Empirical estimation of return period and distribution of duration of wind speed exceedances

D. C. Kiplangat
Department of Statistics & Actuarial Science, Dedan Kimathi Universityof Technology,
Nyeri, Kenya
dennis.cheruiyot@dkut.ac.ke

K. Asokan
Department of Mathematics College of Engineering Thiruvananthapuram
Kerala, India.
asokank@cet.ac.in

G. V. Drisya
Department of Futures Studies, University of Kerala
Kerala, India.
drisyavictoria@gmail.com

K. Satheesh Kumar
Department of Futures Studies, University of Kerala
Kerala, India.
kskumar@keralauniversity.ac.in

***Abstract*—Wind turbines are designed to function within a certain range of wind speed to optimise production cost and minimise damage limited by technical barriers and usually shut down when wind speed crosses these limits. We present an empirical analysis of extreme events of wind speed outside the functional range of wind turbine and compute the return period of such events using a direct empirical approach. We also demonstrate an extrapolation technique to accurately estimate the return period for time series of short duration. Analysis of wind speed data from various locations reveals that the duration of both upper and lower exceedances can be modelled empirically by the exponential family of distributions very closely. The ideas developed would be useful to wind farm managers for making anticipatory calculations based on past data regarding frequency and durations of wind turbine shut-down due to extreme wind speeds.**

***Index Terms*—Wind speed modelling; Generalized Pareto distribution; Peaks over threshold; Duration of exceedances**

I. Introduction

In the recent past, renewable energy sources have attracted increased attention from governments and energy investors as it is one of the fastest growing sources of alternative energy globally. An especially important source of alternate energy is wind, which is abundant in availability and is reasonably cost-effective and environment-friendly. The primary challenge affecting wind power generation is the wind speed variation resulting from numerous meteorological and topographical factors. Without proper modelling and forecasting of these variations, the reliability of the generated power is compromised, and consequently, more spinning reserves would be required to compensate for the unexpected power fluctuations leading to an increase in the overall production cost of wind power. It is also important in wind power production to be able to predict *extreme events* of wind speed, that is wind speeds attaining very large or small magnitudes, which can pose serious problems in wind power production and transmission, especially at power grids where wind energy is connected to energy from other traditional sources.

Wind speeds crossing a high threshold or a low threshold value may be considered as extreme events since both of these conditions force a shut-down of the wind turbine. The power P generated at the wind turbine by a volume of air of density ρ flowing at speed v through area A is given by $P = c_p \frac{\rho}{2} A v^3$ where $c_p \approx 0.4$ is the ratio of the wind power to the power extracted. Thus at higher speeds, the turbines could generate more power, but the cost of maintaining the equipment without damage at those speeds has to be balanced against the cost of energy generated, which determines the upper threshold, called the *cut-out* wind speed. And below the lower threshold, called the *cut-in* wind speed, there would not be enough energy to get the blades moving, and no power would be generated. In both cases, it would be more economical to shut down the turbine. Hence wind turbines are designed to operate only within a certain functional range of wind speeds which, for most common turbines, is between 3 m/s and 25 m/s [1].

Suitable methods to calculate the return periods and durations of the extreme events, which cause shut down of wind power production, would be most useful in optimising the costs associated with managing spinning reserves in such scenarios. In this work, we demonstrate that simple empirical methods can effectively estimate the return period of extreme events of wind speed falling outside the functional range of wind turbine with relatively less amount of data. We also analyse the statistical properties of the *durations of exceedances* of both upper and lower extreme events. The results of our analysis of empirical data show that the duration of both upper and lower extreme events may be modelled by the exponential family of distributions. For the analysis, we have used wind speed measurements of 10 minutes resolution from across 234 different locations for the period from January 2004 to January 2007 available from National Renewable Energy Laboratory (http://www.nrel.gov), USA.

