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”. Your comments have helped to improve the quality of this manuscript. Please find below our replies in blue.

1. Proper extreme distribution may depend on site characteristic. Although authors discussed about only differences of extreme wave height obtained from Gumbel and Weibull distribution, these result itself doesn’t explain the reason that Gumbel distribution is chosen. In other words, why didn’t choose Weibull. It is better to draw raw data used for fitting in Figure A.

   Thank you. The annual maxima data with a Gumbel distribution are added to the manuscript (Figure 4 from the revised manuscript, reproduced below). Confidence intervals are added to the Appendix in Table B1. A one-sample Kolmogorov-Smirnov test was conducted to evaluate whether the data follows a Gumbel cumulative distribution function. The null hypothesis (H_o: the data follows a Gumbel distribution) was not rejected, indicating that, at a 95% confidence level, the Gumbel distribution fits the data adequately. Please also see the response to Question 1 of the Reviewer 1 comments.

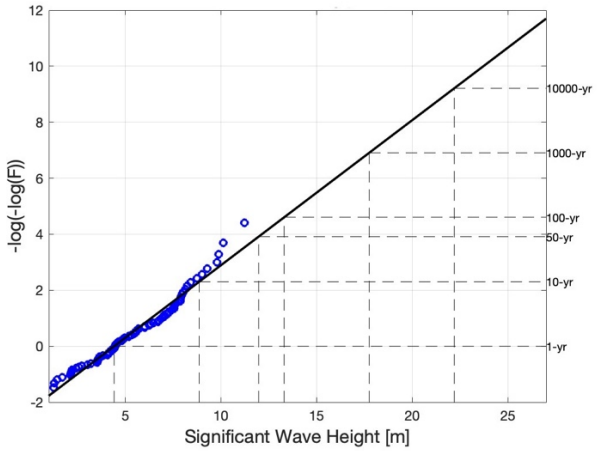


Figure 1: MAB GF-EC Tropical

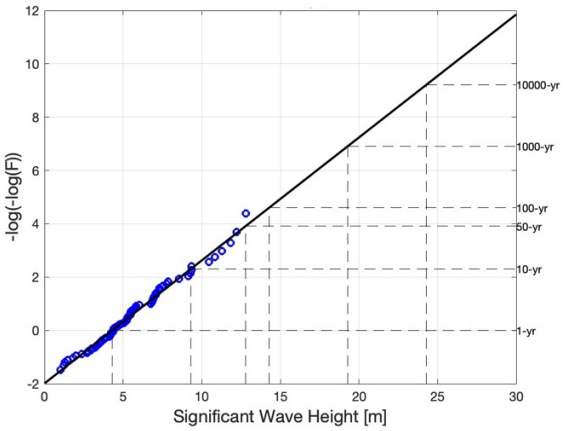


