Proceedings of the ASME 2021 40th International Conference on Ocean, Offshore & Arctic Engineering OMAE2021 June 21-30, 2021, Virtual, Online

OMAE2021-64874

PREPRINT: A SECOND BENCHMARKING EXERCISE ON ESTIMATING EXTREME ENVIRONMENTAL CONDITIONS: METHODOLOGY & BASELINE RESULTS

Ed Mackay
University of Exeter
Penryn, UK

Andreas F. Haselsteiner
University of Bremen
Bremen, Germany

Ryan G. Coe
Sandia National Laboratories
Albuquerque, NM, USA

Lance Manuel
The University of Texas at Austin
Austin, TX, USA

ABSTRACT

Estimating extreme environmental conditions remains a key challenge in the design of offshore structures. This paper describes an exercise for benchmarking methods for extreme environmental conditions, which follows on from an initial benchmarking exercise introduced at OMAE 2019. In this second exercise, we address the problem of estimating extreme metocean conditions in a variable and changing climate. The study makes use of several very long datasets from a global climate model, including a 165-year historical run, a 700-year pre-industrial control run, which represents a quasi-steady state climate, and several runs under various future emissions scenarios. The availability of the long datasets allows for an in-depth analysis of the uncertainties in the estimated extreme conditions and an attribution of the relative importance of uncertainties resulting from modelling choices, natural climate variability, and potential future changes to the climate. This paper outlines the methodology for the second collaborative benchmarking exercise as well as presenting baseline results for the selected datasets.

1 INTRODUCTION

The design of offshore systems is reliant on an understanding of the environmental conditions. Survival, in particular, represents a critical aspect for which characterization of extreme environmental conditions represents a important and challenging step. Many offshore structures are designed for lifetimes of the order of 25 to 50 years [1–4]. Given the expense of making measurements at sea, available data are spatially sparse and are often of a limited duration. Measured data are usually complemented with hindcast data, typically of the order of 20-50 years in length.

In practice, a designer must estimate extreme responses for the structure under consideration, based on limited data. Approaches for estimating long-term extreme responses vary significantly in their computational requirements. Response-based methods (e.g. [5]) require the structural response to be determined over the full range of environmental conditions and extreme responses are estimated directly from the response data rather than environmental conditions. In contrast, in simplified approaches, such as the environmental contour method, the response is only estimated for environmental conditions along a contour. The environmental contour approach is widely recommended by design guidelines and standards [1, 2, 4, 6], as it represents a pragmatic balance between computational requirements and statistical and dynamical rigor.

Given both the wide use of the environmental contour method and the wide variety of methods proposed for producing environmental contours, a benchmarking exercise was recently conducted to compare a variety of environmental contour methods (this exercise is referred to herein as “EC Benchmark 1”) [7]. EC Benchmark 1 provided a set of data and common metrics by which a blind comparison could be made. A total of eleven submissions were made to the exercise; the results of which have recently been compiled [8].[1] The study highlighted several important factors which influence results:

Independence of data: Most contour prediction methods are based on the assumption that each data point is independent. However, there are high levels of serial correlation in metocean conditions over several days [9]. The assumption of independence in the presence of serial correlation can introduce bias into both estimated return values and associated confidence intervals.

Contour method: The method used to construct contours from the estimated joint distribution can lead to significant differences in resulting contours. The inverse first-order reliability method (IFORM) [10] and direct sampling [11] methods define contours in terms of marginal exceedances probabilities under rotations of the coordinate system, either in standard normal space or the original parameter space. In contrast, the inverse second-order reliability method (ISORM) [12] and highest density contours [13] are defined in terms of the total probability outside the contour. These two definitions can lead to contours which include points with marginal exceedance probabilities that differ by orders of magnitude [14].

Statistical model: The choice of statistical model for the joint distribution and method used to estimate the parameters can have a larger influence on the differences between the contours than the choice of method used to construct the contour. It was also found that the fit of the models selected was site-dependent.

Based on the findings from EC Benchmark 1, a second phase of benchmarking exercises for extreme conditions has been designed (referred to herein as "EC Benchmark 2"[2]). Given the importance of model choice and parameter estimation method, highlighted by EC Benchmark 1, EC Benchmark 2 has been designed to further investigate these steps. In this exercise, the focus is on univariate analyses in order to highlight the importance of fitting method and model selection.

A wide range of methods are available for estimating extreme metocean conditions (see [15] for a review). Commonly used approaches include the initial distribution method [4, §3.6.2.1], annual-maxima (e.g. [16, 17]), the r-largest-order statistic method (e.g. [18, 19]), peaks-over-threshold (POT) (e.g. [20–22]), and ACER [23]. For each type of method, practitioners can chose to fit different types of distribution to the data. For example, for the initial distribution method, the Weibull, log-normal or Gamma distributions are commonly used [4, 24] (see [25] for a recent review). For annual maxima the Gumbel distribution is recommended in some standards [4], whilst the generalised extreme value (GEV) distribution is more commonly recommended in statistical texts [26]. For POT, the Weibull, exponential distributions are recommended in some standards [4], whereas the generalised Pareto (GP) distribution is recommended in statistical texts, based on asymptotic arguments [26]. The method used for estimating distribution parameters also influences results [27]. For the GP distribution, a wide range of options have been proposed (see e.g. [28, 29]). For the POT method, the choice of threshold and declustering criteria is also subjective [30, 31]. Various tests for independence between adjacent samples have been proposed, such as the Blum test for block maxima [32, 33] or the extremogram for identifying declustering criteria [34].

Many authors have also proposed models which account for covariate effects such as seasonality or directionality [35–39]. These introduce further modelling choices, such as binning of data, smoothing of parameter variation (using, e.g., splines, Fourier series or Gaussian processes), or methods used to select the optimal 'roughness' penalties for parameter variation. Interference for covariate models can also be far more computationally demanding than for stationary models (see e.g [40]).

Many studies on estimation of extreme conditions tacitly assume that the wave climate is stationary. However, there is evidence that wave heights and wind speeds are dependent on large-scale climate patterns which vary slowly over multiple years, such as the North Atlantic Oscillation (NAO) [41] or the El Niño Southern Oscillation (ENSO) [42]. Moreover, anthropogenic climate change is also reported to have had a measurable effect on metocean conditions (e.g. [43–48]). Some authors have suggested methods for detecting trends in historic extreme events [49] or using non-stationary models to account for trends in historic observations over time [50], whilst others have used predictions of future metocean conditions from outputs of global climate models (GCMs) (e.g. [51–53]).

