Proceedings of the ASME 2020 39th International Conference on Ocean, Offshore & Arctic Engineering
OMAE2020
June 28-July 2, 2020, Fort Lauderdale, FL, USA

OMAE2020-18308

DRAFT FOR OMAE 2020

ESTIMATION OF ENVIRONMENTAL CONTOURS USING A BLOCK RESAMPLING METHOD

Ed B.L. Mackay*
College of Engineering, Mathematics and Physical Sciences
University of Exeter
Cornwall TR10 9FE
United Kingdom
Email: e.mackay@exeter.ac.uk

Philip Jonathan
Department of Mathematics and Statistics
Lancaster University
Lancaster LA1 4YW
United Kingdom
Email: p.jonathan@lancaster.ac.uk

ABSTRACT

A new method for estimating joint distributions of environmental variables is presented. The key difference to previous methods is that the joint distribution of only storm-peak parameters is modelled, rather than fitting a model to all observations. This provides a stronger justification for the use of asymptotic extreme value models, as the data considered are approximately independent. The joint distribution of all data is recovered by resampling and rescaling storm histories, conditional on the peak values. This simplifies the analysis as much of the complex dependence structure is resampled, rather than modelled explicitly. The storm histories are defined by splitting the time series into discrete blocks, with the dividing points defined as the minimum value of a variable between adjacent maxima. Storms are characterised in terms of the peak values of each parameter within each discrete block, which need not coincide in time. The key assumption is that rescaling a measured storm history results in an equally realistic time series, provided that the change in peak values is not large. Two examples of bivariate distribution are considered: the joint distribution of significant wave height (H_s) and zero up-crossing period (T_z) and the joint distribution of H_s and wind speed. It is shown that the storm resampling method gives estimates of environmental contours that agree well with the observations and provides a rigorous method for estimating extreme values.

Keywords: Extremes; Metocean; Joint distribution;

1 INTRODUCTION

Offshore and coastal structures are often designed using the environmental contour method. The method involves estimating combinations of environmental variables (contours) that correspond to a given exceedance probability. The response of the structure is then estimated for various points along the contour and the maximum response along the contour is taken as the design value.

Generally, the process involves two stages: first, the joint distribution of the environmental variables is estimated, and secondly, contours are constructed from the joint distribution. Numerous method have been proposed for both steps. A recent review of methods for estimating environmental contours was presented in [1]. Methods for modelling the joint distribution include hierarchical conditional models [2–5], copula models [6–9], kernel density estimates [10, 11] and conditional extreme value models [12–14]. Methods for estimating contours include the inverse first- and second-order reliability methods (IFORM [4] and ISORM [15]), Monte Carlo methods [16, 17], projection methods [18] and isodensity contours [2, 19, 20]. For further discussion of the differences between the various methods see [1, 20].

For the purpose of estimating contours it is critical to capture the extremal properties of the variables in order to derive design conditions. Some of the methods mentioned above have limitations in this respect. For example, estimates of the tail of a distribution using a kernel density model are highly sensitive to

*Address all correspondence to this author.

the choice of kernel and kernel parameters. Hierarchical conditional models require a trend for the model parameters to be fitted and extrapolated outside the range of observations and there is no guarantee that the observed trend will continue outside the range of observations.

Moreover, many methods fit a model to all observations. The asymptotic models used for estimating the occurrence frequency of extreme events assume that the data are independent and identically distributed (IID). Metocean variables exhibit serial correlation, violating the model assumption that the data are IID. Neglecting serial correlation and fitting a model to all data can lead to a positive bias in the estimated extreme conditions (see e.g. [21]). For example, it was shown in [22] that estimates of the 100-year return values of wave heights that neglected serial correlation in the data can lead to positive biases of the order of 10-20%.

In univariate extreme value analysis, serial correlation is usually dealt with by only modelling peak values that are sufficiently separated in time that they can be considered independent. The selection of peak values can be consider as a form of declustering. For multivariate extreme value analysis this poses two problems. Firstly, a peak value in one variable need not coincide in time with a peak value the other variables. Secondly, the distribution of all observations must be recovered from the model for the distribution of just the storm peak variables.

Derbanne and de Hauteclocque [18] proposed a multivariate declustering method. For two random variables X and Y, the data are projected onto radial lines from the origin of the X-Y plane at various angles. For each angle, the projected univariate data are declustered and a peaks-over-threshold analysis is conducted to estimate exceedance probabilities in that particular direction. The process is repeated at discrete angular steps and the return values for each projection are used to construct contours. There are some limitations with this method. Firstly, as the boundary lines for each angular projection are straight, the resulting contours are convex, which is not appropriate for all combinations of environmental variables. Secondly, since all data are projected to a single line, the method can only be used to estimate an exceedance contour and not the joint distribution. Knowledge of the joint distribution is important for investigating the impact of short-term variability in the response, since the largest response in an N-year period may occur inside the N-year contour.

An alternative approach is to select a dominant environmental variable, X, (usually significant wave height, H_s, for metocean analysis), identify the peak values of X and establish the joint distribution of the peak values of X with concurrent values of the other variables. The distribution of the non-peak data is recovered by simulating peak values from the fitted model, resampling measured data conditional on the peak and rescaling them so that the values match at the time of the peak in X (see e.g. [14, 23]. As the peaks of the variables need not coincide in time, rescaling the measured data based on the values at the times of the peaks of X does not necessarily preserve the distribution of the peak values of the other variables.

In this work we propose a modified resampling approach, which preserves the distribution of the peak values of each variable. The method is illustrated through application to datasets selected for the environmental contour benchmarking exercise [24]. In the examples considered, the datasets contain only two variables. However, the resampling method can be extended without modification to higher dimensions (see e.g. [14]).

The paper is organised as follows. The block resampling method is outlined in Section 2 and the underlying assumptions are discussed. The application of the method to estimating environmental contours for H_s and zero up-crossing period, T_z is presented in Section 3 and the estimation of contours for H_s and wind speed is presented in 4. In both sections we present some novel methods for estimating the joint distribution of the peak variables. Finally, a discussion and conclusions are presented in Section 5.

2 OUTLINE AND ASSUMPTIONS OF METHOD

It is assumed that the time series of environmental variables can be divided into blocks where the peak values in adjacent blocks can be considered independent. The peak values of each variable within the block are not required to coincide in time, but the blocks are assumed to be sufficiently short so that the peak values of each variable are related in some way. A model for the joint distribution of the peak values is then estimated. The distribution of all data is recovered by simulating block-peak values from the joint model and resampling and rescaling the measured blocks so that the peak values from the resampled blocks match the simulated peak values.

