To appear in the *Journal of Hydraulic Research*
Vol. 00, No. 00, Month 20XX, 1–21

Research Paper

On the suitability of the Generalised Pareto to model extreme waves

RUI TEIXEIRA, PhD Researcher, *Department of Civil, Structural & Environmental Engineering, Trinity College, Dublin 2, Ireland*
Email: rteixeir@tcd.ie

MARIA NOGAL, Professor, *Department of Civil, Structural & Environmental Engineering, Trinity College, Dublin 2, Ireland*
Email: nogalm@tcd.ie

ALAN O'CONNOR, Professor, *Department of Civil, Structural & Environmental Engineering, Trinity College, Dublin 2, Ireland*
Email: oconnoa@tcd.ie

Running Head: Extrapolation of significant wave height

On the suitability of the Generalised Pareto to model extreme waves

ABSTRACT

Dealing with extreme events implies working with events that have low probability of occurrence. To characterize these, the Peak-over-threshold method alongside the Generalised Pareto distribution is commonly applied. However, when it comes to significant wave heights, references dissuading this approach are found. In this context, a discussion on the mentioned application is presented based on data collected around the coast of Ireland. A careful choice of threshold takes place, and a new methodology to establish the threshold level is introduced. Five indicators to evaluate the fitting are considered to compare the different statistical models. No evidence was identified to justify the rejection of the Generalised Pareto distribution to model exceedances. Results show that it may be statistically less, equally or more adequate, depending on the Peak-over-threshold implementation. Nevertheless, the Generalised Pareto bounded character is of elementary interest for wave statistics. In some circumstances not considering it might lead to unrealistic significant wave return levels.

Keywords: ocean engineering, hydraulics of renewable energy systems, extremes, statistical theories and models, peak-over-threshold, significant wave height

1 Introduction

The statistical theory of extreme events is a topic of growing interest in all the fields of science and engineering. The changes currently experienced by the world, in all economic and environmental context, have emphasized the importance of dealing with extreme occurrences with improved accuracy.

Under this assumption, the present paper addresses this added need of dealing and modelling extreme events with improved precision, and analyses the methodologies used to statistically characterize the significant wave heights (H_s) in the particular case of offshore engineering.

Characterizing extreme events in a statistical way is commonly conducted through extreme values theory. According to extreme values theory the maxima occurrences from a sample of order statistics with size n coming from a population with a given distribution function follows a known distribution. For the limit case of a sample of independent maxima taken from a population the relations between parent distributions and domains of attractions are fully characterized (Castillo, Hadi, Balakrishnan, and Sarabia, 2005).

However, it is not always of interest to use the limit cases for order statistics. An efficient characterization of the extreme occurrences located in a tail can be obtained by focusing on the analysis of occurrences that exceed a certain threshold value of u. This alternative is of relevance to model extreme occurrences with precision as it may allow the usage of more data points than when using the limit distributions of order statistics.

This is the case when considering significant wave heights (H_s), for which taking into account large values other than the maximum is important to overcome any limitation that may be originated by the high temporal variability naturally subjacent to the ocean waves.

The most common technique to model extremes considering the exceedances over a certain threshold u is the Peak-over-threshold (POT). This technique, which involves truncating independent occurrences that respect the condition $H_s > u$, demands a cautious analysis.

For the case of H_s, application of different statistical distributions can be found in the literature (Castillo, O'Connor, Nogal, and Calvino, 2014; Nogal, Castillo, Calvino, and O'Connor, 2016). One case of particular interest is the preference given in some reference documents to other statistical distributions in relation to the Generalised Pareto (GP) distribution, e.g. DNV (2014b), as the main alternatives to model the exceedances of H_s over a certain threshold u.

For instance, it is appropriate to investigate the reason of discouraging the use of GP with 2 degrees of freedom (DOF), when favouring the application of a GP with 1 DOF (i.e., exponential model), considering that the first is likely to better fit the observed data.

Under this assumption, the current paper proposes then to address the application of the GP to characterize exceedances of H_s over a certain threshold u.

The present paper provides a discussion on the topic under the motivation of setting an unified conclusion about the application of the GP distribution to model H_s exceedances. Therefore, the current work settles the discussion related to the validity of applying the GP distribution to model exceedances of H_s by comparing its application with other widely accepted models such as the 2-parameter Weibull or the Exponential distributions. Records from four oceanographic buoys located in Irish waters were assessed in the study.

To address the proposed challenge five additional sections are defined. Section 2 introduces and identifies the work undertaken in characterizing extreme waves modelling exceedances. Section 3 briefly explains the theoretical background behind modelling the exceedances and evaluating the efficiency of how a statistical model fits a specific set of data. Section 4 and Section 5 present the analysis of the data, the respective results and their discussion. Finally, the most relevant conclusions are compiled in Section 6.

2 POT statistical characterization of H_s

A first approach in reviewing and comparing the different methods to extrapolate wave heights was presented in Muir and El-Shaarawi (1986). Modelling the exceedances of wave heights was addressed and application of the GP distribution to fit the statistical tail of wave data was mentioned as untried and unproven. Although, in this work the GP distribution was not considered as a solution to model independent exceedance over a certain threshold u, modelling exceedance is discussed through the application of compound distributions, such as the Poisson-Rayleigh or the Poisson-Gumbell distributions.

Due to the initial lack of an uniform approach, in 1994 experts in the statistical treatment of wave data were gathered by the International Association for Hydro-Environment Engineering and Research (IAHR) to discuss and standardize the modelling of extreme wave heights. The POT methodology and Weibull distribution were recommended as the reasonable choices for most oceans. Although, it is highlighted in the technical document produced that there is no theoretical argument to indicate a preferred model distribution to fit the observations of maxima. The main findings were published in a recommended practice for extreme wave analysis (Mathiesen, Goda, Hawkes, Mansard, Martín, Peltier, Thompson, and Van Vledder, 1994).

Several standards and practices can also be found to guide the design of offshore structures. While, some of them present generic considerations related to the definition of the extreme H_s occurrences, e.g. emphasizing the need of reliable and robust estimations, other, provide specific recommendations about the techniques to use when modelling extreme H_s. Det Norske Veritas recommended practice on environmental conditions and environmental loads DNV (2014b) previously rejected the application of the GP to model exceedances of H_s. This recommendation, used as reference for offshore standards as DNV (2014a), was recently amended. The application of the GP distribution continues to not be equated at the same level of efficiency of other alternatives, being characterized as of sensitive choice and, little justification is provided to dissuade its use.

Ferreira and Soares (1998) compared the application of the Exponential and the GP distribution to fit H_s POT data and decided on the application of the former.

In opposition to DNV recommendations, Hawkes, Gonzalez-Marco, Sánchez-Arcilla, and Prinos (2008) state that the GP distribution appears to be the optimal regional fit for the extreme wave heights in the North Sea. Regional Frequency Analysis is applied in Van Gelder and Mai (2008) to analyse extreme wave heights in Dutch North Sea Coasts in which five different statistical distribution were evaluated in a fitting process. Wave data are analysed with POT. Results showed that GP distribution fitted better the data analysed when comparing with the other distributions used.

In Méndez, Menéndez, Luceño, and Losada (2006) the POT is used in combination with the GP distribution to evaluate the long term trends in the frequency and intensity of severe storm waves. With the same basis, in Cañellas, Orfila, Méndez, Menéndez, Gómez-Pujol, and Tintoré (2007) the GP distribution is applied to statistically model extreme wave heights in the Balearic Sea.

The motivation that worked as a background for the current paper was to understand the reason behind the identified disparities in application of the GP distribution when the POT methodology is used. As shown, if in some cases the GP distribution is widely applied to model exceedances of H_s in other cases its application is discouraged when compared with other distributions, as the Weibull or the Exponential.

3 Modelling exceedances

The POT methodology assumes that the exceedances over a certain threshold follow a Poisson process. This presumes that the occurrences over a certain u shall follow a purely random process, or in other words, that each occurrence is independent. Extreme H_s over a certain threshold u are very likely to occur in clusters. Therefore, when analysing the data is then mandatory to guarantee that in each event (when the threshold u level is surpassed) the extracted data are independent.

3.1 *Independence of data*

To ensure independence different methodologies are available. The most common techniques consist in setting temporal parameters, usually based on the minimal time lag between two events. In Van Gelder and Mai (2008) a filter of 48 hours is used to extract independent wave data from the POT methodology. A time interval of two to four days is recommended as sufficient to guarantee independent data in Mathiesen et al. (1994). As an alternative, the auto-correlation function can be used for a more meticulous evaluation of the independence between observations. In a review of the framework for dealing with environmental extremes, Bernardara, Mazas, Kergadallan, and Hamm (2014) refer that for the North sea a time lag of 24 hours is sufficient to guarantee independence between storm events. It is also stated that a time lag higher than 24 hours may lead to loss of information, which should be avoided when modelling extreme occurrences. Additional examples of application to extreme H_s are presented in the same work. Van Gelder (2000) uses a value of 48 hours to model exceedances in several North sea locations. It is important to highlight that some authors define the time lag as the minimum time between two storm events, while others define it as the minimum time between two storm peaks. In the present work a time period of 48 hours between two storm events is used to decluster the data. For the proposed goal of fitting a statistical model to POT data, the value of 48 hours ensures that a sufficient margin over the 24 hours is used so that the number of observations above the threshold are independent and the erroneous loss of data is not significant. To confirm the adequacy of the 48 hours to guarantee independence of data, the average values of threshold were compared with the minimum average of 12 hours in between storms. When the decrease in average H_s for the period of 12 hours was significant, the storm was considered finished. As the threshold was increased for further studies, independence was ensured even if two peaks occurred separated by more than 48 hours within the same storm using *a priori* knowledge on the storm duration. If less than 48 hours exist between two events, then they are considered the same storm. A sensitivity analysis of the effect of the decluster time is presented later.

