# Reliability of Extreme Wave Prediction Methods

Aaron Barker
*Doctoral Researcher, MaREI, Environmental Research Institute, University College Cork, Cork, Ireland*

Jimmy Murphy
*Lecturer, MaREI, Environmental Research Institute, University College Cork, Cork, Ireland*

Vikram Pakrashi
*Lecturer, MaREI, Environmental Research Institute, University College Cork, Cork, Ireland*

ABSTRACT: Extreme wave parameters are used for engineering design in our seas and oceans, yet the methods used to determine them are non-standardized and can give highly variable output. With increased commercial activity in the marine sector, the importance of accurate extreme wave parameter determination has become increasingly apparent. This is particularly the case for marine renewable structures where even small over-predictions in design parameters can affect the whole feasibility of the project. This paper addresses the methods of extreme wave prediction currently in use, with a view to selecting the optimal method for the prediction of extreme wave conditions ($H_s$, $H_{max}$ and $T_z$) in coastal Irish waters. The paper identifies pitfalls and drawbacks of current extreme prediction methods, with particular attention given to the use of limited in time buoy data from coastal locations where development is to take place. In addition a new methodology of determining extreme wave periods, that is the wave periods occurring coincidentally with the most extreme wave heights, is established. This is important as the destructive energy of a wave is dependent on the wave period. By estimating the extreme wave energy and significant wave height, it is possible to formulate a method of reliably approximating the likely coincident wave period.

## 1. INTRODUCTION

With the continued exploitation of marine resources, many structures, in particular marine energy devices will be required to be located in very extreme wave environments. Knowledge of the wave climate and design wave conditions are fundamental to any structural design, yet recent analyses have highlighted shortcomings in terms of the input wave data and the analysis methods- notably the work of Jonathan & Ewans (2013). This can impact significantly on the final design; with incorrect estimation of the design wave condition resulting in an over-designed and expensive structure, or a structure that is unsafe. In this work, the authors identify the clear potential for the application of modern mathematical techniques to the estimation of extreme wave parameters. It is the long term goal, to establish a dependable methodology that allows the prediction of extremes of climate. To achieve this, a systematic approach must be applied to determining the factors which have the greatest influence on extreme predictions. These include a number of parameters such as those related to the probability distribution i.e. the shape factor and location factor, the method of curve fit chosen and the criteria for storm event selection.

Extreme prediction methods typically rely on empirical curve fitting and extrapolating from these curves to give estimates of future conditions. These models are often used with theoretical coefficients which do not fit the distribution of the dataset well. The reliability of a prediction made in this manner is questionable and exposes a contractor to significant risk. Further, the choice of the fit is left to the analyst, who may choose an inappropriate distribution to make the prediction. It is clear that the human selection element is undesirable and it would be prudent to have a statistical backing on which to select an appropriate prediction.

There are a number of options available to improve the selection of data and fit of the distribution used to make extreme predictions proposed by You and Callaghan (2013). These will be examined in this paper.

## 2. METHODOLOGY

In the field of extreme analysis threshold selection remains a much debated topic. The selected set of data must cover the variance of the extremes, but the inclusion of too much data below what can be considered an extreme wave will skew the distribution and result in a lower predication of extreme wave height. The work of Mathiesen et al. (1994) describes the recommended practice for extreme wave analysis methods, with particular focus on data selection process. No new work has been done on this topic and the methods suggested therein will be applied in this study. The work of Mazas et al. (2014), however, has examined replacing the distribution used for the analysis (as was encouraged originally by Mathiesen et. Al.). They have replaced the Maximum Likelihood Estimated two-parameter distributions by L-Moments estimated 3 parameter distributions. In this paper, alternative distributions will be used based on the Generalized Extreme Value approach in order to achieve a better fit for the data at the extreme quantiles. In addition, regression and covariate analysis are used to attempt to tie together all sea state parameters at the extremes and give a better understanding of the extreme values of Wave Power and Period.