The data are partitioned into non-overlapping blocks as follows. First, a dominant variable is selected (in both cases considered here we use H_s as the dominant variable). Peak values in the dominant variable are defined as maxima within a window of $\pm t$ days. The choice of t is dependent on the correlation scales of the variables considered. For oceanographic data, $t = 5$ days is sufficient for peak values to be considered independent. The end points of the blocks are defined as the minima between adjacent maxima.

The distribution of the data in the block relative to the peak values of each variable is dependent on the peak values. For example, the distribution of H_s/H_s^{peak} within a block is dependent on the value of H_s^{peak} (see e.g. Fig. 13 in [22]). Similarly, the joint distribution of variables within a block is dependent on the peak values. The blocks that are resampled and rescaled to match the simulated peak values must therefore have peak values that are "sufficiently close" that rescaling the data only results in a small change and results in a realistic storm history. In this work, we define "sufficiently close" in terms of the distance be-

tween the measured and simulated peak values when transformed to Laplace margins

$$r^2 = \left(F_L^{-1}(F_X(X_s^{peak})) - F_L^{-1}(F_X(X_m^{peak}))\right)^2 + \left(F_L^{-1}(F_Y(Y_s^{peak})) - F_L^{-1}(F_Y(Y_m^{peak}))\right)^2, \quad (1)$$

where the subscripts m and s denote measured and simulated values, F_X and F_Y are the marginal distribution functions of X^{peak} and Y^{peak}, and F_L is the Laplace distribution function. The resampled storm is then chosen at random from the n closest blocks, where n acts as a smoothing parameter, with larger values of n leading to more smoothing of the distribution. The X values in the resampled block are rescaled by the ratio X_s^{peak}/X_m^{peak} and the Y values are rescaled by Y_s^{peak}/Y_m^{peak}. The joint distribution of the resampled data is then estimated empirically from the 2D histogram.

A similar method was used in [22] to generate synthetic time series of H_s and T_z with the correct extremal properties of H_s. However, in [22] only the distribution of H_s was modelled and the values of significant steepness were resampled from the measured values within a given range of H_s. The other main differences from previous methods [14,23] are that the peak values are not required to coincide in time and the full block is resampled, rather than just the values for which H_s is above some threshold.

3 JOINT DISTRIBUTION OF H_s AND T_z

The data used in this section is dataset A described in [24], a 10-year time series of H_s and T_z covering the period 1/1/1996 - 31/12/2005. The data were measured by NDBC buoy 44007, located in the Gulf of Maine. The joint occurrence of H_s and T_z for this dataset is shown in Figure 1.

For these variables, the region of the environmental contour of interest includes both high and low values of T_z for a given H_s. In the storm resampling method, the data in each block are rescaled relative to the randomly simulated peak values. If the data are scaled for the maximum T_z then there is no guarantee that the maximum steepness elsewhere in the block will be realistic. Conversely, if we model peak H_s and steepness then the maximum period may not be realistic.

To circumvent this problem we define a variable such that the two frontiers of the H_s-T_z correspond to the highest values of steepness and the new variable. The significant steepness is defined as

$$s = \frac{2\pi H_s}{gT_z^2}. \quad (2)$$

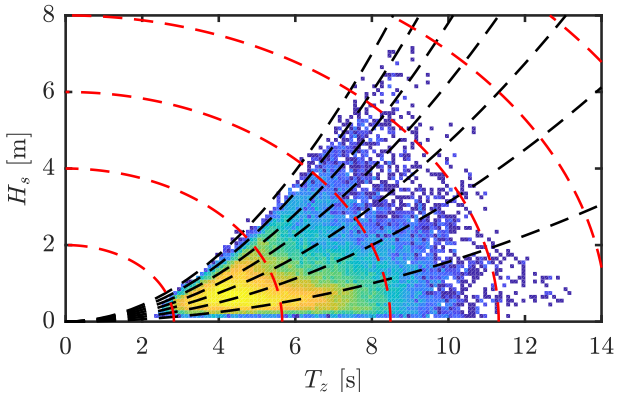

FIGURE 1. JOINT OCCURRENCE OF H_s AND T_z. BLACK DASHED LINES: CONSTANT STEEPNESS. RED DASHED LINES: CONSTANT DISTANCE (EQN. 3).

We define a distance from the origin in the H_s-T_z plane as:

$$d = \sqrt{H_s^2 + T_z^2/2}. \quad (3)$$

The definition of the distance is somewhat arbitrary, but this particular definition has the property that the lines of constant d are orthogonal to the lines of constant s in the H_s-T_z plane, as illustrated in Figure 1. The joint occurrence of s and d is shown in Figure 2. The distance parameter d does not have a physical interpretation, but is a convenient variable to use when examining the joint distribution of H_s and T_z. The use of other, physically based parameters such as wave power ($P \propto H_s^2 T_e$ in deep water, where T_e is the energy period) was examined, but these gave less satisfactory results.

The time series blocks are defined using the peak values of d with a minimum separation time of 5 days. The marginal distributions of s and d are defined as a 2-part model with a kernel density estimate used for the body of the data and a generalised Pareto (GP) model fitted to the tail of the distribution. The threshold for the GP model was selected from inspection of plots of the estimated GP shape parameter and return values against threshold level. The uncertainty from the threshold selection can be estimated by fitting an ensemble of models corresponding to different plausible threshold choices, however this method has not been applied here.

A similar approach is used to estimate the joint distribution of peak s and d. A kernel density model is used for the body of the data and the conditional extremes model of Heffernan and Tawn [25] (referred to as the HT model hereafter) is used to model the extreme values. For pairs of random variables (X,Y) on standard Laplace (or Gumbel) margins, the HT model

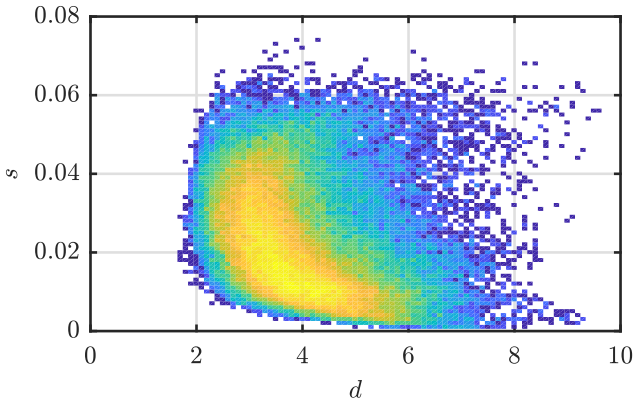

FIGURE 2. JOINT OCCURRENCE OF SIGNIFICANT STEEPNESS AND DISTANCE.

describes the asymptotic form of the distribution of one variable, conditional on the other being extreme. The model form for Y conditional on X under the HT model is

$$(Y|X=x) = \alpha x + x^{\beta} Z, \tag{4}$$

where Z is a residual process, α and β are real-valued, and $\beta < 1$. In the current work, the HT model is estimated by maximising the likelihood under the assumption that Z is normally distributed. After fitting the HT model, the density function of Z is estimated from the residuals $z_i = (y_i - \hat{\alpha} x_i)/x^{\hat{\beta}}$ using a kernel density model.

