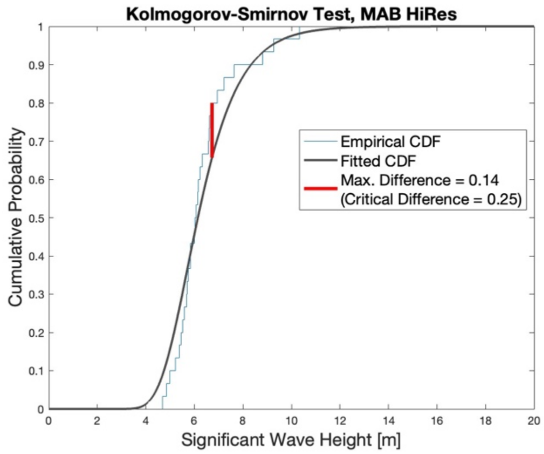
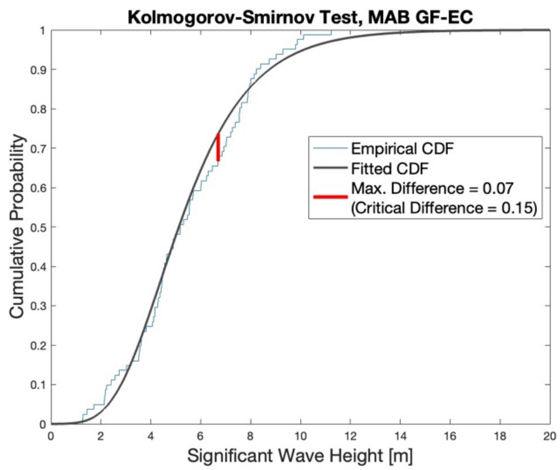
The article looks into differences in estimates of significant wave height extremes due to tropical and extra-tropical storms at two offshore locations in the eastern coast of the United States.

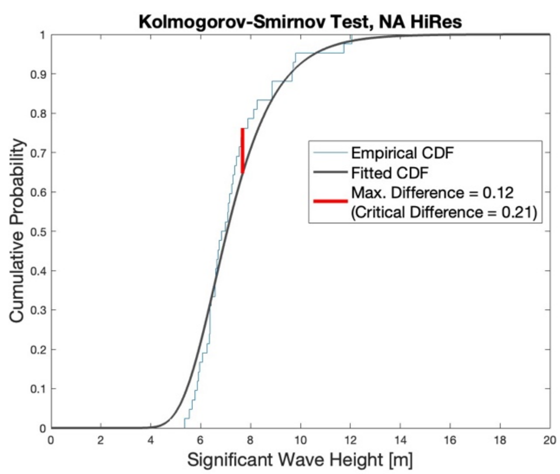
Thank you for your comments on this paper, "Quantifying Tropical Cyclone-Generated Waves in Extreme Value-Derived Design for Offshore Wind". Please find below our replies in blue.

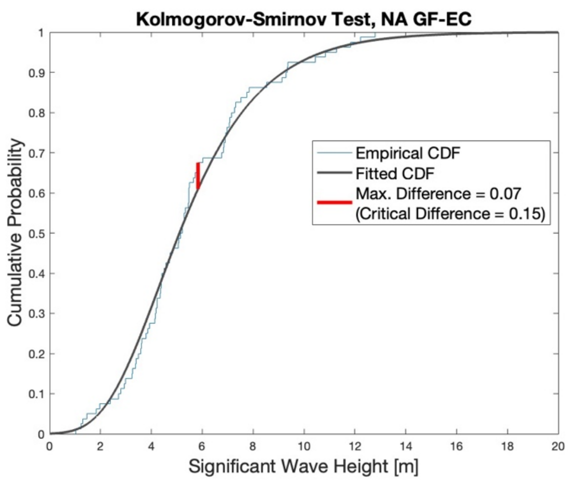
The article is poorly written and some of the analyses are not sound. For example:

