the

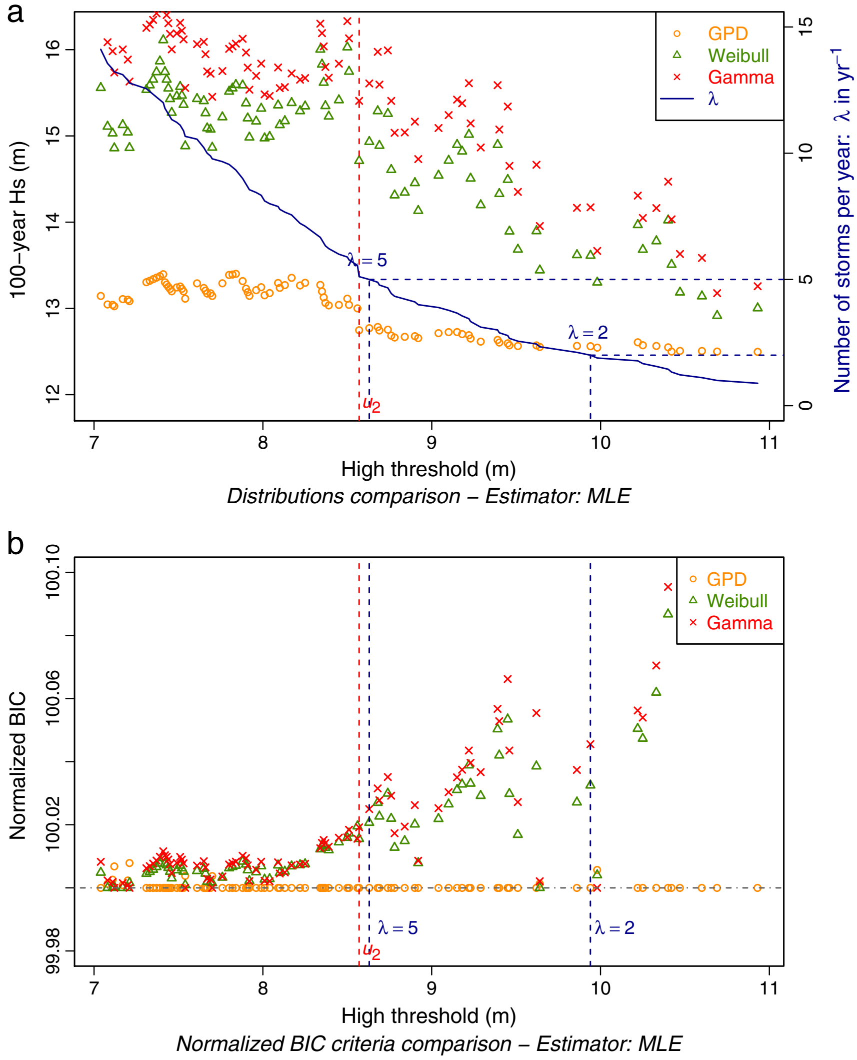

Fig. 4. Haltenbanken: a) stability of the 100-year significant wave height with respect to the high threshold. b) stability of the normalized BIC criterion with respect to the high threshold.