We are interested both in the behaviour of s when d is large and the behaviour of d when s is large. An HT model is therefore fitted for s conditional on large d and d conditional on large s. When both s and d are large there are two potential models to choose from. In this region the two models are blended as illustrated in Figure 3. Suppose the thresholds used for the HT models for $y|x$ and $x|y$ are u_x and u_y respectively. Define an angle $\theta = \text{atan}((y-u_y)/(x-u_x))$. The blended density function at a point $(x \geq u_x, y \geq u_y)$ is given by

$$p(x,y) = \left(1 - \frac{2\theta}{\pi}\right) p_1(x,y) + \frac{2\theta}{\pi} p_2(x,y), \tag{5}$$

$$p_1(x,y) = p(x)p(y|x), \tag{6}$$

$$p_2(x,y) = p(y)p(x|y). \tag{7}$$

The blending of the two density functions in this region does not guarantee that the total probability integrated over the domain is equal to 1. However, this is not problematic for simulation, since the conditional probability of $y|x$ can be normalised to integrate to 1 for each value of x.

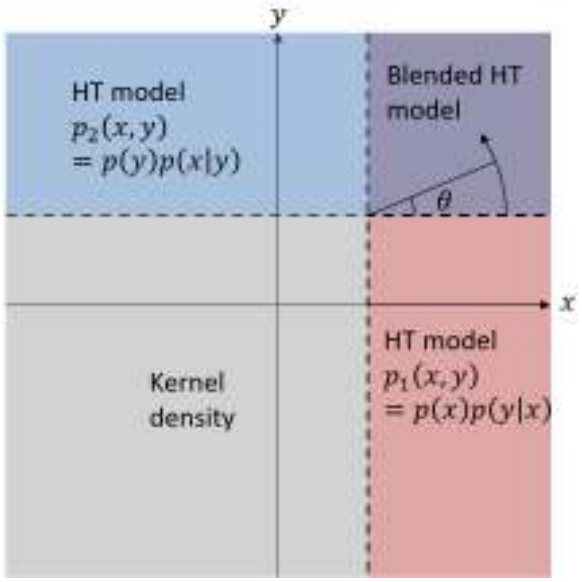

FIGURE 3. MODELS USED FOR RANGES OF VARIABLES.

In the present case, the HT thresholds for both variables were set at a non-exceedance probability of 0.65. The thresholds were selected by fitting the models for a range of threshold levels and selecting the lowest values for which the estimates of α and β tend to stable values. As with the GP threshold, the uncertainty relating to the threshold choice can be estimated by fitting an ensemble of models for ranges of plausible threshold choices, but this method has not been applied here.

Figure 4 shows the 10-years of observed data together with 1, 5 and 50-year contours estimated from 1000 years simulated data from the fitted model. The 1000-year simulation takes less than 1 minute on a standard laptop computer. The contours are calculated from the joint distribution using the IFORM method [4]. The resampled blocks have been drawn from the 10 closest measured blocks, as it was found that this gave a reasonable level of smoothing, without introducing significant changes to the shape of the joint distribution. Visually, the contours appear to be a good fit to the boundaries of the observed data, with several clusters of points outside the 1-year contour and a few points falling outside the 5-year contour. The points falling outside the contours would be expected to cluster, since conditions in individual storms exhibit strong serial correlation. Both the high steepness and high period portions of the contour are in good agreement with the data, implying that the transformation of the data and fitting the model in s-d space has worked well in preserving these features of the data.

The marginal exceedance probabilities for H_s and T_z from the observed and simulated data are shown in Figures 5 and 6. The observed marginal distributions are also well matched by the simulated data. The block resampling method together with

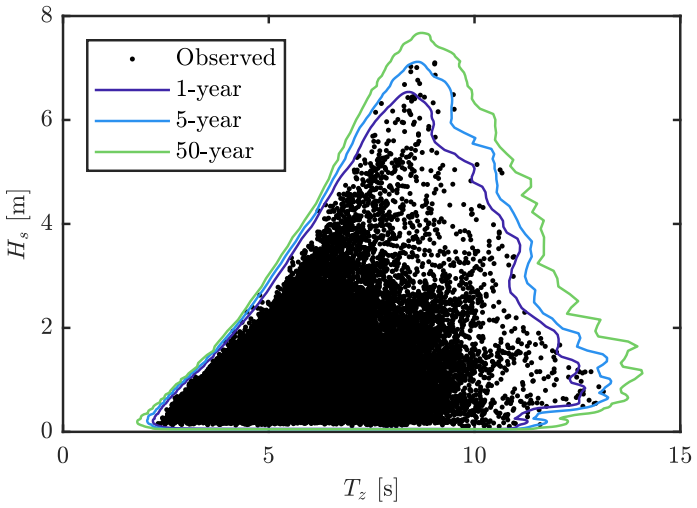

FIGURE 4. 10 YEARS OF OBSERVED H_s AND T_z WITH CONTOURS DERIVED FROM STORM RESAMPLING METHOD.

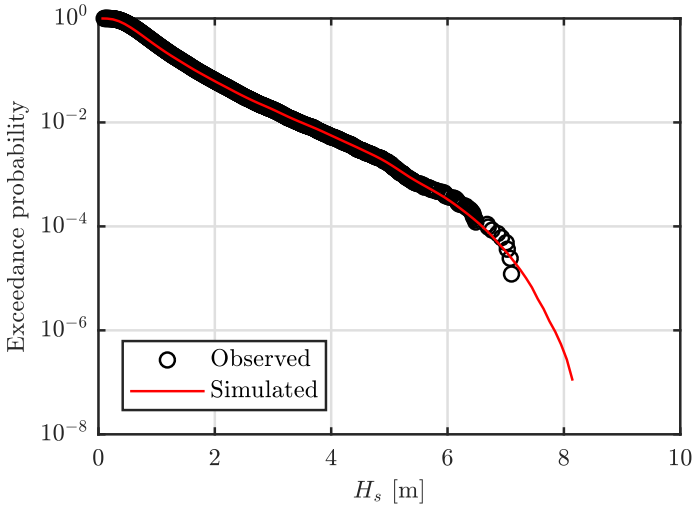

FIGURE 5. MARGINAL EXCEEDANCE PROBABILITY OF H_s FOR OBSERVED AND SIMULATED DATA FROM JOINT H_s-T_z MODEL.

the fitted marginal and joint models has performed well for this dataset and preserved the observed extremal properties of each variable.