3.2 *Statistical models*

The Weibull distribution is a statistical distribution of wide applicability Weibull (1951). In the field of classical extreme value theory it is also one of the specific cases of the Generalised Extreme Value distribution for minima. Its continuous CDF formula is given by

$$F(H_s, u, \sigma, \zeta) = 1 - \exp\left[-\left(\frac{H_s - u}{\sigma}\right)^{\zeta}\right], \tag{1}$$

where ζ and σ are the shape and scale parameters. For the current case where the distribution is truncated and only the independent events are taken into account, the three parameter Weibull distribution is reduced to a two parameter Weibull distribution by setting $H_s - u$ as $H_{s(>u)}$, i.e. the exceedances obtained from the POT methodology.

The GP distribution was presented initially to model the conditional probability of high order statistics, in other words, probability of an observation being greater than x, given the condition that $x \geq u$ Pickands III (1975). Usually it appears as a combined Poisson model, where the exceedance events are assumed to follow a Poisson distribution with expected number of occurrences λ. Its continuous CDF appears in two main forms, depending on the shape parameter ζ.

$$F(H_{s(>u)}, \sigma, \zeta) = \begin{cases} 1 - \left(1 + \frac{\zeta H_{s(>u)}}{\sigma}\right)^{-\frac{1}{\zeta}}, & \zeta \neq 0, \\ 1 - \exp\left(-\frac{H_{s(>u)}}{\sigma}\right), & \zeta = 0, \end{cases} \tag{2}$$

again, ζ and σ are the shape and scale parameters of the function and $H_{s(>u)}$ is the exceedances of H_s over u. When $\zeta = 0$ the GP distribution takes the Exponential distribution form.

The set of statistical distributions presented depends on unknown parameters which need to be estimated from the data.

The fitting evaluating indicators proposed in Section 3.3 provide adequate evaluation of the goodness of fit of the distributions Sheskin (2003). Nevertheless, the estimation of the model parameters of a given dataset and the validation of the fitting, can be also attained by splitting the dataset into two subsets, one for fitting the parameters, and other for validation. Using a cross-validation methodology could be also required, mainly for small sample sizes associated with high values of the threshold.

Different methodologies exist for estimating the unknown parameters of the distributions from a sample of data, thus, different parameter estimating techniques may be applied to determine the statistical models. Several review papers can be found addressing the topic of estimating the parameters of the cited distributions. An extensive discussion of the fitting methodologies is presented in Castillo et al. (2005). Some of these methodologies are extensively discussed for the case of GP distribution in Castillo and Hadi (1997).

Due to the nature of the problem studied, estimation of the statistical model parameters has a key role in the analysis. For comparative purposes the maximum likelihood estimation (MLE) technique is applied to find the unknown parameters of all the Weibull, GP and Exponential distributions. Sánchez-Arcilla, Gomez Aguar, Egozcue, Ortego, Galiatsatou, and Prinos (2008) compares the repercussion on the extrapolation procedure caused by estimation techniques, in the case MLE and a Bayesian approach. No major divergences were found in the return levels estimated.

It is necessary to highlight that the parameter estimates are of major importance when checking a statistical model adequacy for physical modelling.This is the case of unbounded distributions in the right tail, which may not be adequate to model naturally bounded events as H_s. Ortego, Tolosana-Delgado, Gibergans-Báguena, Egozcue, and Sánchez-Arcilla (2012) highlight the importance of modelling waves as a bounded event by using a GP distribution model that is constrained to be upper bounded.

3.3 *Evaluation of the fitting*

The process of evaluating the fitting of a statistical model to a set of data is still a subject of discussion. There are no standards or guidelines that establish in an unquestionable way how this process should be undertaken. Evaluating the quality of the fitting usually compares the real distribution against the fitted theoretical distribution. There are several tests to evaluate the confidence of a statistical fitting frequently denominated as Goodness-of-fit (GoF) tests.

A GoF test consists in evaluating the hypotheses H_0 and H_a that a certain sample of data may come or not from a parent known distribution. Under this assumption the hypothesis may or not be rejected with a specified level of confidence p.

The present work applies therefore five methodologies, three GoF tests and two comparative analyses, to evaluate the fit of the statistical model to the available data. The wide number of comparative measures considered ensures robust results when comparing the fitting of different statistical models. To evaluate the statistical fit the most common tests are the Chi-Squared (χ^2) GoF test and the Kolmogorov-Smirnov (KS) test. The application of the first is limited for the current analysis as in the tail occurrences are scarce and the conditions of the test cannot be fulfilled.

The KS GoF test, in an opposite way to the χ^2 GoF is mainly used to compare continuous data sets. It is based in the evaluation of the i and $i-1$ order statistic of the empirical CDF with the respective i value given by the estimated statistical model. For a double-sided test the KS follows (3).

$$\begin{gathered} KS_{stat} = \max(D^+, D^-), \\ D^+ = |F_{ecdf}(x_i) - F(x_i, \theta)| \quad \text{and} \quad D^- = |F_{ecdf}(x_{i-1}) - F(x_i, \theta)|, \end{gathered} \tag{3}$$

where $F_{ecdf}(x_i)$ is the empirical cumulative distribution and $F(x_i, \theta)$ the statistical cumulative distribution used to describe the empirical data. The terms D^+ and D^- are the differences that represent the test statistics and are calculated from the comparison of the two mentioned models in the i and $i-1$ order statistics.

The KS statistic overestimates the quality of the fitting when the unknown parameters are estimated from the set of data used to evaluate the fitting. Stephens (1970) and Lilliefors (1969) tackle this limitation presenting two distinct methodologies.

The discrepancy between hypothetical distribution and its empirical form can be evaluated also by application of quadratic statistics. Cramér von-Mises (W) GoF test is representative of a test that compares the empirical and the hypothetical distributions by applying quadratic statistics.

$$W = \int_{-\infty}^{\infty} (F_{ecdf}(x) - F(x, \theta))^2 \psi(x) dx, \tag{4}$$

where the $\psi(x)$ is a weight function that in case of the the Cramér von-Mises is equal to 1. A variant of the Cramér von-Mises GoF is the Anderson-Darling (A-D) GoF test, which is given by manipulating the function $\psi(x)$ and forcing it to be more sensible to variations in the tail of the distribution Stephens (1979). Extensive work on these test statistics and the respective asymptotic values of the test statistic depending on the size of the sample is given in the following works. In Stephens (1976) asymptotic results for the exponential distribution with unknown parameters are presented. In Lockhart, O'Reilly, and Stephens (1985) the 3 parameter Weibull distribution is evaluated and the asymptotic values of the test statistic defined. In Choulakian and Stephens

(2001) the limit results for the test statistic are presented for the GP distribution with unknown parameters.

Probability-Probability (P-P) and Quantile-Quantile (Q-Q) plots can be applied to analyse the fitting. Several examples of analysis of fitting with P-P and Q-Q plots are found in Castillo and Hadi (1997).

Measuring the fitting adequacy with quadratic differences is not exclusive to the W GoF test. The root mean square error (RMSE) also evaluates the quadratic differences and evaluate the fitting adequacy by a measure of the dispersion of the results. It is important to highlight that the last two indicators of the quality of the fit presented are not under a hypothesis testing methodology.

4 Wave data

The wave data used in the following study was collected in four locations around the Irish coast by Met Éireann in different periods of time during an interval of 14 years. The approximate and exact(Degrees, Minutes and Seconds) location of the buoys M1, M4, M5 and M6 can be found in Figure 1 and Table 1. The records of wave data start in the year of 2000 and end in the year of

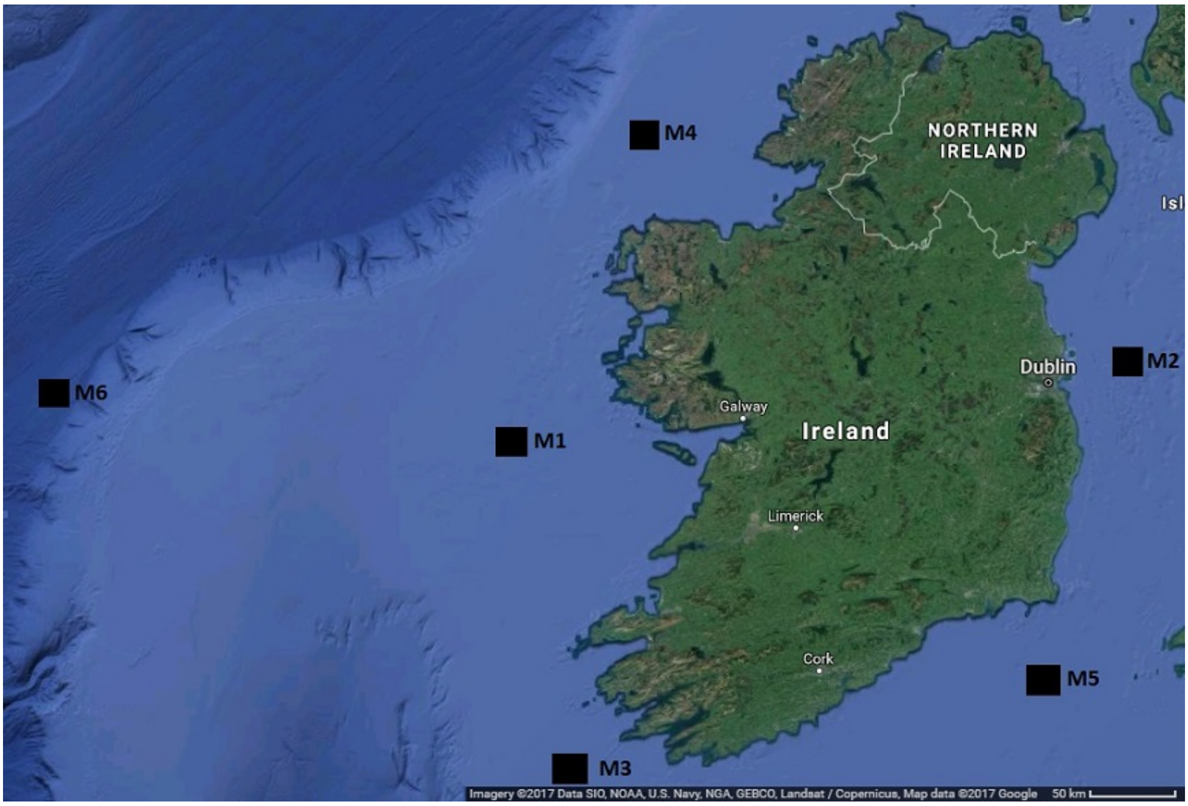

Figure 1 Map with the approximate location of the Met-Éireann oceanographic buoys.Only data from buoys M1, M4, M5 and M6 were used in the analysis.(*Adapted from Google Maps.*)

2015. The periods of time and the availability of the records in the time frame are also presented in the Table 1. The data was collected in hourly averages and registered at HH[1].