- There is no motivation being given for the type I tail assumption being made when fitting the Gumbel instead of the Generalized extreme value distribution to the annual maxima. Furthermore, and as can be read in a wealth of extreme value theory publications, for instance the cited book of Coles, the Weibull and the Gumbel distributions are not the asymptotic distributions of data sampled using the peaks-over-threshold approach. This does not mean that they cannot be used but their use should be justified.
    - Thank you. The Peaks-Over-Threshold method is not the selected methodology for analysis in this paper; however, analysis by POT is presented in the appendices for a data subset to investigate any substantial deviations presented by the chosen method itself. This was first investigated by controlling for the same distribution (Gumbel) to assess the influence of the method. In the revised manuscript, the Generalized Pareto distribution is used. No systemic biases are attributed to the selected method over both locations and all datasets investigated (lines 396 – 404).
    - Thank you, we have added justification to the revised manuscript. A one-sample Kolmogorov-Smirnov test was conducted to evaluate whether each of the four data sets (North Atlantic high-resolution, Mid-Atlantic high-resolution, GF-EC Tropical in the N. Atlantic and GF-EC Tropical in the Mid-Atlantic) follow a Gumbel cumulative distribution function. The null hypothesis (H_0: the data follows a Gumbel distribution) was not rejected, indicating that, at a 95% confidence level, the Gumbel distribution fits the data adequately (see figures on the following page):


Kolmogorov-Smirnov Test, MAB HiRes
Empirical CDF
Fitted CDF
Max. Difference = 0.14
(Critical Difference = 0.25)
Cumulative Probability
Significant Wave Height [m]


Kolmogorov-Smirnov Test, MAB GF-EC
Empirical CDF
Fitted CDF
Max. Difference = 0.07
(Critical Difference = 0.15)
Cumulative Probability
Significant Wave Height [m]


Kolmogorov-Smirnov Test, NA HiRes
Empirical CDF
Fitted CDF
Max. Difference = 0.12
(Critical Difference = 0.21)
Cumulative Probability
Significant Wave Height [m]



- Also, the authors confuse the directional spreading of a wave system or sea state with the variability of the mean wave direction during a storm.
    - Thank you, the details may not have been apparent in the original manuscript, but we will clarify: Directional spreading describes the radial propagation of wave energy, in this context due to applied forcing (wind). Here, the observed features of radial wave propagation—spread—based on the energy flux from winds to ocean surface (generated by a tropical or extra-tropical cyclone) relies heavily on the nature of the flux (storm type), consistent with the physical system described by Forristall and Ewans ("Worldwide Measurements of Directional Wave Spreading", Journal of Atmospheric and Oceanic Technology, 1998). Within the storm fetch, it is not a given that there are many sea states, but instead are highly dominated by local winds. As addressed in the calibration section for the Mid-Atlantic Bight model, wave boundaries with applied ERA-5 wave data on all 3 sides performed worse for wave direction than when forced on only one boundary by ERA-5 and with open boundaries on the remaining two sides. This is attributed to the better performance of modeled wind-driven waves than prescribed directly by the global reanalysis dataset. This investigation was therefore considered necessary when gauging the physical representation by the models.

Specific comments

Lines 82-83: What is the rational for fitting the Weibull (of minima, I assume) and the Gumbel distribution to POT data? Can you justify why you are deviating from the Generalized Pareto distribution?

- The peaks-over-threshold method is not focus of paper, but as described above, the inclusion of this detail is to gauge result sensitivity to the paper method, rather than to determine the single best fit for each data set. As GP is traditionally the best selection for POT analysis, Figures A2a and A2b have been updated with return value estimates from GP distribution over the threshold indicated in the legend.

Lines 136-138: Please indicate whether there is corresponding between the storms leading to the annual maxima in both GF and hindcast datasets.

