On defining storm intervals: Extreme wave analysis using extremal index inferencing of the run length parameter

C.L.G. Oikonomou a,∗, M. Gradowski b, C. Kalogeri c, A.J.N.A. Sarmento d

a IDMEC, Instituto Superior Técnico, Universidade de Lisboa, Av. Rovisco Pais 1, 1049-001, Lisbon, Portugal
b Bombora Wave Power, Cleddau Reach, Pembroke Dock, SA726UJ, Wales, UK
c University of Athens, Department of Physics, University Campus, Bldg. PHYS-V, Athens 15784, Greece
d WavEC Offshore Renewables, Edificio Diogo Cão, Doca de Alcântara Norte, 1350-352, Lisbon, Portugal

ARTICLE INFO

Keywords:
Extreme wave analysis
Runs de-clustering
Extremal index
Pareto distribution
Return periods

ABSTRACT

Extreme wave analysis is essential for the design and deployment of marine structures. Since extremes in natural phenomena tend to occur in clusters, it is necessary to de-cluster them in order to form a dataset of independent samples. There are several algorithms used to identify independent storms (clusters of significant wave height extremes), most of which have the disadvantage of relying on an arbitrarily selected de-clustering parameter. In this paper, an existing statistical method for systematic cluster size inferencing is used with runs de-clustering, and applied for the first time to extreme wave analysis. The Generalised Pareto Distribution (GPD) is fitted to an extreme wave dataset, and the return periods of significant wave height extremes are calculated using the resulting model function. The methodology proposed in this paper is illustrated using hindcast data for the winter months of two locations: one that is exposed to the long Atlantic swell off the west coast of France, and another in the North Sea that is characterised by short fetch. This work demonstrates how extremal index estimation may be used in conjunction with the well-known runs de-clustering algorithm to predict the return periods of significant wave height extremes.

1. Introduction

Offshore and coastal engineering requires accurate estimation of the wave height at a potential deployment location. A variety of sources may be used, including (but not limited to) hindcast data, buoy measurements, and satellite observations; general guidelines are discussed in Mathiesen et al. (1994). Predicting the occurrence of extreme waves is particularly important for the survivability of offshore platforms (Raed et al., 2020), marine renewable energy deployments (e.g. Wave Energy Converters (WECs)) (Falcão, 2010; Mackay, 2012, 2017), breakwaters (Marzeddu et al., 2020), and mooring systems (Correia da Fonseca et al., 2016; Barrera et al., 2019). The loading damage of mooring connections can be assessed based on the wave period, as reported in Thies et al. (2014), DNV Recommended Practice RP-C205 (2010) and DNV GL Offshore Standard DNVGL-OS-E301 (2015). Physical modelling of WECs and their mooring components requires modelling of extreme waves at the location for which the technology has been optimised, as was performed in Correia da Fonseca (2014). Lack of knowledge on the extreme nature of the wave climate can significantly increase the risk associated with a project (Greaves and Iglesias, 2018). A WEC must be designed to operate efficiently under typical wave climate conditions (Morim et al., 2019a), which are determined via resource assessment at the deployment site (Iglesias and Carballo, 2010a,b; Sierra et al., 2017). It should also be capable of surviving in stormy seas. Designing for the survivability of marine deployments involves statistical modelling of extreme waves, where the goal is to predict the extremes of the future based on the extremes of the past.

Extreme wave analysis for engineering purposes may be undertaken by fitting a statistical distribution model to significant wave height records (Goda, 2010; Castillo and Sarabia, 1994; Mackay and Johann, 2018). A comparison of different extreme wave analysis methods being applied at a particular location was performed by Sartini et al. (2015), Li et al. (2012) and Al-Mashan et al. (2019) and a complete review on extreme wave analysis techniques was described by Jonathan and Evans (2013).

Some extreme value distributions rely on the definition of an extreme threshold, the selection of which was elaborated upon in Dupuis (1998). If the threshold is exceeded at a particular instance in time, an extremogram may be used to quantify the length of time for which the probability of observing subsequent exceedances becomes stable,

https://doi.org/10.1016/j.oceaneng.2020.107988

Received 30 April 2020; Received in revised form 2 August 2020; Accepted 20 August 2020
Available online 11 September 2020
0029-8018/© 2020 Elsevier Ltd. All rights reserved.
providing insight into the time needed to consider a storm as being independent (Mackay and Johanning, 2018).

Due to the tendency of extremes in natural phenomena to exhibit serial correlation, the total dataset is typically divided into separate groups, before forming a new, presumably independent dataset of extreme values from the maxima of each group. Automated ways of doing this have been discussed in Smith and Weissman (1994) and Shao et al. (2020). De-clustering the data helps to identify independent storms, which collectively characterise the extreme nature of the waves in a given location.

