Investigation of wave height and period distributions in coastal environments

Jemerson P James and Vijay Panchang

Department of Ocean Engineering, Texas A&M University, College Station, TX 77843, USA. (jemersonpjames@tamu.edu, corresponding author)

Abstract

The short-term statistics of wave conditions in coastal waters around the UK have been investigated using over 40,000 half-hour long sea state records with significant wave height greater than 3 m. The extensive data set facilitates an assessment of various wave height and period distribution models in shallow and intermediate waters. The results reveal that the relative wave height H_s/D (where H_s is the significant wave height and D the water depth) can serve as a key indicator in choosing the distribution with least error in a given sea state. The Naess model is found to be the most accurate in describing the tail of the wave height distribution in a sea state for low relative wave heights ($H_s/D < 0.2$), and the depth-dependent van Vledder model for high relative heights ($H_s/D > 0.4$). In between these sea states, a transition in the performance of the deep-water and the depth-dependent models is visible. When details of the spectrum are not available, the Weibull distribution is the most accurate amongst applicable models, in spite of considerable variability in its parameter values. While the spectral bandwidth appears to have minimal impact on the distribution of wave heights in a sea state, it does appear to influence the distribution of wave periods. Wave period relationships based on measurements are found to deviate from empirical relationships proposed, for example, by the US Army Coastal Engineering Manual. Improved formulas that incorporate the spectral width are therefore proposed.

PLAIN LANGUAGE SUMMARY The distribution of wave heights and periods is usually obtained by a number of theoretical models. The validity of these models is examined using a large data set off the coast of the UK for sea states indicated by significant wave heights greater than 3 m. A large data set, exceeding 40,000 sea states, enables a comprehensive assessment of these models, which is performed, for the first time, as a function of the relevant sea-state characteristics. The wave height distribution models with the lowest errors are identified. For wave periods, commonly used empirical formulas are shown to deviate from observations and improvements are recommended.

1. Introduction

Certain "characteristic" ocean wave heights (i.e. averages relating to various percentiles in a time series) are frequently employed in the planning of various coastal and maritime operations. In many cases, one relies on wave hindcast/forecast models or in situ measurements (e.g. Panchang and Li 2006; Panchang et al. 2013; Singhal et al. 2010) for obtaining the needed wave information. Although these sources are based on spectral calculations or time series of water surface elevations, only summary information, consisting typically of the mean or peak wave period, the mean direction, and the significant wave height (H_s, usually defined as $4\sqrt{m_0}$ where m_n= nth spectral moment), is generally available. However, this information may not be sufficient since other quantities such as $H_{1/10}$, $H_{1/100}$ etc. are frequently needed for practical applications, e.g. engineers require $H_{1/10}$ for the design of breakwaters using Hudson's formula. (Here, $H_{1/n}$ represents the average of highest n% of the wave heights in a typical, say 20-minute, sea-state). The different percentile averages may be estimated from H_s using several available distributions. These include the well-known Rayleigh distribution (Longuet Higgins, 1952) and modifications to it, as well as a number of others, for instance, those developed by Forristall (1978); Naess (1985); Boccotti (1989); Tayfun (1990); Battjes and Groenendijk (2000); Mendez et al. (2004); and Wu et al. (2016).

Several authors have examined the reliability of some of these distributions by comparing their results to field data. However, the field observations used were often fairly limited both in terms of the data duration as well as in the number of locations examined. For example, Cartwright and Longuet-Higgins (1956) examined the validity of Rayleigh model using only two wave records measured by a shipborne instrument. Earle (1975) analyzed ten hours of wave data during Hurricane Camille in the Gulf of Mexico; Forristall (1978), Nolte and Hsu (1979), and Longuet-Higgins (1980) examined additional wave records for this hurricane. Tayfun and Fedele (2007) reviewed the performance of three distributions using two datasets, again quite short in length (viz. eight and nine hours in duration), from one location in the North Sea. Other efforts based on a specific storm event or relatively short measurement programs include Chakrabarti and Cooley (1977), Dattatri et al. (1979), Larsen (1981), Amrutha and Kumar (2015), and Nayak and Panchang (2016). Relative to these efforts, Vinje (1989) used a significantly larger data set, viz. eighty months of data at one location in the North Sea, to assess the performance of the Raleigh, Longuet Higgins (1980), and Naess (1985) models.