II. Modelling of extreme wind speeds

There are two popular models for studying extreme events, namely the *block maxima method* and the *peak over threshold method*. The block maxima method is based on the classical

The authors are grateful to UGC, New Delhi, for their financial assistance (no. F. 42-30/2013(SR)). The first author (DCK) and third author (GVD) would like to acknowledge Indian Council for Cultural Relations (ICCR) and DST-PURSE programme, University of Kerala, respectively for the financial assistance. They wish to thank National Renewable Energy Laboratory (http://www.nrel.gov), USA, for making their data available.

extreme value theory which says that if $X_1,\, X_2, \ldots X_n$ is a sequence of independent and identically distributed random variables and M_n is their maximum, then, for large n, M_n have an approximate distribution within the so called family of Generalised Extreme Value (GEV) distributions. The cumulative distribution function of GEV family is given by

$$G(x) = \begin{cases} \exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\} & \text{if } \xi \neq 0 \\ \exp\left[-\exp\left(\frac{x-\mu}{\sigma}\right)\right] & \text{if } \xi = 0 \end{cases} \tag{1}$$

where ξ, σ and μ are the shape, scale and location parameters, respectively. If $\xi = 0$, it is called *Gumbel distribution (Type I GEV)*; If $\xi > 0$, it's *Frechet distribution (Type II GEV)* having a long right tail; and if $\xi < 0$, it's *Weibull distribution (Type III GEV)* having a short tail which is usually used in modelling extreme wind speeds [2].

In practical applications, the extreme events, here represented by the block maxima, must be statistically independent to get reliable results [3]. According to [4] one needs to use at least 20 years data for annual block maxima method to provide reliable results.

1) Peaks Over Threshold method: In the POT, we consider all values beyond a specific threshold, called *exceedances*, to form a sequence of extreme values and attempt to fit a statistical distribution to this sequence. Under suitable conditions, the behavior of the exceedances $X_i, (X_i > u)$ over a suitable threshold u is described by some member of an asymptotic distribution, called the *Generalized Pareto Distribution family (GPD)*. This has the cumulative distribution

$$H(x) = \begin{cases} 1-\left[1+\frac{\xi}{\sigma}(x-u)\right]^{\frac{1}{\xi}} & \text{if } \xi \neq 0 \\ 1-\exp\left[-\left(\frac{x-u}{\sigma}\right)\right] & \text{if } \xi = 0 \end{cases} \tag{2}$$

where the random variable x satisfies $x > u$. ξ and σ are respectively the shape and scale parameters.

Fitting a given data into the GPD family involves several steps, of which the first is finding the optimal value of the threshold. The threshold should be high enough so that the exceedances follow a Poisson law [5], which is necessary for the extremes to converge to a GPD. On the other hand, setting too high a value for the threshold will leave an insufficient number of extremes, making parameter estimation difficult. A commonly used method for aiding the selection of appropriate threshold involves constructing a *mean residual graph*, which is a plot of the mean of threshold excesses $X_i - u, (X_i > u)$, as a function of possible threshold values u, for $u < \max_i\{X_i\}$ [6], [7]. The lowest value of u, above which the graph is approximately linear, is then an estimate of a suitable threshold.

Once the threshold value has been fixed, the parameters of the distribution may be found, the recommended technique for which is the method of maximum likelihood estimators [8], [9].

Estimates of the extreme events are commonly expressed in terms of the quantile value X_T, which is the maximum value exceeded, on the average, every T years. Then T is called the *return period* and X_T the *return level*. In the case of GPD the return level corresponding to return period T can be estimated as

$$X_T = u + \frac{\sigma}{\xi}\left[(T\lambda)^{\xi} - 1\right]$$

where $\lambda = P(X_i > u)$ is the probability of exceedance, which may be estimated as $\lambda = m/N$ where N is the total number of observations and m is the number of observations that exceed the threshold u [10]. Several diagnostic plots are used to assess the goodness of fit, such as probability plots, quantile plots, return level plots, density plots. The *probability plot* is a comparison of empirical and model values of the distribution function for each exceedance, while the *quantile plot* compares the quantiles of the exceedances.