There are various algorithms for de-clustering a wave record, and the performance of several of these techniques was explored by Soukissian and Arapi (2011). The problem with most de-clustering algorithms is that they depend on the arbitrary selection of one or more de-clustering parameters, leading to potential invalidity of the independence criterion, or to statistical bias in the results. Fortunately, these parameters may be inferred from the data itself using extremal index estimators (Ancona-Navarrete and Tawn, 2006; Laurini and Tawn, 2003; Ferro and Segers, 2003), which quantify the clustering tendency of the data as a function of extreme wave threshold. In this paper, the well-known runs de-clustering method is coupled with an extremal index inferencing scheme proposed by Ferro and Segers (2003), and is used to find model distributions to represent significant wave height extremes.

The following sections begin with a review on extreme wave distributions and de-clustering algorithms. Special attention is drawn to how they may be coupled with the extremal index of Ferro and Segers (2003). In the Methodology, a formulation is laid out for how to fit the distribution parameters to a de-clustered extreme wave dataset, using a region of parameter stability as a function of threshold. This procedure is applied to two locations as a case study, with a brief description of the models that were used to generate the data. In the Results and Discussion section, the clustering and separation tendencies of significant wave height extremes are presented for the studied locations. A single set of distribution parameters and extreme wave threshold are used to define a distribution model and calculate extreme wave return periods. Finally, the effectiveness of using the runs method with extremal index inferencing is compared to that of existing de-clustering algorithms.

2. Background information

A fundamental requirement for performing an extreme wave analysis is that the extreme value dataset be independently and identically distributed (i.i.d.). This is a dataset that is characterised by independence and homogeneity. The autocorrelation of an unsorted extreme value dataset may be used to test for independence by ensuring that the autocorrelation of lags greater than the zero lag does not exceed independence and homogeneity. The autocorrelation of an unsorted, extreme dataset is the scale parameter (where \( \sigma > 0 \)) where

\[
\exp\left(-\frac{x - \mu}{\sigma}\right) \quad \xi \neq 0
\]

or

\[
\exp\left(-\frac{x - \mu}{\sigma}\right) \quad \xi = 0
\]

where \( \xi \) is the shape parameter, \( \mu \) is the location parameter, and \( \sigma \) is the scale parameter (where \( \sigma > 0 \)). This is a family of Cumulative Distribution Functions (CDF) that encompasses the Gumbel, Fréchet, and Weibull distributions (Coles, 2001). The problem with describing extremes with the GEV distribution family is that a change in the estimate of its location parameter \( \mu \) may significantly alter its parameters \( \xi \) and \( \sigma \), such that the model changes from one of the distributions that the GEV encompasses to another. Asymptotic theory says that for extremes exceeding some high enough threshold \( u \), the GEV distribution family may be approximated by the Generalised Pareto Distribution (GPD). Provided that \( x > u \),

\[
F(x) = \begin{cases} 
1 - \exp\left(-\frac{x - \mu}{\sigma}\right) & \xi \neq 0 \\
1 - \exp\left(-\frac{x - \mu}{\sigma}\right) & \xi = 0
\end{cases}
\]

where \( \xi \) is the shape parameter of the GEV on the interval \( -\infty < \xi < \infty \), \( \sigma' \) is related to the GEV scale parameter by \( \sigma' = \sigma + \xi (u - \mu) \) (where \( \sigma' > 0 \)), and \( u \) is the extreme significant wave height threshold. From here on in the notation \( \sigma' \) will be replaced with \( \sigma \). The benefit of modelling extremes with the GPD is that its parameters \( \xi \) and \( \sigma \) are less sensitive to the choice in threshold \( u \), provided that \( u \) is high enough.

2.2. De-clustering algorithms

In a time-series of data that describe a phenomenon in nature, extremes tend to cluster together. To ensure that two extreme wave samples are not part of the same storm, it is important to de-cluster the dataset. When fitting an extreme value dataset to the GEV, it is common practice to de-cluster the data using the Standard Storm Length (SSL) method (Tawn, 1988) or the De-Clustering Algorithm (DeCA) proposed
by Soukissian and Kalantzi (2011). SSL assumes that all storms last for a duration $T$. Furthermore, it assumes that all wave height values before and after a time interval of $T/2$ from the storm’s peak are independent from it. DeCA constitutes a more physically meaningful method, since it depends on the wave energy density (proportional to $H_i^2T_i$ in deep water) of a storm, and only considers wave height values to be independent after a sufficient reduction in its energy content.

When fitting the GPD to an extreme value dataset, it is possibly best practice to de-cluster using the runs method, where a new cluster of threshold exceedances is considered to start after a separation of at least $r$ non-exceeding datapoints from the previous cluster. Contrary to SSL, the runs method considers how long it takes the sea to calm down rather than how long it remains stormy. Nevertheless, a short-coming of this approach is the arbitrary selection of the run length parameter $r$.