4 JOINT DISTRIBUTION OF H_S AND WIND SPEED

The data used in this section is dataset F described in [24], a 50-year time series of H_s and wind speed at 10m above sea level, W, covering the period 1/1/1965 - 31/12/2014. The data is from

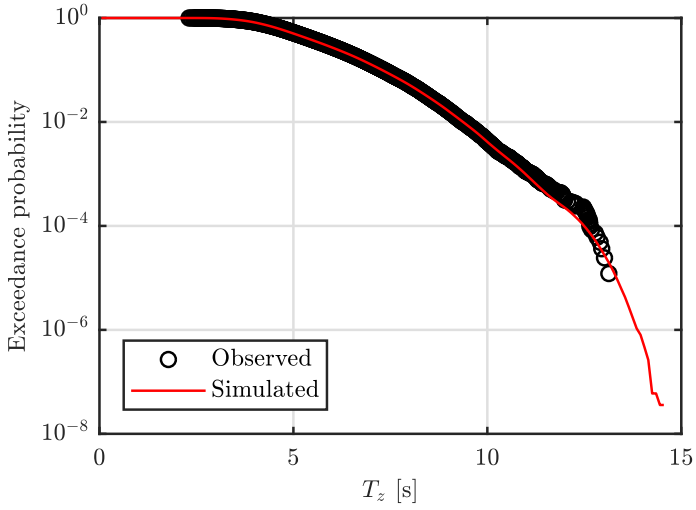

FIGURE 6. MARGINAL EXCEEDANCE PROBABILITY OF T_z FOR OBSERVED AND SIMULATED DATA FROM JOINT H_s-T_z MODEL.

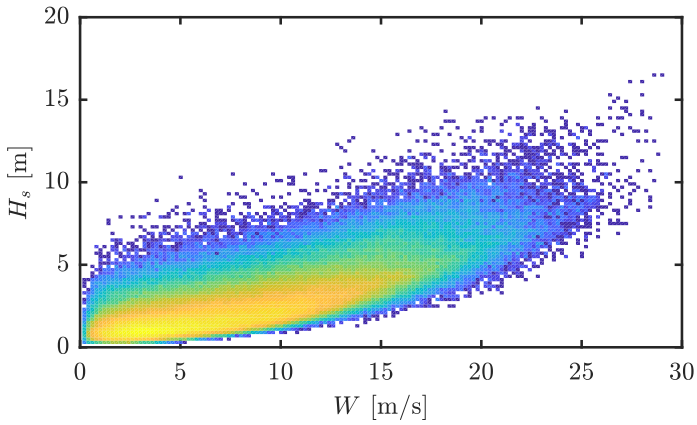

FIGURE 7. JOINT OCCURRENCE OF H_s AND WIND SPEED.

a hindcast model [26] for a North Sea location off the Norwegian coast. The joint occurrence of H_s and W is shown in Figure 7. The design of offshore wind turbines requires an estimate of the 50-year joint contour of H_s and W [27]. Values of H_s along the 50-year contour around the rated wind speed of the turbine are important for assessment of design loads. The strategy presented above for modelling H_s and T_z transformed the data to a space where the two frontiers of interest could be modelled using the HT model. The HT model is only appropriate for modelling the conditional distribution of a variable when the other is extreme, so if no transformation is applied to H_s and W then we require a different modelling strategy to estimate the extreme values of H_s when W is not extreme. A similar transformation as used for

H_s and T_z could be applied here, however a different approach has been implemented that does not require a transformation of variables.

One approach to model the tail of the distribution of H_s^{peak} conditional on W^{peak} would be to bin the data in small ranges of W^{peak} and fit a GP model in each bin. The bin size would be a trade-off between the sample sizes in each bin and the approximation that the distribution is stationary in each bin. This type of analysis can lead to significant bias and high variance (see [28, 29]). To avoid these problems and provide a model where the underpinning assumptions are better matched by the data, we use a penalised piecewise-linear (PPL) covariate model. The PPL model is described in detail in [30] and a brief overview is given below.

In the PPL model the values of H_s^{peak} above some local threshold are assumed to follow a GP distribution conditional on W^{peak}. The threshold and scale parameters of the GP model are estimated at a number of equally spaced values of W^{peak}, referred to as nodes, denoted w_k, $k = 1,2,...,m$. The threshold and scale parameters at intermediate positions vary linearly between the values at the nodes. Due to the difficulty in estimating the GP scale parameter, ξ, it is assumed that ξ is constant (but unknown) across the covariate domain. The model is similar to the penalised piecewise-constant (PPC) model described in [31], but with the key difference that the model parameters are assumed to vary linearly across each bin in the PPL model (where a bin is defined as the range of values between adjacent nodes), whereas the parameters are assumed to be constant in each bin in the PPC model.

The GP threshold at the nodes, $u(w_k)$, is intended to be set so that the local non-exceedance probability is a constant, $\psi \in (0,1)$. In practice, this is approximated by estimating the values of $u(w_k)$ using a simplex search optimisation method [32] (implemented using the MATLAB function '*fminsearch*') with a target of achieving a constant proportion, ψ, of observations in each bin below the threshold. The first guess for the optimisation is defined as the empirical quantiles at a non-exceedance value ψ in each bin, interpolated/extrapolated to the node positions.

Once the threshold has been set, the parameters ξ and $\{\sigma(w_k)\}$ are estimated by maximising the predictive performance of a roughness-penalised model using a cross-validation procedure. For a sample of observations of storm peak H_s and W, denoted $\{(h_i, w_i)\}_{i=1}^n$, the GP likelihood under the piecewise-linear model is

$$\mathscr{L} = \prod_{i:\, h_i > u(w_i)} \frac{1}{\sigma(w_i)} \left[1 + \frac{\xi}{\sigma(w_i)} \left[h_i - u(w_i)\right]\right]^{-1/\xi - 1}, \quad (8)$$

where the values $u(w_i)$ and $\sigma(w_i)$ are linearly interpolated from the values at the nodes. The negative log likelihood, penalised for the roughness of $\{\sigma(w_k)\}$, is then

$$\ell^* = -\log \mathscr{L} + \lambda_\sigma \sum_{k=1}^{m-2} \left[\sigma(w_{k+2}) - 2\sigma(w_{k+1}) + \sigma(w_k)\right]^2, \quad (9)$$

where λ_σ is the roughness penalty. The roughness is defined here as the change in the gradient of $\sigma(w)$ between adjacent bins. If $\lambda_\sigma = \infty$ then the model has two degrees of freedom (DOF) for σ, since the gradient of σ is forced to be constant across the domain. If $\lambda_\sigma = 0$ then the fitted model has m DOF for σ. For intermediate values, the "effective" number of DOF for σ is at some intermediate value.