Suppose that $y_1 \leq y_2 \leq \cdots \leq y_m$ are the exceedances. Then an empirical estimate for the cumulative distribution $H(y_i)$ is $\tilde{H}(y_i) = i/(m+1)$. We can get a model estimate $\hat{H}(y_i)$ of the distribution by substituting the parameter values into eq. (2). Thus a probability plot consists of the points $(\tilde{H}(y_i), \hat{H}(y_i)), i = 1, 2, \ldots m$ while a quantile plot consists of the points $(\hat{H}^{-1}(i/(m+1), y_i), i = 1, 2, \ldots m$. If the exceedances fits into GPD, the points should be approximately linear and close to the diagonal in both the plots.

The *return level plot* is a plot of the return period T versus the corresponding return level X_T, for larger values of T. Finally, a *density plot* compares the fitted GPD model with the histogram of the exceedances.

The empirical estimates $\tilde{H}(y_i)$ for cumulative probabilities are called plotting positions, for which there are many formulae available in the literature. Choosing the right plotting formula that leads to unbiased quantile estimates has been a subject of much discussion and controversy and are reviewed by [11]–[14]. For the peak over threshold method, the Weibull formula used above is generally recommended.

III. Results and discussion

We first consider wind speed crossing the upper limit of the functional range of wind turbine, and all measurements above 25 m/s can be considered as extreme events in this case. The maximum number of such exceedances registered was 251, which was observed at only one site, while 11 sites did not have any exceedance above 25 m/s. The statistics of the distribution of the number of exceedances normalised to a period of one year is plotted in Fig. 1. It may be noted that only less than 10% of the sites considered have annual exceedances 30 or more. If the data of extremes is insufficient, it cannot be properly fitted to standard extreme event models, and we will have to rely on empirical estimates for the calculation of return periods and duration of exceedances.

A. Return period estimates

We will now calculate the T-year return level X_T, which is exceeded once every T years on the average, for wind speed observations at the above locations. The POT model is more suited to our study as our aim is to estimate return period of wind speed crossing the functional range of the wind turbine.

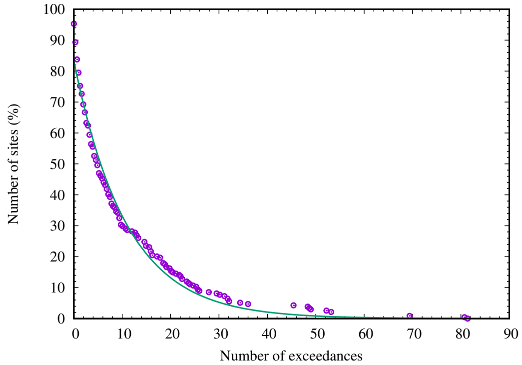

Fig. 1. The plot of number of values exceeding 25 m/s normalised to one year period versus percentage of the total number of sites. The data were fitted with function $f(x) = ae^{-bx}$ where $a = 83.2498 \pm 1.085$ and $b = 0.0920862 \pm 0.002059$.

The main advantage of this approach, compared to the block maxima method, is that a shorter time series can also yield more reliable estimates. Patlakas et al. have used five-year 3-hourly wind speed data for their analysis using POT [15]. Coles & Walshaw have opined that, statistically, 5 to 6 years of data is adequate for successful application of POT method compared to the 20 year period of block maxima method [16].

A typical example of the procedure explained in previous section is illustrated for the location with latitude $34.98420°$ and longitude $-104.03971°$. The threshold estimated by the mean residual graph method is 21.23 m/s. Theoretically, this means that if we use the higher value 25 m/s (which is the wind power cut-out limit) as the threshold, we could still fit the data into a suitable member of the GPD family. The parameters of the GPD model for the exceedances above 25 m/s were obtained from the maximum likelihood method. The diagnostic plots generated from the model is portrayed in Fig. 2. The fact that probability and quantile plots are nearly linear and close to the diagonal, and that the density plot of the model fits the histogram of the data fairly well, shows that the chosen GPD with threshold $u = 25$ fits the data well. From the return level plot, the expected return period for exceedances above the 25 m/s mark for the specific site is approximately 979 data points *(6.8 days)*.