2.2.1. Runs De-clustering

In a time-series of $n$ datapoints, the runs method considers the termination of a cluster to occur when at least $r$ consecutive non-exceedances of a threshold $u$ follow the final exceedance in the cluster. In extreme wave analysis, the run length parameter $r$ may be interpreted as the minimum required storm interval (period of calm) needed to achieve sample independence in the de-clustered extreme dataset. The total cluster count $Z$ is incremented every time an exceedance in the $i$th position of the time-series is followed by $r$ consecutive non-exceedances. A parameter $W_i$ with a binary value describes whether or not the $i$th observation exceeds the threshold. An exceedance corresponds to $W_i = 1$, and a non-exceedance corresponds to $W_i = 0$. This is to say,

$$N = \sum_{i = 1}^{n} W_i,$$

$$Z = \sum_{i = 1}^{n} W_i(1 - W_{i+1})... (1 - W_{i+r-1}),$$

where $N$ is the total number of exceedances (Smith and Weissman, 1994). The extremal index estimator is then defined as the reciprocal of the mean cluster size $M$, and is given by

$$\hat{\theta} = \frac{Z}{N}. \tag{5}$$

The parameter $\theta$ is an estimate of the true extremal index $\theta$. The extremal index ranges from 0 to 1 (Davis and Mikosch, 2009), and describes the clustering tendency of the data. In the extreme case of $\theta = 1$, all clusters consist of only a single exceedance. The other extreme is the case of $\theta = 0$, where the data consists of nothing but exceedances (i.e. the selected threshold is too low).

2.3. The extremal index

Before applying the runs method, the extremal index $\theta$ should be estimated so that a run length parameter $r$ that is optimal for the given dataset may be extracted. This indicator $\theta$ has been presented in Laurini and Tawn (2003) and Ferro and Segers (2003). The value of $r$ in Eq. (4) should not be chosen arbitrarily; it should be chosen based on the clustering tendency of the data. The extremal index estimator adopted throughout this paper is the one suggested by Ferro and Segers (2003)

$$\hat{\theta}(u) = \frac{2\sum_{j=1}^{N-1}(T_j - 1)^2}{(N - 1)\sum_{j=1}^{N-1}(T_j - 1)(T_j - 2)}, \text{ if } \max T_j > 2, \tag{6}$$

$$\hat{\theta}(u) = \frac{2\sum_{j=1}^{N-1}T_j}{(N - 1)\sum_{j=1}^{N-1}T_j^2}, \text{ if } \max T_j \leq 2, \tag{7}$$

whose first order bias is zero. The above equation considers a set of datapoints with $N$ exceedances of a threshold $u$, where $n > 1$, $N \leq n$, and observation $i \in [1,n]$. The extremal index estimator $\hat{\theta}$ takes into consideration the inter-exceedance times $T_j = S_{j+1} - S_j$, where $S_i$ represents the position of an exceedance and $j \in [1,N - 1]$ if $N > 1$. The estimator of Eqs. (6) and (7) is illustrated in Fig. 1 for a set of observations and threshold $u$. In addition to the extreme wave analysis presented in this paper, the estimator has also been used to de-cluster datasets in extreme water level (Arns et al., 2013) and extreme rainfall (Acero et al., 2011) studies.

3. Methodology

Runs de-clustering with an inferred run length parameter was used to identify clusters of independent storms, on which the POT method could then be applied to form an extreme value dataset. The GPD was chosen as the model function for predicting extreme wave return values. When a fixed threshold $u$ is considered, the distribution parameters of the GPD exhibit relatively low sensitivity to changes in threshold selection, compared to those of the GEV distribution. (A variable threshold approach to modelling extremes with the GPD is discussed by Tancredi et al., 2006). For a given threshold $u$, the GPD parameters were fitted to the resulting extreme wave dataset. The mean excess was plotted against the threshold, as were the shape and scale parameters. The threshold $u$ should be chosen from a region where the mean excess plot is linear with increasing threshold, and where the shape and scale parameters fitted using Maximum Likelihood Estimation (MLE) remain constant (Coles, 2001). (A brief description of MLE is presented in Appendix A). A subtlety of this approach is that the
higher the threshold \( u \), the smaller the resultant extreme value dataset becomes. With that in mind, the lowest value of \( u \) was chosen where the linearity of the mean excesses of \( u \) and the constancy in \( \xi \) and \( \sigma \) began to be fulfilled, in order to keep sample variance to a minimum. This chosen set of \( \xi_q \) and \( \sigma_q \), for the corresponding threshold \( u_q \), was then used as the best set of parameters for the GPD to model the location’s extreme occurrences in significant wave height. Quantile–quantile (QQ) and probability–probability (PP) plots can provide a visual check as to whether or not the extreme value dataset is well represented by the chosen model function. If the empirical data align well with the modelled data, then it is likely that the chosen model function is a good representation of the true extreme distribution of significant wave heights. The run length parameter was inferred using the extremal index estimator of Ferro and Segers (2003). This process is summarised in Fig. 2.