- Thank you, the question is not entirely clear. If the reviewer means "correspondence", the investigated peaks are associated with historical tracks for each storm. For each storm, lines 179 – 182 of the original manuscript (lines 185 – 186 of the revised manuscript) state: *"To preserve the independence criterion in this study, only the peak significant wave height is retained in a period of 98 hours during an identified storm."*

Lines 159-161: Please rephrase of remove. The estimates are obtained using the likelihood method? If so, it suffices to state it.

- The values are determined by maximum likelihood; the text is updated (line 168 of the revised manuscript) for clarity.

Section 2.2.1: Please motivate why the Gumbel instead of the Generalized extreme value distribution is being fitted to the data.

- A one-sample Kolmogorov-Smirnov (KS) test was conducted to evaluate whether the data follows a Gumbel cumulative distribution function. The null hypothesis (H_0: the data follows a Gumbel distribution) was not rejected at the 5% significance level, indicating with 95% confidence that the Gumbel distribution fits the data adequately. (Please refer to the plots on pages 2 and 3 of this reply.)

  Additionally, calculated GEVD parameters for the four main datasets in this work have a shape factor, k, that is close to or effectively 0:

	location - mu	**scale - sigma**	**shape - k**
NA GF-EC: *GEVD*	4.3879	2.122	-0.0498
NA GF-EC: *Gumbel*	4.31	2.17	0
NA HiRes: *GEVD*	6.564	0.92	0.192
NA HiRes: *Gumbel*	6.57	1.34	0
MAB GF-EC: *GEVD*	4.624	2.53	-0.2168
MAB GF-EC: *Gumbel*	4.42	1.93	0

MAB HiRes: *GEVD*	5.7588	0.7988	0.1351
MAB HiREs: *Gumbel*	5.49	1.34	0

This data is also provided in Appendix B, revised Table B1.

Both of these factors support the use of a Gumbel distribution for this study.

Lines 170-171: Why is *p* called "probability **period**"? Please add that when computing the return values p is substituted by 1/*n* with *n* being the return period in years. In the tables and text only *n* is being given, not *p*.

- The term has been updated to "Annual exceedance probability", which is p = 1/n

Appendix B. The associated return value, x_p, for annual exceedence probability p are calculated as:

170 $$x_p = \mu + \sigma(-\ln(-\ln(1-p))), \quad for \quad -\infty < x_p < \infty$$

Line 178: Why does the storm list given in Appendix C only starts in 1991?

- The storm list is provided for the hindcast period with overlapping observations. The MAB hindcast starts in 1990 and the first observations in region are in 1991. Observations do not occur in NA until 2009. Lines 93 – 94 of the revised manuscript are updated to: *"A list of significant storms during the hindcast period, available observations, and storm events used for model calibration is provided in Appendix C, beginning with the first available buoy-based observations."*

Figure 3: 1)The data to which the distributions were fitted need to be added to the figure. (If not possible in absolute scale, then in relative scale as in Figure 4.) 2) Preferably also the 95% confidence intervals of the estimates should also be given. 3)The legend should contain for each of the lines the periods covered by the data or the sample size (number of considered annual maxima).

- Thank you. The fits are for the entire period outlined in Tables 1 and 2 unless otherwise indicated in the legend (i.e., only for Figures 13 and 14 in the revised manuscript). Lines 179 – 185 of the revised manuscript state: "*The GF-EC MAB site data includes 81 tropical cyclone events. The GF-EC NA site data includes 80 tropical cyclone events. In the event that a tropical cyclone did not pass nearby the analysis location during the season, no maxima are recorded for that year in the GF-EC dataset. In the "tropical cyclone only, high-resolution" NA and MAB datasets, a smaller, non-extra-tropical cyclone is picked as a maxima for a year without nearby tropical cyclones activity. Annual maxima analysis for the "high-resolution" model datasets account for 42 storms (tropical or extra-tropical* *cyclones) at the NA location and 30 storms (tropical or extra-tropical cyclones) at the MAB location."*