With the passage of time and improvements in technology, long wave records at multiple locations have become more readily available, enabling a comprehensive inter-comparison of the distributions. Attempts in this direction include the work of Casas-Prat and Holthuijsen (2010), who utilized data (varying between three and twelve years in length) from four buoys in water depths of 45-74 m off the coast of Spain. They report that some "Raleigh-like" models and the Forristall (1978) models were the most accurate. Kvingedal et al. (2018) used 7.6 years of data from one location in the North Sea (depth = 190 m) and compared the Raleigh, Naess and Forristall distributions. They found the Rayleigh distribution to be a "conservative upper bound" and the Forristall distribution to have high accuracy in most sea-states. For very severe sea states (H_s > 9.5 m), though, they found all three distributions to be deficient. At ten locations shallower than those in the aforementioned two studies, Karmpadakis et al. (2020) used about six years of measurements in the North Sea (water depths mostly varying between 23.3 m and 45 m). They found the Boccotti (1989) model to be the most reliable for describing the observed wave height statistics for the case of low relative wave heights (defined as H_s/D, where D = water depth), and the depth-dependent models to be optimal for larger relative wave heights. They also found model performance to be influenced by the sea state (i.e., H_s/D, the relative depth K_pD (where K_p = wave number corresponding to the peak period), the spectral width, and the steepness).

The present paper is a continuation of these efforts, with an emphasis on coastal regions. We use approximately five to eight years of data from 34 locations around the entire UK coast (Fig. 1). The majority of the locations have depths of around 10 m, which is shallower than those in the previous three studies; and, since 70% of the locations are less than 5 km from the coast and nearly 90% less than 8 km, the present effort may be viewed as an investigation of the distributions in the "coastal" regime. Finally, since studies involving wave period distributions are relatively few, we examine the relationships between the maximum period T_{max}, mean period ($\overline{T}$), and the average of the highest one-third ($T_{1/3}$) using a database of approximately 41,120 sea states.

The major questions that motivate the current study relate to the extent to which different theoretical models predict the observed wave height and period distributions in intermediate and shallow waters. In particular, based on recent research noted above, we examine the performance of various distributions relative to the sea state characteristics. Their performance must also be considered in the context of practicality (that can influence model selection). To be specific, some

of the distributions require only bulk parameters such as the significant wave height, while others require more detailed spectral information. The paper is organized as follows. Section 2 provides the details of the data used along with the quality control procedures and identifies key sea state characteristics that enable subdividing the data into appropriate groups. The different theoretical distributions considered here are described in Section 3, and Section 4 provides the results of the wave height and period comparisons. The paper ends with concluding remarks in Section 5.

2. Data

The wave data used in this study are obtained from the Channel Coastal Observatory (CCO) in the UK, which operates a network of coastal measurement programs. Santos et al. (2017) and Dhoop and Mason (2018) have used these data for studying spatial variations in relation to meteorological forcing and to estimate extreme wave height climatology. Dhoop and Mason (2018) also provide a detailed description of the measurements; therefore, here we only describe the salient features. The wave data, collected using Datawell wave rider buoys deployed at about 40 locations (Fig. 1) in 30-minute segments, are available on the CCO website (https://coastalmonitoring.org).

The Datawell MKIII wave rider buoys use vertical accelerometers mounted on a gravity-sensitive platform to measure heave. The analog output from the sensor is subjected to a low pass filter with a cut off frequency of 1.5 Hz and then sampled at 3.84 Hz; it is then subjected to a high pass filter with a cutoff of 30 seconds and converted to a sample rate of 1.28 Hz. The 1.28 Hz water surface elevation data as well as the computed spectra are provided on the website. Some details regarding the wave buoys are summarized in Table 1. For most locations, the 1.28Hz water surface elevation data are available from 2014. However, for some locations, the available data covers a shorter span.

Relating to these measurements, two issues must be noted. First, it is generally recognized that measurements obtained from wave rider buoys tend to “linearize” the wave heights owing to their hydrodynamic characteristics, i.e. they exhibit a tendency to “flatten” the peaks and/or to be dragged through the peaks; see for example, Casas Pratt and Holthuijsen 2010; Kvingedal et al. 2018. However, as noted by these researchers and by Tayfun and Fedele (2007), nonlinear effects appear to influence wave crest measurements; their effect on wave height measurements is minimal. Second, the sampling interval has an impact on the accuracy with which the wave heights and crests are measured (Zheng et al. 2006; Tayfun 1983) and some recent efforts are based on

high sampling rates, of the order of 4 Hz or better (e.g. Kvingedal et al. 2018; Karmpadakis et al. 2020). In the present study, the raw data were obtained at 3.84 Hz, however the spectra and time series were subsequently based on a 1.28 Hz resampling. We presume this is due to the logistics of maintaining a very large network as part of an ongoing program. To address any possible limitations that may arise, we have followed the recommendations of Tayfun (1983) and modified the time-series-based estimates of $H_{1/n}$ in the following manner:

$$\frac{H'_{1/n}-H_{1/n}}{H_{1/n}} = -\frac{\pi^2}{6}\left(\frac{\Delta}{\bar{T}}\right)^2 \quad (1)$$

where $H'_{1/n}$ is based on the measured data and $H_{1/n}$ is the true value. Δ is the sampling period and mean period, $\bar{T}$, is estimated as $T_{m01} = m_0/m_1$.