Figure 2: NA GF-EC Tropical

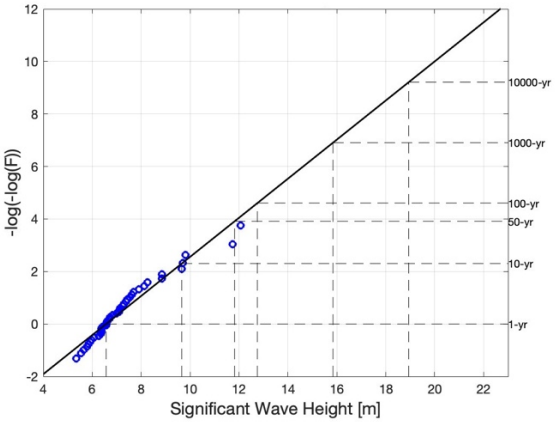


Figure 3: NA high-resolution model

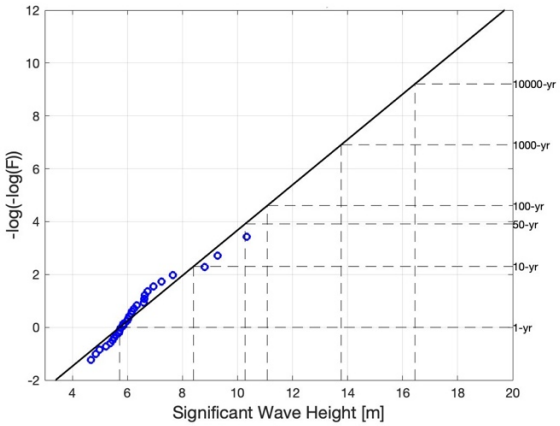
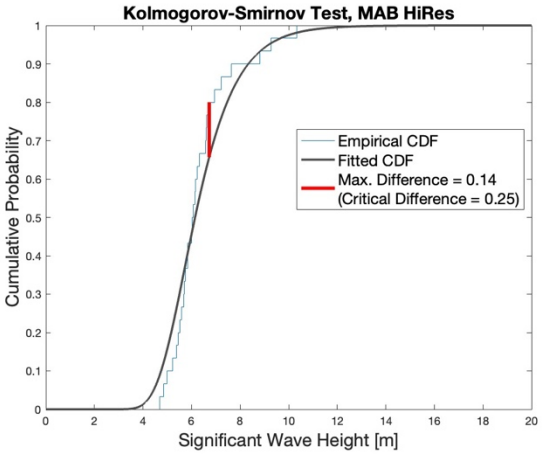
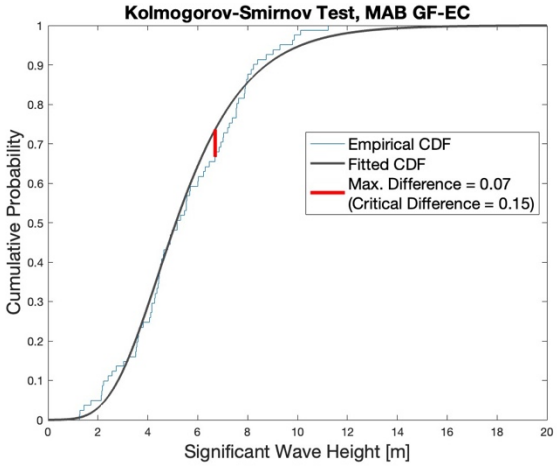
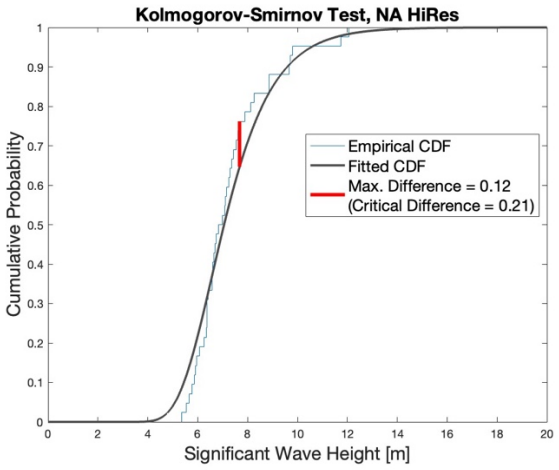


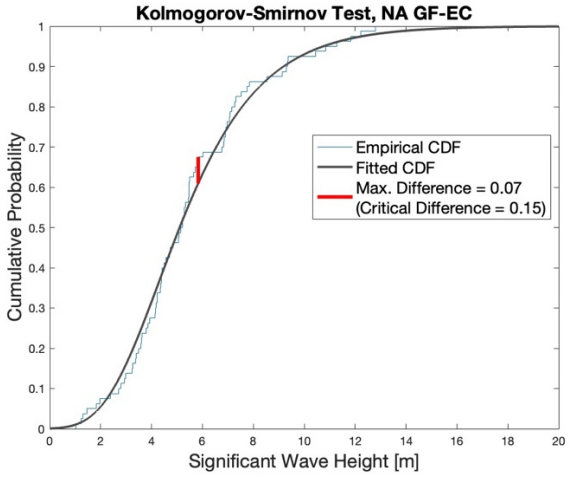
Figure 4: MAB high-resolution model

2. The authors say “block maxima method was considered suitable for this study” by referring two papers, however, the reason is not clearly mentioned. What part of these papers are referred? Need explanation.
    a. Papers by Bhaskaran (*Comparison of Extreme Wind and Waves Using Different Statistical Methods in 40 Offshore Wind Energy Lease Areas Worldwide*, Energies, 2023), Barthelmie (*Extreme Wind and Waves in U.S. East Coast Offshore Wind Energy Lease Areas,* Energies, 2021), and Jonathan (*Statistical modelling of extreme ocean environments for marine design: A review*, Ocean Engineering, 2013) are referenced as published examples of suitable applications of Annual Maxima for significant wave height extreme value analysis. Additional figures for the Block Maxima-Gumbel distribution are provided in the revised Figure 4, and in the appendix (KS one-sample tests).