A challenge in assessing the goodness-of-fit of statistical models for extreme conditions is that most measurement or hindcast datasets are relatively short, meaning that there is high sampling variability associated with the largest observations. To address this issue, in EC Benchmark 2, we makes use of several long datasets of wave parameters generated from the FIO-ESM v2.0 global climate model [54, 55]. These datasets include a 700-year dataset for a quasi-steady state climate, a 165-year historical dataset, three 85-year datasets for future climates under various emissions scenarios, and two 150-year CO_2 sensitivity experiments.

The use of data from a GCM allows us to address several important questions in estimation of extreme metocean conditions. In EC Benchmark 2 we propose two exercises. The first makes use of the data from the 700-year quasi-steady state cli-

[1] The results and code for replicating the analyses in EC Benchmark 1 are available online: https://github.com/ec-benchmark-organizers/ec-benchmark

[2] In EC Benchmark 1, EC referred to "environmental contours", in EC Benchmark 2 the same acronym is used for continuity, to refer to "extreme conditions".

mate runs to assess the accuracy of estimates of extremes from relatively short datasets, without the influence of anthropogenic climate change. The second exercise investigates the estimation of extremes in a changing climate, making use of the historical data and future scenarios. Participants can choose to address one or both of these exercises.

The goal of EC Benchmark 2 is to examine differences in estimates of extremes resulting from the wide range of modelling choices available to practitioners. For the first exercise, the availability of the 700-year dataset will allow a quantitative assessment of the predictions of extremes at lower return periods (e.g., 50-year return values, which are used in the design of offshore renewable energy structures [1, 2]). As with EC Benchmark 1, an additional objective of the exercise is to prompt further development of the state of the art in estimation of extreme metocean conditions and promote discussion and collaboration between researchers in this area.

2 DESCRIPTION OF BENCHMARKING EXERCISE

2.1 Datasets

As noted in the introduction, the data used for EC Benchmark 2 is taken from the FIO-ESM v2.0 climate model [54, 55]. The 700-year quasi-steady state runs are based on forcing from pre-industrial conditions, with 1850 as the reference year for the climate-forcing variables. The model was integrated for 1000 years, reaching a quasi-equilibrium state after 300 years. The data used in EC Benchmark 2 are from the last 700 years in quasi-steady state. The historical simulations cover the period 1850-2014 and are based on climate-forcing variables from CMIP6 [56]. The three future scenarios are for low, medium and high future forcing pathways (scenarios SSP126, SSP245 and SSP585 in the CMIP6 experiments) and cover the period 2015-2100.

The datasets contain values of significant wave height H_s, peak period T_p, zero up-crossing period T_z, and mean wave direction D_m at 3 hour intervals for a duration of 700-years. The wave model used a grid with approximately 1° resolution, so it is not expected to be able to accurately replicate the spatial variability in extreme wave conditions close to shore and may not be able to properly resolve extreme conditions generated by tropical cyclones.

In EC Benchmark 1, datasets were provided for three locations in the North Sea, two locations on the northern and southern ends of the US East Coast and one in the Gulf of Mexico (GOM). As the GOM and south-eastern coastline of the US are affected by tropical cyclones, these locations were not selected for analysis in the second exercise. Also, since the FIO-ESM v2.0 wave model has a relatively coarse resolution, the two locations in the central and southern North Sea were not used in EC Benchmark 2.

Three locations were chosen for analysis in EC Benchmark 2. Sites have been selected at locations where reference data are available for comparison (either from wave buoys or hindcasts) and the model is expected to perform reasonably well (i.e., sufficiently far from the coast and in areas not affected by tropical cyclones). The locations of the sites selected for analysis are illustrated in Figure 1 and listed in Table 1, together with details of the reference datasets used for comparison.

Site 1 is located off the Norwegian coast, at the location of one of the datasets used in EC Benchmark 1. The reference data is taken from the coastDat2 hindcast [57] and covers the 50-year period 1965-2014. The site is in approximately 265 m water depth and is close to the Utsira Nord offshore wind farm lease area [58].

Site 2 is located close to the edge of the continental shelf on the US East Coast, approximately 100 km southeast of Nantucket. The site is in approximately 75 m water depth and is close to the Massachusetts offshore wind farm lease area. The location corresponds to the site of NDBC buoy 44008, where data are available from 1982 to the present day.

Site 3 is located off Oregon on the US West Coast, in approximately 160 m water depth. The site is close to the PacWave wave energy test site. The reference data are taken from NDBC buoy 46050, where measurements are available from 1982 until the present day.

To justify the use of data from the climate model, a brief comparison of data from the historical runs with the reference datasets is presented in Appendix A. The comparisons indicate that the FIO-ESM data gives realistic distributions of H_s and its extremes, from which meaningful results can be inferred. Participants should note that the reference data listed in Table 1 is only used for comparison with the FIO-ESM data to justify its use, and is not used directly in the benchmarking exercises.

2.2 Exercises

The focus of the exercises is on estimating return values of H_s and associated confidence intervals. Return values can be defined in a number of ways that are asymptotically equivalent for large return periods (see, e.g., [59]). A common definition of the T-year return value, is the value that is exceeded with probability $1/T$ in a given year. That is, if $F_A(x)$ is the cumulative distribution function (CDF) of the annual maximum H_s, the T-year return value $H_{s,T}$ is defined as the solution of

$$F_A(H_{s,T}) = 1 - \frac{1}{T}. \qquad (1)$$

However, if it is assumed that the annual distribution is non-stationary, either due to 'natural' long-term variability in the climate or due to anthropogenic climate change, then F_A will be time-dependent. In this situation, it can be clearer to quantify extremes in terms of the CDF of the maximum value in a specified

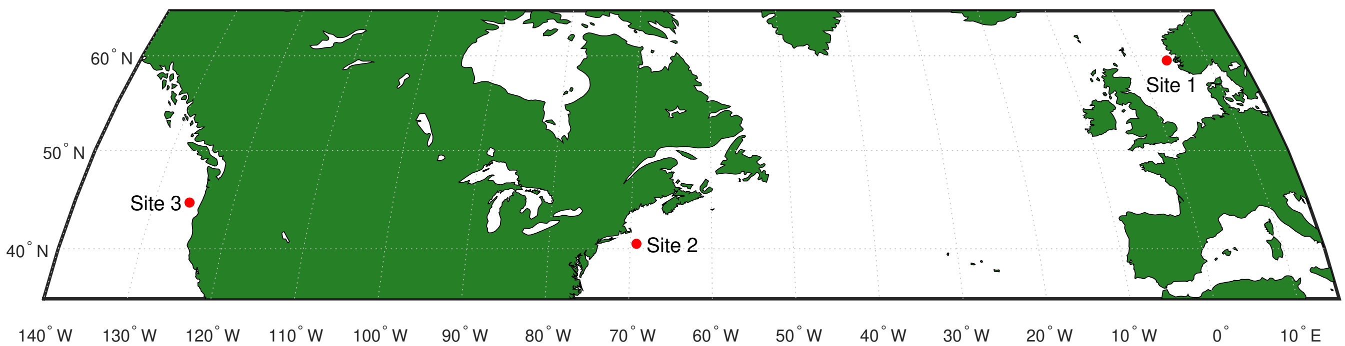

FIGURE 1. LOCATIONS OF DATASETS USED IN BENCHMARKING STUDY.