Figure 4 of the revised manuscript now presents the data and Gumbel distribution for the four main datasets investigated:

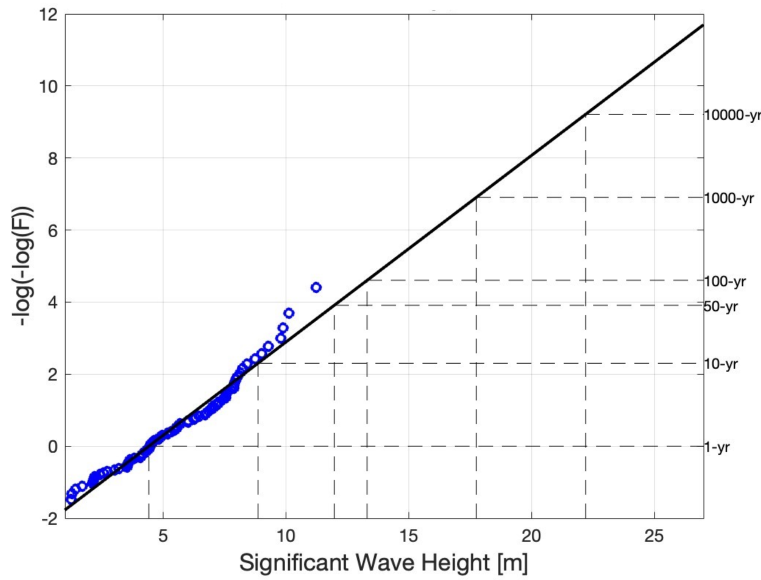


Figure 1: MAB GF-EC Tropical

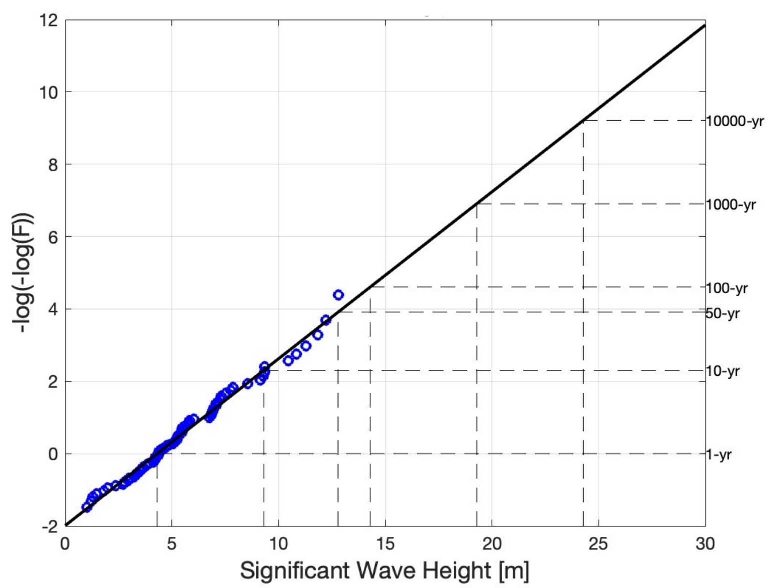


Figure 2: NA GF-EC Tropical

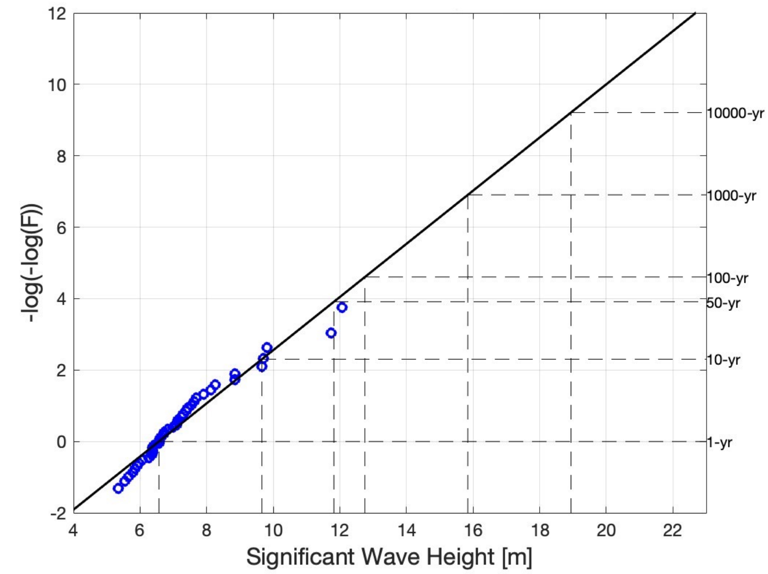


Figure 3: NA high-resolution model

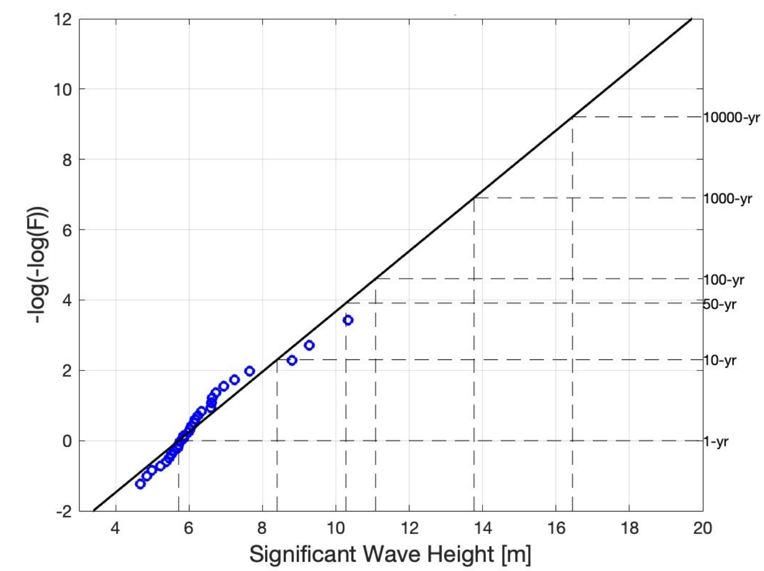


Figure 4: MAB high-resolution model

A one-sample Kolmogorov-Smirnov (KS) test was conducted to evaluate whether the data follows a Gumbel cumulative distribution function. The null hypothesis (H_0: the data follows a Gumbel distribution) was not rejected at the 5% significance level, indicating with 95% confidence that the Gumbel distribution fits the data adequately. (Please refer to the plots on pages 2 and 3 of this reply.)

Line 91: There are only 15 samples in the 'GF-EC. Trop.' fit? Please comment on the uncertainty of the estimates.

- We agree that 15 events are a small sample size for calculating extreme values. There are 15 identified hurricanes during the "NA high-resolution" hindcast period (1979 – 2021). There are 80 in the NA GF-EC Tropical period (1924 – 2021).

Section 3.1.2: In my opinion this section can be removed. What is its purpose? Why are the plots of the significant wave height (even if normalised) not shown?

- The performance of the numerical models in this section are investigated: *Is poor numerical representation of the sea state a cause of differences in return values?* Given that some global reanalysis datasets have been shown to under-represent the highest winds from tropical cyclones, Section 3.1.2 shows that, through overall representation Hs, the partitioned sea states, and wave trajectory over the course of the storm, there is fairly good consistency despite potentially high variability for tropical cyclones as compared to extra-tropical cyclones. Numerical performance is an essential piece in assessing how extremes are represented, and to what degree calibration can improve validation (Neary (2020), Caires (2005), Stephens (2006)), with the goal of estimating more suitable return values. It may be surprising to some readers that generous calibration of values was not enough to close the gap between the trend and magnitude of estimated return values between the data sets, even over the same time period (storm sample size). This has nontrivial implications for the design of metocean models in hurricane-prone regions, as direct modeling of TCs or synthetic representation of TCs are not required by standards for ocean models. Out of consideration for the total number of figures in the paper, a selection of events with measurements were originally chosen. Bob (no observations) and Dorian for the North Atlantic have been added to the revised manuscript Figure 6.