3.1. Run length inference by extremal index estimation

From a given wave record of significant wave heights \( H_s \), a threshold \( u_q \) is selected to define an extreme event. Before the runs method can be applied to de-cluster the data, the run length parameter \( r_q \) should be chosen in a way that is not arbitrary, but rather, so that it is inferred by the data itself. This can be accomplished by means of the extremal index \( \hat{\theta}_q \), which is based on the inter-exceedance times \( T_j \) beneath the threshold \( u_q \) (see Eqs. (6) and (7)), and physically represents the reciprocal of the mean cluster size \( M_q \). Knowing the total number of exceedances \( N \) and knowing what the optimal extremal index should be from Eqs. (6) and (7), the run length parameter \( r_q \) can be inferred from Eq. (4). Threshold exceedances may then be de-clustered using the runs method, and the highest value in each cluster may be extracted using the POT method, forming a dataset of extreme values. Finally, MLE may be used to fit the GPD to the extreme value dataset and obtain the parameters \( \xi_q \) and \( \sigma_q \) corresponding to the threshold \( u_q \).

3.2. Parameter stability and threshold selection

The choice of threshold parameter is undertaken with the aid of graphical procedures, as suggested by Coles (2001). These are parameter stability plots and mean excess plots (also known as mean residual life plots) (Scarrot and MacDonald, 2012).

If the GPD is a valid model for exceedances over some high enough threshold \( u_0 \), then it will be valid for exceedances over all thresholds \( u > u_0 \). The expected value of the threshold exceedances is given by

\[
E[X-u|X > u] = \frac{\sigma_u + \xi_u u}{1-\xi}.
\]

For all \( u > u_0 \), \( E[X-u|X > u] \) is simply the mean value of the threshold exceedances, and is a linear function of \( u \). This linear relationship facilitates the use of a graphical procedure for identifying a suitably high threshold for modelling extremes via the GPD. The point where the threshold exceedances become linear with threshold \( u \) marks the beginning of the region of validity, where the GEV distribution may be approximated by the GPD.

In the mean excess linear region where the threshold choice is valid, the GPD parameter estimates should also remain constant. In fitting the shape and scale parameters of the GPD using MLE, there will be a region where the returned estimator values remain approximately constant with \( u \). Constant GPD parameter estimates, along with a linear region in the mean excess plot, were used to justify the selection of
the threshold value for the GPD, above which significant wave height values are considered to be extreme.

The size of the resulting extreme value dataset decreases with increasing threshold \( u \); it should be noted that too high of a threshold would lead to a lower number of data, and consequently, a higher statistical variance. On the other hand, too low of a threshold could violate the validity of the GPD approximating the GEV distribution. Although the techniques presented in this section are well-established methods for the selection of the threshold \( u \), they inherently contain a small degree of subjectivity. The quality of the model function's fit to the empirical data may be assessed by using a methodology suggested by Coles: QQ plots (Coles, 2001) (described in Appendix B).

### 3.3. Estimation of the return period

The runs method involves extracting independent clusters from the dataset that exceed an extreme value threshold \( u \), and hence after applying the POT method to the independent clusters, the remaining i.i.d. dataset is extreme by definition. After performing MLE on the resulting extreme value dataset and obtaining GPD parameter estimates for \( \xi \) and \( \sigma \), the extreme significant wave height that is expected to occur only once every \( R \) years is given by

\[
H^*_R = u + (\sigma/\xi) \left( RZ/\lambda^\xi - 1 \right) \quad \text{for} \quad \xi \neq 0, \tag{9}
\]

where the notation \( * \) denotes an extreme quantity, \( \lambda \) is the number of years spanned by the original dataset, and \( Z \) is the number of i.i.d. exceedances that make up the extreme value dataset.

### 4. Application of methodology

To illustrate the proposed methodology, this section presents a case study comparing two ocean basin locations with different characteristics. Both locations exhibit considerable wave energy resource, and hence they may be regarded as favourable areas for exploitation (Kalogeri et al., 2017), provided that the extreme occurrences in significant wave height are small enough to ensure the survivability of a converter.

Runs de-clustering, with an inferred run length parameter, was performed on significant wave height hindcast data to form an extreme value dataset for each location. The proposed methodology was followed to obtain a GPD model for predicting the return period of an extreme occurrence in significant wave height.

#### 4.1. Data used in the study areas

The first site selected for applying the proposed methodology is situated on the northeastern part of the French west coast, approximately 10 km offshore, with coordinates 47° 42′ 24.63″ N and 4° 23′ 28.57″ W, at a water depth of 85 m. This location is exposed to the long Atlantic fetch, and hence the extremes are dominated by swell.