3. The word “Calibration”, which also appear in text many time is ambiguous. The authors have to explain the detail methodology or procedure.
    a. The calibration and “model skill’ (validation) discussions have been revised with statistics and more details on the validation locations, time periods, storms used, and calibration parameters. Calibration of the North Atlantic model is discussed in “Model Skill”, line 143: *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. After this calibration, validation was conducted over a 10-year period at five observation locations throughout the model domain.*

       More details on the Mid-Atlantic calibration are also provided in Appendix C.
4. Appendix B -> Appendix C
    a. Thank you, updated.
5. These wave models need not only lateral boundary conditions but also bathymetry or sea surface boundary conditions. Table 1 may be used for sea surface boundary conditions for wave models in Table 2, however, it is difficult to understand it because no

explanation made here. Computational area (i.e. domain) for each model is also important information. The authors have to explain about these modeling configurations.

a. Thank you. The bathymetric input to each model is added to the model descriptions in sections 2.1.1, 2.1.2, and 2.1.3, along with domain extents, which appear in Figure 3 in the revised manuscript:

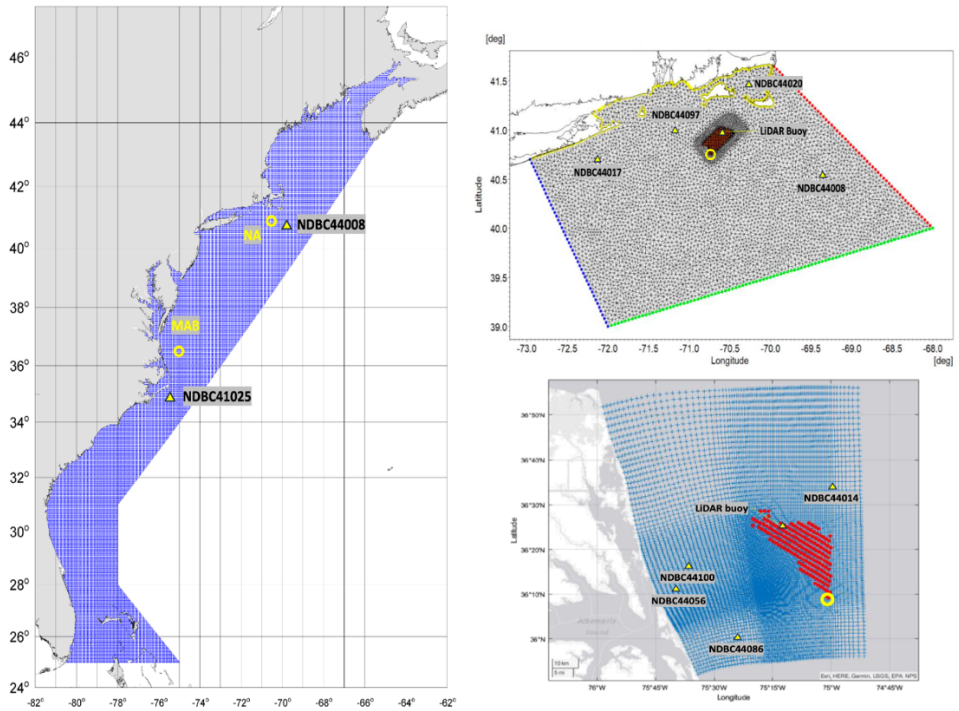


Figure 3. Analysis locations are indicated by circles and validation locations are indicated by triangles. Clockwise from the left: (a) The structured-grid GF-EC domain spans from 25 to 45.85 °N, and 82 to 64.3 °W. (b) The unstructured-grid NA "high-resolution" domain spans from 39 to 41.5 °N, and 73 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.

