General Comment

This article proposes a methodology to find extreme wave values in regions where both tropical and extra-tropical cyclones occur. The motivation given is for the more accurate prediction of design values for offshore wind turbine farms. Long term return values (up to 10,000 year return periods) are calculated using statistical methods. Overall, I find that this article has scientific merit; however, the findings/methods are somewhat obfuscated/unclear. I believe that by clarifying the methodologies used and providing more context for why specific scientific decisions were made the article would be significantly improved.

Thank you for the time and effort you dedicated to reviewing this paper, "Quantifying Tropical Cyclone-Generated Waves in Extreme Value-Derived Design for Offshore Wind". Please find below our replies in blue to your comments.

Specific Comments

Line 73: You mention "validated and calibrated" models. Have these calibrations/validations been published elsewhere? If so, please include the citation to the appropriate papers. If not, please include a subsection with an overview of both the calibration (what was calibrated, how parameters were selected, etc.) and some of the validation data.

- Thank you. The referenced model reports are not public (and the references have been removed), however, the relevant details are re-created and expanded in the revised manuscript "Model Skill" section (*"The North Atlantic model calibration was determined from a range of cap to friction velocity values and nonlinear growth coefficients for overall performance during 1) a mixed set of storms, and 2) during the entire year of 2012"*)

  and in Appendix C (*"The Mid-Atlantic model was calibrated against LiDAR buoy observations of "Hurricane Isais" in July 2020. The friction coefficient, whitecapping parameterization (Westhuysen or Komen), time step, and wave boundary conditions were adjusted during the calibration process. The boundary condition calibration (ERA-5 wave forcing on three boundaries, or ERA-5 waves on the eastern boundary with open condition on the remaining boundaries) resulted in notable differences in the mean wave direction compared to observations during hurricane passage. This is attributed to improved treatment of wind-driven waves during this event, and ERA-5 forcing only on the eastern boundary was chosen for the final configuration. Significant wave heights were similar to observations for both boundary condition configurations."*):

	Hs [m]			*T02 [s]*			*Tp [s]*		
Location and Duration	**Correlation Coefficient [%]**	**Scatter Index**	**Bias [m]**	**Correlation Coefficient [%]**	**SI**	**Bias [m]**	**Correlation Coefficient [%]**	**SI**	**Bias [m]**
NDBC44097 40.97 °N, 71.12 °W 2009–2020	91	0.25	0.03	82	0.13	-0.52	73	0.23	-0.01
NDBC44008 40.50 °N, 69.25 °W 1982–2020	93	0.24	-0.03	82	0.12	-0.39	75	0.17	0.06
NDBC44017 40.69 °N, 72.05 °W 2002–2020	92	0.25	0.10	83	0.11	-0.38	77	0.20	0.13
NDBC44020 41.50 °N, 70.28 °W 2009–2020	81	0.35	0.10	61	0.07	-0.65	44	0.11	-0.15
LiDAR Buoy 41.07 °N, 70.48 °W 2018–2020	91	0.26	0.08	84	0.11	-0.21	71	0.22	0.09

Table 3. Validation statistics throughout the North Atlantic model domain against continuous measurements, 2010 - 2020.

	Hs [m]			*Tp [s]*		
Location	**Correlation Coefficient [%]**	**Scatter Index**	**Bias [m]**	**Correlation Coefficient [%]**	**SI**	**Bias [m]**
Combined 20 Buoys	0.93	0.24	0.13	0.59	0.28	-0.07
NDBC44008 40.50 °N, 69.25 °W	0.90	0.21	0.04	0.47	0.32	-0.95
NDBC41025 35.01 °N, 75.45 °W	0.90	0.23	0.20	0.59	0.24	-0.27

Table 4. Selected validation statistics for the GROW-Fine East Coast model, overall and in the North and Mid-Atlantic regions.

Whitecapping	**Time step**	**Chezy coefficient**	**Jonswap coefficient**	
Calibration Range	Westhuysen or Komen	2min–0.2 min	65–85	0.25–0.85
Selected	Westhuysen	Variable, reduced from 2 min during certain events	65	0.67

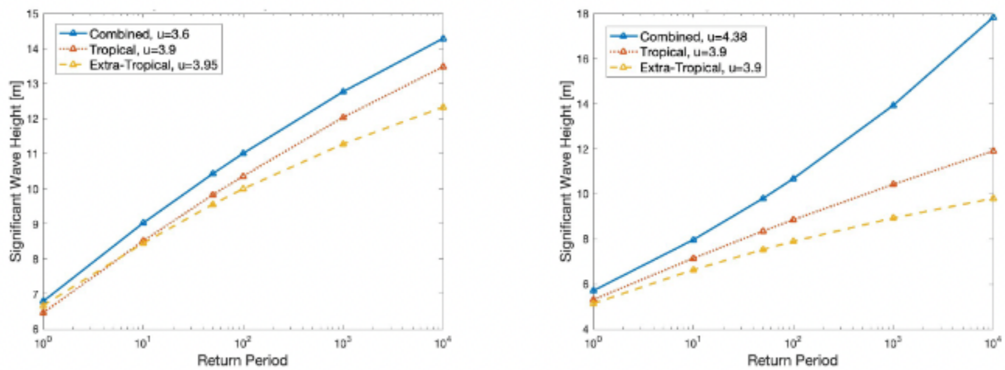
Table C1. Calibration values for the MAB "high-resolution" model.