TABLE 1. LOCATIONS OF SITES USED IN EC BENCHMARK 2 AND REFERENCE DATA SOURCES USED FOR COMPARISON.

	Location	Lat / Lon	Water depth [m]	Reference data source	Reference coverage
Site 1	Norwegian Coast	59.500 N, 4.325 E	265	coastDat-2 hindcast	01/01/1965 - 31/12/2014
Site 2	Massachusetts, US East Coast	40.504 N, 69.248 W	75	NDBC buoy 44008	18/08/1982 - present
Site 3	Oregon, US West Coast	44.669 N, 125.546 W	160	NDBC buoy 46050	16/11/1991 - present

period of T years, denoted $F_T(x)$. Under the assumption of a stationary climate, the distribution of the maximum H_s in a T-year period is given by $F_T(x) = (F_A(x))^T$. Substituting the definition (1) into this expression we have

$$F_T(H_{s,T}) = (F_A(H_{s,T}))^T = \left(1 - \frac{1}{T}\right)^T. \tag{2}$$

Since $(1-1/T)^T \to \exp(-1)$ as $T \to \infty$, the T-year return value can be defined as the solution of

$$F_T(H_{s,T}) = \exp(-1) \approx 0.3679. \tag{3}$$

In the two exercises described below, we make use of both definitions (1) and (3). Participants are free to estimate $F_A(x)$ and $F_T(x)$ using any method of their choice and are not restricted to using annual maxima methods. For example, under the assumption of a stationary climate, if $F_P(x)$ is the distribution of peaks-over-threshold, with m independent peaks expected per year on average, then $F_A(x) = (F_P(x))^m$. Non-parametric approaches may also be employed to predict return values and associated confidence intervals, and participants are not required to use solely parametric models such as Gumbel, GEV, etc. applied to the data.

The following subsections present the two exercises for EC Benchmark 2 in detail. The first exercise focuses on estimating return values of H_s for the quasi-steady state climate and the second exercise focuses on predicting future climate based on historic climate.

2.2.1 Exercise 1 - Estimation of extremes in a steady state climate:

In this exercise, participants will be provided with a 25-year times series for each of the three sites, taken from the pre-industrial control dataset. The datasets contain values of H_s, T_p, T_z, and D_m. A central portion of the 700-year pre-industrial control dataset has been used, running from 1st July in year 350 to 30th June in year 374 (see Figure 2). Note that these years do not correspond to historical years, but time since the start of the quasi-steady state period. The summer-to-summer period has been chosen so that a full winter period is available for both the start and end years. Participants to the exercise should note that all years are a fixed length of 365 days, with no leap years.

Participants are asked to provide estimates of return values together with a 95% confidence interval (CI), for return periods of 5, 50, 500 years, for each of the three sites. As mentioned in the introduction, even in a nominally steady-state climate, it is possible that there will be long-term trends in the wave climate, related to large-scale climate patterns. In this case, the distribution of the annual maximum, $F_A(x)$, could be modelled as varying from year to year, dependent on a large-scale climate index. In this exercise, we ask participants to estimate the long-term average value of $F_A(x)$ and use definition (1) to calculate return values. Participants are free to use any method of their

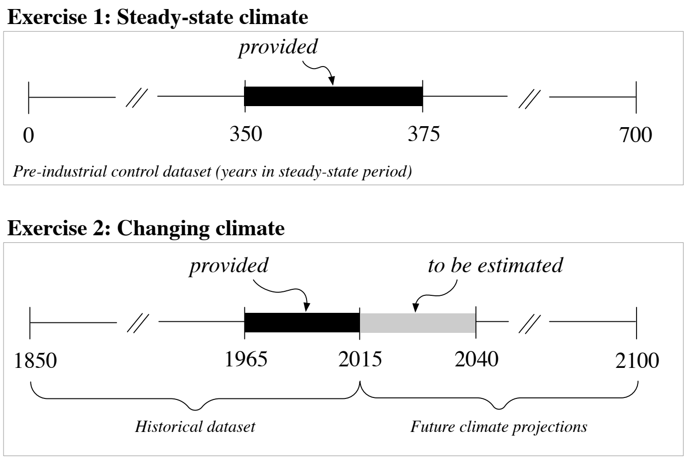

FIGURE 2. SECTIONS OF TIME SERIES FROM FIO-ESM DATASETS USED IN EACH EXERCISE.

choice for estimating $F_A(x)$ and the confidence interval for the return values.

For graphical comparison, participants are also asked to provide quantiles of F_A and an associated 95% CI at non-exceedance probabilities, $P = 1 - 1/N$ for $N \in [1, 1000]$ at 100 equally spaced logarithmic increments, corresponding to return periods between 1 and 1000 years (e.g. using MATLAB, N would be defined as: `N=logspace(0,3,100);` or in Python `N=numpy.logspace(0,3,num=100)`). Finally, we also ask participants to provide a description of the method used to estimate the annual distribution, $F_A(x)$ and associated CI.

2.2.2 Exercise 2 - Estimation of extremes in a changing climate:

This exercise follows a similar format to Exercise 1, the difference being that the provided datasets are drawn from the historical runs of the FIO-ESM model. Participants are provided with a 50-year time series covering the period 01/01/1965 - 31/12/2014 from the historic run of the FIO-ESM model, for each of the three sites. Participants are asked to estimate the distribution of the maximum H_s in the subsequent 25 year period, (01/01/2015 - 31/12/2039), denoted $F_{25}(x)$. This metric is used in favour of return values defined in terms of the annual distribution (1), due to potentially non-stationary trends in the climate. The exercise is intended to be representative of a typical design problem, where conditions over the lifetime of a proposed deployment must be estimated based on historic data (see Figure 2).

Participants are free to use any method of their choice for estimating F_{25} and the associated CI. It is optional whether participants choose to model historic or future trends in the extreme wave climate. As the future CO_2 emissions are unknown, participants are asked to provide an estimate of F_{25} and associated CI, which reflects this uncertainty. Participants could choose to provide different estimates for different scenarios or a single estimate for all future scenarios.