From these data, we consider only the sea states with significant wave height $H_s \geq 3$m, since we are interested in assessing how well the distributions predict sea states that are of practical interest to coastal or offshore applications. This limit aligns roughly with the lower range of wave heights specified for a Beaufort 6 sea state, where large waves start forming. A few locations (e.g. Felixstowe and Minehead) did not have any sea state that exceeded this criterion during the period considered and are hence not included in Table 1. At most of the other locations, during April–October, less than 5% of the sea states recorded met this criterion. On the other hand, during December–February, Wave Hub experiences such sea states more than 50% of the time. In general, the largest H_s values were of the order of 10 m at Wavehub and 7-8 m at six locations (Chesil, Westbay, Bideford Bay, Porthleven, Looe Bay, and Perranporth) lying around the southwest of UK. Twelve other locations had largest *Hs* values in the range of 5-6.7 m. The largest mean period was approximately 13.5 seconds, and the vast majority of the spectra showed one dominant peak.

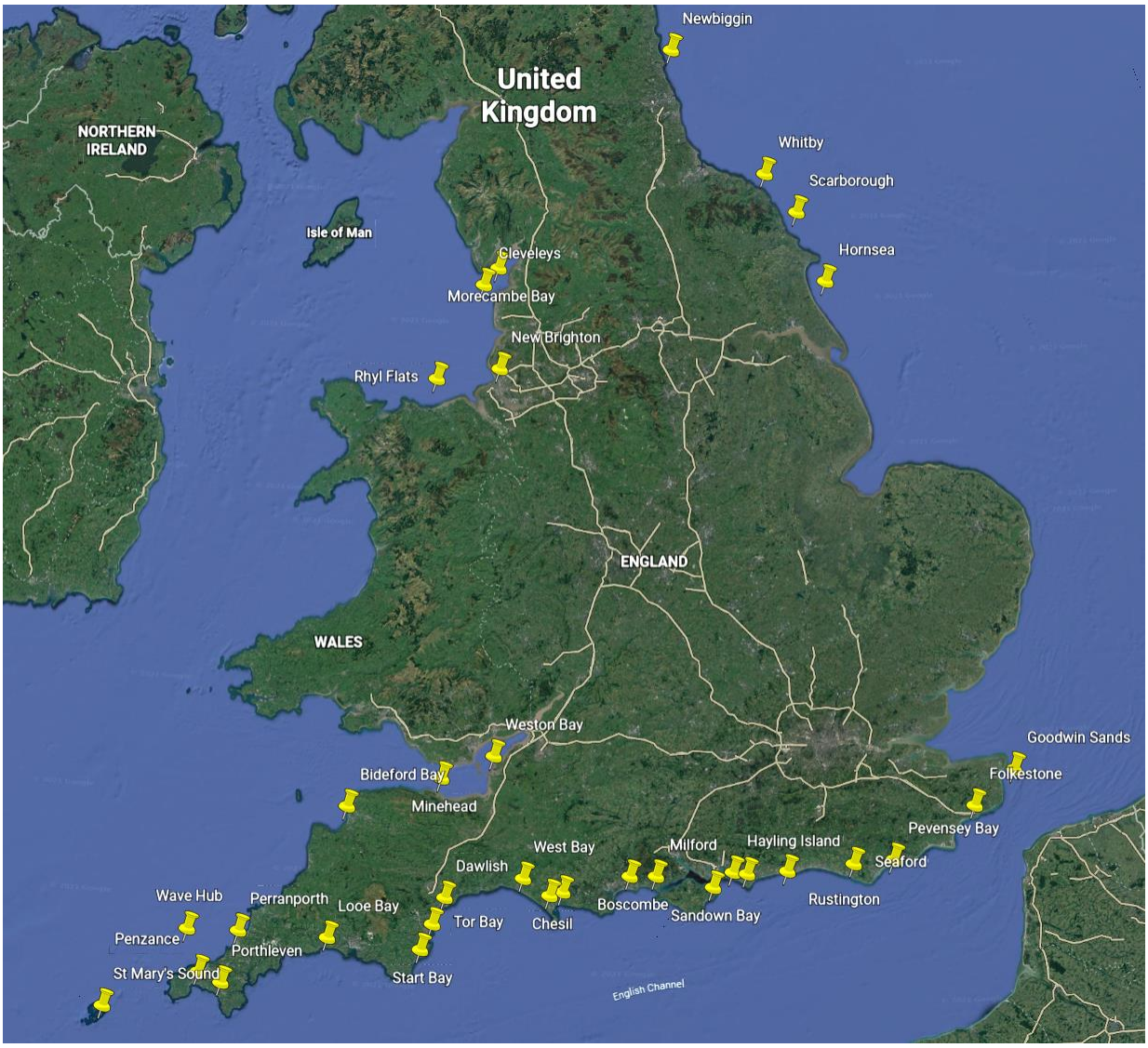

Fig. 1. Locations of Datawell wave rider buoys around the UK (courtesy: Channel Coastal Observatory).

Table 1 Locations of Wave Buoys

Location	Depth (m)	Period of data considered	No. of sea states with H_s > 3 m (after QC)	Shortest distance from shore (km)
Bideford Bay	11	2014/ 01 - 2020/ 07	4117	3.24
Blakeney Overfalls	23	2017/ 02 - 2020/ 12	34	10.91
Boscombe	10.4	2014/ 01 - 2020/ 12	76	0.91
Bracklesham Bay	10.4	2014/ 01 - 2020/ 12	325	2.34
Chesil	12	2014/ 01 - 2020/ 11	2249	0.39
Cleveleys	10	2014/ 01 - 2020/ 12	824	8.73
Dawlish	11	2014/ 01 - 2020/ 12	110	2.35
Folkestone	12.7	2014/ 01 - 2020/ 12	15	0.98
Goodwin Sands	10	2014/ 01 - 2020/ 12	4	5.77
Gwynt Y Môr	10	2016/ 04 - 2020/ 11	346	16.28
Happisburgh	10	2017/ 03 - 2020/ 12	58	0.82
Hayling Island	10	2014/ 01 - 2020/ 12	162	5.03
Hornsea	12	2014/ 01 - 2020/ 11	428	5.84
Looe Bay	10	2014/ 01 - 2020/ 11	1076	2.65