Following the customary extreme value analysis, we computed 1-year, 2-year, 5-year, 10-year, 20-year and 50-year return levels of extreme wind speeds using the peaks over threshold method based on GPD, and a sample figure for 1-year is shown in Fig. 3. These results are for exceedances above 25 m/s using wind speed data for three years, plotted for locations with enough data points to ensure convergence of the distribution. The results of the generalised Pareto distribution method are sensitive to fitting accuracy as noted by [15]. The GPD model has been found to fit the data fairly well at sites with a sufficiently large number of extreme values (approximately 30 or more), and the return level predictions would be more accurate in those locations. However, in other locations with fewer instances of peaks, wind speed data of at least 5-6 years is required for more accurate results. As

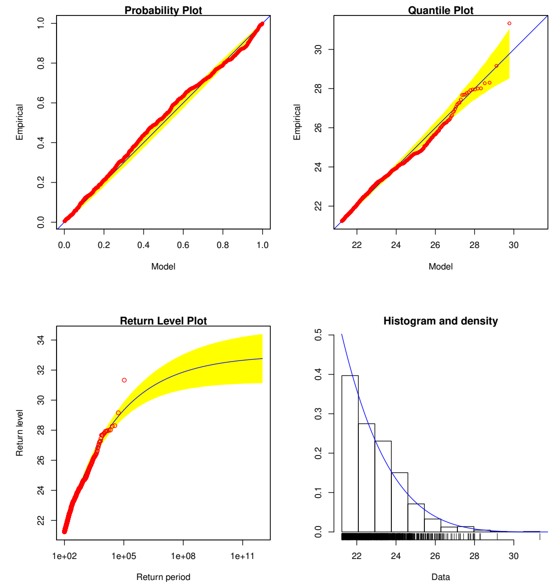

Fig. 2. The diagnostic plots for the generalized Pareto distribution with the corresponding return level plot and the histogram for the exceedances for the location with latitude $34.98420°$ and longitude $-104.03971°$.

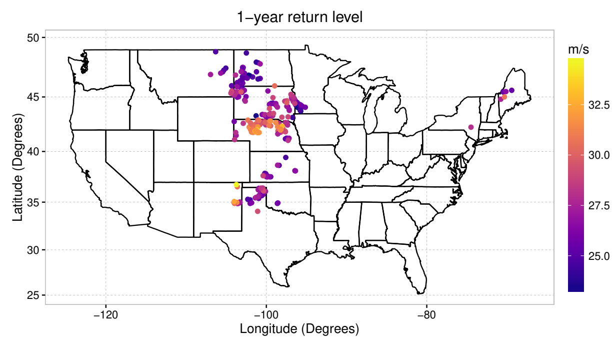

Fig. 3. The expected return levels at different return periods obtained using the generalized Pareto distribution.

observed above, the GPD model fits fairly well in locations where the number of threshold exceedances is 30 or more which, from Fig 1, is about 10% of the sites considered here. In other locations with insufficient exceedance data points, the fitting and the return level estimates are less accurate. We will therefore use an empirical method to calculate the return periods, which gives better estimates than the GPD model when the number of threshold excess data points are lesser. The return period R of a particular event is the inverse of the probability that the event will be exceeded in any given year. Thus, if λ is the probability of exceedance, then $R = \frac{1}{\lambda}$.

Method	Proponent (%)
$P = (i - 0.5)/m$	Hazen (1913) [17]
$P = i/(m + 1)$	Weibull (1939) [18]
$P = (i - 0.31)/(m + 0.38)$	Beard (1951) [19]
$P = (i - 0.44)/(m + 0.12)$	Gringorten (1963) [20]

TABLE I
DIFFERENT FORMULAS FOR COMPUTING PLOTTING POSITIONS EMPIRICALLY.