Participants are asked to provide quantiles of $F_{25}(x)$ at non-exceedance probabilities $(1 - 1/N)^{1/25}$ for $N = 5$, 50 and 500 and an associated 95% CI. Under the assumption of a stationary climate, these would correspond to the 5, 50 and 500-year return values defined using definition (1). Similar to Exercise 1, for graphical comparison, participants are also asked to provide quantiles of F_{25} and an associated 95% CI at non-exceedance probabilities, $P = (1 - 1/N)^{1/25}$, for $N \in [1, 1000]$ at 100 equally spaced logarithmic increments (see preceding subsection). Finally, we also ask participants to provide a description of the method used to estimate $F_{25}(x)$ and associated CI.

2.3 Benchmark comparison metrics

The contributions to the benchmarking study will be compared graphically using plots similar to those shown in Section 3. For Exercise 1, the estimated annual distribution, F_A, will be compared against empirical estimates of F_A derived from the annual maxima of the full 700-year pre-industrial control time series for each site. The 95% CI will also be compared between contributions and used to assess whether the CI's contain the empirical distribution from the 700-year dataset over the range of return periods considered.

Contributions for Exercise 2 will be compared in a similar manner. However, in this case there is no long-term reference data to compare the entries with, and assess the accuracy of the predictions. Instead, data from the first 25 years of the three future scenario experiments (scenarios SSP126, SSP245 and SSP585) will be used for comparison. As the future emissions are not known, the three datasets represent three possible future scenarios, but it is not known which is more likely to occur. Moreover, these three datasets represent particular stochastic realisations of a theoretically infinite population of future climates under each scenario, so there is sampling variability associated with the extreme conditions in each 25-year dataset, as well as uncertainty related to the model physics. Nevertheless, these datasets will be used for comparative reference, rather than quantitative assessment of the accuracy of contributions.

2.4 How to participate

This paper represents the official announcement of EC Benchmark 2. Interested participants can enter the benchmark by following the instruction described on a dedicated GitHub repository, which is available at `https://github.com/ec-benchmark-organizers/ec-benchmark-2`.

The repository contains the time line for the organization of the benchmark, the provided datasets and instructions on how to prepare your results for submission. Similar to the previous EC Benchmark 1, we intend to publish all submitted results on this repository. Participants are encouraged to write stand-alone papers on their submission if novel methods are used. The comparative analysis and discussion of the benchmark's results will leave only limited space for the description of individual contributions such that referencing any such stand-alone papers could be valuable in the write-up of the comparative analysis of all submissions.

In the course of developing the scope of the two exercises for EC Benchmark 2, several additional options were discussed, such as the use of spatial data and provision of additional covariates (e.g. climate forcing variables). The decision was made to analyse time series for single locations, as this is a more common problem in the design of an offshore structure for a particular location. However, participants are free to download additional spatial data or covariates from the FIO-ESM v2.0 climate model [54, 55] and use this in their analysis if they wish.

3 BASELINE RESULTS

In this section, we provide a set of results for Exercises 1 and 2 that can serve as a baseline for comparison against submissions to the benchmarking exercise. The baseline results are calculated using the FIO-ESM data which will be provided to participants, rather than reference data listed in Table 1 (the reference data is used solely for validating the FIO-ESM model outputs, to justify their use).

To provide baseline results for the benchmarking exercise, we have estimated return values of H_s following one of the procedures currently recommended in DNVGL-RP-C205 [4, §3.6.2.4]. The same approach has been used for both exercises, under the baseline assumption of a stationary wave climate and independent annual maxima. The annual maxima of H_s over the 25-year period were fitted with a Gumbel distribution, which has CDF

$$F_A(x) = \exp\left(-\exp\left(-\frac{x-u}{k}\right)\right). \tag{4}$$

The location and scale parameters, u and k, were estimated using the method of moments, where the values are related to the mean, μ, and standard deviation, σ, of the annual maxima by

$$k = \frac{\sqrt{6}}{\pi}\sigma, \tag{5}$$

$$u = \mu - \gamma k, \tag{6}$$

and $\gamma \approx 0.577216...$ is the Euler–Mascheroni constant. Confidence intervals for parameters and return values have been estimated using the bootstrap method. In this method, each dataset is randomly re-sampled (with replacement) and the parameters of the Gumbel distribution and return values are estimated for each random resampling. This procedure is repeated 1000 times and the results are then used to estimate 95% confidence intervals.

Comparisons of the exceedance probabilities of the observed annual maximum H_s from the pre-industrial control data for Exercise 1 and the fitted distributions are shown in Figure 3, for each of the three sites. The empirical non-exceedance probability for the ordered observations $x_{(1)} \leq x_{(2)} \leq ... \leq x_{(n)}$ is defined as $\tilde{P}(x_{(k)}) = k/(n+1)$, where $n = 25$ is the number of annual maxima that the distribution is fitted to. Figure 3 indicates that the Gumbel distribution is a reasonable fit to each of the datasets, with the fitted model broadly following the observations. However, for all three cases there are observations in the right tail of the distribution which fall outside the estimated 95% CI, and thus suggests that the Gumbel distribution may not be an optimal model for these datasets, given our interest in extremes. The estimated Gumbel distribution parameters and 5, 50 and 500-year return values for the three sites are listed in Table 2, together with a 95% CI, shown in brackets.

Comparisons of the observed annual maxima and fitted distributions for Exercise 2 are shown in Figure 4, with parameter estimates and return values listed in Table 3. For the baseline results, as we have made the assumption of a stationary climate, we present the results in terms of the distribution of annual maxima rather than the distribution of the maximum H_s in a 25 year period, since under the assumptions of the model, $F_{25}(x) = F_A^{25}(x)$. The results follow a similar pattern to those for Exercise 1. For Site 1, the estimated return values from the historic data are slightly lower than those for the pre-industrial control data, although the central estimates from the historic data fall within the 95% CI estimated from the pre-industrial control data. For sites 2 and 3 the estimated return values are remarkably similar for Exercises 1 and 2, although there is little reduction in the width of the CI for the 50-year historic dataset. If the assumptions of the model were correct (a stationary climate with independent, Gumbel distributed annual maxima), then asymptotically the variance in parameter and return value estimates would decrease as $1/n$, where n is the number of independent observations. The similarity of the width of the CI for the 25-year pre-industrial control data and 50-year historic data may indicate that some of the model assumptions are not appropriate for the data.