Buoy M6 is representative of Atlantic West deep waters, while the buoy M1 and M4 is likely to represent a similar wave climate but in shallower waters. M4 is located north of M1 and M6, where the wave climate is more energetic. Buoy M5 is representative of the wave climate of the Irish

[1]Averaged records from 15 minutes before the hour; e.g. 01:00 PM record will correspond to the previous 15 minutes average.

Table 1 Met Éireann buoys location, periods of operation and availability.

Buoy	Degree Minutes Seconds (DMS)	General Location	Period of Activity	Availability	Sample Size (no. of points)
M1	53° 07' 36" N 11° 12' 00" W	Off the Galway coast	2000-2007	89%	47065
M4	55° 00' 00" N 10° 00' 00" W	Off the Donegal coast	2007-2015	69%	51343
M5	51° 41' 24" N 06° 42' 16" W	Off the south Wexford coast	2004-2015	76%	81667
M6	52° 59' 09" N 15° 52' 00" W	Deep Atlantic	2006-2015	78%	58187

Sea, being therefore representative of less energetic wave states. In all the cases the more energetic events are mainly confined to two periods of the year, January-April and October-December.

Modelling measured wave data requires a careful analysis due to its statistical and physical nature. (Sánchez-Arcilla et al., 2008) and (Vanem, Huseby, and Natvig, 2012) model H_s using a logarithm transformation. The logarithm transformation removes the heteroscedasticity common to H_s data and is expected to improve the trend of analysis for the extreme sample.

In addition to the modelling of H_s using the different statistical models presented it is then of interest to model wave data using a logarithmic transformation. Results from modelling wave data using a logarithm transformation and a GP distribution using the criteria presented in (Vanem et al., 2012) are presented to complement the comparison of the different statistical models.

5 GP as an alternative to model POT exceedances

When dealing with exceedances in the POT, the first step is to analyse the value of the threshold u that should be used to truncate the data.

5.1 *Choice of threshold u*

Setting an optimal threshold of a POT methodology is a very complex and difficult task (Embrechts, Resnick, and Samorodnitsky, 1999). Several methodologies and rules of thumb exist to this end.

More than one approach can be used to select the threshold level. It can depend on physical considerations, such as setting the H_s level that can be considered as a threat, or based on straight mathematical considerations (Lang, Ouarda, and Bobée, 1999). An extensive discussion of the choice of the threshold to model exceedances over a specified u is presented with specific application to environmental data and with specific application to the GP distribution in Bommier (2014).

In Scarrott and MacDonald (2012) the process of choice of threshold is reviewed, with reference to the multiple threshold approaches and automated threshold methodologies.

In Sánchez-Arcilla et al. (2008) a careful selection of the threshold is proposed. Implementing a Bayesian approach, it is shown that uncertainty can be reduced and support to the management of coastal infrastructures improved.

Two of the most common graphical methodologies to choose the threshold value are; the parameter stability plots and the mean residual life (MRL) plots. The first, applied to the GP distribution, is used to choose the threshold in combination with the $\boldsymbol{\theta}$ vector of parameters. The threshold is then picked as a function of the variation of $\boldsymbol{\theta}$ with u. The results of the choice of threshold based on the stability of $\boldsymbol{\theta}$ are shown in Figure 2(a), with the respective 95% confidence intervals for the estimation. Due to the decreased number of data extracted as the value of u increases the stability plots became unstable for high values of u, and the uncertainty in the estimation rises significantly. The same occurs when the value of u approaches 0.

As a graphical alternative methodology the MRL plot, Figure 2(b), can be used to set the threshold value (Davison and Smith, 1990). It consists in picking the value of u which represents the lowest level where all the higher threshold based sample mean excesses are consistent with a straight line (Scarrott and MacDonald, 2012). Despite their wide application, graphical methodologies can

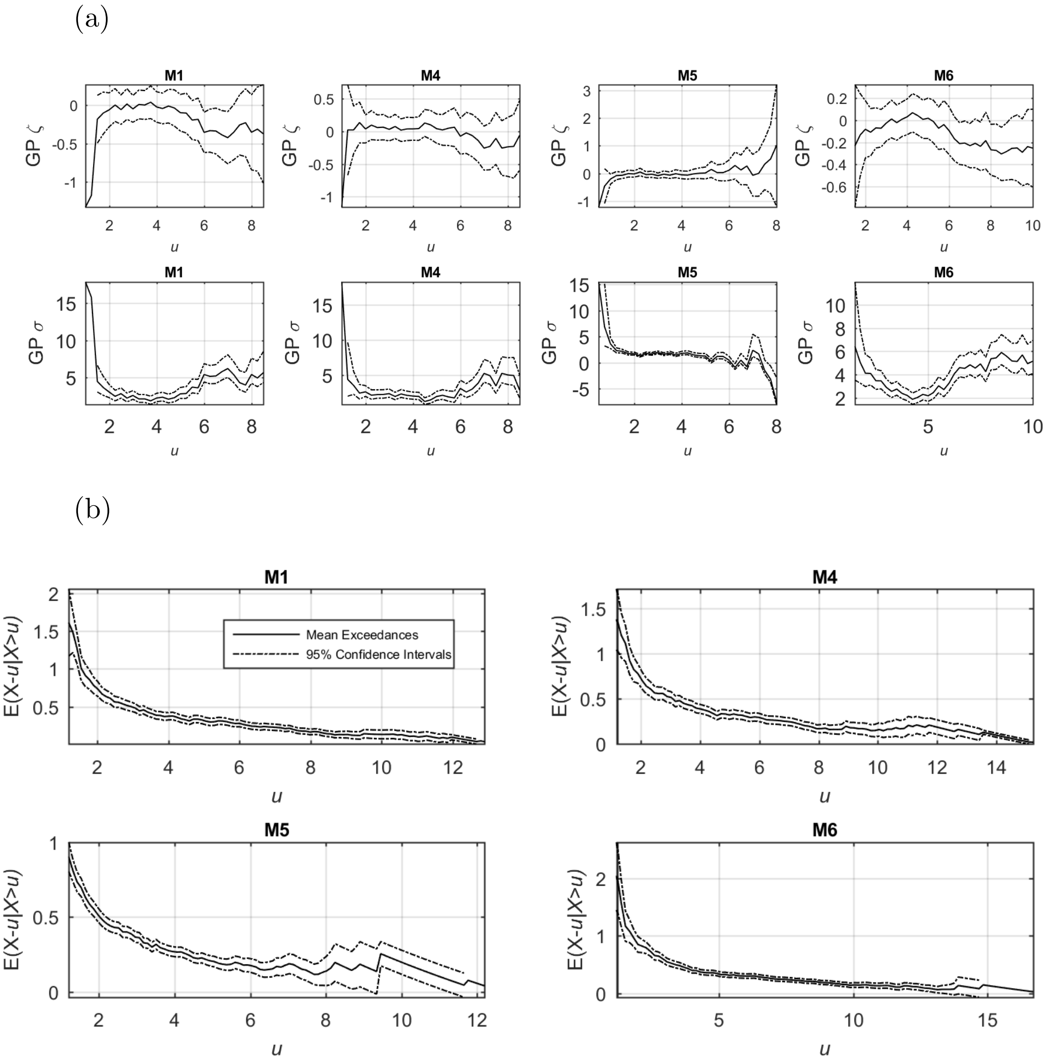

Figure 2 Choice of threshold results for the parameter stability of the GP distribution (a) and the MRL plot (b).

be challenging to interpret. Comparing the previous graphical representations it is possible to infer that different plateaus for u can be obtained. For instance, if for the MRL plot this choice falls into a multiple threshold category with minimum value between 3 and 4; for the stability of the parameters the parameter u is almost impossible to set without high uncertainty.

A very common and widely applied rule of thumb when choosing the threshold is considering only the occurrence over a certain order statistic, e.g. the 90% quantile. The 10% upper tail increases the threshold value to 5 in M1, M4, M6 and to 3.5 in the case of M5. Although of immediate implementation, this methodology was proven theoretically non-adequate (Scarrott and MacDonald, 2012).