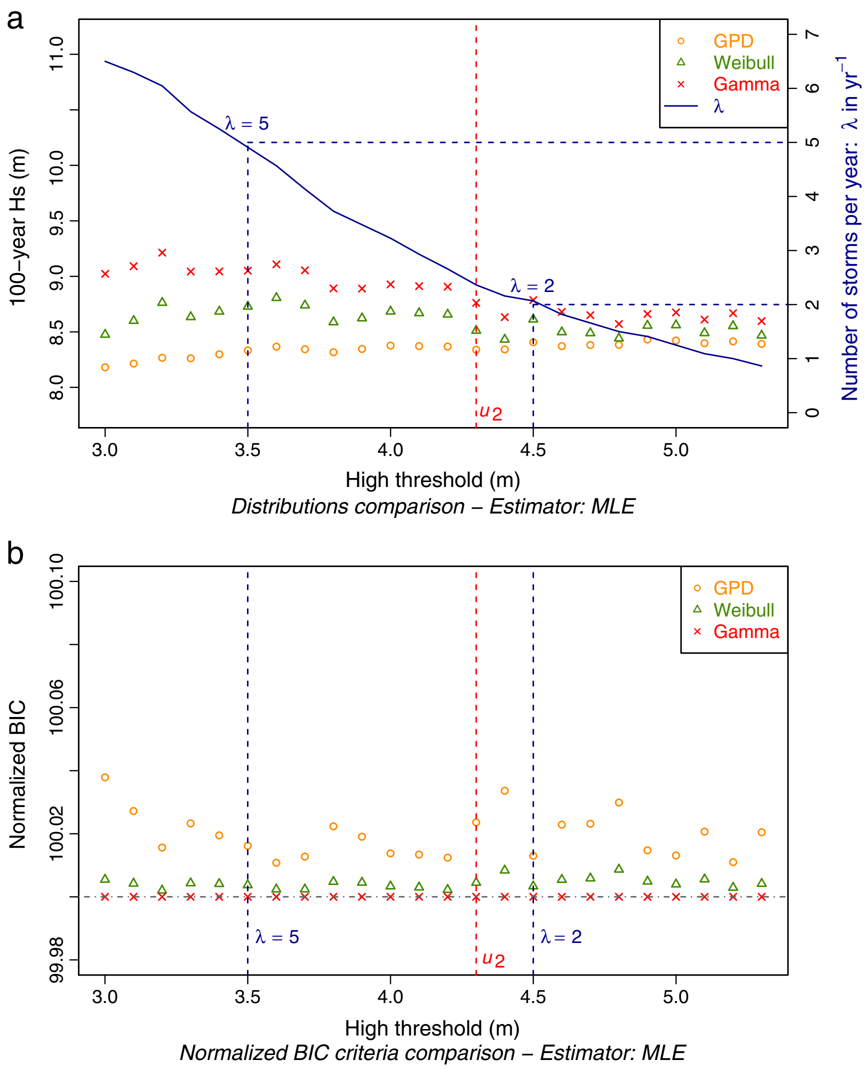

Fig. 5. Gibraltar: a) stability of the 100-year significant wave height with respect to the high threshold. b) stability of the normalized AIC criterion with respect to the high threshold.

change in return values and goodness of fit (i.e. in BIC/AIC criteria) can provide interesting information for validating the results or, conversely, reconsidering the choice of threshold.

Let us draw similar plots to those showing the stability of the GPD parameters, but this time with Hs_{100} and BIC/AIC criteria for each distribution. As for the criteria, it is actually necessary to normalize