4 CONCLUSIONS

The proposed benchmarking exercise has been designed to address several key questions related to the estimation of extreme conditions. The first of two proposed exercises in EC Benchmark 2 addresses the question of how accurately extreme conditions can be predicted in a stationary climate, based on a relatively short dataset of 25-years in length. In contrast, the second exercise addresses the question of how accurately extremes can be predicted in a changing climate, if a relatively long dataset of 50 years is available. For the first exercise, the availability of the

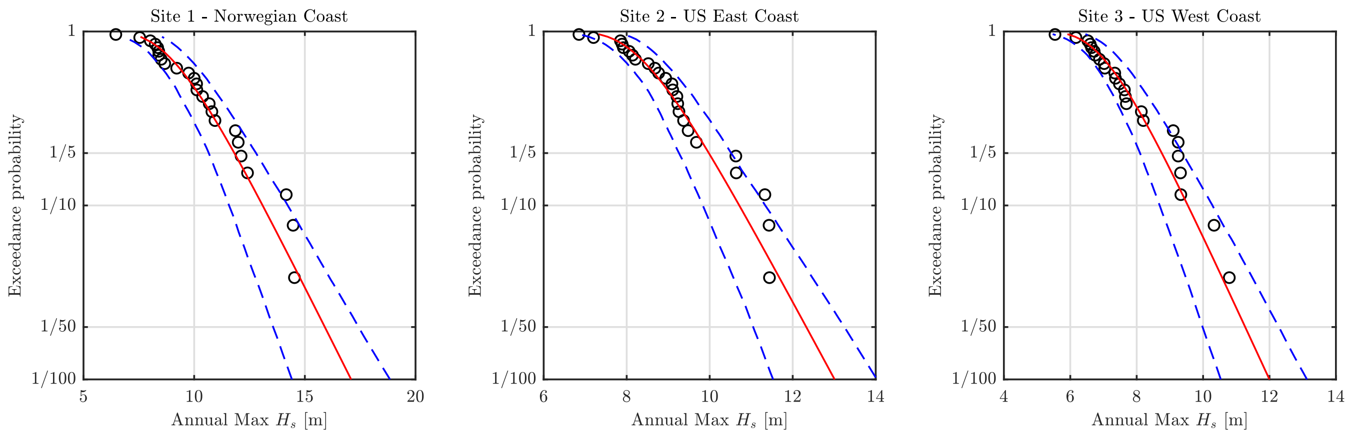

FIGURE 3. BASELINE RESULTS FOR EXERCISE 1: COMPARISON OF EXCEEDANCE PROBABILITIES FOR OBSERVED ANNUAL MAXIMUM H_s FOR 25-YEAR PRE-INDUSTRIAL CONTROL DATA (CIRCLES) AND FITTED GUMBEL DISTRIBUTION (RED LINES) FOR THE THREE SITES. A 95% CONFIDENCE INTERVAL FOR THE FITTED DISTRIBUTION IS SHOWN BY DASHED LINES.

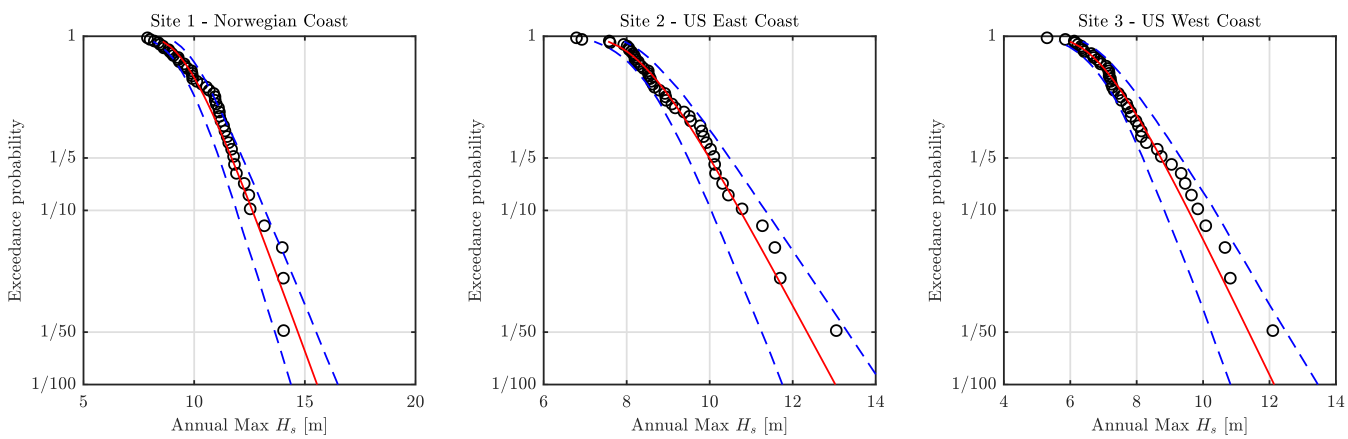

FIGURE 4. BASELINE RESULTS FOR EXERCISE 2: COMPARISON OF EXCEEDANCE PROBABILITIES FOR OBSERVED ANNUAL MAXIMUM H_s FOR 50-YEAR HISTORICAL DATA (CIRCLES) AND FITTED GUMBEL DISTRIBUTION (RED LINES) FOR THE THREE SITES. A 95% CONFIDENCE INTERVAL FOR THE FITTED DISTRIBUTION IS SHOWN BY DASHED LINES.

TABLE 2. BASELINE RESULTS FOR EXERCISE 1: GUMBEL PARAMETER ESTIMATES, RETURN VALUES AND 95% CONFIDENCE INTERVAL (IN BRACKETS) FOR THE 25-YEAR PI-CONTROL DATA FOR THE THREE SITES.

	Site 1 - Norway	Site 2 - US East	Site 3 - US West
Location parameter, μ	9.26 (8.54, 10.1)	8.51 (8.10, 8.99)	7.19 (6.76, 7.70)
Scale parameter, σ	1.70 (1.21, 2.06)	0.98 (0.69, 1.17)	1.05 (0.75, 1.28)
5-year return value, $H_{s,5}$ [m]	11.8 (10.7, 12.8)	9.98 (9.33, 10.5)	8.76 (8.05, 9.41)
50-year return value, $H_{s,50}$ [m]	15.9 (13.7, 17.5)	12.3 (11.1, 13.2)	11.3 (9.86, 12.4)
500-year return value, $H_{s,500}$ [m]	19.8 (16.6, 22.3)	14.6 (12.7, 15.9)	13.7 (11.6, 15.3)

TABLE 3. BASELINE RESULTS FOR EXERCISE 2: GUMBEL PARAMETER ESTIMATES, RETURN VALUES AND 95% CONFIDENCE INTERVAL (IN BRACKETS) FOR THE 50-YEAR HISTORICAL DATA FOR THE THREE SITES.