The second location is situated in the North Sea, with coordinates 55° 8′ 60″ N and 3° 27′ 0″ E, at a water depth of 29 m. The North Sea is a semi-enclosed basin with a short fetch area. The wave field is primarily wind-driven. Wave extremes are caused by the passage of extra-tropical cyclones and polar lows (Kalogeri et al., 2017).

The wave dataset used in this analysis is the product of hindcast simulations of the 3rd generation spectral ocean wave model WAM. The simulations were performed under the framework of the FP7 MARINA platform project (https://www.msp-platform.eu/projects/marina-platform) by the Atmospheric Modelling and Weather Forecasting group of National and Kapodistrian University of Athens. The data span the years 2001–2010 (inclusive), at an hourly interval, providing information for the main met-ocean parameters needed to describe the wave field of an area. The data cover the entire European coastline with a resolution of 5 km. The atmospheric model SKIRON provided the forcing wind field to WAM with the same high spatial resolution and an hourly frequency (Kallos et al., 1997; Spyroul et al., 2010; Kallos et al., 2006). The modelling system has been implemented in high resolution, in a wide enough region, to depict the main storm activity of the Atlantic Ocean that is responsible for remotely generated Atlantic swell, which reaches and affects the local wave field of the western coasts of Europe. The atmospheric and wave systems used in this model have assimilated available altimeter measurements to correct their initial conditions (Kalogeri et al., 2017; Patlakas et al., 2016, 2017). The general performance of the modelling system has been evaluated in numerous research and operational projects (Edwards et al., 2014a, b; Larsén et al., 2015; Kalogeri et al., 2017) and its output has been also used for extreme value analysis (Patlakas et al., 2016, 2017).

More details on the dataset and modelling approach may be found in Kalogeri et al. (2017). Altimeter data have already been assimilated into WAM, so to validate the output of the model, wave buoy measurements were used. Various statistical indices show that the dataset is in close agreement with these observations. In particular, for the areas under study, a correlation coefficient greater than 0.8 between the modelled and measured significant wave heights indicates a strong linear relationship between them. The data show low systematic error, returning bias and Root Mean Squared Error (RMSE) values both less than 0.4 m and scatter indices less than 0.4. However, it should be noted that buoys are not always reliable when recording the heights of extreme waves (Janssen, 2002). Therefore, buoy measurements cannot guarantee a perfectly reliable source for evaluating the ability of a model to accurately predict extreme waves.

#### 4.2. Results and discussion

For both locations, the 10-year hourly datasets of significant wave height are presented in Fig. 3, along with box plots depicting the seasonal nature of their wave resource. If seasonality were successfully removed from a dataset, an extreme wave phenomenon coming from the same parent distribution could be modelled. In an effort to satisfy the homogeneity criterion of extreme value theory, only the winter values were considered. (In the Northern Hemisphere, waves are usually most energetic during the winter. In this work, winter values were defined as those spanning the months of December, January, and February.) In the box plots of Fig. 3, \( H \) medians and quartile ranges – denoted by the red lines and box edges respectively – are highest during the winter months of both locations.

For both 10-year winter datasets of significant wave height, two extreme value datasets were formed using runs de-clustering above a threshold \( u \), where the run length was inferred by the estimator \( \hat{\theta} \) given by Eqs. (6) and (7). In Fig. 4(a) and (b), the mean cluster size \( M \) and run length parameter \( r \) are plotted as a function of the threshold \( u \), respectively, providing insight into the clustering and separation tendencies of storm events at both locations. The mean cluster size \( M \) initially decreases rapidly with increasing threshold \( u \), before reaching a point where the rate of decrease slows down significantly — possibly at a value \( u \) that is high enough to make the GPD a valid approximation to the GEV distribution. The physical interpretation is that for low thresholds \( u \), a considerable portion of the significant wave heights that exceed \( u \) are either part of the wave climate’s natural variability or represent a storm that is developing or calming down. Conversely, for thresholds \( u \) that are high enough, exceeding significant wave heights generally represent developed storms. For very high extreme value thresholds, a point is reached where the run length parameter begins to increase rapidly. This corresponds to the situation where very few clusters of exceeding significant wave heights remain in the original time-series, thereby causing the minimum required storm interval to increase significantly.

2. The scatter index is equal to the RMSE divided by the mean.
MLE was applied to the extreme value datasets in order to extract estimates of the parameters $\xi_q$ and $\sigma_q$. This procedure was repeated for many thresholds $u_q$, so that the threshold diagnostics described in the methodology could be used to choose the best values for defining the occurrence of an extreme event.

To compare the stability of the GPD parameters, the shape and scale parameters $\xi$ and $\sigma$ are plotted against the threshold $u$ in Fig. 5(a)–(d). The corresponding mean excesses are also plotted against the threshold $u$ in Fig. 5(e) and (f). These plots helped determine the appropriate thresholds $u_q$ for the GPD model of the extreme wave occurrences in the two locations. The reported methodology was repeated for thresholds $u_q$ ranging from 2.50 – 8.00 m for the west coast of France and 1.00 – 6.50 m for the North Sea, both with a step size of 0.25 m.