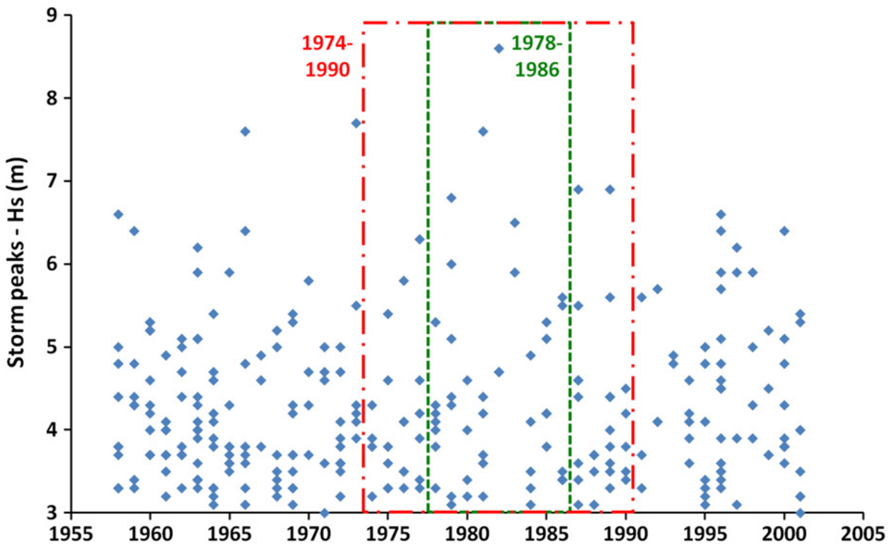

Fig. 6. Gibraltar: storm peaks above the low threshold u_1 with respect to the calendar years. Choice of the 9-year period 1978–1986 (dashed line) and of the 17-year period 1974–1990 (dashed-dotted line).

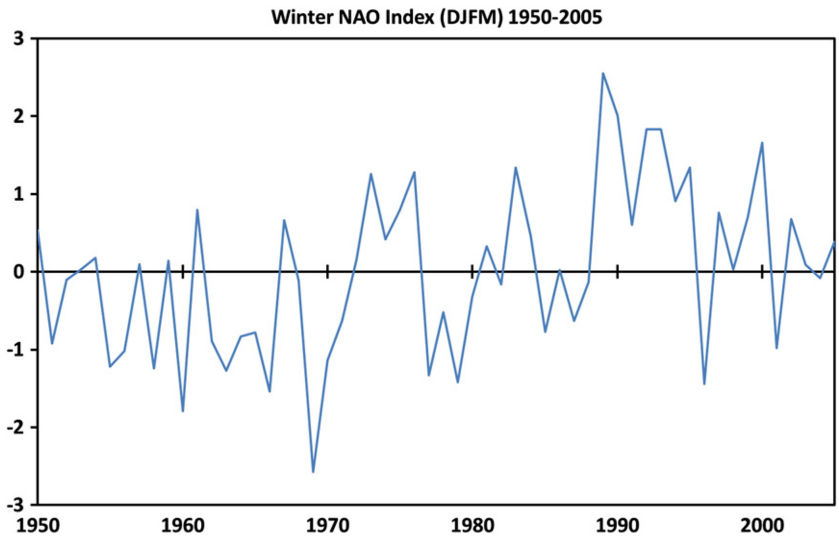

Fig. 7. PC-based winter NAO index from 1950 to 2005.

them in order to have clear graphs. Indeed, they will depend on the sample size and so direct comparison between all the tested thresholds is difficult. For each threshold, the criteria are computed for the three distributions studied, and the minimum criterion is set at 100. The relative difference of the other criteria with respect to the minimum are then added to 100.

5. Results

For the Haltenbanken dataset, we can see that from 8.57 m, that is the high threshold we chose previously, the GPD quantile is extremely stable, which is quite a remarkable result (Fig. 4a). The Weibull and Gamma distributions are much more unstable and seem to tend downwards towards the GPD value. The change in the normalized BIC criterion (Fig. 4b) shows that the GPD is almost always the best-fitting distribution in this case. Thus, the choice of u_2 on the basis of the stability of the GPD parameters appears to be particularly relevant here. It is even probable that the asymptotic domain starts at 8.57 m, thus giving further argument for choosing the GPD.

As in the case of the Gibraltar dataset, Fig. 5a shows that the three distributions converge towards a common value of Hs_{100}. Similarly to Haltenbanken, the GPD quantiles are below those of the other distributions. This pattern was observed in many tests: the GPD value is clearly not conservative compared to other distributions. As a matter of fact, the Weibull and Gamma distributions often behave well when no saturation (i.e. no “flattening” of the highest peaks) is observed in the data, whereas the GPD with a strongly negative shape parameter generally fits well when saturation occurs. Here, the (normalized) AIC criterion shows that the Gamma distribution is always the best-fitting one (see Fig. 5b). Once again, the choice of threshold appears to be relevant.

These stability plots for Hs_{100} and normalized criteria are a very helpful way of checking the results obtained previously. If it seems obvious that the return value is not at all representative, the choice of threshold will have to be reconsidered.