Section 3.1.3: The contents of this section are incorrect. First, how can the authors not be aware that the waves in the roses in Figure 7 are from the coast and therefore not realistic. Second, the authors present the variation in the mean wave directions during the consider storms (may wave systems, sea states) and analyse with reference to the article of Forristall and Ewans on directional spreading of a wave system or sea state.

- Thank you for pointing this out. During the submission process, we identified this error and posted a WESD comment to this effect when the manuscript became public (Nov 4th). A scripting error in the rose plot code shifted measurements clockwise by 90 degrees, and the images have been revised. Note that the analysis location is ~90 km offshore from the west. Your 2nd comment is answered on page 1 of this reply sheet.

Section 3.2: This section needs also to be completely redone. When comparing statistical estimates the sample sizes and confidence intervals should be given.

Furthermore, when making assumption in terms of the tail of the data these should be justified.

- Thank you, these points are addressed in replies to your previous comments. Figures of the data and fit, are added to the manuscript (revised Figure 4), with 95% confidence intervals in Table B1.

Technical corrections

Line 46: Please specify which are the variables being considered in the univariate and bivariate analyses you are referring to. Why is this relevant for this article?

- The paper focuses on univariate extreme significant wave height, which is an approach commonly found in the literature. However, it can be more realistic to consider seas in terms of environmental contours—bi-variate analysis—especially when it comes to a small, focused storm event like a hurricane. While investigation of the actual differences in return values between uni- and bi-variate return value assessments during different types of events is worthwhile, it is beyond the scope of this paper. This is mentioned in the paper only to clarify and ensure that uni-variate analysis does not misrepresent significant wave height as it is projected to larger return periods. Specification that the return values in this paper are determined with a univariate approach is specified in the revised manuscript (Section 2.2).

Line 69: Specify which 3 models and models of what?

- Line 69 in the original manuscript (line 70 of the revised manuscript) refers to the 3 numerical ocean models, called out for both regions in Figure 1, and described in Tables 1 (wind) and 2 (waves + hydrodynamics) in the column “Model”.

Line 78: What does “Return period results” mean? Should it be “Return value estimates”?

- “Results” has been changed to “estimates”.

Line 96: You mean Appendix C instead of B?

- Thank you, this error has been updated.

Line 104: Is the magnitude of Cds correct? Please introduce the meaning of the symbols it wanting to give the values.

- Thank you, the coefficient of surface drag, C_ds, was missing a factor of 10^(-2) and is updated. The names of these wave model parameterizations are added.

Line 110 and elsewhere: Explain what you mean with “most “at-risk” turbine location” and how this has been defined.

- This is identified for the present study as the position most exposed to waves from Atlantic Basin combined with deepest points in the Wind Energy Area. Given that this is not a significant detail for this investigation, it has been updated to “turbine location of interest” in the revised manuscript (lines 106, 118).

Lines 114-118, …: Provide references for SWAN, Delft3D, Westhuysen, WAM, OWI3G,…

- These citations have been added to the revised manuscript.

Line 117: Define acronyms throughout the text. For instance, what does YSU mean?

- YSU is the “Yonsei University” scheme, a non-local turbulence closure planetary boundary layer model in WRF. This specification has been added to the revised manuscript in line 113.

Lines 124: State also model depth for the considered output location.

- The depths of the points of interest/analysis are given in the original manuscript model descriptions, lines 109 and 123 (revised manuscript lines 105, 117). It is specified in the revised manuscript that the nearby GF-EC analysis points are located at the same depths (lines 107, 119).