5.2 *Selection of threshold using the derivatives of the empirical PDF*

Due to the lack of uniformity in the methods identified a new methodology is introduced here to calculate the value of u. For it, we assume that a tail of a distribution starts in the point where the curvature of the PDF reaches its maximum positive value after the peak of the PDF distribution is attained. At a certain point there will be an inflection in the curvature of the PDF due to the approach of the tail and due to the fact we are moving away from the "bulk" of the data. Here the maximum value of the curvature, after the inflection point, is picked.

The idea that supports this approach is not only the need to assure the negative slope required for the tail, but also to guarantee that these points follow the same curvature trend as the distribution used to fit the POT data set. The calculation is then performed through the combination of the empirical cumulative function and a finite differences scheme to calculate the higher order derivatives. Results are shown in Figure 3. The values of u for the different buoys are pointed in

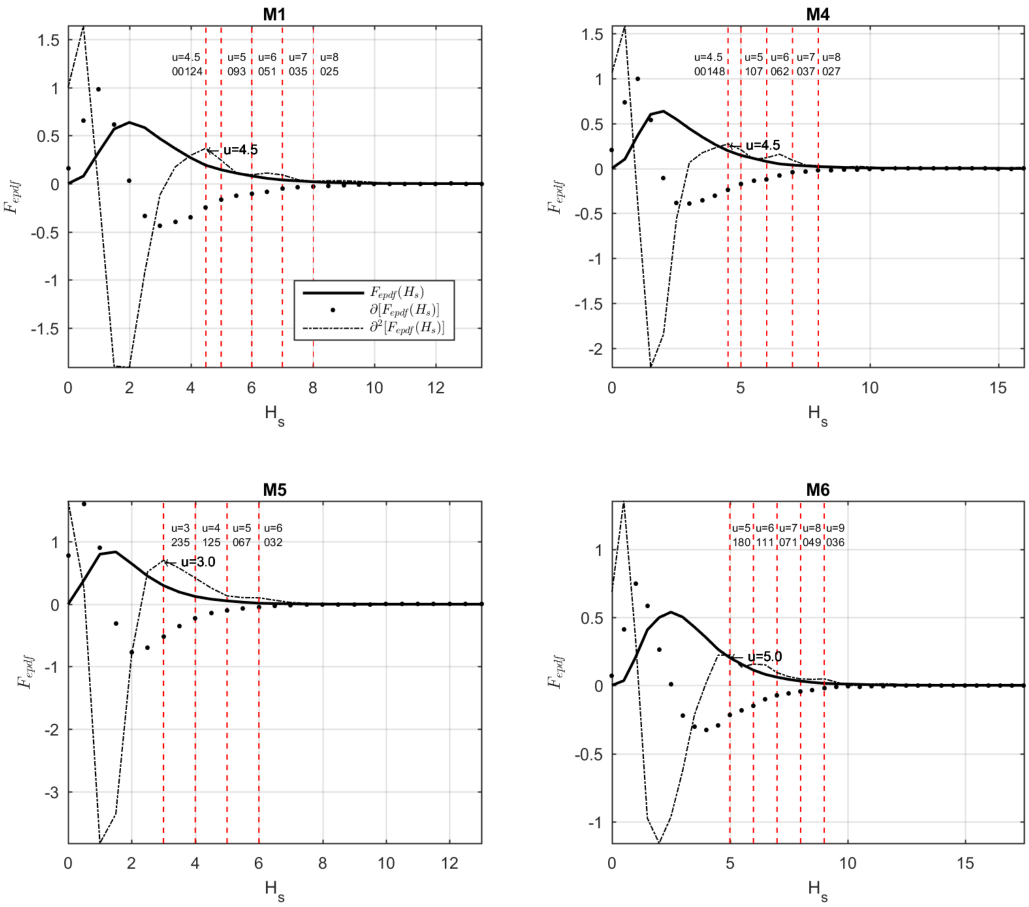

Figure 3 Calculation of u for the four studied buoys based on the analysis of the second derivative ($\delta^2[\mathrm{eF_{PDF}(H_s)}]$) of the empirical density function PDF.

the graphic by the maximum of the second order derivative that correspond to the beginning of the tail region. The u levels obtained show good agreement with other methodologies analysed. In addition, the direct usage of the empirical PDF function ensures an explicit analysis of the tail region, ensuring a better comprehension of the process and reducing the risk of erroneously selecting a u level that is inappropriate for the shape of the empirical PDF function tail. When analysing other graphical measures studied it is possible to infer that a significant advantage is to get a quantified u value with minimum bias.

For a more understandable analysis of the following results, the different levels of threshold considered in the present study of GoF and the respective number of points after truncation by the POT methodology are represented by the trimmed (red) lines.

The following subsection presents the central research of the work undertaken and investigates to which extent the GP distribution is less appropriate than other commonly applied distributions, the two parameter Weibull and the Exponential, to be used when a POT methodology is applied to H_s data.

5.3 *Goodness of Fit results*

The minimum value of u for analysis of the statistical fitting was set to be variable depending on the analysis of Section 5.2. In addition, more robust results can be achieved by using a wide range of thresholds in the analysis. Table 2 and Table 3 present the results for the statistical fitting parameters for each case and the results for the three GoF and the two fitting indicators considered. The KS GoF gives a measure of the biggest deviation in the fit, being very useful to detect outliers. Considering that it aims to identify the biggest difference between the real and the estimated curves, in the present case its application is well complemented by the tests that work with mean values and weighted averages of the error, which is the case of the remaining tests applied. An increased value of the test statistic is representative of a smaller $p - value$, and consequently a worse fit to the data.

Table 2 Statistical parameters obtained from the fitting proccess for the different distributions and levels u of truncation.

$u(m)$	**Weibull**		**Exponential**	**GenPareto**		**log() GenPareto**	
	σ	ζ	σ	ζ	σ	ζ	σ
			Buoy M1				
4.5	2.009	0.987	2.021	-0.009	2.039	-0.322	0.442
5	2.149	1.04	2.115	-0.096	2.32	-0.381	0.446
5.5	2.273	1.077	2.209	-0.192	2.646	-0.453	0.455
6	2.599	1.277	2.411	-0.353	3.28	-0.597	0.508
6.5	2.369	1.179	2.247	-0.336	3.024	-0.566	0.442
7	2.498	1.437	2.261	-0.423	3.242	-0.651	0.445
7.5	2.003	1.191	1.883	-0.256	2.385	-0.463	0.315
8	2.021	1.245	1.878	-0.352	2.585	-0.507	0.326
			Buoy M4				
4.5	1.963	0.93	2.033	0.13	1.774	-0.19	0.39
5	2.193	0.975	2.218	0.043	2.122	-0.257	0.412
5.5	2.202	0.972	2.23	0.05	2.12	-0.245	0.38
6	2.519	1.072	2.449	-0.072	2.627	-0.349	0.422
6.5	2.398	0.988	2.41	-0.063	2.563	-0.336	0.387
7	2.918	1.242	2.717	-0.256	3.426	-0.527	0.472
7.5	2.379	1.016	2.363	-0.106	2.619	-0.361	0.348
8	2.583	1.109	2.492	-0.236	3.103	-0.503	0.39
			Buoy M5				
3	1.683	1.188	1.581	-0.082	1.709	-0.314	0.496
3.5	1.61	1.161	1.523	-0.065	1.621	-0.28	0.417
4	1.547	1.1	1.492	-0.054	1.571	-0.255	0.362
4.5	1.465	1.129	1.397	-0.019	1.423	-0.206	0.299
5	1.419	1.117	1.356	0.014	1.336	-0.166	0.257
5.5	1.437	1.051	1.407	0.027	1.369	-0.154	0.242
6	1.323	1.018	1.311	0.115	1.161	-0.059	0.192
6.5	1.456	0.985	1.466	0.123	1.288	-0.07	0.199
7	1.856	1.107	1.782	-0.069	1.907	-0.314	0.28
7.5	1.375	0.908	1.45	0.331	1.011	0.138	0.14
			Buoy M6				
5	2.297	0.976	2.322	0.019	2.277	-0.295	0.442
5.5	2.504	1.038	2.466	-0.072	2.645	-0.35	0.453
6	2.457	0.984	2.474	-0.094	2.71	-0.356	0.427
6.5	2.784	1.138	2.66	-0.211	3.229	-0.428	0.45
7	2.777	1.178	2.624	-0.23	3.232	-0.432	0.421
7.5	2.7	1.181	2.549	-0.236	3.155	-0.426	0.387
8	2.692	1.189	2.542	-0.268	3.223	-0.445	0.372
9	2.433	1.185	2.297	-0.254	2.885	-0.412	0.303

The KS GoF for a single sample was initially developed for the case of a completely specified $\boldsymbol{\theta}$. The need for a correction of the KS statistic is justified by the fact that since the $\boldsymbol{\theta}$ vector was obtained from the sample it is very likely for the estimated model to approximate better the real distribution when compared with the case where $\boldsymbol{\theta}$ is not estimated from the data. The

Table 3 GoF results for the three test statistic K-S, W and A-D and the two GoF indicators studied for the Weibull, Exponential and GP Distribution. The p-values in % are given for the test statistics performed. Underlined values correspond to the appropriate level u as specified in Section 5.1. Results for the log transformation are shown for the GP distribution between brackets.