	Site 1 - Norway	Site 2 - US East	Site 3 - US West
Location parameter, μ	9.82 (9.43, 10.28)	8.52 (8.23, 8.85)	7.08 (6.79, 7.47)
Scale parameter, σ	1.25 (1.01, 1.44)	0.98 (0.75, 1.21)	1.10 (0.81, 1.35)
5-year return value, $H_{s,5}$ [m]	11.69 (11.16, 12.19)	9.99 (9.51, 10.46)	8.73 (8.15, 9.32)
50-year return value, $H_{s,50}$ [m]	14.68 (13.69, 15.56)	12.34 (11.35, 13.33)	11.37 (10.17, 12.51)
500-year return value, $H_{s,500}$ [m]	17.57 (16.08, 18.85)	14.61 (13.09, 16.14)	13.92 (12.05, 15.62)

700-year dataset will allow quantitative assessment of the estimated return values from the short dataset.

The baseline results presented here follow a very simple methodology and set of assumptions. Participants to the exercise are encouraged to use any method of their choice, both for estimating distributions and confidence intervals. It is hoped that participants will make contributions to the exercises, based on the wide variety of methods proposed to date, and potentially develop new methodologies, furthering the state of the art. Comparison between the contributions will illustrate the uncertainties resulting from modelling choices made by participants as well as uncertainties inherent in estimating extreme conditions in a variable and changing climate.

ACKNOWLEDGMENT

EM was funded by the EPSRC Supergen Offshore Renewable Energy Hub [grant no: EP/S000747/1] Flexible Fund project “Improved Models for Multivariate Metocean Extremes (IMEX)”.

LM acknowledges funding from the UT Austin Portugal Program Project, ”Sustainable Reuse of Decommissioned Offshore Jacket Platforms for Offshore Wind Energy” (Grant No. UTA-EXPL/IET/0111/2019).

RGC acknowledges funding from the US Department of Energy’s Water Power Technologies Office. Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy’s National Nuclear Security Administration under contract DE-NA0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

APPENDIX A: COMPARISON OF FIO-ESM DATA WITH REFERENCE DATASETS

Comparisons of the mean parameters from the FIO-ESM climate model and the 99th percentile of H_s with ERA5 hindcast data over the entire globe are presented in [55]. In this comparison, we focus on the distributions of H_s at the three sites. For the comparison, the buoy and hindcast data were averaged to 3-hour values, the same as the FIO-ESM data. The FIO-ESM for each site was restricted to the same time period as the reference sites, removing any periods where data are missing from the buoy records. Note that the CMIP6 historical period runs until 31/12/2014, so buoy data after this time were not used in the comparison.

It is important to note that climate models forced with historic input data are not the same as hindcasts commonly used for offshore applications. Hindcasts models calculate the wave field based on historic wind fields, and are directly comparable to measurements at a specific time and location. In contrast, as climate models calculate the evolution of the entire atmosphere (and a wide range of other systems), even when they are forced with historic inputs (such as solar radiation, atmospheric gas concentrations, etc.), the outputs are only comparable to historic observations in a statistical sense. We therefore compare the distributions of 3-hour and annual maximum H_s between the FIO-ESM data and reference data in terms of quantile-quantile (QQ) plots, as shown in Figures 5 and 6.

The distribution of 3-hour H_s is reasonably well replicated at all three sites. The FIO-ESM data are slightly lower than the reference data for Sites 1 and 3 and at Site 2, the FIO-ESM data is slightly higher than the buoy data at high H_s. There are similar trends in the annual maxima. However, the QQ plots show a reasonably linear trend overall, indicating that the distribution of extremes is reasonably well-matched to the reference data.

REFERENCES

[1] International Electrotechnical Commission (IEC), 2019. *Wind energy generation systems - Part 3-1: Design re-*