An empirical estimate for λ may be obtained by one of the plotting formulas in Table I. The Weibull estimate gives $\lambda = m/(N+1)$ where m is the number of exceedances above 25 m/s, and N is the total number of data points. Fig. 4 shows a log-log plot of the return period versus the number of exceedances computed for all the 234 sites considered. For each site, the return period calculated by the various empirical formula as well as the GPD model are shown in the figure. It is evident that, when the number of exceedances is larger (approximately 30 or more), the return period estimates by both the methods are in fairly good agreement. However, for sites with fewer extreme values, the GPD model estimates show marked variations both from the empirical estimates and also among the sites with the same number of exceedances. The empirical estimates, by the various formulae, were fitted by the inverse law $f(x) = \frac{K}{x}$ with $K = 51232.860 \pm 2.887$ which gives the average return period as a function of the annual number exceedances for wind speed data measured at 10-minute intervals.

As we have seen, the empirical estimate of the return period coincides with the estimation by GPD method. In this work, we propose an extrapolation technique for estimating the return period of the extreme events for shorter time series. From a given time series decreasing sequence of sub time series are obtained to estimate the return period using Gringorten method empirically. Then we plot the return periods versus the inverse of the number of extreme events in each sub-series. The return period is then extrapolated as the inverse of exceedances tends to zero (corresponding to the number of exceedances tends to infinity). The procedure is illustrated in Fig. 5. The data considered for the figure has 214 number of exceedance where the empirical estimate is nearly equal to the GPD estimate of the return period. So we constructed several subsequences of sub-series with lesser number of exceedance. The return period for each sub-series is considered as a function of the inverse of the number of exceedance. The dotted line represents the extrapolation fitted through the data points in Fig. 5. In this case, the extrapolated value nearly coincides with GPD estimates of the original time series demonstrating the usefulness of the procedure for finding the accurate estimate of the return period for shorter time series.

An equally important factor in wind turbine applications is the *duration* of exceedance, that is, the time it takes for the extreme wind speeds beyond 25 m/s to return to the safe operating speed of the wind turbine. The analysis of the data of the duration of extreme events is important since it can help in determining how long backup power is required to compensate for the power loss during the turbine shut-down before it restarts operation at speeds below 25 m/s.

Fig. 6 portrays the quantile-quantile plot of the empirical distribution matched to a best-fit exponential distribution. The plot shows a reasonably good fit and suggests the suitability of exponential distribution as a model for exceedance durations in general. It also indicates that the distribution of extreme event durations at the individual sites could be governed by an exponential family. However, the lack of sufficient data prevents verification of this claim at many of the individual sites with fewer extreme events. But at sites with a moderately larger number of extreme values, the durations could be fitted by suitable exponential models with reasonable accuracy.

For many turbines, the cut-in speed is 3 m/s, and if the wind speed drops below the cut-in speed, the wind turbine ceases to produce electric power. We now carry out a similar analysis of these lower extreme events, that is, wind speed realizations that fall below 3 m/s. The return periods for the lower extreme

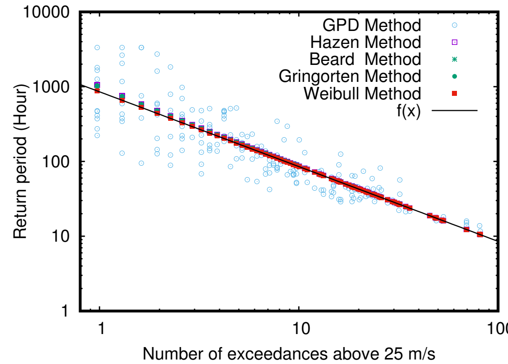

Fig. 4. Log-log plot of return period versus the number of wind speed exceedances above 25 m/s, normalised to a period of one year, calculated from the GPD model and the various plotting formulas. The results of the plotting formulas are fitted with $f(x) = K/x$ with $K = 51232.860 \pm 2.887$.