For given ψ and λ_σ, estimates for ξ and $\{\sigma_k\}$ are found by minimising ℓ^*. The minimisation is conducted using a simplex search method [32]. The search is initialised using first guess of $\hat{\xi} = 0$ and the moment estimates of σ in each bin, interpolated/extrapolated to the node positions. The optimisation is constrained to give $\hat{\xi} \geq -0.5$ and $\max\{x_i\} \leq \hat{u}(w_i) - \hat{\sigma}(w_i)/\hat{\xi}$ when $\hat{\xi} < 0$. A random 10-fold cross-validation is then used to select the value $\hat{\lambda}_\sigma$ and corresponding $\hat{\xi}$, $\{\hat{\sigma}_k\}$ which maximise predictive performance (i.e. the value that gives the minimum value of ℓ^*). The cross-validation groups are defined so that the observations in each bin are split approximately equally between each group. The use of cross-validation ensures the best predictive rather than descriptive performance and avoids over-fitting.

The PPL model requires a sufficient number of observations in each bin to accurately estimate the parameters. There must be a sufficient number of observations around the first and last nodes that the estimates of $u(w_i)$, $\sigma(w_i)$, $i = 1,m$ are properly constrained by the data. This means that the PPL model can only be applied to the central range of W^{peak}. For the lowest values of W^{peak} and when H_s^{peak} is below the PPL threshold we estimate the joint distribution using a kernel density model, since these ranges do not influence the extreme values (note that all values of H_s and W in a block are, by definition, less than or equal to the peak values). For the upper range of W^{peak}, we estimate the conditional distribution of H_s^{peak} using the HT model. This also avoids issues with needing to extrapolate from the PPL model into unobserved ranges of W^{peak}.

The threshold for the HT model is defined to be strictly less than the maximum node position of the PPL model. In the overlapping region, the two models are blended linearly. In practice, this is achieved by simulating conditional values of H_s^{peak} from both models when $w_{HT} < W^{peak} < w_m$ (where w_{HT} is the threshold for the HT model) and selecting the value simulated from the PPL-kernel density model with probability $(w_m - w)/(w_m - w_{HT})$. The models applied for ranges of each variable are illustrated in Figure 8.

For the current dataset, the PPL model has been fitted using 6 equally-spaced nodes between $W^{peak} = 10$m/s and $W^{peak} =$

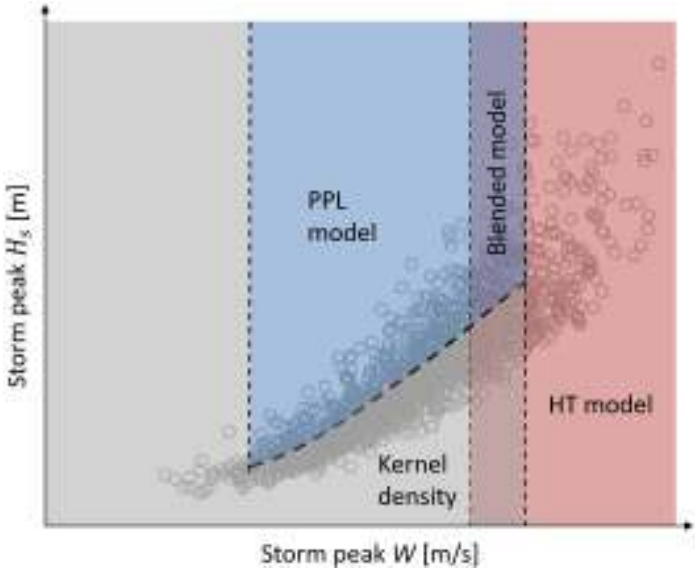

FIGURE 8. MODELLING STRATEGY USED FOR ESTIMATING JOINT DISTRIBUTION OF STORM PEAK H_s AND W.

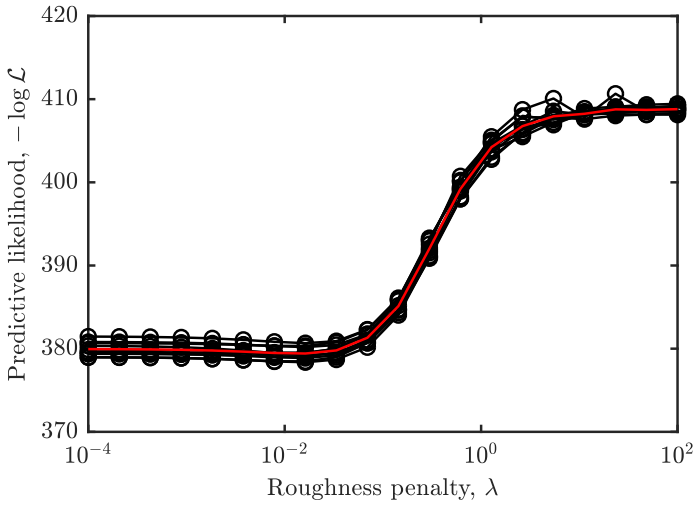

FIGURE 9. PREDICTIVE LIKELIHOOD AGAINST ROUGHNESS PENALTY. BLACK LINES: RESULTS FOR 10 RANDOM PARTITIONINGS OF DATA. RED LINE: AVERAGE OVER 10 REPEATS.

25m/s and a GP threshold non-exceedance probability $\psi = 0.7$. The predictive likelihood for $10^{-4} < \lambda_\sigma < 10^2$ is shown in Figure 9. There is some variability due to the random selection of cross-validation groups, so the cross-validation procedure was repeated 10 times. The results for each random partitioning into groups is shown, together with the average over the 10 repeats. For the present dataset, there is no clear minimum to the predictive likelihood, with any value $\lambda_\sigma < 5 \times 10^{-2}$ leading to roughly the same predictive likelihood. This indicates that there is sufficient information in the sample to estimate the six values of the scale parameter at the nodes without roughness-penalisation. For a smaller sample size or larger number of nodes we would expect to see a defined minimum in the predictive likelihood at some value of λ.