Defining storm duration must take into account that most extreme predictions require the data to be discreet and independent. For this reason storm events must have at least 3 days between them to qualify as independent in this analysis.

Though this paper will not delve deeply into each of these topics, it will give an insight into the controllable factors which can be adjusted to create the best predictions at a given location.

Current extreme prediction methods typically do not consider covariate analysis using wave period, the maximum sustainable slope typically seen in the climate which will govern, to an extent, the wave period that accompanies the storm condition.

The analysis undertaken focuses primarily on an investigation of:

- Selection of the best distribution for a particular dataset using autoregressive techniques, homoscedatic/heteroscedatic assessment techniques, regression analysis and with the aid of the Ameva statistical toolbox (IH Cantabria, 2014) version 1.32 and version 1.41. Matlab was utilized extensively for Extreme analysis and data selection methods.
- The use of Generalized Extreme Value (GEV) analysis to determine $H_s$ values, $H_{max}$ and Wave Power at 5, 10 and 20 year return periods.
- The Anomaly Index (AI), the ratio of $H_{max}$ to $H_s$. Establishing the relationship between these parameters is critical for safe design in extreme conditions.
- Wave Steepness values at the extreme values of $H_s$. This was done using regression techniques and the best fit 3 part regression model was used to estimate the likely slope at the predicted extreme wave heights.
- The most appropriate $T_z$ value to use at extreme Hs values. These were estimated using regression analysis and with a new method based on the estimated wave power at the extremes.
- Period Correlation to wave height at extremes. Extreme predictions provide more definitive information on wave height than on wave period. (Mercier J. A., 1982). It is normal to determine the associated wave period by assuming a wave slope. However the slope of extreme waves will often be markedly different to the average climate, as will the wave with maximum period. Where there is a dynamic response involved the wave period can in some cases be equally or more important than wave height. Thus a joint probability of height and period was assessed

## 3. DATA SOURCE

The data used for the analysis was acquired from the Marine Institute, and consists of waverider buoy measurements acquired at Berth B of the Atlantic Marine Energy Site (AMETS) at Belmullet in Ireland. It spans a time period from December 2009 to August 2013 and consists of 50173 records at 30 minute intervals.

Spectral records are logged approximately every 30 minutes by the waverider buoy with an output of the full spectrum logged 6 minutes later. This output also includes the timeseries output of $H_{1/3}$, $H_{1/10}$ and $H_{max}$. Spectral data includes the significant wave height ($H_{m0}$) and mean wave period ($T_z$). Missing and invalid entries due to buoy or logging error were removed from the dataset along with their associated values at that timestamp. Such anomalies are not ideal but in this case consist of 1588 records (3%) of the dataset and were mostly confined to more benign sea states. It is expected that dealing with false/missing data in this manner provides optimal error prevention without compromising the outcome of the extreme analysis. An Occurrence scatter plot for the dataset used is given in figure 1.

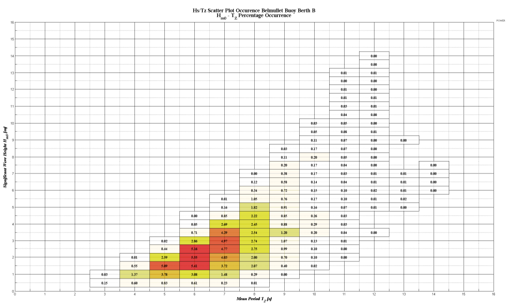

*Figure 1: Hs/Tz Occurrence Scatter Plot at AMETS Berth B*

## 4. ANALYSIS

### *4.1 Identification of best fitting distribution for Wave Parameters $H_S$ 98th percentile data*

The $H_s$, $H_{max}$, and wave power parameters were fitted to an optimum distribution function for the dataset using auto-regression to identify the shape, location and scale parameters which best fit the data. The data selection process began by selecting the 98th percentile values of each dataset to remove the effect of lower values on the fit of the distribution.