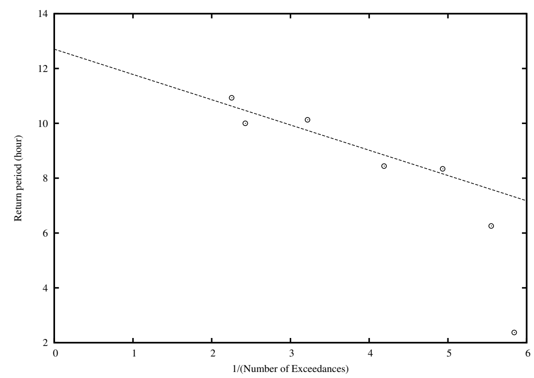

Fig. 5. Extrapolation of the return period versus the inverse of number of wind speed exceedances above 25 m/s, calculated for the location latitude $45.00238°$ and longitude $-70.31773°$.

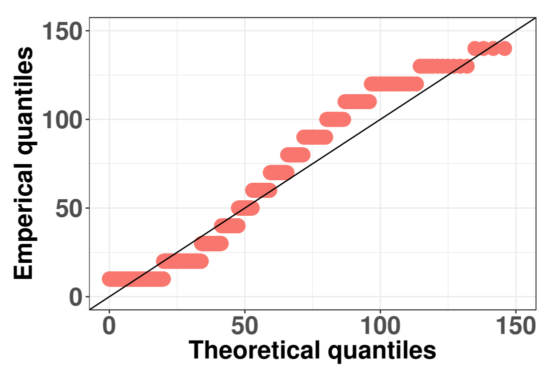

Fig. 6. QQ-Plot of the durations (< 150 minutes) of upper extreme events from all locations, fitted to a suitable exponential distribution.

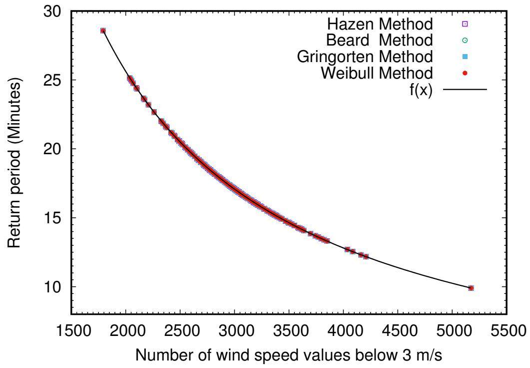

Fig. 7. Plot of return period (recurrence interval) calculated by the various plotting formulas versus the number of wind speed values below 3 m/s normalised to one year period. The data is fitted with $f(x) = K/x$ with $K = 51232.9 \pm 5.792$.

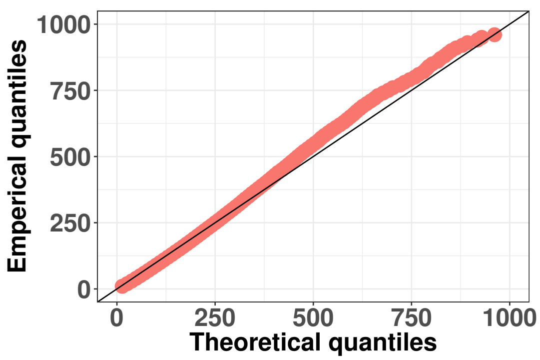

Fig. 8. QQ-Plot of the durations (< 1000 minutes) of lower extreme events from all locations, fitted to a suitable exponential distribution.

events calculated empirically by the different plotting formulas in Table I are shown in Fig. 7 where the points are fitted by the curve $f(x) = K/x$ with $K = 51232.9 \pm 5.792$, which expresses the return period as a function of the count of lower extremes. Wind farms can make use of the function $f(x)$ to estimate the average return period of wind speeds below 3 m/s from the wind speed data of 10-minute resolutions irrespective of the geographical location. Compared to the upper extreme events, the lower extreme events are more frequent with return periods ranging from 10 minutes to 30 minutes.

Fig 8 shows a quantile-quantile plot for the data set of durations (less than 1000 minutes) from all sites against a best-fit exponential distribution and suggests that the exponential model fits the data very closely. This further indicates, as in the case of upper extremes, that an exponential family could model the durations of lower extreme wind speeds at the individual locations. The empirical and theoretical quantiles are more or less in close agreement and support the suitability of the exponential model. agreement and support the suitability of the exponential model.