As outlined in the Methodology, a region with little change in the distribution parameter estimates and a region of linearity in the mean excesses has been identified in the runs de-clustering plots between $u_q = 4.25$ m and $u_q = 5.75$ m for the west coast of France and between $u_q = 2.00$ m and $u_q = 3.25$ m for the North Sea, shown by the
highlighted areas in Figs. 5(a)–(f). The lower part of this range was preferred in order to reduce the variance, and although the selection was still slightly arbitrary, a threshold of \( u_q = 4.25 \) m was chosen as the definition of an extreme event occurrence for the west coast of France, and a threshold of \( u_q = 2.75 \) m was chosen for the North Sea. With these thresholds, 88 and 156 independent storm values were de-clustered for the west coast of France and the North Sea, which were then used to form the extreme value datasets. The corresponding shape and scale parameters, \( \xi_q = -0.13 \) m and \( \sigma_q = 2.50 \) m, were selected for the west coast of France and the North Sea, which were then used to fully specify their respective GPD functions, as given by Eq. (2).

With the GPD models determined, their representation of the hindcast data could be described qualitatively with the aid of PP and QQ plots, illustrated in Fig. 6. If the fit is good, the dots in these figures should closely follow the straight, dashed red lines, which is the case here. This means that the GPD, with the fitted parameters from MLE, was a good extreme wave model for the hindcast data of the studied locations. For the open sea area (French coast), an extreme value threshold of \( u_q = 4.25 \) m was selected, with shape and scale parameters \( \xi_q = -0.13 \) m and \( \sigma_q = 2.50 \) m. For the semi-enclosed basin (North Sea), the extreme value threshold of \( u_q = 2.75 \) m was selected, with shape and scale parameters \( \xi_q = -0.32 \) m and \( \sigma_q = 2.09 \) m. As a final check for independence between the samples of selected extreme value datasets, their autocorrelation tendencies were examined. As shown in Fig. 7(a) and (b), no noticeable pattern exists in the autocorrelations of the extreme datasets, and the majority of lags after the zero lag do not
Fig. 6. Probability–probability and quantile–quantile plots of the fitted GPD model, for the French coast (left) and the North Sea (right). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

These GPD models were then used to predict the return periods of extreme significant wave height occurrences, as shown in Fig. 8. In both locations, the first 20 years show a significant increase in the extreme return value with increasing return period. There is little difference in the predicted return values between a 50-year return period and a 100-year return period.

Although the QQ and PP plots shown in Fig. 6 suggest a good fit of the modelled data to the empirical data, it is clear from the range of 90% confidence intervals in Fig. 8(a) that, in the case of the west coast of France, extreme wave occurrences were simply not enough data to keep statistical uncertainty within reasonable limits. (To study the extreme wave resource of this location with higher accuracy, a time-series of significant wave height spanning more years would need to be obtained (Sartini et al., 2015; Surendran et al., 2005; Kumar et al., 2009)). This was not the case for the North Sea (Fig. 8(b)), where a larger extreme wave sample of 156 occurrences gave much lower uncertainty in the predicted extreme wave return values. The smaller sample of extremes in the long fetch area (French coast) can be attributed to the fact that its wave-field is swell-dominated. The swell formations are the result of remotely generated weather systems that can travel hundreds of kilometres with little attenuation. They tend to persist for longer periods of time compared to the wind–sea wave fields of the North Sea, which are coupled to the local weather conditions and closely follow their changes (Barber and Ursell, 1948; Munk et al., 1963; Cavaleri et al., 2007; Ardhuin et al., 2009). Therefore, during the same 10-year sample period, the swell-dominated location exhibited fewer independent storms than the wind–sea location. Consequently, in the two original datasets of the same length, the one pertaining to the west coast of France had fewer independent storm data points, because individual storms would last longer than in the North Sea. This lead to a smaller de-clustered extreme value dataset for the west coast of France, and a consequently higher statistical uncertainty in its predicted extreme wave values compared to the North Sea.

4.3. Comparative analysis with other de-clustering algorithms

Diagnostic plots for MLE fitted GPD parameters were also generated for the two alternative de-clustering algorithms described in Section 2.2 (i.e. SSL (Tawn, 1988) and DeCA (Soukissian and Kalantzis, 2006)), to compare the effectiveness of the proposed methodology with algorithms that currently exist in the literature. The following investigations were made regarding the stability of the GPD shape and scale parameters:

- Using storm length and energy reduction parameter values that are frequently suggested in the literature, for SSL and DeCA respectively.
- Using the extremal index estimator of Eqs. (6) and (7) and the total number of threshold exceeding significant wave heights, \( N \), to calculate the optimal number of values \( Z \) in the extreme dataset. Then, finding the storm length and energy reduction parameter values for SSL and DeCA that would make the dataset be of this size.