The best fitting distributions was found to be the Weibullmin distribution with the following parameters;

| | |
|---|---|
| Mean | 8.245 |
| Log likelihood= | -1329.6 |
| Location factor | 6.9 ± 6.504e-08 |
| Shape Factor | 1.14 ± 0.079 |
| Scale Factor | 0.87 ± 0.0459 |

*Table 1 Table of distribution descriptors for $H_s$.*

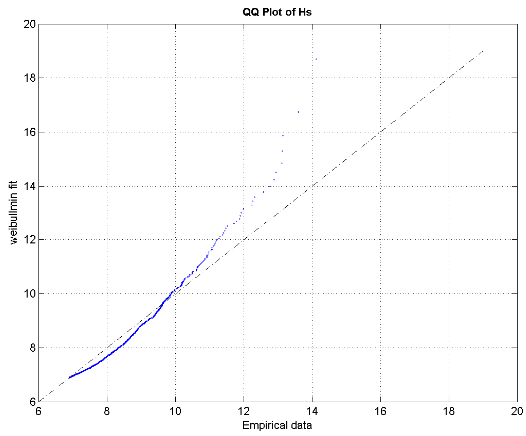


*Figure 2: Quantile-Quantile Plot Hs 98th Percentile Values*

The Quantile-Quantile plot which shows the probabilistic fit of the distribution to the fitted curve displays an upward concavity at higher quantiles, displaying a tendency to overestimate extreme values. This is evidence that a more rigorous data selection process is needed to identify only the storm condition maximum values which contribute to the extreme wave estimate. Therefore the methods proposed by Mathiesen et al. (1994) are used, selecting the monthly maximum values. For this approach storms are not considered unless separated by 3 days. This ensures discrete data and heteroscedasicity of the model.

### *4.1. Monthly Maximum Hs data*

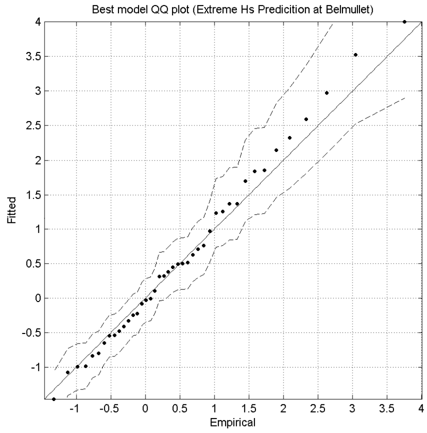


*Figure 3: Quantile-Quantile plot of adjusted dataset shows improved correlation at extremes*

The model fitted is a GEV distribution whose cumulative distribution takes the form:

$$F(x;\mu;\sigma;\xi) = \exp(-\left[1 + \xi\left(\frac{x-\mu}{\varphi_1}\right)\right]^{\frac{1}{\xi}})$$

Where $\mu \in R$ *is the Location parameter*
$\sigma > 0$ *The Scale parameter.*
$\xi \in R$ *The Shape parameter.*
And where Scale parameter of the model is assumed to follow the physical description:

$$\sigma(t) = exp\left(\alpha_o \sum_{i=1} [a_{2i-1}\cos(i\omega t) + a_{2i}\sin(i\omega t)] + \beta_{T2} t + \sum_k \varphi_k\, n_{k,t}\right)$$

With

$$\mu(t) = \beta_0 = 5.844$$
$$\alpha_0 = 0.726$$

The extreme $H_s$ analysis was then conducted with the newly reduced dataset. The maxima encountered in each month show the highly seasonal nature of the test site and the influence that seasonality has on the prediction of extremes. The results show that the annual maximum data contribute the most to the extreme prediction. As expected the most extreme events occur during winter months and make up the largest component of the maximum wave estimation.

| Return Period (Year) | Expected $H_s$ (m) at RP | 95% Confidence interval bounds for $H_s$ (m) |
|---|---|---|
| 5 | 14.1 | [11.8-16.4] |
| 10 | 15.6 | [13.0-18.3] |
| 20 | 17.1 | [14.1-20.1] |

*Table 2: Extreme $H_s$ Values and associated uncertainty at 5, 10, 20 yr. Return Period*

Wide separation for 95% confidence intervals is still observed. Benign data can be seen to be contributing significantly to this uncertainty in the estimate, with summer monthly maximums seen to be much lower and therefore affecting the fit. It is evident then that even with great care taken in the data selection process, a limited dataset in a seasonal site will still present a challenge with regards to the certainty of the prediction.