6. Is Wrenger (2022) publicly available report? If not, the authors have to explain the relevant part in the report in Annex or somewhere.
    a. Thank you. Unfortunately, the project report is not a publicly-available document, however, relevant details concerning modeling, validation, and study reproducibility are incorporated in the revised paper. The reference is removed.
7. “vertically nested domain...” is correct? Horizontal nesting to perform locally high-resolution simulation is more common way to use WRF. Explanation about computational domain is needed.
    a. Thank you, this is a typo and has been updated to “horizontally-nested”.
8. “real lateral boundary condition...” mentioned here may be CFSR according to Table 1. However, it is difficult to understand that. It is recommended to mention as text clearly.
    a. “Real lateral” is a WRF boundary condition option, as opposed to an idealized applied value; it is specified that this refers to the CFSR model data (line 113).

9. Is Georgas(2023) publicly available report? If not, the authors have to explain the relevant part in the report in Annex or somewhere.
    a. Thank you. Unfortunately, the project report is not a publicly-available document, however, relevant details concerning modeling, validation, and study reproducibility are incorporated in the revised paper. The reference is removed.
10. Because 62m and 40m are shallow water region, simulated wave height by wave model is very sensitive to water depth, especially in high wave height. Also, there are geographical distances between buoy and model grid. The authors have to explain the differences between real and modelled water depth.
    a. The real depth at buoy 44097 is 49.4m. The modeled depth at the NA high-resolution analysis point is 62m. The real depth at buoy 44014 is 49.1m. The modeled depth at the MAB high-resolution analysis point is 38m. The manuscript has been updated to reflect this (section "Model descriptions"), with bathymetric sources.
11. Explanation of abbreviation OWI3G is needed.
    a. The "Oceanweather 3rd generation" description is added to the text (line 124).
12. "100 years of tropical storms and 75 years of extra-tropical storms..." Use of as long as data has aspect to improve extreme value, however, old data may have quality problem. The authors have to discuss about data quality issue.
    a. Thank you, this is an important point. The model is both validated and verified for post-1979 events. Selected events are then simulated with the verified model based on available observations, however limited, or to events which were known to have a significant impact on coastal areas. The text is updated to reflect this.
13. There is no explanation about temporal resolution about buoy observation. Also, the authors have to explain how handled or corrected differences of temporal resolutions between each model and buoy observation.
    a. Buoy temporal resolution is added to the text. Note that the quantile-quantile plots (formerly Appendix A) have been replaced by validation statistics (Tables 3, 4, and C2 in the revised manuscript) for all three models. Please see replies to Reviewers 1 and 3 for examples of these validation tables.
14. Unit is needed for the RMSE values.
    a. RMSE unit of m is added. The figures have been replaced with tabulated values from multiple buoys in Tables 3, 4, and C2 in the revised manuscript. Please see replies to Reviewers 1 and 3 for examples of these validation tables.
15. The authors have to explain the real and modelled water depth. According to line 109 and 123, water depths are 62m and 40m. Because these depths are shallow, simulated wave heights by wave models are very sensitive to water depth, especially for high wave heights. Also, according to Table 2 model resolutions are 400m for NA and 600m for MA, however, grid point 29km away from buoy are used for validation. The authors have to explain the reason.
    a. All buoys in the region have been used for validation and a selection for calibration, as shown in the new Figure 3 (see comment 5). High-resolution geophysical survey data for each turbine location (that is, both analysis points

discussed here) were an input to the metocean model, and supplemented by GEBCO data outside of the project area, which has been specified in the model descriptions.

16. font of “x” in formula and text are different.
    a. Updated formatting.
17. “exp” and “ln” should not be italic letters.
    a. Updated formatting
18. “empirical estimation” is not clear explanation. The authors have to explain more detail methodology.
    a. The text now clarifies that the estimation procedure is maximum likelihood (line 169)
19. The authors explain “annual largest value” is used distribution fit. Does tropical cyclone occur and approach to site of interest every year? If not, authors have to explain how handled annual maximum value derived by tropical cyclone for zero tropical cyclone years.
    a. In the Mid-Atlantic Bight, tropical cyclones are an annual occurrence. In the North Atlantic, there are a handful of years where tropical cyclones did not occur. In the post-processed continuous models, where extra-tropical events have been removed, a smaller, non-TC, non-ETC event is picked as a maxima (TCs and ETCs are the strongest storms in the region). In the GF-EC dataset, no maxima is selected for the year. The text is updated to reflect this (line 173).
20. “Extreme Value Theory assumes that extremes are independent variables.” I could understand that what the authors want to say but this sentence may be difficult to understand for some reader. it is suggested to explain a bit detail by changing “extremes are” to other word.
    a. Updated phrase to “extreme values are”
21. What we can understand from figure 3 is only that extreme distribution obtained from GF models and high-resolutions model show qualitatively close values or distributions for extra-tropical cyclone. Because both these are obtained model, nothing explains storm physics are represented or not.
    a. Yes, Figure 3 only represents the statistical distributions. However, the clear difference in trend and magnitude motivates further investigation into how these storms are represented (i.e., if physical processes are well/poorly resolved) in the following sections.
22. Drawing annual maxima used for fitting of extreme distribution in Figure 3 is suggested.
    a. Thank you. This has been addressed in previous comments and is added to the manuscript.
23. Dolan-Davis scale and Saffir-Simpson scale are probably US specific. References to explain about these scales are needed.
    a. The text has been updated with originating references.
24. Duplicated “the”.
    a. Thank you, this is updated.