For the first investigation, commonly used storm length values of 24, 48, 72, 96, and 120 h were considered for the SSL algorithm (Zachary et al., 1998; Soukissian and Arapi, 2011). Similarly, commonly used wave energy reductions of 80%, 85%, 90%, and 95% were considered for the DeCA algorithm (Soukissian and Arapi, 2011). Threshold exceedances were de-clustered using SSL and DeCA, while considering these typical de-clustering parameter values. From the
clusters that were identified this way as being storms, the peak values were taken from each cluster to form an extreme value dataset – one for every \( u \) examined. With SSL, the GPD parameter stability plots of Fig. 9 show that both the west coast of France and the North Sea exhibited higher (i.e. less negative) shape parameters and lower scale parameters, for shorter storm lengths. In Fig. 10, DeCA exhibited the same tendency when lower energy reductions were considered. For the GPD shape parameter in particular, it is clear from Fig. 10(a) and (b) that the region of stability varies significantly with choice of de-clustering parameter. In Fig. 9(a) and (b), the GPD parameter response is comparatively more consistent with changing de-clustering parameters than in the case of Fig. 10(a) and (b). However, there seems to be a definite downward shift in the shape parameter responses and upward shift in the scale parameter responses, with increasing
storm length parameter. These results suggest that unless the use of a particular storm length or energy reduction can be justified for SSL and DeCA, arbitrarily choosing these values and using them for all thresholds could lead to a region of stability that poorly represents the extreme nature of a given wave climate.

The second investigation made use of the extremal index estimator, to infer optimal storm length and energy reduction values for SSL and DeCA at every threshold \(u_q\). This allowed a fair comparison of GPD shape and scale parameter stabilities to be made with those obtained using the runs method. Knowing the number \(N\) of significant wave heights exceeding \(u\), and estimating \(\theta\) from Eqs. (6) and (7), the number of de-clustered values \(Z\) forming the extreme dataset may be calculated using Eq. (5). The de-clustering parameter of a given algorithm is hence inferred as being the one to separate the original \(H_s\) time-series into \(Z\) clusters, each of them assumed to represent an independent storm event containing (on average) \(M\) consecutive extreme values of \(H_s\). In runs de-clustering, the run length parameter \(r\) is inferred as being the one to make Eq. (4) hold, and this was found by trial and error. Similarly in the SSL and DeCA algorithms, the storm length and energy reduction parameters were found by trial and error to be those which would separate the \(H_s\) time-series into \(Z\) clusters.

Although it is likely that DeCA is a very robust algorithm in terms of ensuring independence in the samples of an extreme value dataset, its two stages of monotonicity filtering inherently leave very few remaining samples in the resulting \(H_s\) time-series. For a given threshold \(u_q\), when the number of de-clustered exceedances \(Z_q\) that was suggested by the extremal index estimator \(\theta_q\) turned out to be higher than the number of samples remaining after DeCA de-clustering – even with an energy reduction of 1% – the DeCA extreme value dataset was simply formed using the remaining \(H_s\) extremes. This was the case for threshold values in Fig. 11 where the fitted GPD parameters of DeCA deviated noticeably from those of SSL and runs de-clustering. For both the west coast of France and the North Sea, the runs method seems to have performed best when coupled with extremal index based inferencing of its de-clustering parameter.

5. Conclusions

This paper presented a new methodology for the estimation of return periods of extreme significant wave height occurrences. The methods presented in this work could be of particular interest when designing for survivability in a marine environment. Modelling of extreme events relies on the formation of an extreme dataset of i.i.d. values. Since the extremes of natural phenomena tend to cluster, de-clustering techniques are used to identify independent clusters, from which the POT method is used to pick out their maxima. Although current bibliography encourages the use of runs de-clustering for extreme wave analysis, the choice of run length parameter has remained arbitrary. Therefore, an extremal index inferencing method from statistics literature - one that is based on the inter-exceedance times of the data - was proposed to infer the optimal run length. Once the data has been de-clustered and subjected to the POT method, the newly formed extreme dataset may then be fitted to a model distribution. In this paper, the GPD was chosen due to the low sensitivity of its parameters to the choice of extreme value threshold. For a given threshold, MLE was used to estimate the corresponding GPD shape and scale parameters.
C.L.G. Oikonomou et al.