Monthly maximums at this site include waves with a height as low as 2.8m should perhaps not be considered extreme waves. Analysis was then performed by introducing an additional wave height threshold to determine storm events. This was run after the selection of monthly maximum values. The results showed that an imposition of an additional wave height limitation in a short duration dataset served to decrease confidence in the upper estimates of extreme wave height due to decreased sample size. Thus for this case it can be said that a human selection element was not necessary – and acted to reduce certainty in the prediction. This result displays the importance of threshold selection and it was found that the selection of the most extreme waves by monthly maximum selection was superior to the quantile method for this particular dataset and location. A point that will be examined in subsequent work is whether the distribution that gives the best fit with the data provides the best prediction of extremes. The expected $H_s$ values indicated in Table 2 are higher than what has been normally used at the AMETS site so further analysis is required to understand the sensitivity of the extreme predictions to distribution type, threshold selection and data length.

### 4.2. $H_{max}$ Data

Using monthly maximum data from the wave buoy, the $H_{max}$ extremes were determined using the same methods as the $H_s$ extreme values, by selecting the monthly maximum values, to provide comparable results. These extreme $H_{max}$ values will be used in the determination of AI values in section 5.4

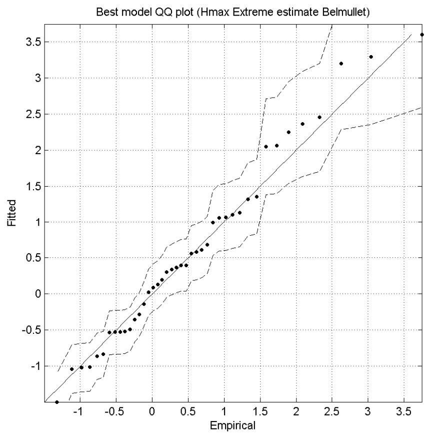


*Figure 4: Quantile-Quantile plot of distribution fit for Hmax Extreme estimation*

The distribution selected for $H_{max}$ fits quite well at the average conditions but begins to show variability at the upper quantiles. Nevertheless, the distribution and results were determined using the following GEV model:

$$F(x;\mu;\sigma;\xi) = \exp(-\left[1 + \xi\left(\frac{x-\mu}{\varphi_1}\right)\right]^{\frac{1}{\xi}})$$

$$\sigma(t) = exp\left(\alpha_o \sum_{i=1}[a_{2i-1}\cos(i\omega t) + a_{2i}\sin(i\omega t)] + \beta_{T2}t + \sum_k \varphi_k\, n_{k,t}\right)$$

With

$$\mu(t) = \beta_0 = 9.849$$
$$\alpha_0 = 1.1216$$

| Return Period (year) | Expected $H_{max}$ at RP (m) | 95% Confidence interval bounds |
|---|---|---|
| 5 | 23.3 | [19.6-27.0] |
| 10 | 25.8 | [21.6-30.1] |
| 20 | 28.3 | [23.4-33.1] |

*Table 3: Extreme $H_{max}$ values*

These $H_{max}$ values are bounded by wide confidence intervals due to the limited sample size available and the variability of the maxima encountered from month to month

### 4.3. Analysis of Anomaly Index from Hmax/Hs

#### 4.3.1. AI to $H_S$ Correlation (from $H_{max}/H_s$ AI regression analysis)

Correlation between AI and $H_s$ measured using a 3 parameter regression analysis and were found to trend towards mean AI values governed by the following equation.

$$\mu(t) = p1 + p2t + p3t^2$$

*With p1= 2.453, p2= -0.166, p3= 0.007*

Confidence in the mean diminishes with increasing wave height due to decreased data availability at this level. This is a significant drawback of the limited nature of the dataset. Future work should look at the behaviour over a much longer period to establish a firmer relationship. Figure 5 does show that as $H_s$ increases the variability of AI decreases.