25. Although the authors explain “neither model...”, lack of representation of high frequency wave could be caused by frequency range of wave model. Add information about the highest frequency in Table 1.
    a. Thank you, this was erroneously left out in the original manuscript. The maximum frequency for each model is added to Table 2 of the revised manuscript:

Model	Tool	Resolution	Boundary Conditions	Coupling	Spectral Parameterization
NA HiRes	MIKE21	600m wave (2D) 600m hydro (2D) 1-hour	DHI East Coast Waves (waves) DHI East Coast (hydro)	1-way, hydro to waves	36 directions 32 freq. bins 0.033 Hz min 0.667 Hz max
MAB HiRes	SWAN+DELFT3D	400m wave (2D) 400m hydro (3D) 1-hour	ERA5 (waves) HYCOM (hydro)	2-way, waves and hydro	36 directions 24 freq. bins 0.05 Hz min 0.448 Hz max
GROW-Fine East Coast	OWI3G+ADCIRC	5.5km wave (2D) 5.5km hydro (2D) 15-minute	GROW2012 (waves) Prevost ’08 (hydro)	No dynamic coupling. Reanalysis of each modeled storm.	48 directions 26 freq. bins 0.029 Hz min 0.322 Hz max

Table 2. Wave and hydrodynamic parameterization for the three investigated models in the North Atlantic, Mid-Atlantic Bight, and along the US Atlantic coast (GROW-Fine East Coast). All model bathymetries are derived from GEBCO. The NA and MAB model bathymetries are supplemented with 1m geophysical survey measurements within the project area.

26. Why are wave height axis normalized? In general, higher wave height more difficult to simulate. For this reasons, magnitude of wave height is very important information and recommended not to be normalized.
    a. We understand that absolute values are important for quantifying the quality of a modeled event. However, as a condition for use of the GF-EC dataset for academic investigation (safeguarding the intellectual property of the GF-EC product) the normalized results are presented here, to preserve scale between model results. We believe that this observable difference in scale is nonetheless valuable, as there are no other comparable datasets publicly available.
27. Need explanation why SWAN + WRF shows poor Tp resolution.
    a. The quantile-quantile plot of Tp (MAB “high-resolution”) this comment refers to has been replaced by a set of validation statistics in Tables 3, 4, and C2 of the revised manuscript. Please see the replies to Reviewers 1 and 3 for the revised validation tables.
28. The authors have to explain how inside/outside of storm fetch was defined in this manuscript. Also, explanations about closest approach distance and radius of maximum wind speed of cyclones are need as general information to judge inside/outside.
    a. Inside/outside categorization was taken here simply as within or beyond of 200 km from the storm eye. This was identified for the closest point on the storm IBTRaCS record to the analysis location. Two similarly-scaled events with clear differences in distance were selected for this purpose.
29. Y axis of Figure 5a and 5b are cut off. Also, Tp=18 on Y axis in Figure 5c is missing.
    a. Thank you, the formatting is updated.
30. In general, wave periods in inside of storm are dominated by wind-wave and those for outside are significantly affected by wind field both inside and outside of cyclone. Is simulation period enough long, or simulation area enough large? It is suggested that

draw wind field and wave height field and add explanation about simulation period about this cyclone. These may help to understand this phenomenon.

   a. Thank you. More details about the model domains are added in Figure 3 of the revised manuscript. The high-resolution models are continuous on an hourly basis. The GF-EC model timescale varies between individual events and concludes when the storm peak decreases to pre-storm levels; for example, the GF-EC full storm period for the three events is shown in Figure 7 of the revised manuscript. While the GF-EC model may only cover a snapshot of the storm as it passes by a location, the domain extent allows the full storm extent on the east coast to be modeled, from growth through decay.