$u(m)$	3	3.5	4	4.5	5	5.5	6	6.5	7	7.5	8	9
						Weibull Distribution						
						M1						
$K-S(p\%)$	-	-	-	67.8	58.0	55.5	72.8	57.8	85.3	98.4	90.7	-
$W(p\%)$	-	-	-	58.2	46.3	46.2	65.7	62.4	57.3	65.2	49.1	-
$A-D(p\%)$	-	-	-	11.1	13.9	14.7	44.0	21.0	38.9	47.6	40.1	-
$P-P$ (°)	-	-	-	44.99	45.01	45.13	45.06	45.2	44.64	44.93	44.91	-
$RMSE$	-	-	-	0.027	0.034	0.038	0.031	0.038	0.037	0.032	0.047	-
						M4						
$K-S(p\%)$	-	-	-	79.8	94.8	90.2	75.0	87.2	96.5	91.8	87.3	-
$W(p\%)$	-	-	-	56.7	62.1	54.1	49.7	54.9	59.7	62.0	55.9	-
$A-D(p\%)$	-	-	-	13.0	35.7	28.5	28.7	32.1	46.5	39.7	33.4	-
$P-P$ (°)	-	-	-	44.83	44.86	44.91	44.88	45.2	44.64	45.08	44.82	-
MSE	-	-	-	0.022	0.022	0.027	0.032	0.033	0.030	0.035	0.039	-
						M5						
$K-S(p\%)$	17.3	13.7	56.4	46.3	27.2	75.4	77.7	72.4	61.5	66.4	-	-
$W(p\%)$	62.5	60.7	66.1	58.7	34.8	37.3	22.0	35.7	14.2	11.4	-	-
$A-D(p\%)$	< 1.0	1.6	4.7	12.5	< 1.0	9.2	8.1	18.4	10.6	9.3	-	-
$P-P$ (°)	44.69	44.66	44.69	44.38	44.11	44.15	43.68	43.82	43.77	43.72	-	-
$RMSE$	0.024	0.028	0.028	0.029	0.046	0.045	0.057	0.064	0.010	0.104	-	-
						M6						
$K-S(p\%)$	-	-	-	-	56.9	46.1	70.9	96.4	92.3	87.9	85.0	77.4
$W(p\%)$	-	-	-	-	52.2	44.8	55.0	76.7	68.3	59.5	71.3	61.2
$A-D(p\%)$	-	-	-	-	11.6	10.5	10.7	61.6	51.9	37.2	50.5	45.2
$P-P$ (°)	-	-	-	-	45.02	45.03	45.33	45.15	45.1	45.12	45.26	45.26
$RMSE$	-	-	-	-	0.023	0.027	0.025	0.017	0.023	0.029	0.026	0.035
						Exponential Distribution						
						M1						
$K-S(p\%)$	-	-	-	52.3	48.3	29.6	7.1	7.4	2.9	86.2	70.1	-
$W(p\%)$	-	-	-	65.6	70.9	70.8	47.9	53.6	35.0	82	68.9	-
$A-D(p\%)$	-	-	-	26.1	40.3	39.5	15.2	22.3	3.3	48.9	42.3	-
$P-P$ (°)	-	-	-	44.97	45.07	45.25	45.41	45.43	45.29	45.34	45.49	-
$RMSE$	-	-	-	0.028	0.029	0.034	0.056	0.058	0.081	0.035	0.046	-
						M4						
$K-S(p\%)$	-	-	-	21.7	81.4	71.2	86.9	81.4	32.6	80.8	51.6	-
$W(p\%)$	-	-	-	34.3	65.3	56.8	70.9	66.2	40.7	72.2	57.3	-
$A-D(p\%)$	-	-	-	6.4	45.3	37.3	50.4	47.1	26.1	56.1	39.5	-
$P-P$ (°)	-	-	-	44.61	44.79	44.82	45.08	45.17	45.32	45.14	45.17	-
$RMSE$	-	-	-	0.036	0.025	0.032	0.027	0.034	0.053	0.035	0.050	-
						M5						
$K-S(p\%)$	< 1.0	6.5	6.2	28.6	12.2	50.2	63.1	66.5	33.1	40.1	-	-
$W(p\%)$	73.2	72.5	64.3	70.1	58.6	53.0	43.6	49.7	40.7	18.9	-	-
$A-D(p\%)$	< 1.0	1.4	7.9	10	7.1	25.4	26.9	35.9	30.3	18.0	-	-
$P-P$ (°)	44.82	44.81	44.8	44.6	44.37	44.31	43.76	43.74	44.29	43.1	-	-
$RMSE$	0.037	0.036	0.040	0.039	0.049	0.049	0.055	0.065	0.096	0.121	-	-
						M6						
$K-S(p\%)$	-	-	-	-	29.5	56.7	53.7	79.7	76.4	69.2	55.6	55.7
$W(p\%)$	-	-	-	-	54.9	71.5	64.8	67.1	61.9	62.1	53.2	69.0
$A-D(p\%)$	-	-	-	-	20.4	37.7	25.5	44.3	36.9	37.7	39.0	47.1
$P-P$ (°)	-	-	-	-	44.98	45.12	45.29	45.47	45.51	45.53	45.64	45.53
$RMSE$	-	-	-	-	0.027	0.022	0.027	0.027	0.031	0.032	0.041	0.040
						Generalised Pareto Distribution						
						M1						
$K-S(p\%)$	-	-	-	50.6 (53.9)	36.1 (40.3)	53.0 (60.5)	68.2 (62.7)	60.9 (56.6)	28.8 (25.8)	88.3 (81.4)	70.3 (62.0)	-
$W(p\%)$	-	-	-	54.5 (51.3)	43.8 (45.2)	45.0 (49.0)	67.8 (60.3)	61.9 (55.5)	37.1 (25.7)	61.9 (55.0)	41.6 (30.0)	-
$A-D(p\%)$	-	-	-	13.7 (12.1)	18.3 (20.1)	22.5 (26.1)	51.2 (39.6)	38.7 (32.4)	9.5 (2.8)	46.6 (39.9)	34.0 (23.9)	-
$P-P$ (°)	-	-	-	44.95 (44.83)	44.86 (44.75)	44.82 (44.7)	44.61 (44.46)	44.63 (44.45)	44.21 (43.95)	44.62 (44.42)	44.42 (44.09)	-
$RMSE$	-	-	-	0.028 (0.029)	0.035 (0.033)	0.037 (0.033)	0.028 (0.028)	0.034 (0.035)	0.053 (0.051)	0.035 (0.039)	0.051 (0.057)	-
						M4						
$K-S(p\%)$	-	-	-	67.0 (67.8)	93.6 (90.3)	84.3 (82.7)	77 (81.0)	71.7 (73.2)	63.0 (60.1)	88.5 (88.4)	81.3 (83.6)	-
$W(p\%)$	-	-	-	64.0 (60.5)	66.9 (65.0)	58.2 (55.4)	58.1 (55.8)	54.3 (51.8)	42.7 (29.8)	60.2 (53.2)	39.5 (24.8)	-
$A-D(p\%)$	-	-	-	28.4 (23.7)	48.2 (45.2)	40.5 (37.0)	40.9 (36.8)	33.3 (30.6)	34.2 (22.9)	44.1 (37.7)	31.7 (21.0)	-
$P-P$ (°)	-	-	-	45.01 (44.91)	44.92 (44.83)	44.98 (44.86)	44.85 (44.73)	44.96 (44.81)	44.43 (44.19)	44.76 (44.55)	44.3 (43.96)	-
$RMSE$	-	-	-	0.021 (0.021)	0.021 (0.021)	0.027 (0.027)	0.029 (0.028)	0.035 (0.034)	0.038 (0.044)	0.035 (0.037)	0.047 (0.054)	-
						M5						
$K-S(p\%)$	1.7 (< 1.0)	12.6 (4.2)	12.5 (5.0)	31.7 (22.8)	11 (8.8)	43.9 (40.2)	45.6 (44.7)	50.8 (51.8)	40.3 (44.7)	76.2 (75.2)	-	-
$W(p\%)$	74.0 (52.8)	70.2 (56.7)	61.1 (47.5)	62.7 (55.1)	49.0 (44.4)	42.2 (37.2)	43.8 (41.0)	56.3 (52.7)	23.4 (15.7)	41.5 (35.7)	-	-
$A-D(p\%)$	27.8 (< 1.0)	3.6 (58.3)	3.8 (1.7)	2.2 (< 1.0)	< 1.0 (2.3)	14.4 (9.7)	20.9 (17.4)	35.7 (32.2)	19.1 (14.0)	35.5 (30.6)	-	-
$P-P$ (°)	44.70 (44.69)	44.69 (44.70)	44.69 (44.69)	44.55 (44.55)	44.42 (44.42)	44.41 (44.42)	44.23 (44.22)	44.29 (44.24)	44.01 (43.76)	44.47 (44.39)	-	-
$RMSE$	0.029 (0.027)	0.030 (0.028)	0.034 (0.033)	0.037 (0.033)	0.049 (0.042)	0.051 (0.052)	0.055 (0.055)	0.055 (0.056)	0.098 (0.097)	0.086 (0.090)	-	-
						M6						
$K-S(p\%)$	-	-	-	-	38.5 (38.5)	33.0 (47.1)	39.8 (47.4)	98.3 (95.6)	97.9 (98.3)	96.2 (97.5)	95.9 (96.3)	86.8 (85.8)
$W(p\%)$	-	-	-	-	50.1 (49.7)	47.4 (56.4)	43.8 (51.5)	78.5 (81.3)	75.0 (76.9)	70.6 (73.6)	77.9 (75.4)	70.1 (71.9)
$A-D(p\%)$	-	-	-	-	15.2 (15.3)	16.8 (22.9)	3.9 (8.7)	73.0 (70.1)	64.2 (59.7)	55.3 (54.3)	71.2 (67.1)	62.2 (61.5)
$P-P$ (°)	-	-	-	-	45.03 (44.91)	44.92 (44.87)	45.02 (44.98)	44.88 (44.96)	44.88 (44.95)	44.91 (44.96)	44.98 (45.02)	44.96 (44.94)
$RMSE$	-	-	-	-	0.025 (0.024)	0.028 (0.023)	0.031 (0.026)	0.016 (0.014)	0.019 (0.017)	0.022 (0.020)	0.020 (0.020)	0.028 (0.026)

Lilliefors test $p - values$ are then expected to decrease for the same asymptotic values of the test statistic. For the current implementation it is expected that the results of the KS will hold relevant comparative results of GoF. Although it is important to understand that for practical applications of not rejecting any of the statistical models as valid for representation of the empirical distribution, the GoF results need to consider the estimation of $\boldsymbol{\theta}$.