IV. Conclusion

We have carried out a detailed analysis, based on empirical data, of the wind speed events falling outside the functional range of the wind turbine using extreme value theory. The functional range of the most popular wind turbines is between 3 m/s and 25 m/s. However, the methodology adopted in this work can be applied to any different functional range as well. For locations with limited data points of extreme events, the calculation may be less accurate due to the poor convergence of the distributions. However, even for a moderate number of data points, the expected return periods can be calculated using empirical methods, expressing the return periods as a function of the average number of exceedances per year. These results have been presented for both lower and upper exceedance events. The analysis was also carried out for the duration of exceedances, that is, the period for which each exceedance remain outside the functional range of wind turbines.

References

[1] C. G. Justus, W. Hargraves, and A. Mikhail, "Reference wind speed distributions and height profiles for wind turbine design and performance evaluation applications.[usa]," Georgia Inst. of Tech., Atlanta (USA). School of Aerospace Engineering, Tech. Rep., 1976.

[2] X. Lin, "Statistical modelling of severe wind gust," in *International congress on modelling and simulation, Townsville*, vol. 2. Citeseer, 2003, pp. 620–5.

[3] J. Palutikof, B. Brabson, D. Lister, and S. Adcock, "A review of methods to calculate extreme wind speeds," *Meteorological applications*, vol. 6, no. 02, pp. 119–132, 1999.

[4] N. Cook, "The designers guide to wind loading of structurespart 1, bre," 1985.

[5] J. Palutikof, T. Holt, B. Brabson, and D. Lister, "Methods to calculate extremes in climate change," 2003.

[6] A. C. Davison, "Modelling excesses over high thresholds, with an application," in *Statistical extremes and applications*. Springer, 1984, pp. 461–482.

[7] W. Ledermann, *Handbook of applicable mathematics*, ser. Handbook of Applicable Mathematics. Wiley, 1990, no. v. 7.

[8] E. B. Mackay, P. G. Challenor, and A. S. Bahaj, "A comparison of estimators for the generalised pareto distribution," *Ocean Engineering*, vol. 38, no. 11, pp. 1338–1346, 2011.

[9] C. Anderson, D. Carter, and P. Cotton, "Wave climate variability and impact on offshore design extremes," *Report commissioned from the University of Sheffield and Satellite Observing Systems for Shell International*, 2001.

[10] S. Coles, J. Bawa, L. Trenner, and P. Dorazio, *An introduction to statistical modeling of extreme values*. Springer, 2001, vol. 208.

[11] C. Cunnane, "Unbiased plotting positionsa review," *Journal of Hydrology*, vol. 37, no. 3-4, pp. 205–222, 1978.

[12] E. Castillo, "Extreme value theory in engineering," 1988.

[13] C. Folland and C. Anderson, "Estimating changing extremes using empirical ranking methods," *Journal of Climate*, vol. 15, no. 20, pp. 2954–2960, 2002.

[14] I. Jordaan, *Decisions under uncertainty: probabilistic analysis for engineering decisions*. Cambridge University Press, 2005.

[15] P. Patlakas, G. Galanis, N. Barranger, and G. Kallos, "Extreme wind events in a complex maritime environment: Ways of quantification," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 149, pp. 89–101, 2016.

[16] S. G. Coles and D. Walshaw, "Directional modelling of extreme wind speeds," *Applied Statistics*, pp. 139–157, 1994.

[17] A. Hazen, "Storage to be provided impounding reservoirs for municipal water supply," in *Proceedings of the American Society of Civil Engineers*, vol. 39, no. 9. ASCE, 1913, pp. 1943–2044.

[18] W. J. Weibull, "A statistical theory of strength of materials," 1939.

[19] L. R. Beard, "Statistical analysis in hydrology," in *Proceedings of the American Society of Civil Engineers*, vol. 68, no. 7. ASCE, 1951, pp. 1077–1088.

[20] I. I. Gringorten, "A plotting rule for extreme probability paper," *Journal of Geophysical Research*, vol. 68, no. 3, pp. 813–814, 1963.