Table 4
Gibraltar: fit characteristics and 100-year significant wave height with 90% confidence intervals for the 9-year, 17-year and 44-year periods centered on 1982.

	9 years (1978–1986)	17 years (1974–1990)	44 years (1958–2001)
u_2 (m)	3.1	3.3	4.3
N (−)	44	79	104
λ (yr^{-1})	4.89	4.65	2.37
Min BIC/AIC distribution	Gamma	Gamma	Gamma
Hs_{100} GPD (m) 90% CI	9.3	9.1	8.3
	7.3–12.0	7.4–11.2	7.6–9.2
Hs_{100} Weibull (m) 90% CI	10.0	9.4	8.5
	8.1–12.3	8.1–11.0	7.8–9.3
Hs_{100} Gamma (m) 90% CI	10.5	9.6	8.8
	8.6–12.6	8.4–11.0	8.0–9.6

6. Sample duration

6.1. Purpose of the analysis

We have studied two datasets, one very short compared to the usual available duration (around 20 years) and the other quite long. It is likely that the very short duration of the Haltenbanken dataset (9 years) is the cause of the huge differences between the return values given by the three statistical distributions.

The Gibraltar dataset provides an opportunity to test the sensitivity of the return values with respect to the sample duration. Fig. 6 shows the westerly storm peaks above the low threshold u_1 with respect to the calendar years.

6.2. Link with the North Atlantic Oscillation

We also know that the Atlantic storm tracks are related to the North Atlantic Oscillation (NAO), i.e. the oscillation of the atmospheric pressure gradient between the Iceland low and the Azores high around a long-term mean (Hurrell, 1995). Bacon and Carter (1993) were among the first to suggest such a link. Recently, Dodet et al. (2010) used a 57-year hindcast (1953–2009) to quantify this link. They found correlation coefficients between the winter NAO index on the one hand, and the 90% *Hs* percentile, the mean wave direction winter-means and the peak period winter-means on the other hand. As for *Hs*, the Pearson correlation coefficients are close to 1 off the British Isles, close to zero off Galicia and become negative off the Moroccan coast. The coefficient is around −0.3 at the western entry of the Strait of Gibraltar. Taking account of the winter NAO index for this location thus provides valuable information.

If we look at the change in the (PC-based) winter NAO index from 1950 to 2005 (see Fig. 7), we see that from 1950 to approximately 1980 (except for 1973 to 1975), the index is mostly negative, which means the Atlantic storm tracks go preferentially southwards. After 1980, the index is mostly positive and storms sweep preferentially over northern Europe.

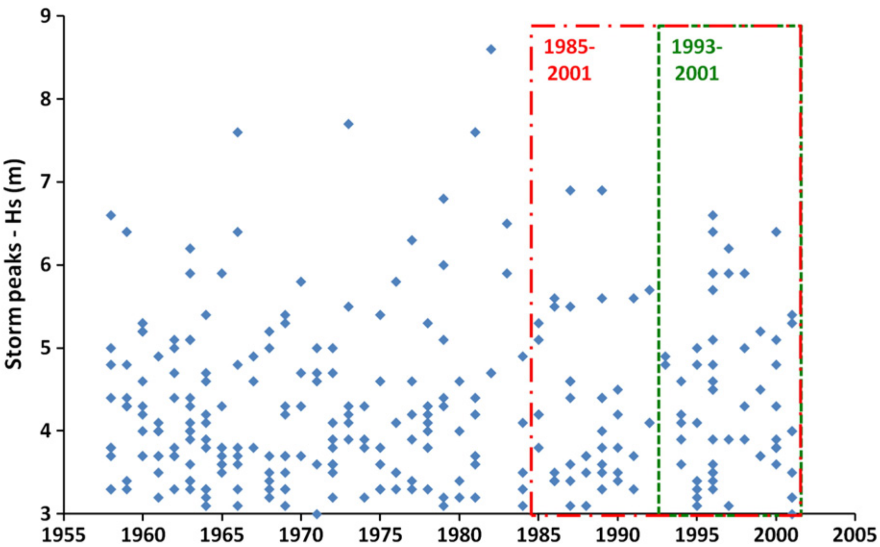

Fig. 8. Gibraltar: storm peaks above the low threshold u_1 with respect to the calendar years. Choice of the 9-year period 1993–2001 (dashed line) and of the 17-year period 1985–2001 (dashed-dotted line).