The fit of the PPL model is illustrated in Figures 10 - 12. Figure 10 shows the sample together with the threshold and quantiles from the fitted PPL model. It appears that the quantiles are a good match to the data with a roughly-evenly spread number of exceedances of the 0.99 quantile. We can define a normalised threshold exceedance as $z_i = (h_i - u(w_i))/\sigma(w_i)$. Under the assumption of a constant shape parameter, the normalised exceedances, z_i should have a stationary distribution with W^{peak}. Values of z_i against w_i are shown in Figure 11. Visually, it appears that distribution is approximately stationary, indicating that the assumption of a constant shape parameter is reasonable. Finally, under the model assumptions, the normalised threshold exceedances follow a GP distribution with $\sigma = 1$ and shape parameter $\hat{\xi}$. Figure 12 shows the empirical exceedance probabilities of the normalised threshold exceedances together with the GP

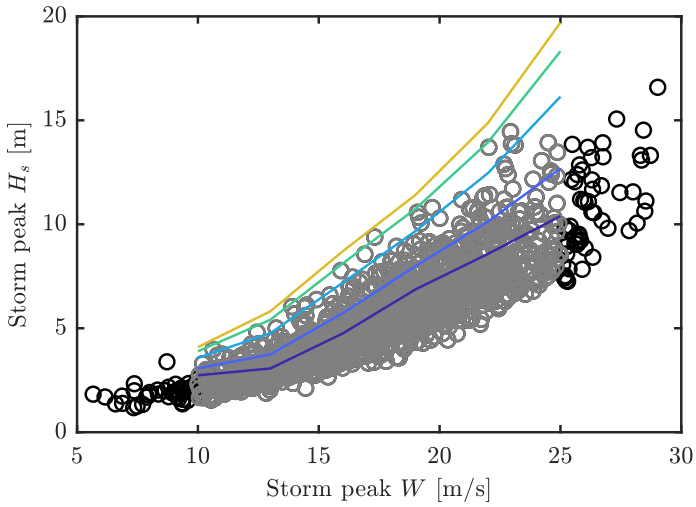

FIGURE 10. CIRCLES: OBSERVED VALUES OF STORM PEAK H_s AND W (OBSERVATIONS USED FOR FITTING THE PPL MODEL SHOWN IN GREY). LINES: ESTIMATED QUANTILES OF PPL MODEL AT EXCEEDANCE PROBABILITIES 0.3 (THRESHOLD LEVEL), 10^{-1}, 10^{-2}, 10^{-3} AND 10^{-4}.

model described above. The fitted model appears to be a good fit for the data with a relatively small scatter of the largest observations about the fitted model.

The threshold for the HT model was set at $W = 21$m/s and was selected in the same way as described in the previous section.

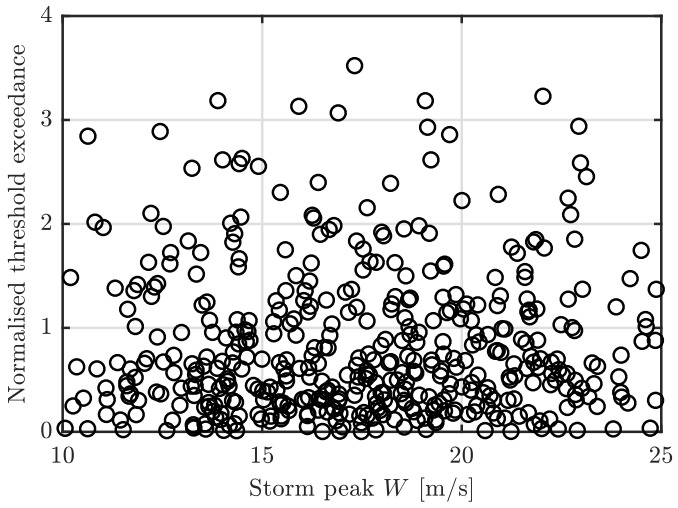

FIGURE 11. NORMALISED THRESHOLD EXCEEDANCES AGAINST STORM PEAK WIND SPEED.

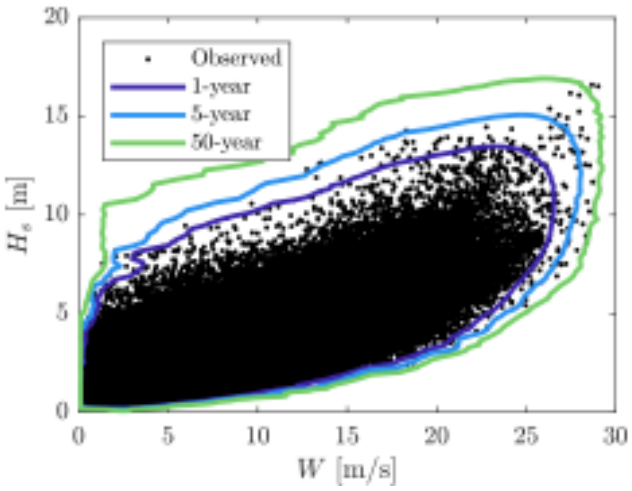

FIGURE 13. 50 YEARS OF OBSERVED H_s AND W WITH CONTOURS DERIVED FROM STORM RESAMPLING METHOD.

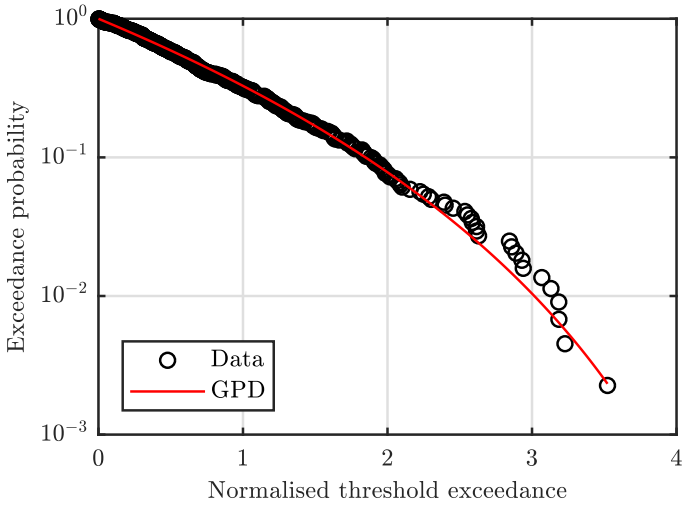

FIGURE 12. OBSERVED AND FITTED NORMALISED THRESHOLD EXCEEDANCES FROM PPL MODEL.