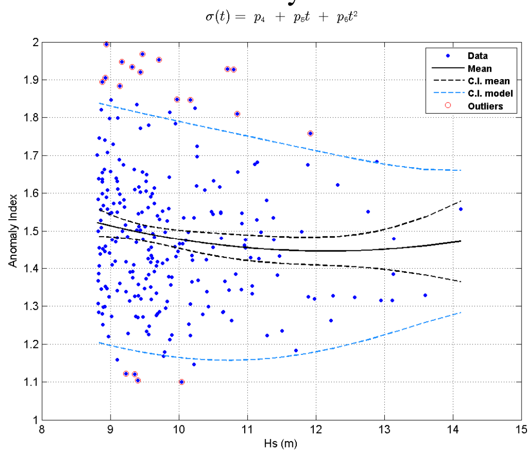


*Figure 5: Regression analysis of AI using $H_s/H_{max}$ data*

This gives the following estimates for AI values at the 5, 10, 20 year return periods.

| Return Period (year) | $H_S$ (Table 2) (m) | Expected AI at RP |
|---|---|---|
| 5 | 14.1 | 1.503 |
| 10 | 15.6 | 1.569 |
| 20 | 17.1 | 1.662 |

*Table 4 Regression derived AI values at extremes of $H^s$*

### *4.3.2. AI to HS Correlation obtained at predicted extreme values*

The maximum predicted $H_s$ and $H_{max}$ as shown in tables 2 & 3 respectively were used to determine values of AI. The results in all cases show the AI Value to be relatively constant with a value of about 1.65. This agrees relatively well with the values determined in Table 4. It is likely that the Table 4 values are less accurate given the sensitivity of the equation to the limited available data at the high values of $H_s$.

While it is difficult to draw firm conclusions, it appears that the industry accepted 1.65 $H_{max}/H_s$ ratio appears realistic for the determination of maximum $H_{max}$ values, and that ratios of 1.87 and 2.0 as are often recommended would give a large factor of safety.

| Return Period (Year) | Expected AI at RP | 95% Confidence interval bounds |
|---|---|---|
| 5 | 1.654 | [1.661-1.650] |
| 10 | 1.652 | [1.659-1.648] |
| 20 | 1.651 | [1.657-1.646] |

*Table 5 AI index values determined using extreme $H_{max}/H_s$ values*

### *4.4. Determination of wave steepness at the extremes*

Regression analysis was performed for the 98th percentile $H_s$ and accompanying wave steepness values to determine a link between $H_s$ and wave steepness at the extremes. This will enable a better estimate of the $T_z$ associated with the extreme $H_s$ to be determined.

It is evident from Figure 6 that there is an a very obvious trend of convergent wave steepness values at the extreme, suggesting that there is a limiting wave steepness which governs the behavior of the wave at the extremes. Tabulated results of this regression analysis:

| | |
|---|---|
| Mean | 18.8707 |
| Std | 2.6898 |
| Log likelihood | -2122.1926 |

*Table 6 Tabulated results of wave steepness regression analysis*

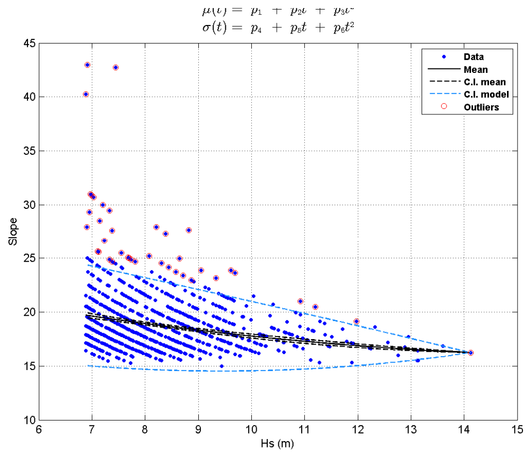


*Figure 6: Wave steepness Regression Analysis against $H_s$*

According to the regression analysis, the mean steepness for the 98th percentile waves is given by the following equation.