Méndez et al. (2006) use a non-stationary POT model to take into account the NAO index by allowing the GPD parameters to be time-dependent. However, it is not easy to determine the period of such an index, and the more parameters the distribution has, the more uncertainties there are in the final result. We thus kept on working with a stationary POT model.

7. Results

As the highest peak is reached in 1982 (not far from the switch between the negative and positive indices), this will be taken as a pivotal year. We will therefore study a 9-year period (1978–1986) and a 17-year (1974–1990) period, both centered on 1982, and the results will be compared with the 44-year dataset. The results are given in Table 4.

The main conclusion that may be drawn is that the return values are lower when the duration of the dataset increases. The extreme peak of 1982 (most probably very close to the 100-year wave height) clearly plays a role in this phenomenon, as its weight is greatly enhanced in the 9-year dataset compared to the 44-year one. The confidence intervals are therefore much wider in the short duration dataset. Another important fact is that the Gamma distribution is considered the best-fitting distribution by both the BIC and AIC criteria for the three datasets.

It is also noteworthy that the deviation between the 100-year wave heights for the three distributions is only 11% in the case of the 9-year sample and no more than 5–6% in that of the 17-year and the 44-year datasets.

The interest of a long dataset is clear, as the deviation between the best 100-year wave heights is around 9% between the 17-year and 44-year datasets. This interest is all the greater when the presence of an outlier is evident, as is the case here. A long period also allows cyclical regional climatic patterns with long (decadal or multidecadal) periods, such as the NAO, to be taken into account. However, engineers usually work with datasets whose duration is rather 15 to 20 years. Special attention should therefore be paid to such climatic patterns, as the dataset could cover periods with storm peaks that are lower or higher than a long-term mean.

If we choose now to study a 9-year and a 17-year period with weaker storms, the results are quite different. Let us carry out the analysis for the periods 1993–2001 (9 years) and 1985–2001 (17 years), when the NAO index is mostly positive and the storms rather weak (see Fig. 8). Results are given in Table 5.

BIC and AIC criteria now select the GPD for these two periods, though it gives far lower return values than the Weibull and Gamma distributions, whose return values remain by chance quite constant. Surprisingly, the GPD confidence intervals are narrower for the 9-year period whereas they are wider for the other two distributions.

Actually, it seems that when there is an outlier in a short-duration dataset, the GPD is much less sensitive to it than the Gamma and Weibull distributions. But when such a short duration corresponds to a calmer than usual period, the GPD returns wave heights that are too low, although it fits the data very well. This may well be the case for the Haltenbanken dataset, where no outlier appears. In spite of the likelihood-based criteria discrimination, a conservative choice for Haltenbanken would be to choose another distribution, say the Weibull one, with Hs_{100} around 15 m, which would be in accordance with observations (see for instance Magnusson et al., 2006).

It is clear that 9 years is definitely too short a period for a robust extreme wave heights analysis. When storms are thought to be stronger than usual in this period (or if the dataset contains an outlier), the GPD-Poisson model gives good results and seems quite stable. In contrast, if storms are thought to be rather weaker than usual, the GPD-Poisson model may produce return values that are too low, in spite of a very good statistical fit. In such a case, it is imperative to extend the duration of the dataset. If it is not possible, choosing the most conservative distribution could be safer than relying on the BIC/AIC criteria for design purposes.