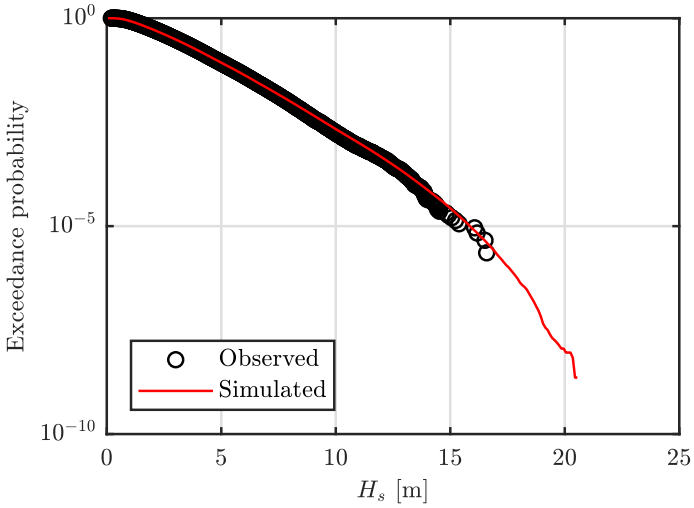

FIGURE 14. MARGINAL EXCEEDANCE PROBABILITY OF H_s FOR OBSERVED AND SIMULATED DATA FROM JOINT H_s-W MODEL.

Figure 13 shows the 50 years of observed H_s and W together with the 1-, 5- and 50-year contours derived from the block resampling method. In this plot the 20 nearest blocks are used for resampling and the total length of data is 10,000 years. The contours appear to follow the observed trends in the data well, with a reasonably even distribution of points exceeding the 1-year contour. There may be slightly fewer observations than expected where the value of H_s exceeds the 5-year contour at lower wind speeds, indicating that the model may be slightly conservative in this range. However, the marginal distribution of H_s and W are well matched between the observed and simulated data (see Figures 14 and 15), indicating that the model is performing well overall.

5 DISCUSSION AND CONCLUSIONS

A new method has been presented for estimating the joint distribution of environmental variables, which accurately captures both the marginal and joint the extremal characteristics of

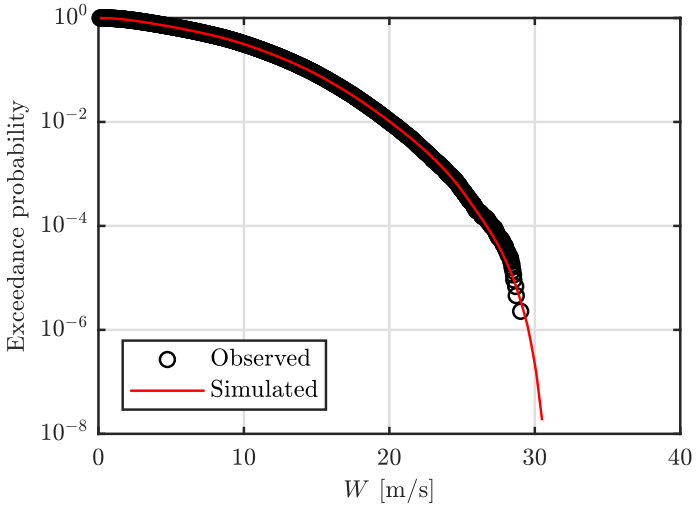

FIGURE 15. MARGINAL EXCEEDANCE PROBABILITY OF W FOR OBSERVED AND SIMULATED DATA FROM JOINT H_s-W MODEL.

the observations. The key assumption is that rescaling a measured storm history results in an equally realistic time series, provided that the change in peak values is not large. It was shown that the method, together with careful modelling of the joint distribution of peak values, can produce contours that closely follow the observed range of the data and accurately reproduce the marginal distributions of the data.

An aspect of the method that has not been explored in the present work is how it performs when extrapolating far outside the observed range of the data. The distribution of the data within each block is dependent on the peak values of each variable, so if measured blocks are rescaled to have significantly larger peak values, then this may result in unrealistic storm profiles. This could possibly be dealt with by characterising the joint distributions of the data within each block and examining how this varies with the peak values. Accounting for this effect may be important for long-tailed distributions in which there are rare occurrences far from the observed range, or when deriving contours for return periods orders of magnitude larger than the length of observations.

Another key feature of the block resampling is that it preserves the joint temporal variation in the variables within in each block. This means that arbitrary lengths of realistic time histories can be generated from the model. The values at the block boundaries will not match, but these values are defined to be non-extreme, so this is not important if the interest is in the extreme response of a structure. Neglecting serial correlation effects in the data is known to lead to positive bias in estimates of return values of individual wave and crest heights (see e.g. [22, 33, 34]). Using the resampling method to generate data from which extreme structural responses can be calculated could possibly lead to less conservative designs of offshore structures - something that may be particularly important for offshore renewable energy devices.

The use of block resampling can also give a more realistic estimate of the uncertainty in contour estimates than resampling all observations at random. Vanem et al [35] investigated the uncertainty in estimates of environmental contours by resampling the data under the assumption that each observation was independent. This method neglects the serial correlation in the data and the clustering of observations at the edges of the joint distribution. Neglecting this effect could potentially lead to non-conservative estimates of the uncertainty in the contours. However, further work would be required to quantify the impact of this effect.

ACKNOWLEDGMENT

We thank Andreas Haselsteiner for productive discussions that motivated this work. EM was funded under the Engineering and Physical Sciences Research Council (EPSRC) grant EP/R007519/1 and as part of the Tidal Stream Industry Energiser Project (TIGER), which has received funding from the European Union's INTERREG V A France (Channel) England Research and Innovation Programme, which is co-financed by the European Regional Development Fund (ERDF).

REFERENCES

[1] Ross, E., Astrup, O. C., Bitner-Gregersen, E., Bunn, N., Feld, G., Gouldby, B., Huseby, A., Liu, Y., Randell, D., Vanem, E., and Jonathan, P., 2019. "On Environmental Contours for Marine and Coastal Design". *Ocean Engineering,* ***In press***.

[2] Haver, S., 1985. "Wave climate off northern Norway". *Applied Ocean Research,* ***7***, pp. 85–92.

[3] Mathisen, J., and Bitner-Gregersen, E., 1990. "Joint distributions for significant wave height and wave zero-up-crossing period". *Applied Ocean Research,* ***12***, pp. 93–103.

[4] Winterstein, S. R., Ude, T. C., Cornell, C. A., Bjerager, P., and Haver, S., 1993. "Environmental parameters for extreme response: Inverse FORM with omission factors". In 6th International Conference on Structural Safety & Reliability (ICOSSAR).