$$\mu(t) = p1 + p2t + p3t^2$$

Where the $p$ values were determined to be as follows:

| *p* | Value | Upper | Lower |
|---|---|---|---|
| *p1* | 26.380 | 28.658 | 24.103 |
| *p2* | -1.207 | -0.767 | -1.647 |
| *p3* | 0.035 | 0.054 | 0.015 |

*Table 7 Equation parameters determined for wave steepness from regression analysis*

The resulting wave steepness predicted using regression analysis shows that they converge to a value of about 16 (Table 8). From this result it is proposed that a limiting wave slope at the extremes of $H_s$ can be applied. This means that a methodology can be developed to determine related parameters such as $T_z$ given an extreme $H_s$ value. From this work we can say that a wave steepness of 16 is a reasonable assumption at the extreme values of $H_s$ in Eastern Atlantic Waters

| Return Period (Years) | Predicted $H_s$ (m) | Wave Steepness (Reg. Model) |
|---|---|---|
| 5 | 14.1 | 16.324 |
| 10 | 15.6 | 16.065 |
| 20 | 17.1 | 15.975 |

*Table 8 Wave Steepness values at extreme Hs values determined from regression analysis*

### *4.5. Calculation of extreme $T_z$ using regression analysis*

Linear Regression analysis was performed on the $T_z$ and $H_s$ concurrently. Outliers above a 95th percentile threshold were removed and confidence intervals were set at the 90th percentile of values (Figure 7).

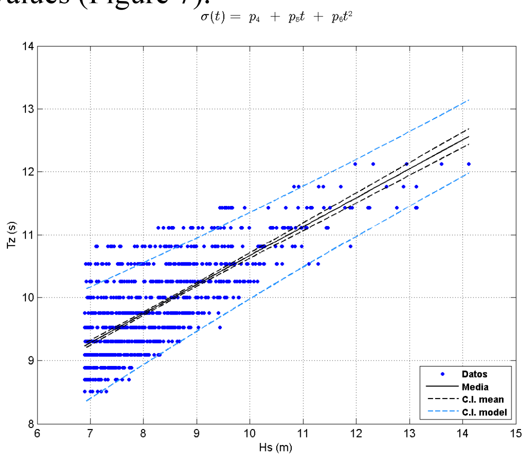


*Figure 2: Regression analysis to determine $T_z/H_s$ relationship at AMETS*

With P-Values governing the plotting of the curve as follows:

$$\mu(t) = p1 + p2t$$

| *p* | Value | Upper | Lower |
|---|---|---|---|
| *p1* | 6.026 | 6.196 | 5.856 |
| *p2* | 0.465 | 0.485 | 0.445 |

*Table 9: Equation parameters determined for wave period determined from regression analysis*

Using this, the values of $T_z$ for extreme $H_s$ values at 5, 10 and 20 year return periods were estimated.

| Return Period (Years) | $H_s$ (m) | $T_z$ (s) Reg. Model | Wave Steepness |
|---|---|---|---|
| 5 | 14.1 | 12.5 | 15.57 |
| 10 | 15.6 | 13.3 | 15.22 |
| 20 | 17.1 | 13.9 | 14.93 |

*Table 10: $T_z$ values predicted by regression*

The Wave steepness results as shown in Tables 8 & 10 broadly agree and indicate that a relatively reliable prediction can be achieved. Therefore the wave periods will continue to increase with the Hs value for higher return periods. This result is significant in terms of putting more certainty on the determination of the most appropriate $T_z$ value.