Location	**RMSE**	**Mean Absolute Error**
LiDAR Buoy 06/2020 - 07/2020	0.26m	0.20m
NDBC44099 2009–2019	0.25m	0.18m

Table C2. Validation statistics for the MAB "high-resolution" model.

Line 80: Why was the Block Maxima with Gumbel Fit selected over, for example, a Generalized Pareto DIstribution with peaks over threshold? Just below this line you mention that you performed a sensitivity analysis using different methods but you do not explain why you ended up highlighting the BM with Gumbel results.

- Thank you, this is important to clarify. The two main methods for evaluating extremes for offshore applications are considered to be POT and BM (ref. Jonathan 2013). As elaborated in the following reply, POT was not considered suitable for accurately representing the GF-EC data set. However, the POT sensitivity study was updated to include return estimates by Generalized Pareto in the revised manuscript:



(a) Return values at the North Atlantic site from the high-resolution model ("Combined") and two post-processed subsets of the high-resolution model based on storm type.

(b) Return values at the Mid-Atlantic Bight site from the high-resolution model ("Combined") and two post-processed subsets of the high-resolution model based on storm type.

Figure A2. Return values by Peaks-Over-Threshold for the "high-resolution" return estimates with a Generalized Pareto distribution and selected thresholds, u.

A one-sample Kolmogorov-Smirnov 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, indicating that, at a 95% confidence level, the Gumbel distribution fits the data adequately. While the Generalized Extreme Value Distribution is the most comprehensive distribution choice for BM analysis, the GEVD shape parameter is near or effectively 0 for the four datasets considered in this paper and priority was given to the treatment of the distribution tail. The GEVD parameters are added to the manuscript appendix:

	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

Line 89: Could you explain more about why a POT method is not appropriate for "only storm events"? Given that POT assumes the events are independent (which I would say applies to individual storm events) and that threshold selection, whether through graphical or automated methods, relies upon the fact that for any threshold that produces an adequate fit a threshold larger than that should produce the same fit (when using a generalized Pareto distribution), I fail to see why the lack of "normal sea states" precludes the use of a POT methodology.

- Yes, thank you—determining an “adequate fit” is key. The GF-EC data sets are non-continuous time series of storms over roughly 8-day periods. Despite covering 100 years of events, in the case of the tropical cyclone model, the relatively short individual storm period meant threshold and clustering time selection were based on this limited range (the growth and decay of a storm peak) challenging whether the assessed fits were “adequate”. (This detail is added to the revised manuscript in line 169.) While trying to maintain a sufficient data sample for fitting, the calculated return values by this method for the GF-EC model data were suspiciously large, suggesting that the fitted distribution did not fully characterize the site.

Section 2.1: I find the description of the numerical models to be lacking in detail. As you mention in multiple locations, the location and derivation of boundary conditions can greatly change the results of a numerical model. Despite this, there is no description of the model domains, i.e., does the model cover the entire North Atlantic basin? Does it only cover the insets from Figure 1? You also mention again that the models are "locally validated". Where can I find this validation data? You mention the Commonwealth Wind metocean report (Wrenger, 2022) at line 108. Using the information in your works cited, I was unable to locate this report. There is the same issue with the Georgas (2023) report you cite for the Mid-Atlantic model (line 122). For the GROW-Fine East Coast model we are simply referred to Oceanweather inc. Please either provide the validation statistics in your work or, if possible, provide open-source and easily accessible reports showing why we should trust these models.

- Thank you, these are important details for assessing model applicability and validity for this study. Validation statistics have been added to the manuscript (Tables 3, 4, and C2 shown above). The model domains and validation locations have been added in Figure 2 to the revised manuscript:

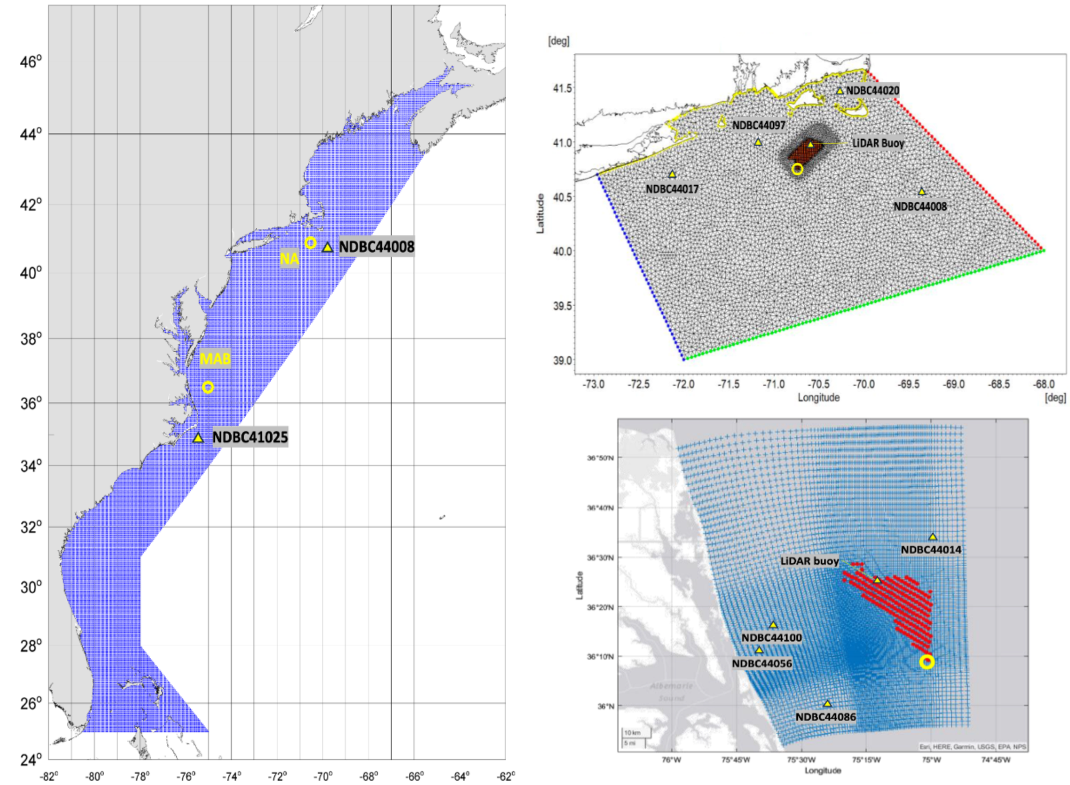


Figure 3. Analysis locations are indicated by circles and validation locations are indicated by triangles. Clockwise from the left: (a) The structured-grid GF-EC domain spans from 25 to 45.85 °N, and 82 to 64.3 °W. (b) The unstructured-grid NA "high-resolution" domain spans from 39 to 41.5 °N, and 73 to 68 °W. The wave boundary conditions are taken from a regional spectral wave model that spans 28 to 46 °N, and 82 to 58 °W, covering 16 directions and 25 frequencies from 1 to 33s. (c) The structured-grid MAB "high-resolution" domain spans from 35.83 to 37 °N, and 75.58 to 74.83 °W.

As the model reports are not publicly available, and these models were designed and developed for internal industrial use, they are removed from the references.

Section 2.1.4: I see here some mention of model validation. Consider moving some part of Appendix C into the body of the text. Especially given you specifically refer to the figures and error values in the appendix it seems appropriate that it would be part of the main text.

- Yes, there was some back-and-forth about the best location for this. The quantile-quantile plots that previously appeared in Appendix A have been replaced by NA and GF-EC model validation statistics in the “Model Skill” section (revised Tables 3 and 4) and MAB validation statistics in Appendix C (Table C2).

Figures 11 and 12: What are the confidence intervals of the return periods you calculate here? Given the use of m such a short time series for the estimation of very long return periods I would expect to see relatively large confidence intervals.

- Thank you. For clarity, Figures 11 and 12 are maintained as presented in the original manuscript. 95% confidence intervals are added to the appendix in Table B1.

There are other locations where the methodology could be clarified and greatly improve this manuscript. At the moment, I find that the experiments herein would be very difficult for another researcher to reproduce, greatly limiting the usefulness of the findings.

- Thank you, the manuscript has been updated to better reflect these details in your above comments, and if any important details remain missing, we would be happy to further revise.

  Regarding your question on the reproducibility of our results and methods, this paper attempts to elucidate industry-standard methods and tools, which are often kept proprietary, for public discussion and scrutiny. Furthermore, the "high-resolution" models presented here are examples we've selected to represent typical tools and robust methods in common practice today. We believe this is important for improving the state of the art in the industrial, standards-development, and academic domains (where numerous recent publications have calculated Hs extremes for this region from a single "mixed-type" sample, for example). The full wave timeseries for the NA and MAB locations, in addition to nearby observational data, are publicly available at https://doi.org/10.5281/zenodo.13884957. The GROW-Fine East Coast model is the only long-duration, direct-hindcast tropical and extra-tropical model we are aware of that allows direct comparison; while we are unable to publish long timeseries or peak absolute values from this dataset, we believe that the clear trends discussed in this work are useful for current and future model development and infrastructure design activities.