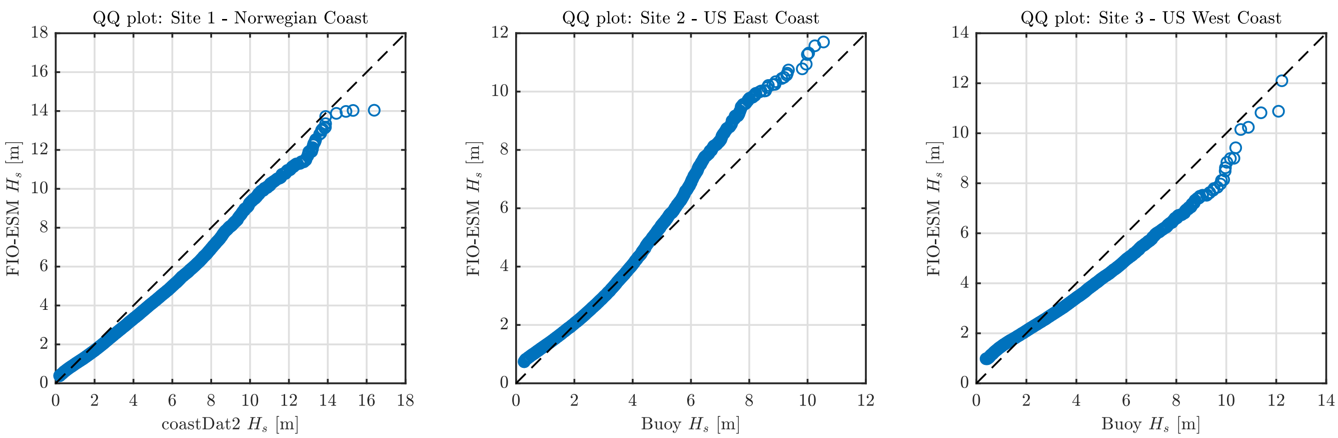

FIGURE 5. QQ-PLOTS OF 3-HOUR H_s FROM FIO-ESM HISTORIC DATA AND REFERENCE SOURCES LISTED IN TABLE 1

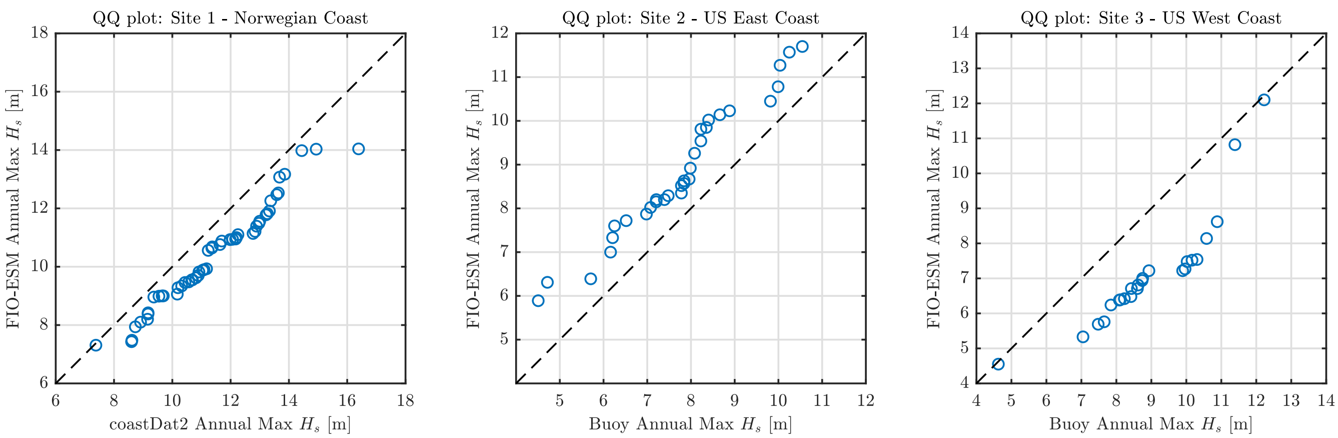

FIGURE 6. QQ-PLOTS OF ANNUAL MAXIMA OF H_s FROM FIO-ESM HISTORIC DATA AND REFERENCE SOURCES LISTED IN TABLE 1

quirements for fixed offshore wind turbines. IEC 61400-3-1:2019.

[2] International Electrotechnical Commission (IEC), 2016. *Marine energy - Wave, tidal and other water current converters - Part 2: Design requirements for marine energy systems. IEC TS 62600-2:2016*.

[3] for Standardization (ISO), I. O., 2015. *Petroleum and Natural Gas Industries—Specific Requirements for Offshore Structures—Part 1: Metocean Design and Operating Considerations: ISO 19901-1:2015*. Geneva, Switzerland.

[4] DNVGL, 2019. *Recommended Practice DNVGL-RP-C205: Environmental conditions and environmental loads*. Norway, September 2019.

[5] Vanem, E., Guo, B., Ross, E., and Jonathan, P., 2020. "Comparing different contour methods with response-based methods for extreme ship response analysis". *Marine Structures*, ***69***.

[6] NORSOK, 2017. *Actions and action effects (N-003), Edition 3*. Norway, January 2017.

[7] 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.

[8] Haselsteiner, A. F., Coe, R. G., Manuel, L., Chai, W., Leira, B., Clarindo, G., Soares, C. G., Hannesdóttir, Á., Dimitrov, N., Sander, A., Ohlendorf, J.-H., Thoben, K.-D., de Hauteclocque, G., Mackay, E., Jonathan, P., Qiao, C., Myers, A., Rode, A., Hildebrandt, A., Schmidt, B., Vanem, E., and Huseby, A., 2021. "A benchmarking exercise for environmental contours". *In preparation*.

[9] Ochi, M. K., 2005. *Ocean Waves: The stochastic approach*, Vol. 6. Cambridge University Press.

[10] Winterstein, S. R., Ude, T. C., Cornell, C. A., Bjerager, P., and Haver, S., 1993. "Environmental parameters for ex-

treme response: Inverse FORM with omission factors". In Proceedings of the 6th International Conference on Structural Safety & Reliability (ICOSSAR), Vol. 93, International Association for Structural Safety and Reliability, pp. 77–84.

[11] 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.

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

[13] 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.

[14] Mackay, E., and Haselsteiner, A. F., 2021. "Marginal and total exceedance probabilities of environmental contours". *Marine Structures,* ***75***, p. 102863.

[15] Jonathan, P., and Ewans, K., 2013. "Statistical modelling of extreme ocean environments for marine design : A review". *Ocean Engineering,* ***62***, pp. 91–109.

[16] van Vledder, G., Goda, Y., Hawkes, P., Mansard, E., Martin, M. J., Mathiesen, M., Peltier, E., and Thompson, E., 1993. "Case Studies of Extreme Wave Analysis: A Comparative Analysis". In Ocean Wave Measurement and Analysis, O. Magoon and J. Hemsley, eds., American Society of Civil Engineers, pp. 978–992.

[17] Guedes Soares, C., and Scotto, M., 2001. "Modelling uncertainty in long-term predictions of significant wave height". *Ocean Engineering,* ***28***(3), pp. 329–342.

[18] Sobey, R. J., and Orloff, L. S., 1995. "Triple annual maximum series in wave climate analyses". *Coastal Engineering,* ***26***(3-4), pp. 135–151.

[19] Guedes Soares, C., and Scotto, M. G., 2004. "Application of the r largest-order statistics for long-term predictions of significant wave height". *Coastal Engineering,* ***51***(5-6), pp. 387–394.

[20] Naess, A., 1998. "Statistical Extrapolation of Extreme Value Data Based on the Peaks Over Threshold Method". *Journal of Offshore Mechanics and Arctic Engineering,* ***120***(2), may, pp. 91–96.

[21] Ferreira, J. A., and Guedes-Soares, C., 1998. "An application of the peaks over threshold method to predict extremes of significant wave height". *Journal of Offshore Mechanics and Arctic Engineering,* ***120***(3), aug, pp. 165–176.

[22] Caires, S., and Sterl, A., 2005. "100-year return value estimates for ocean wind speed and significant wave height from the ERA-40 data". *Journal of Climate,* ***18***(7), apr, pp. 1032–1048.

[23] Naess, A., and Haug, E., 2010. "Extreme value statistics of wind speed data by the pot and acer methods". *Journal of Offshore Mechanics and Arctic Engineering,* ***132***(4), sep.

[24] Ferreira, J. A., and Guedes Soares, C., 1999. "Modelling the long-term distribution of significant wave height with the Beta and Gamma models". *Ocean Engineering,* ***26***(8), pp. 713–725.

[25] Haselsteiner, A. F., and Thoben, K. D., 2020. "Predicting wave heights for marine design by prioritizing extreme events in a global model". *Renewable Energy,* ***156***, pp. 1146–1157.

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

[27] Soukissian, T. H., and Tsalis, C., 2018. "Effects of parameter estimation method and sample size in metocean design conditions". *Ocean Engineering,* ***169***, pp. 19–37.

[28] de Zea Bermudez, P., and Kotz, S., 2010. "Parameter estimation of the generalized Pareto distribution-Part I". *Journal of Statistical Planning and Inference,* ***140***(6), pp. 1353–1373.

[29] Mackay, E. B. L., Challenor, P. G., and Bahaj, A. S., 2011. "A comparison of estimators for the generalised Pareto distribution". *Ocean Engineering,* ***38***(11-12), pp. 1338–1346.

[30] Tancredi, A., Anderson, C., and O'Hagan, A., 2006. "Accounting for threshold uncertainty in extreme value estimation". *Extremes,* ***9***(2), pp. 87–106.

[31] Ragan, P., and Manuel, L., 2008. "Statistical extrapolation methods for estimating wind turbine extreme loads". In Journal of Solar Energy Engineering, Transactions of the ASME, Vol. 130, American Society of Mechanical Engineers Digital Collection, pp. 0310111–03101115.