The $p - value$ was also used as a reference to compare the GoF results for the W and A-D statistics. Extensive tables that evaluate the W and the A-D GoF are available in the literature.

The calculation of the $p - value$ for each test statistics involves interpolating and extrapolating techniques. A 3^{rd} degree polynomial function was fitted to the asymptotic points to calculate the $p - values$, as their variation with the test statistic is clearly non-linear. The variation of the confidence of the fit with the test statistic for the considered distributions can be found in Stephens (1976,7,7) and Spineili and Stephens (1987).

The 2-parameter Weibull distribution, when the location parameter is known, needs to be tested against a type I GEV distribution (Lockhart and Stephens, 1993). A transformation of variable needs then to be applied to the exceedance values to test against a type I GEV.

To allow a quantified comparison of the P-P plots, the slope of the curves was used as a measurement of the GoF. Ideally, the fit should generate a curve with a slope of 45(°). Forcing the curve to go through the origin and calculating its deviation when compared to the ideal slope is applied in the current work as a measure of the fit. With the $F(x, \boldsymbol{\theta})$ in the x axis; if the deviation is positive, the fitted model is underestimating $F_{ecdf}(x)$; if the deviation is negative, the fitted model is overestimating $F_{ecdf}(x)$.

It is interesting to notice that for the buoy M4 in the reference threshold the GP distribution produces better fitting results in four of the five indicators used for comparison. Nevertheless, the GP distribution is unbounded for this case as $\zeta > 0$. For M5, the same happens in two of the five indicators. In the case of buoys M1 and M6, the two-parameter Weibull and the Exponential distributions achieved, respectively, better fitting results in three and two of the five indicators used for comparison.

One of the concerns highlighted when using the Generalised Pareto to model physically limited events, is to ensure that the distribution is bounded. As mentioned, this happens when the shape (ζ) parameter is smaller than 0. In the present case, no significant improvement of the fitting indicators occurs when a logarithmic transformation is applied. Still, one of the interesting points of using a logarithmic transformation is that it ensures the ζ parameter to be negative, thus bounded for the Generalised Pareto. As example, GP fits for non transformed data with unbounded domain occur for values of u close to the reference value in the case of the buoys M4 and M6. This should be avoided on a design basis approach.

It is important to notice that the Weibull and Exponential distributions have support in the interval $[0,+\infty)$. Therefore, due to the physical character of the waves, even considering the following comparison of GoF the GP should be applied primarily when its right tail is bounded.

As regard of the other threshold values considered, cases where the GP distribution presents better indicators than the other two statistical distributions can be found, e.g. all the cases above the threshold of 6.5m meters in the buoy M6, where the GP distribution is bounded.

Nonetheless, relevant levels of u where the GP has no better value of fitting in the five indicators used can be identified; instances are M1 with $u = 4.5$ or M6 $u = 5$, among other examples.

Under the discussed fluctuation in the results it is then of additional interest to understand, in order to identify any limitation from the use of the GP distribution as a statistical model to fit the exceedances of H_s, how the results vary with different decoupling times in the application of the POT.

5.4 *Influence of the decoupling time*

The analysis of the decoupling process that guarantees independent data is of major importance to ensure the validity of the study and the robustness of the results obtained. It was seen that the methodology to ensure independence involved the application of a minimum period of time between storm events.

One GoF test and a fitting indicator were considered here for studying the influence of the decoupling time in the GoF results, i.e., Kolmogorov-Smirnov and the RMSE. The choice of these two test statistics is justified by the two quantities evaluated in the tests, while the KS test evaluates the maximum deviation from the real function and calculates the confidence of the fit based on that value, the RMSE considers an averaged value of deviation from the real distribution, taking into account all the points of the CDF. Figures 4 and 5 show, respectively, the results for the variation of the GoF with the variation of the decoupling time.

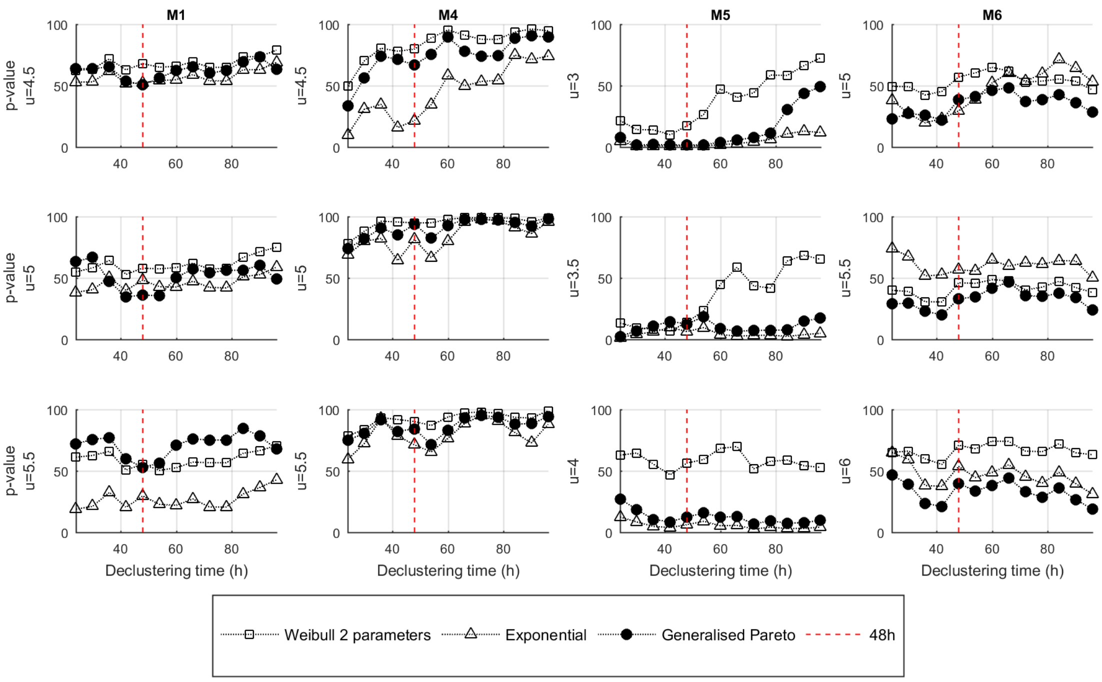

Figure 4 Influence of the decoupling time in the KS results. The trimmed (red) line represents the reference value of 48 hours.

The results show similar trends for all the statistical distributions. In M1 the $p - values$ are stable to the point that the acceptance of the fitting is never doubted (never inferior to 25%). In M4, the increase of the decoupling time increases the quality of the fit in the KS test. The GP has lower averaged deviation from the empirical distribution. In M5, the results indicate that probably a more detailed analysis of choice of the u may be needed as the quality of the fit in average seems to improve substantially with the decoupling time. This buoy is particularly distinct from the other buoys due to its location. It has a low number of points with H_s bigger than 8 meters and high number of points with H_s between 3 and 5 meters. For M5, as the decoupling time increases the GP averaged deviation decreases. For the decoupling time used, the Weibull presents better average deviation results. Finally, in M6 the quality of the fit is more variable with the decoupling time but not enough to question the acceptance of any of the distributions studied. The RMSE results for M6 are do not change significantly for different decoupling times.

The quality of the fit fluctuates and all of the three distributions compared may produce a better

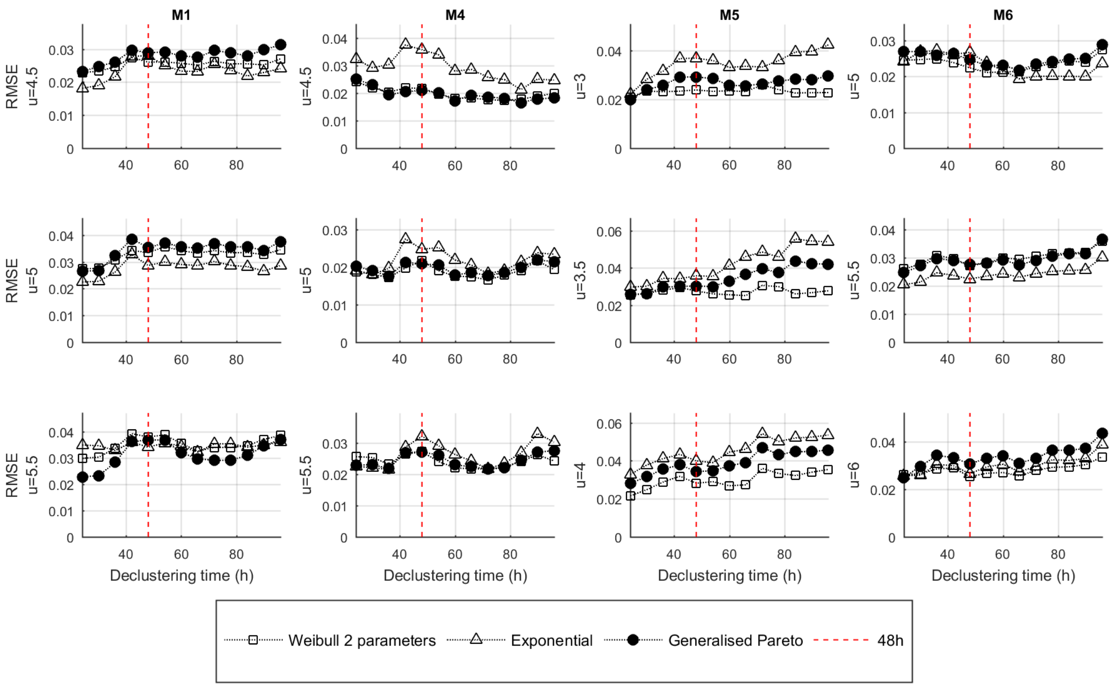

Figure 5 Influence of the decoupling time in the $RMSE$ results. The trimmed (red) line represents the reference value of 48 hours.

fit than the others under specified circumstances.