Two conclusions may be drawn from this analysis when working with very short datasets (less than 10 years). Firstly, the interest of extending the GPD-Poisson model to other statistical distributions is manifest as we may obtain return values that are too low with this

Table 5
Gibraltar: fit characteristics and 100-year significant wave height with 90% confidence intervals for the calmer 9-year, 17-year and 44-year periods.

	9 years (1993–2001)	17 years (1985–2001)	44 years (1958–2001)
u_2 (m)	3.9	4.2	4.3
N (−)	32	43	104
λ (yr^{-1})	3.55	2.53	2.37
Min BIC/AIC distribution	GPD	GPD	Gamma
Hs_{100} GPD (m) 90% CI	6.6	7.0	8.3
	6.3–6.8	6.6–7.3	7.6–9.2
Hs_{100} Weibull (m) 90% CI	7.9	7.9	8.5
	6.9–9.2	7.1–8.9	7.8–9.3
Hs_{100} Gamma (m) 90% CI	8.7	8.4	8.8
	7.4–10.1	7.4–9.5	8.0–9.6

model. Secondly, in this case criteria based upon likelihood fail since the data are not fully representative of the local climate: engineers have to keep in mind that the ultimate goal is to provide safe design criteria and not the "purest" statistical fit.

To conclude with regard to the duration of the dataset, it may be said that a limit to the ratio between T and K, i.e. the return period and this duration, is necessary but not enough (this ratio is generally close to 5). This analysis has shown that K must be large enough with respect to the local climate in order to avoid covering only particularly weak or strong periods. We believe twenty years is a minimum period for a reasonably robust extreme wave analysis.

8. Conclusions

We carried out a complete review of a rigorous method for determining extreme wave heights using the GPD-Poisson model, in particular for choosing the high threshold. Although objective methods exist, it is clear that the choice may still be difficult. Parameters such as N and λ should be kept in mind when choosing u_2. Even so, several thresholds sometimes need to be tested.

It was also seen that although the GPD has the best theoretical justification for being selected as the asymptotic law, other distributions may give better results. Criteria for selecting the best-fitting distribution are presented. They are based upon the fit likelihood. However, analysts must always be very careful about the location of the high threshold with respect to the first exceedances, as instabilities can occur for both Gamma and Weibull distributions. We recommend using only high thresholds equal to the data values, but a better understanding of this purely mathematical phenomenon is necessary.

Sensitivity analyses for the return value and/or the criteria with respect to high thresholds are very helpful for post-checking the relevance of the choice made for u_2. However, these graphs should only be used as a verification tool, and not for decision-making.

This method was tested for two locations. As for the Haltenbanken dataset, the GPD-Poisson model had the best behavior and led to significantly lower 100-year wave heights than those calculated by the IAHR Working Group in 1993 (van Vledder et al., 1994), probably due to too short dataset duration. As for the Gibraltar location, the Gamma distribution was considered the best in relation to both the BIC and AIC criteria. Graphs illustrating the sensitivity analyses reinforced these estimations.

The interest of working on long duration datasets was also demonstrated. This interest is enhanced when the presence of outliers is suspected or when decadal or multidecadal climatic patterns may play a role. The multi-distribution approach appears to be necessary for very short datasets, although the means for discriminating the best-fitting distribution requires improvement. Indeed, we have shown that the GPD-Poisson model can lead to dangerously low return values when the analysis is carried out for a very short and rather calm period. In this particular case, if the period cannot be extended, BIC/AIC criteria may be put aside and the most conservative results may be chosen. Consequently, a dataset covering at least 20 years is strongly recommended.

We have thus a robust enlargement of the stationary GPD-Poisson model. It may be useful for engineers wishing to cover a wide range of situations. The choice of distributions proposed here is not exclusive, and others may be used. Engineers should remember that their aim is to determine safe design criteria rather than perfect statistical fits, so they must always be careful to ensure that the available data are fully representative of the local climate.

Acknowledgements

The authors wish to thank Dr. Philippe Garat from Pierre-Mendes-France University (Grenoble, France) for his help and advice.