### *4.6. Calculation of extreme Wave Power*

The analysis also examined the determination of the extreme wave power values, but proved difficult due to the nonlinearity of wave power with increasing wave height. It was concluded that there was no added benefit to doing such an analysis in this case as the output was substantially different to power values determined using the predicted $H_s$ and $T_z$ values.

## 5. RELATED WORK

Below are listed a few studies that are relevant to the work in this paper.

- Jonathan, Philip & Ewans, Kevin (2013) raised the issue of covariate effects in extreme prediction in their paper "Statistical modelling of extreme ocean environments for marine design: A review,"
- They commented that interest may lay in estimating EV models for each variable independently (marginal modelling) or in joint modelling. The specific case given raises the desire to have associated values for $T_p$ at the extreme value of $H_s$.
- In the work they cite, "Dependence Measures for Extreme Value Analyses" - (Coles et al. 2000) on dependence in extreme value data, it is said that a standard method for multivariate extremes is based on distributions for which the variables are asymptotically independent. Given this asymptotic independence, we have strived to find a method to reliably predict the accompanying $T_z$ value for an extreme $H_s$ value.
- Bell (1972) found that the measured period of the highest individual wave has been found to always be longer but to have a wide range of values relative to $T_z$.
- Carter and Chaellenor (1990) found that Fischer-Tippett Type 1 and Weibull 1 give a good fit in British waters.

## 6. CONCLUSIONS AND FURTHER WORK

From this work the following methodology is proposed for the optimization of the prediction of extreme conditions at an Eastern Atlantic Site:

- Select the monthly extremes of $H_s$, $H_{max}$ values and the accompanying $T_z$ and $H_{max}$ values for the $H_s$ results.
- Determine the best fitting distribution for the data.
- Using regression analysis, identify trends in AI, Slope, and Wave Period.
- Estimate $H_s$ and $H_{max}$ extreme values using GEV or similar techniques.
- Estimate corresponding extreme $T_z$ and AI values using the results of the regression analysis.
- Ensure correlation between results using wave steepness at the extremes as a guiding value.

The method proposed above is an extension of the type of analysis already undertaken to determine extremes but is designed to provide more certainty in terms of the predicted values. What is most significant is that the $H_{max}$ and $T_z$ magnitudes can be calculated based on a particular methodology rather than just by assuming values of the anomaly index and wave steepness. This paper only outlines the work on one particular data set but this is now being extended to other measured and numerically generated datasets in order to validate the approach. The motivation from this work comes from the offshore renewable energy sector, whose sustainability is dependent on reducing both cost and risk. By the provision of more accurate design information there can be more certainty in terms of the survivability of a structure and thus this would help de-risk the project.

## 7. ACKNOWLEDGEMENTS

The authors acknowledge Marine Research Energy Ireland (MaREI), grant no. 12/RC/2302, a Science Foundation Ireland (SFI) supported project.

## 8. REFERENCES

Coles, S., Heffernan, J. & Tawn, J., 2000. Dependence Measures for Extreme Value Analyses. , pp.339–365.

Jonathan, P. & Ewans, K., 2013. Statistical modelling of extreme ocean environments for marine design: A review. *Ocean Engineering*, 62, pp.91–109. Available at: http://linkinghub.elsevier.com/retrieve/pii/S002980181300019X [Accessed November 27, 2014].

Mathiesen, M. et al., 1994. Recommended practice for extreme wave analysis. *Journal of Hydraulic Research*, 32(6), pp.803–814. Available at: http://dx.doi.org/10.1080/00221689409498691.

Mazas, F., Garat, P. & Hamm, L., 2014. Questioning MLE for the estimation of environmental extreme distributions. *Ocean Engineering*, 92, pp.44–54. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0029801814003679 [Accessed December 10, 2014].

You, Z.-J. & Callaghan, D., 2013. Discussion of "Modelling significant wave height distributions with quantile functions for estimation of extreme wave heights" [Ocean Eng. 54 (2012) 119–131]. *Ocean Engineering*, 70, pp.208–210. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0029801813002059 [Accessed December 10, 2014].