The analysis of the decoupling time is a methodology that trains the statistical model to some extent. Even though, on a design basis context the training the methodology implemented as presented in Section 3.2 may complement the analysis of fitting. This can contribute to more robust results and confidence in the fitting procedure underpinned.

5.5 *Return periods of H_s*

Frequently the necessity to fit a statistical model to a known set of data is created by the need to extrapolate the occurrence of a certain event to a pre-specified return period T_r. This is the case for offshore wind turbines, where some of the standards that regulate the design of these equipments e.g. IEC (2009), demand the extrapolation of the environmental conditions to return periods of between 20 and 50 years.

Extrapolating long-term occurrences is achieved by knowing the yearly average number of events λ. With the knowledge of λ it is possible to estimate the probability of exceeding in average a certain H_s level in 20 or 50 years of occurrences. No higher return levels were considered due to the amount of data available in the records. As a rule of thumb in the industry applications, it is commonly considered that the minimum amount of the data to extrapolate a certain T_r corresponds to a period of time of at least its value divided by four. The 50 years period, which is out of this limit, is considered due to its importance in most of the offshore sectors.

Different values of H_s with return level T_r depending on u are shown in Table 4. The reference u level for each buoy is highlighted in the table. The relative difference to the GP extrapolated non-transformed data value is shown between brackets for the respective 20 and 50 year extrapolations. As expected, in the majority of the cases the return periods do not vary substantially.

Table 4 20*yr* and 50*yr* return levels of H_s as a function of u and the statistical model applied. Shadowed values correspond to the appropriate level u as specified in Section 5.1.

Return Level (m)	**Weibull**		**Exponential**		**GenPareto**			
					H_s		$\log(H_s)$	
	T_{r20}	T_{r50}	T_{r20}	T_{r50}	T_{r20}	T_{r50}	T_{r20}	T_{r50}
M1 $u = 4.5$	17.12 (+2.7%)	19.03 (+3.3%)	16.90 (+1.4%)	18.75 (+1.7%)	16.67	18.43	14.68	15.42
M1 $u = 5$	17.03 (+9.3%)	18.79 (+12.3%)	17.68 (+13.5%)	19.62 (+17.27%)	15.58	16.73	14.29	14.8
M1 $u = 5.5$	17.17 (+16.2%)	18.87 (+21.7%)	18.37 (+24.4%)	20.39 (+31.5%)	14.77	15.50	13.99	14.32
M1 $u = 6$	16.18 (+21.2%)	17.44 (+15.2%)	19.79 (+40.8%)	22.0 (+52.9%)	14.05	14.39	13.65	13.85
M4 $u = 4.5$	18.51 (-14.0%)	20.74 (-18.3%)	17.16 (-20.3%)	19.02 (-25.1%)	21.52	25.40	18.72	20.7
M4 $u = 5$	18.87 (-4.0%)	21.03 (-5.4%)	18.39 (-6.4%)	20.43 (-8.1%)	19.65	22.22	17.71	19.02
M4 $u = 5.5$	19.21 (-4.0%)	21.40 (-6.0%)	18.70 (-7.0%)	20.74 (-8.9%)	20.10	22.77	18.06	19.44
M4 $u = 6$	18.96 (+2.9%)	20.86 (+4.5%)	20.18 (+9.5%)	22.43 (+12.4%)	18.43	19.96	17.11	17.88
M5 $u = 3$	10.84 (-4.3%)	11.80 (-3.5%)	12.84 (+13.3%)	14.29 (+16.8%)	11.33	12.23	11.63	12.30
M5 $u = 3.5$	11.13 (-4.5%)	12.11 (-4.0%)	12.77 (+13.3%)	14.17 (+16.8%)	11.65	12.62	11.82	12.57
M5 $u = 4$	11.71 (-1.5%)	12.8 (-0.9%)	12.73 (+7.1%)	14.10 (+9.1%)	11.89	12.92	12.01	12.83
M5 $u = 4.5$	11.32 (-7.0%)	12.29 (-7.8%)	12.44 (+2.2%)	13.72 (+2.9%)	12.44	13.33	12.21	13.21
M5 $u = 5$	11.36 (-8.5%)	12.33 (-10.3%)	12.25 (-1.4%)	13.49 (-1.9%)	12.42	13.75	12.45	13.62
M6 $u = 5$	20.13 (-0.6%)	22.39 (-1.1%)	19.62 (-3.1%)	21.74 (+3.9%)	20.25	22.63	17.68	18.69
M6 $u = 5.5$	19.93 (+6.8%)	21.99 (+9.0%)	20.69 (+10.9%)	22.94 (+13.7%)	18.66	20.17	17.25	17.98
M6 $u = 6$	21.37 1(+15.2%)	23.73 (+19.3%)	21.02 (+13.4%)	23.29 (+17.1%)	18.54	19.89	17.25	17.98
M6 $u = 6.5$	19.74 1(+15.6%)	21.54 (+18.7%)	22.19 (+27.7%)	24.63 (+35.7%)	17.38	18.15	17.34	18.02

Results show that the logarithmic transformation tends to stabilize the extrapolation with the GP. This may be connected to the shape parameter that, for the logarithmic transformation, Table 2, sets the GP to be bounded.

In buoy M1, despite the fact that the GP distribution has no better fitting results for the reference level of u, the values extrapolated are similar for the non-logarithmic results. However, when the value of the threshold increases, the two-parameter Weibull distribution and the Generalised Pareto distribution show a trend to decrease the values extrapolated for the return levels of H_s. This decrease was discussed in Gōda (2010) as one of the points of criticism of the POT. As the threshold increases the sample standard variation decreases. The fact that these two distributions have one more DOF than the Exponential distribution might be connected to this decreasing trend. This implies an additional capacity to adapt to the data set to be fitted. This may suffice further research on the homogeneity of data when comparing with the logarithm transformed results. Considering that the maximum value of H_s in the data set is of 13.5m it is very likely that the logarithmic data set is extrapolating the results more accurately in the whole range of threshold while the Exponential distribution is overestimating the return level.

Nonetheless, the reference values are in accordance with the ones given by HSE (2002) indicating that the extrapolation for the reference level may hold valid results for all the cases.

One case of particular concern happens in the reference u level of M4. Here, the values of H_s extrapolated are substantially higher with the GP statistical model. Accordingly to Table 3 and the discussion developed previously, this is one of the cases where the GP distribution has better fitting indicators in four of the five used. Although the GP is, as the other two distributions, unbounded for this case and despite the better fitting indicators, a unrealistic extrapolation level may have been attained. Again, using a logarithmic transformation avoids this unrealistic estimate while maintaining better fitting indicators for all the measures used for comparison.

Important to notice that, considering that the the logarithmic transformed GP produces more realistic results, the Exponential extrapolation results are non-conservative for the reference level and thus, of high risk on a design basis approach.

In M5, the results of both GP cases and the two-parameter Weibull do not deviate much from each other. The Exponential presents higher estimates of the return level. M5 is the other case where the GP distribution demonstrates better fitting indicators while maintaining its bounded character, see Tables 2 and 3.

In the case of M6, the same decreasing trend was found in the value extrapolated for H_s as the u increases. In this case, the fitness indicators calculated point to the application of the two-parameter Weibull or the Exponential distributions. As the maximum occurrence registered is of 17.2m, the usage of the logarithmic transformation may results in underestimating the real return level and the usage of the Weibull and the Exponential in overestimating it. This again may be related to the support of these distributions.

The analysis held focuses substantially on the descriptive comparison of the different statistical models. Nevertheless, the overall goal of the going through the procedure implemented has the ultimate goal of being implemented in a design basis approach. Vrijling and Van Gelder (1999) presents a rationale on how to approach the optimal design basis when considering the need to extrapolate values. Important to highlight that the practical risks of implementing an analysis such as the one presented need to be addressed in the design approach.

Generalization of the results may be obtained using the methodology implemented in Van Gelder (2000), where a sampling process is implemented to study the selection of a probability distribution to model probability exceedance comparing five 3-parameter distributions using samples from a GP distribution.

Considering the results of the return periods when different conditions are met (e.g. increase in u), it is unreasonable to discard the GP as a potential solution to fit exceedance data. In fact, the GP distribution, considering both, transformation and no transformation, should be always considered to use with the POT methodology. More critically even when physically limited data is being modelled.

6 Conclusion

The main purpose of the current work was to investigate on the rejection of the Generalised Pareto distribution over the preferential application of the two-parameter Weibull and the Exponential distributions for the extrapolation of significant wave heights.

Results of the GoF analysis and extrapolation showed that there is no evidence to reject the GP distribution over the two-parameter Weibull and the Exponential distributions. Additionally, the support of the GP distribution and its bounded character is expected to contribute to a physically more realistic extrapolation process. Nevertheless, attention should be given to the selection and fitting process when applying POT to model H_s exceedances. Results showed that when using a POT approach the user should be careful during the selection of the u level that will be used to truncate the H_s data.