References

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. Second International Symposium on Information Theory 267–281 Akademiai Kiado, Budapest.
Bacon, S., Carter, D.J.T., 1993. A connection between mean wave height and atmospheric pressure gradient in the North Atlantic. Int. J. Climatol. 13, 423–436.
Castillo, E., Sarabia, J.M., 1992. Engineering analysis of extreme value data: selection of models. J. Waterw. Port Coast. Ocean Eng. 118 (2), 129–146.
Coles, S., 2001. An Introduction to Statistical Modelling of Extreme Values. Springer, London.
Coles, S., Stephenson, A., 2006. ismev: An Introduction to Statistical Modelling of Extreme Values(Original S functions) (R port and R documentation files) http://www.maths.lancs.ac.uk/stephena/2006R package version 1.2.
Dodet, G., Bertin, X., Taborda, R., 2010. Wave climate variability in the North-East Atlantic Ocean over the last six decades. Ocean Model. 31, 120–131.
Efron, B., Tibshirani, R.J., 1993a. An Introduction to the Bootstrap. Chapman & Hall, London.
Efron, B., Tibshirani, R.J., 1993b. An Introduction to the Bootstrap, Monographs on Statistics and Applied Probabilities 57. Chapman & Hall, New York.
Embrechts, P., Kluppelberg, C., Mikosh, T., 1997. Modelling Extreme Events. Springer, New York.
Goda, Y., 2000. Random Seas and Design of Maritime Structures, 2nd Edition. Advanced Series on Ocean Engineering, Vol. 15. World Scientific.
Hawkes, P., Gonzalez-Marco, D., Sánchez-Arcilla, A., Prinos, P., 2008. Best practice for the estimation of extremes: a review. J. Hydraul. Res. Vol. 46 (Extra Issue 2), 324–332.
Hosking, J.R.M., Wallis, J.R., 1987. Parameter and quantile estimation for the generalized Pareto distribution. Technometrics vol. 29, 339–349.
Hurrell, J.W., 1995. Decadal trends in the North Atlantic Oscillation: regional temperatures and precipitations. Science 269, 676–679.
Magnusson, A.K., Reistad, M., Breivik, O., Myklebust, R., Ash, E., 2006. Forecasting a 100-year wave event. Proceedings of the 9th International Workshop on Wave Hindcasting and Forecasting, Victoria, Canada, 24th to 30th September 2006.
Mathiesen, M., Goda, Y., Hawkes, P.J., Mansard, E., Martín, M.J., Peltier, E., Thompson, E.F., Van Vledder, G., 1994. Recommended practice for extreme wave analysis. J. Hydraul. Res. 32 (6), 803–814.
Méndez, F.J., Menéndez, M., Luceño, A., Losada, I.J., 2006. Estimation of long-term variability of extreme significant wave height using a time-dependent Peaks Over Threshold (POT) model. J. Geophys. Res. 111, C07024. doi:10.1029/2005JC003344.
Panchang, V.G., Gupta, R.C., 1989. On the determination of three-parameter Weibull MLE's. Commun. Statist. Simula. 18 (3), 1037–1057.
Pickands, J., 1975. Statistical inference using extreme order statistics. Ann. Statist. 3, 119–131.
R Development Core Team, 2009. R: A Language and Environment for Statistical Computing. 3-900051-07-0. http://www.R-project.org, Vienna, Austria
Schwarz, G., 1978. Estimating the dimension of a model. Ann. Statist. 6, 461–464.
Thompson, P., Cai, Y., Reeve, D., Stander, J., 2009. Automated threshold selection methods for extreme wave analysis. Coastal Eng. 56, 1013–1021.
Van Vledder, G., Goda, Y., Hawkes, P., Mansard, E., Martin, M.J., Mathiesen, M., Peltier, E., Thompson, E., 1994. Case studies of extreme wave analysis: a comparative analysis. Proceedings of the Second Symposium on Ocean Wave Measurement and Analysis, New Orleans, Louisiana, USA, pp. 978–992.