[32] Blum, J. R., Kiefer, J., and Rosenblatt, M., 1961. "Distribution Free Tests of Independence Based on the Sample Distribution Function". *The Annals of Mathematical Statistics,* ***32***(2), pp. 485–498.

[33] Fogle, J., Agarwal, P., and Manuel, L., 2008. "Towards an improved understanding of statistical extrapolation for wind turbine extreme loads". *Wind Energy,* ***11***(6), pp. 613–635.

[34] Davis, R. A., and Mikosch, T., 2009. "The extremogram: A correlogram for extreme events". *Bernoulli,* ***15***(4), pp. 977–1009.

[35] Méndez, F. J., Menéndez, M., Luceño, A., Medina, R., and Graham, N. E., 2008. "Seasonality and duration in extreme value distributions of significant wave height". *Ocean Engineering,* ***35***(1), pp. 131–138.

[36] Ewans, K., and Jonathan, P., 2008. "The effect of directionality on northern north sea extreme wave design criteria". *Proceedings of the International Conference on Offshore Mechanics and Arctic Engineering - OMAE,* ***2***, pp. 479–488.

[37] Randell, D., Feld, G., Ewans, K., and Jonathan, P., 2015. "Distributions of return values for ocean wave characteristics in the South China Sea using directional-seasonal ex-

treme value analysis". *Environmetrics,* ***26***(6), pp. 442–450.

[38] Zanini, E., Eastoe, E., Jones, M., Randell, D., and Jonathan, P., 2020. "Flexible covariate representations for extremes". *Environmetrics*.

[39] Mackay, E., and Jonathan, P., 2020. "Assessment of Return Value Estimates from Stationary and Non-Stationary Extreme Value Models". *Ocean Engineering,* ***207***, p. 107406.

[40] Raghupathi, L., Randell, D., Ewans, K., and Jonathan, P., 2016. "Fast Computation of Large Scale Marginal Extremes with Multi-Dimensional Covariates". *Computational Statistics & Data Analysis,* ***95***, pp. 243–258.

[41] Woolf, D. K., Challenor, P. G., and Cotton, P. D., 2002. "Variability and predictability of the North Atlantic wave climate". *Journal of Geophysical Research C: Oceans,* ***107***(10), pp. 9–1.

[42] Bromirski, P. D., Cayan, D. R., and Flick, R. E., 2005. "Wave spectral energy variability in the northeast Pacific". *Journal of Geophysical Research C: Oceans,* ***110***(3), pp. 1–15.

[43] Young, I. R., Zieger, S., and Babanin, A. V., 2011. "Global trends in wind speed and wave height". *Science,* ***332***(6028), apr, pp. 451–455.

[44] Reguero, B. G., Losada, I. J., and Méndez, F. J., 2015. "A global wave power resource and its seasonal, interannual and long-term variability". *Applied Energy,* ***148***, pp. 366–380.

[45] Aarnes, O. J., Reistad, M., Breivik, Ø., Bitner-Gregersen, E., Ingolf Eide, L., Gramstad, O., Magnusson, A. K., Natvig, B., and Vanem, E., 2017. "Projected changes in significant wave height toward the end of the 21st century: Northeast Atlantic". *Journal of Geophysical Research: Oceans,* ***122***(4), apr, pp. 3394–3403.

[46] Reguero, B. G., Losada, I. J., and Méndez, F. J., 2019. "A recent increase in global wave power as a consequence of oceanic warming". *Nature Communications,* ***10***(1), dec.

[47] Cattrell, A. D., Srokosz, M., Moat, B. I., and Marsh, R., 2019. "Seasonal intensification and trends of rogue wave events on the US western seaboard". *Nature,* ***9***(1), dec.

[48] Mosquera-Mosquera, I. A., Simão, M. L., Videiro, P. M., and Sagrilo, L. V., 2020. "Evaluating the impact of climate change on offshore structures design: A practical case study". *Applied Ocean Research,* ***94***.

[49] De Leo, F., De Leo, A., Besio, G., and Briganti, R., 2020. "Detection and quantification of trends in time series of significant wave heights: An application in the Mediterranean Sea". *Ocean Engineering,* ***202***, apr.

[50] Vanem, E., 2015. "Non-stationary extreme value models to account for trends and shifts in the extreme wave climate due to climate change". *Applied Ocean Research,* ***52***, aug, pp. 201–211.

[51] Mori, N., Shimura, T., Yasuda, T., and Mase, H., 2013. "Multi-model climate projections of ocean surface variables under different climate scenarios-Future change of waves, sea level and wind". *Ocean Engineering,* ***71***, oct, pp. 122–129.

[52] Morim, J., Hemer, M., Wang, X. L., Cartwright, N., Trenham, C., Semedo, A., Young, I., Bricheno, L., Camus, P., Casas-Prat, M., Erikson, L., Mentaschi, L., Mori, N., Shimura, T., Timmermans, B., Aarnes, O., Breivik, Ø., Behrens, A., Dobrynin, M., Menendez, M., Staneva, J., Wehner, M., Wolf, J., Kamranzad, B., Webb, A., Stopa, J., and Andutta, F., 2019. "Robustness and uncertainties in global multivariate wind-wave climate projections". *Nature Climate Change,* ***9***(9), pp. 711–718.

[53] Meucci, A., Young, I. R., Hemer, M., Kirezci, E., and Ranasinghe, R., 2020. "Projected 21st century changes in extreme wind-wave events". *Science Advances,* ***6***(24), pp. 1–10.

[54] Bao, Y., Song, Z., and Qiao, F., 2020. "FIO-ESM Version 2.0: Model Description and Evaluation". *Journal of Geophysical Research: Oceans,* ***125***(6), pp. 1–21.

[55] Song, Z., Bao, Y., Zhang, D., Shu, Q., Song, Y., and Qiao, F., 2020. "Centuries of monthly and 3-hourly global ocean wave data for past, present, and future climate research". *Scientific Data,* ***7***(1), pp. 1–11.

[56] Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E., 2016. "Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization". *Geoscientific Model Development,* ***9***(5), pp. 1937–1958.

[57] 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.

[58] Norway Opens Two Offshore Wind Areas, 2020. *www.offshorewind.biz/2020/06/12/breaking-norway-opens-two-offshore-wind-areas-4-5-gw-available/, accessed 17/01/2021*.

[59] Jonathan, P., Randell, D., Wadsworth, J., and Tawn, J., 2021. "Uncertainties in return values from extreme value analysis of peaks over threshold using the generalised Pareto distribution". *Ocean Engineering,* ***220***, p. 107725.