The u level is the variable of major influence in the outcome results of the POT application. The method used to ensure independence and the number of the field points do not show the same decisive influence in the results as the level u. A contribution to the choice of threshold was added through the introduction of a new methodology to define the threshold level in a straightforward way. It consists in settling the u value by analysing the three derivatives of the cumulative density function.

For the studied data set, it is impossible to generalize the findings from which of the three distributions is going to produce a better fit. The process should be evaluated in conjunction with a sensitivity analysis of the parameters of the study; level u, decoupling time, and fitting methodology; so that a definitive conclusion can be drawn. With the variation of these parameters, no indicators to reject the Generalised Pareto were found. On the contrary, strong evidence was found that backs its application. In particular when analysing the statistical support of the models analysed.

Using a logarithmic transformation of the data when applying the Generalised Pareto distribution to model a physical limited phenomena was adequate for the present case, and therefore must be considered for future studies of wave data.

The analysis of H_s associated to specific the return periods showed that special attention should be given when fitting the tail with a multi-parameter distribution as the return levels might be underestimated or overestimated. The current work is representative of a data set of the Irish coast, which in turn is representative of the north Atlantic and the Irish Sea. It is of interest to extended the presented analysis to other areas.

To conclude, it is important to emphasize the interest of the presented probabilistic based methodology to extrapolate the values of H_s and its relevance for highlighting the role of the different parameters in the extrapolation process. The methodology implemented can be used to produce robust results on a design basis approach.

Acknowledgements

Acknowledgement is given to Met-Éireann, who had the courtesy of providing the significant wave data analysed in the current paper.

Funding

This project has received funding from the European Union Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No. 642453.

Notation

A-D	= Anderson-Darling
CDF	= Cumulative distribution function
DOF	= Degree-of-Freedom
DNV	= Det Norske Veritas
exp	= Exponential function
F	= Cumulative distribution function
F_{ecdf}	= Empirical cumulative distribution
F_{epdf}	= Empirical probability density distribution
GEV	= Generalised extreme value distribution
GoF	= *Goodness-of-fit*
GP	= Generalised Pareto
H_s	= Significant wave height (m)
$H_{s(>u)}$	= Exceedances of significant wave height over threshold level (m)
i	= Generic indicator of order statistic
KS	= Kolmogorov-Smirnov GoF
max	= maximum
ML	= Maximum Likelihood
MRL	= Mean residual life
MRL	= Mean residual life
PDF	= Probability density function
POT	= Peak-over-threshold
$P-P$	= Probability plot
p	= $p-value$ significance indicator (%)
RMSE	= Root mean square error (-)
u	= Threshold level (m)
T_r	= Return period (years)
W	= Cramér von-Mises
x	= Generic data variable
λ	= Average number of storms in reference interval
ζ	= Distribution function model shape paremeter (-)
$\boldsymbol{\theta}$	= Distribution function model parameter vector
σ	= Distribution function model scale parameter (m)
ψ	= Weight function

References

P Bernardara, F Mazas, X Kergadallan, and L Hamm. A two-step framework for over-threshold modelling of environmental extremes. *Natural Hazards and Earth System Sciences*, 14(3): 635–647, 2014.

Esther Bommier. Peaks-Over-Threshold Modelling of Environmental Data. Technical report, U.U.D.M. Project Report 2014:33, 2014.

Bartomeu Cañellas, Alejandro Orfila, Fernando J Méndez, Melisa Menéndez, Lluís Gómez-Pujol, and Joaquín Tintoré. Application of a POT model to estimate the extreme significant wave height levels around the balearic sea (Western Mediterranean). *Journal of Coastal Research*, 50:329–333, 2007.

Enrique Castillo and Ali S Hadi. Fitting the Generalized Pareto distribution to data. *Journal of the American Statistical Association*, 92(440):1609–1620, 1997.

Enrique Castillo, Ali S Hadi, Narayanaswamy Balakrishnan, and José-Mariá Sarabia. *Extreme* *value and related models with applications in Engineering and Science*. Wiley Hoboken, NJ, 2005.

Enrique Castillo, Alan J O'Connor, Maria Nogal, and Aida Calvino. On the physical and probabilistic consistency of some engineering random models. *Structural Safety*, 51:1–12, 2014.

V Choulakian and MA Stephens. Goodness-of-fit tests for the Generalized Pareto distribution. *Technometrics*, 43(4):478–484, 2001.

Anthony C Davison and Richard L Smith. Models for exceedances over high thresholds. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 393–442, 1990.

DNV. Design of offshore wind turbine structures. Offshore Standard DNV-OS-J101, DNV GL AS, 2014a.

DNV. Environmental condition and environmental loads. Recommended Practice DNV-RP-C205, DNV GL AS, 2014b.

Paul Embrechts, Sidney I Resnick, and Gennady Samorodnitsky. Extreme value theory as a risk management tool. *North American Actuarial Journal*, 3(2):30–41, 1999.

JA Ferreira and C Guedes Soares. An application of the peaks over threshold method to predict extremes of significant wave height. *Journal of Offshore Mechanics and Arctic Engineering*, 120(3):165–176, 1998.

Yoshimi Gōda. *Random seas and design of maritime structures*. World scientific, 2010.

Peter J Hawkes, Daniel Gonzalez-Marco, Agustin Sánchez-Arcilla, and Panayotis Prinos. Best practice for the estimation of extremes: A review. *Journal of Hydraulic Research*, 46(S2): 324–332, 2008.

HSE. Environmental considerations. Technical report, Health & Safety Executive, 2002.

IEC. Wind turbines part 3: Design requirements for offshore wind turbines, 2009.

M Lang, TBMJ Ouarda, and B Bobée. Towards operational guidelines for over-threshold modeling. *Journal of hydrology*, 225(3):103–117, 1999.

Hubert W Lilliefors. On the Kolmogorov-Smirnov test for the exponential distribution with mean unknown. *Journal of the American Statistical Association*, 64(325):387–389, 1969.

R. A. Lockhart and M. A. Stephens. Estimation and tests-of-fit for the three parameter Weibull Distribution. Technical report, Office of Naval Research. Technical Report No. 476, 1993.

Richard A Lockhart, F O'Reilly, and M A Stephens. Tests for the extreme value and Weibull distributiuons based on normalized spacings. Technical report, Office of Naval Research, 1985.

Martin Mathiesen, Yoshimi Goda, Peter J Hawkes, Etienne Mansard, María Jesús Martín, Eric Peltier, Edward F Thompson, and Gerbrant Van Vledder. Recommended practice for extreme wave analysis. *Journal of hydraulic Research*, 32(6):803–814, 1994.

Fernando J Méndez, Melisa Menéndez, Alberto Luceño, and Inigo J Losada. Estimation of the long-term variability of extreme significant wave height using a time-dependent peak over threshold (POT) model. *Journal of Geophysical Research: Oceans*, 111(C7), 2006.

Langley R Muir and AH El-Shaarawi. On the calculation of extreme wave heights: a review. *Ocean Engineering*, 13(1):93–118, 1986.

Maria Nogal, Enrique Castillo, Aida Calvino, and Alan J O'Connor. Coherent and compatible statistical models in structural analysis. *International Journal of Computational Methods*, 13 (02):1640008, 2016.

MI Ortego, R Tolosana-Delgado, J Gibergans-Báguena, JJ Egozcue, and A Sánchez-Arcilla. Assessing wavestorm hazard evolution in the nw mediterranean with hindcast and buoy data. *Climatic Change*, 113(3):713–731, 2012.

James Pickands III. Statistical inference using extreme order statistics. *The Annals of Statistics*, pages 119–131, 1975.

Agustin Sánchez-Arcilla, Jesus Gomez Aguar, Juan Jose Egozcue, MI Ortego, Panagiota Galiatsatou, and Panagiotis Prinos. Extremes from scarce data: The role of bayesian and scaling techniques in reducing uncertainty. *Journal of Hydraulic Research*, 46(S2):224–234, 2008.

Carl Scarrott and Anna MacDonald. A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT–Statistical Journal*, 10(1):33–60, 2012.

David J. Sheskin. *Handbook of Parametric and nonparametric statistical procedures.* CRC Press LLC, 2003.

John J Spineili and Michael A Stephens. Tests for exponentiality when origin and scale parameters are unknown. *Technometrics*, 29(4):471–476, 1987.

Michael A Stephens. Use of the Kolmogorov-Smirnov, Cramér-von Mises and related statistics without extensive tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 115–122, 1970.

Michael A Stephens. Asymptotic results for goodness-of-fit statistics with unknown parameters. *The Annals of Statistics*, pages 357–369, 1976.

Michael A Stephens. Goodness of fit for the extreme value distribution. *Biometrika*, 64(3):583–588, 1977.

Michael A Stephens. The Anderson-Darling Statistic. Technical report, U.S. Army Research Office, 1979.

PHAJM Van Gelder. Statistical methods for the risk-based design of civil structures. *Communications on Hydraulic and Geotechnical Engineering Report (Netherlands)*, 2000.

PHAJM Van Gelder and Cong V Mai. Distribution functions of extreme sea waves and river discharges. *Journal of Hydraulic Research*, 46(S2):280–291, 2008.

Erik Vanem, Arne Bang Huseby, and Bent Natvig. Modelling ocean wave climate with a bayesian hierarchical space–time model and a log-transform of the data. *Ocean dynamics*, 62(3):355–375, 2012.

JK Vrijling and PHAJM Van Gelder. Uncertainty analysis of water levels on lake ijssel in the netherlands: A decision-making analysis. In *Proceedings of the 9th Annual Conference Risk Analysis: Facing the New Millennium, Rotterdam, The Netherlands*, 1999.

Waloddi Weibull. A statistical distribution function of wide applicability. *Journal of applied mechanics*, 103:293–297, 1951.