[5] Li, L., Gao, Z., and Moan, T., 2013. "Joint environmental data at five European offshore sites for design of combined wind and wave energy devices". In 32nd International Conference on Ocean, Offshore and Arctic Engineering, pp. OMAE2013–10156.

[6] Montes-Iturrizaga, R., and Heredia-Zavoni, E., 2015. "En-

vironmental contours using copulas". *Applied Ocean Research,* ***52***, pp. 125–139.

[7] Montes-Iturrizaga, R., and Heredia-Zavoni, E., 2016. "Multivariate environmental contours using C-vine copulas". *Ocean Engineering,* ***118***, pp. 68–82.

[8] Manuel, L., Nguyen, P. T., Canning, J., Coe, R. G., Eckert-Gallup, A. C., and Martin, N., 2018. "Alternative approaches to develop environmental contours from metocean data". *Journal of Ocean Engineering and Marine Energy,* ***4***, pp. 293–310.

[9] Fazeres-Ferradosa, T., Taveira-Pinto, F., Vanem, E., Reis, M. T., and das Neves, L., 2018. "Asymmetric copula-based distribution models for met-ocean data in offshore wind engineering applications". *Wind Engineering,* ***42***, pp. 304–334.

[10] Eckert-Gallup, A., and Martin, N., 2016. "Kernel density estimation (KDE) with adaptive bandwidth selection for environmental contours of extreme sea states". In OCEANS 2016 MTS/IEEE, pp. 1–5.

[11] Haselsteiner, A., Ohlendorf, J.-H., and Thoben, K.-D., 2017. "Environmental contours based on kernel density estimation". In 13th GermanWind Energy Conference DEWEK.

[12] Jonathan, P., Flynn, J., and Ewans, K. C., 2010. "Joint modelling of wave spectral parameters for extreme sea states". *Ocean Engineering,* ***37***, pp. 1070–1080.

[13] Jonathan, P., Ewans, K., and Randell, D., 2014. "Non-stationary conditional extremes of northern North Sea storm characteristics". *Environmetrics,* ***25***, pp. 172–188.

[14] Hansen, H. F., Randell, D., Zeeberg, A. R., and Jonathan, P., 2020. "Directional–seasonal extreme value analysis of North Sea storm conditions". *Ocean Engineering,* ***195***, p. 106665.

[15] Chai, W., and Leira, B. J., 2018. "Environmental contours based on inverse SORM". *Marine Structures,* ***60***, pp. 34–51.

[16] Huseby, A. B., Vanem, E., and Natvig, B., 2013. "A new approach to environmental contours for ocean engineering applications based on direct Monte Carlo simulations.". *Ocean Engineering,* ***60***, pp. 125–135.

[17] Jonathan, P., Ewans, K., and Flynn, J., 2014. "On the estimation of ocean engineering design contours". *Journal of Offshore Mechanics and Arctic Engineering,* ***136***, p. 41101.

[18] Derbanne, Q., and de Hauteclocque, G., 2019. "A new approach for environmental contour and multivariate declustering". In 38th International Conference on Ocean, Offshore and Arctic Engineering, p. OMAE2019/95993.

[19] Haver, S., 1987. "On the joint distribution of heights and periods of sea waves". *Ocean Engineering,* ***14***, pp. 359–376.

[20] Haselsteiner, A. F., Ohlendorf, J.-H., Wosniok, W., and Thoben, K.-D., 2017. "Deriving environmental contours from highest density regions". *Coastal Engineering,* ***123***, pp. 42–51.

[21] Coles, S., 2001. *An Introduction to Statistical Modeling of Extreme Values*. Springer.

[22] Mackay, E., and Johanning, J., 2018. "Long-term distributions of individual wave and crest heights". *Ocean Engineering,* ***165***, pp. 164–183.

[23] Feld, G., Randell, D., Wu, Y., Ewans, K., and Jonathan, P., 2014. "Estimation of storm peak and intra-storm directional-seasonal design conditions in the North Sea". *33rd International Conference on Ocean, Offshore and Arctic Engineering*, pp. OMAE2014–23157.

[24] Haselsteiner, A. F., Coe, R. G., Manuel, L., Nguyen, P. T. T., Martin, N., and Eckert-Gallup, A., 2019. "A benchmarking exercise on estimating extreme environmental conditions: Methodology & baseline results". In 38th International Conference on Ocean, Offshore and Arctic Engineering, pp. OMAE2019–96523.

[25] Heffernan, J. E., and Tawn, J. A., 2004. "A conditional approach for multivariate extreme values". *J. R. Statist. Soc. B,* ***66***, pp. 497–546.

[26] Groll, N., and Weisse, R., 2017. "A multi-decadal wind-wave hindcast for the North Sea 1949-2014: CoastDat2". *Earth System Science Data,* ***9***, pp. 955–968.

[27] International-Electrotechnical-Commission, 2019. "Wind energy generation systems - Part 3-1: Design requirements for fixed offshore wind turbines". *IEC 61400-3-1:2019*.

[28] Mackay, E. B., Challenor, P. G., and Bahaj, A. B. S., 2010. "On the use of discrete seasonal and directional models for the estimation of extreme wave conditions". *Ocean Engineering,* ***37***, pp. 425–442.

[29] Mackay, E., and Jonathan, P. "Assessment of Return Value Estimates from Stationary and Non-Stationary Extreme Value Models". *Ocean Engineering (submitted)*.

[30] Mackay, E. B. L., and Jonathan, P., 2020. "A piecewise linear extreme value model for variables with covariate dependence". *(In preparation)*.

[31] Ross, E., Sam, S., Randell, D., Feld, G., and Jonathan, P., 2018. "Estimating surge in extreme North Sea storms". *Ocean Engineering,* ***154***, pp. 430–444.

[32] Lagarias, J. C., Reeds, J. A., Wright, M. H., and Wright, P. E., 1998. "Convergence Properties of the Nelder-Mead Simplex Method in Low Dimensions". *SIAM Journal of Optimization.,* ***9***, pp. 112–147.

[33] Tromans, P. S., and Vanderschuren, L., 1995. "Response based design conditions in the North Sea: application of a new method". *Offshore Technology Confernence, Houston (OTC–7683)*.

[34] Forristall, G., 2008. "How should we combine long and short term wave height distributions?". In 27th International Conference on Offshore Mechanics and Arctic En-

gineering, pp. OMAE2008–58012.

[35] Vanem, E., Gramstad, O., and Bitner-Gregersen, E. M., 2019. "A simulation study on the uncertainty of environmental contours due to sampling variability for different estimation methods". *Applied Ocean Research,* ***91***, p. 101870.