The proposed methodology was applied to hindcast data between the years 2001–2010, for two locations with different climatological characteristics. Diagnostic plots assisted with the selection of an appropriate threshold for both locations: the mean threshold excesses, as well as the stability of the shape and scale parameters, as a function of the threshold. Once a threshold had been selected along with its corresponding shape and scale parameters, the respective quantiles and probabilities of the empirical data and the GPD model were plotted together on QQ and PP plots. The resulting GPD model was used to predict the return periods of extreme significant wave height occurrences. A considerable spread in the predicted return periods was obtained for the location off the west coast of France, and a significantly smaller spread was obtained for the North Sea location. This was likely attributed to the fact that the swell formations reaching the coasts of France are more persistent systems than the wind-driven waves of the North Sea. For the same time-series length there were fewer independent storms from which the POT method was able to form an independent set of extreme values, and a smaller sample implies higher variance. Finally the performance of the proposed methodology was compared to that of existing algorithms: SSL and DeCA were first tested using storm length and energy reduction values that had frequently been used in the literature, then the de-clustering parameters of SSL and DeCA were also inferred using extremal index estimation. Qualitatively speaking, it was found that runs de-clustering with a run length parameter that is inferred via extremal index estimation exhibited the best GPD parameter stability, for the two locations analysed in this work.

CRediT authorship contribution statement

C.L.G. Oikonomou: Conceptualisation, Methodology, Formal analysis, Writing - original draft, Visualisation, Investigation, Project administration. M. Gradowski: Software, Methodology, Formal analysis, Data curation, Writing - review & editing, Investigation. C. Kalogeri: Data curation, Formal analysis, Writing - review and editing, Resources. A.J.N.A. Sarmento: Supervision, Writing - review & editing, Resources.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was partially supported by the Portuguese Foundation for Science and Technology (FCT), through IDMEC, under LAETA, project UIDB/50022/2020. The wave data for the west coast of France and the North Sea were provided by Professor George Kallos from the University of Athens. The authors would like to thank Professor Jonathan Tawn from the department of Mathematics and Statistics of Lancaster University for his constructive suggestions, and Dr. Christopher Ferro from University of Exeter for kindly answering their questions on his extremal index estimator. The authors have also benefited from discussions with Dr. Takvor Soukissian from the Hellenic Centre for Marine Research.
The GPD parameters are fitted to the extreme data by using Maximum Likelihood Estimation (MLE). The purpose of MLE is to find the distribution parameters of a representative model function so that the probability of obtaining the de-clustered i.i.d. extreme dataset is maximised. Since the joint probability of a series of independent events $x_k$ is equal to the product of the events' individual probabilities $p(X = x_k)$, a likelihood function may be defined as the joint probability of having obtained the resulting i.i.d. extreme dataset for a given set of GPD parameters $\xi$ and $\sigma$

$$L = \prod_{k=1}^{Z} p(X = x_k | \xi, \sigma).$$  

The parameter $Z$ had previously been used to denote the number of independent clusters resulting from runs de-clustering. Since the POT method picks out only one peak value from each cluster, $Z$ will be used from hereon to represent the size of the de-clustered extreme dataset such that $k \in [1, Z]$. The objective of MLE is to find the optimal set of model function parameters such that the likelihood function is maximised.

### Appendix B. Quantile–quantile and probability–probability diagnostic plots

Ordered statistics lie at the core of this methodology. It is of interest to investigate how well the empirical quantiles and probabilities correspond to their respective modelled counterparts. It begins with ranking the values of the extreme dataset, where the highest rank (i.e. $m = 1$) corresponds to the greatest significant wave height. The empirical quantiles are simply given by the values of the extreme dataset, whilst the empirical cumulative probabilities are given by

$$F_{\text{emp}} = 1 - \frac{m}{\delta + 1},$$  

where $\delta$ is the number of independent storms and $m$ is the rank of the storm, and lies on the interval $m \in [1, \delta]$. The theoretical quantiles are given by solving for $x$ in Eq. (2), where parameters $\xi$ and $\sigma$ are those returned by the MLE method for the threshold $u$, and $F(x)$ is the empirical cumulative probability given by Eq. (B.1). Specifically,

$$x_{\text{mod}} = u + \sigma \left( \left( 1 - \frac{F_{\text{emp}}}{\xi} \right)^{-\frac{1}{\xi}} - 1 \right),$$  

where $x_{\text{mod}}$ is the modelled quantile. Similarly, the modelled cumulative probability may be obtained by solving Eq. (2) for the empirical quantiles of $x$ and the distribution parameters $\xi$ and $\sigma$ returned by MLE. This is to say,

$$F_{\text{mod}} = \begin{cases} 
1 - \left[ 1 + \xi \left( \frac{x_{\text{emp}} - u}{\sigma} \right) \right]^{-\frac{1}{\xi}}; & \xi \neq 0 \\
1 - \exp \left( \frac{x_{\text{emp}} - u}{\sigma} \right); & \xi = 0
\end{cases},$$

where $x_{\text{emp}}$ are the empirical values from the extreme dataset. If the selection of the GPD with the returned MLE parameters $\xi_q$ and $\sigma_q$ for the threshold $u_q$ is a wise choice for the studied data, then the empirical and modelled values on the QQ and PP plots will consist of data points close to the unit diagonal (Li et al., 2012).