31. Add reference height of "storm winds (90 knots, or 46 m/s)"
    a. The value is taken from IBTRaCS; the US National Hurricane Center defines the surface wind speed as 1-minute sustained average at a 10m height. The manuscript is updated to reflect this (line 255).
32. "Wave buoy measurements occurred on a 30-minute cycle, ..." Is buoy observation available for Hurricane Bob. If so, the authors have to show comparison with observation and modeled value such as Figure 4 and Figure 6b.
    a. Unfortunately not. Observations began at (NA location) NDBC 44097 in 2009. Additionally, there is no recorded data at (MAB location) NDBC 44014 during the passage of Bob.
33. Although the authors considered "primarily to be a function of fetch or duration representation", development of cyclone or error of track depends on model horizontal resolution of weather simulation, or wind data etc. The authors have to understand and explain only wave models are "high resolution" in this study, not for wind models, which were used for input of wave model.
    a. Thank you, this is a helpful clarification. The "high-resolution" nomenclature was one way to identify a commonality for investigating wave model performance. A specification that the high-resolution moniker only applies to waves is added in the introduction. Indeed, it's an important point that these traditional coupled "high-resolution" models are forced with atmospheric models with resolutions that are too low to assess tropical cyclone track and core features.
34. There is no explanation about what Cd model used in this study. Although the authors show wind stress values in Figure 10, these values are strongly affected by Cd models or formula, and each wave models may use different Cd model. Need explanation.
    a. Thank you. The text is updated to reflect that the coefficients of surface drag in the "high-resolution" models are from the Charnock formulation (line 103).
35. It is not clear that the meaning of "higher-than-average tropical cyclone activity". If it means that annual occurrence is higher than usual, I comment that it is not affect extreme wave height because the authors use only annual maxima.
    a. This is based on multi-decadal assessment. See the histogram below, for example, for the Mid-Atlantic region. This figure was not originally included in the manuscript for brevity, and can be included if considered important.

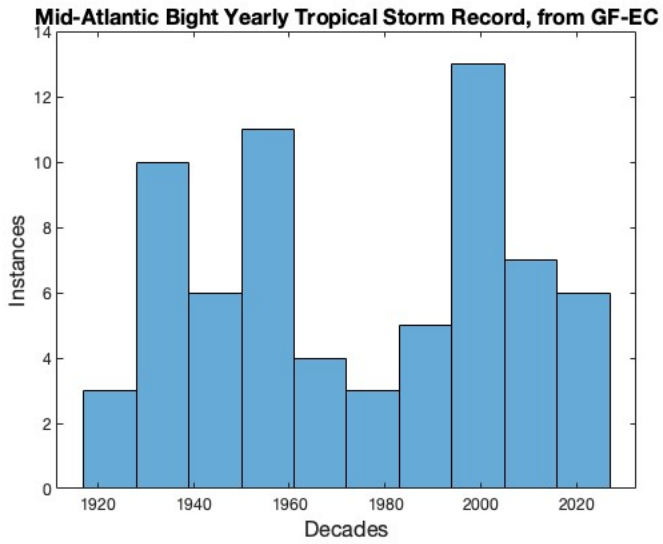


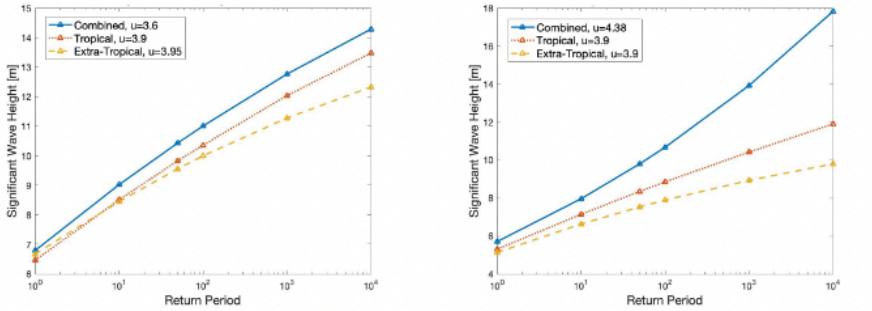
Figure 5: number of tropical cyclone events at the Mid-Atlantic location, 1924 - 2020

36. The meaning of sentence “in fact...” is not clear. Are extreme distributions in Figure 11a and 12a based on “original high-resolution data set” or “post-processed highresolution data set”? If the result mentioned in the sentence is not shown in graph, it is better to add word “not shown in graph” in the text.
    a. Thank you. This sentence refers to the difference between the solid blue lines and the red dashed lines in revised Figure 5; this reference has been added to the text.
37. It is suggested to draw annual maxima use for fitting of each extreme distribution to understand reasonability, trend etc. of each distribution.
    a. These plots have been added to the revised manuscript in Figure 4.
38. “ERA5-boundary conditions” -> “ERA5 wave boundary conditions” is suggested.
    a. Thank you, this has been updated.
39. It is understood that these four are main conclusions of this study, however, 1) it is questionable that how “200km or less” in bullet 1 is quantified (e.g. cyclone has radius more then 200km exist.)? There is no detail discussion or deep insight about this. 2) bullet 2-4 are already explained in IEC 61400-1 Annex J. Although the standard is about only for wind, not mentioned for wave, the authors should at least explain and refer in somewhere in this manuscript, then have to explain the differences or originality of this paper.
    a. The 200km threshold was selected as an engineering rule of thumb based on the sizes of hurricanes in the mid-Atlantic and north Atlantic sites. It is certainly true that there are larger tropical storms than this. However, as no distinction is currently made in standards/industrial practices between inside/outside of storm fetch, a value is proposed as a first step.
    b. Local wave features are due to a combination of wind forcing and additional factors, therefore this investigation is considered separate from and in addition to the implications of IEC 61400-1 Annex J. Annex J of IEC 61400-1 covers Monte-

Carlo simulations (synthetic hurricane modeling), and indeed it does identify the separated analysis of storm types. This has not been followed in a number of publications—i.e., use of datasets such as ERA5 and CFSR for determining extremes. However, this is not specified for the wave and ocean environment, and a key motivation of this paper is: given that wave growth is driven by a number of factors, does that matter? We suggest in this paper that it does, and more work is required to quantify which features are important for future ocean modeling and extreme events, be it directly hindcasted or by coupling to synthetic wind fields. Many publications to date use reanalysis datasets that are time-limited to 30-40 years (i.e., CFSR-, HYCOM-derived) and this caveat has implications for use of these datasets for extremal analysis. The authors have not seen mention of this time duration as insufficient for hurricanes but sufficient for winter storms in the standard referenced. As obvious as it may seem, there do not appear to be any standards/requirements that ocean extremes be quantified by the same wind fields as those used to determine wind extremes. This paper is an attempt to highlight some of these gaps in current offshore wind metocean methods. Indeed, there are plenty of public opinions that the current "high-resolution" model approach remains the state of the art for offshore wind design on the US east coast, including when these results were presented at the NAWEA conference.

Finally, the paper lays out the metocean models and analysis process in use in the offshore wind industry today for wider scrutiny, discussion, and replication. Normally, this is not presented in the public domain.

40. What value are used for threshold "u"? Need explanation.
    a. The values of threshold for each analysis are added to legend in revised Figure A2, Appendix A.



(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.

41. Unit for RMSE is needed. Also, definition of NRMSE is needed.
    a. RMSE unit of m is added. NRMSE is changed to Scatter Index, SI. The figures have been replaced with tabulated values from multiple buoys.

42. Because the authors decided to use annual maxima for fitting of extreme value distribution, the validation should be made for annual maximum wave heights. Otherwise, readers don't understand the reasonability of all results and conclusions.
    a. Thank you. Please see the K-S test results in the reply to Question 2, and the distribution fit in reply to Question 1 for goodness of fit assessments of the annual maxima data from each model. Nearby observations to the NA site (NDBC 44097) cover 10 years of data, which is not considered a long enough period to conduct extreme value analysis. However, quantile-quantile plots of annual maxima values between observations and model data, at both sites (note the 40-50km distance between), are presented below for the high-resolution model datasets. Note that due to the differences in dataset duration, some model annual maxima events did not occur during the measurement period.