Lowestoft	20	2017/ 06 - 2020/ 12	47	3.78
Milford	10	2014/ 01 - 2020/ 12	29	1.48
Morecambe Bay	10	2014/ 01 - 2020/ 12	78	5.95
Newbiggin	18	2014/ 01 - 2020/ 11	1242	1.21
Penzance	10	2014/ 01 - 2020/ 11	336	1.23
Perranporth	14	2014/ 01 - 2020/ 11	8717	1.10
Pevensey Bay	9.8	2014/ 01 - 2020/ 12	292	4.98
Porthleven	15	2014/ 01 - 2020/ 11	2449	1.14
Rhyl Flats	7.2	2016/ 05 - 2018/ 10	17	7.93
Rustington	9.9	2014/ 01 - 2020/ 11	546	7.34
Sandown Bay	10.7	2014/ 01 - 2020/ 12	59	1.24
Scarborough	19	2014/ 01 - 2020/ 11	2201	4.33
Seaford	11	2014/ 01 - 2020/ 11	261	1.26
St Mary's Sound	53	2014/ 05 - 2020/ 11	2630	0.95
Start Bay	10	2014/ 01 - 2020/ 11	373	1.59
Tor Bay	11	2014/ 01 - 2020/ 12	94	2.33
Wave Hub	50	2015/ 03 - 2018/ 05	8485	16.37
West Bay	10	2014/ 01 - 2020/ 11	1421	1.22
Weymouth	10.6	2014/ 01 - 2020/ 12	20	1.60
Whitby	17	2014/ 01 - 2020/ 11	2031	1.52

2.1 Quality control

CCO conducts quality control (QC) tests monthly and annually and publishes an annual dataset for each location in addition to the real-time 1.28Hz data. Sea states that fail the QC tests, which consist of "out of range" and "jump" checks for H_s, peak and mean wave periods, direction, spread, etc. are flagged. Additional details may be found in CCO's QC manual (Mason and Dhoop, 2017). The annual data also includes flags for periods with missing data. We discarded all sea states that were flagged by CCO. However, even in the remainder, we found a number of suspicious measurements and hence applied additional checks on the data. For example, the 1.28 Hz data were found to contain some abnormally large wave heights which are artefacts of the buoy riding down the crest of large breaking waves, or of waves breaking over the buoy and inflicting a "shock" on the accelerometers (Dhoop and Mason, 2018). Following Casas-Prat and Holthuijsen (2010) and Karmpadakis et al. (2020), the flags listed below were adopted and applied to sea states that passed CCO's QC tests. Among these, flags 2, 3 and 6 address the issue discussed above.

Flag 1: Five or more consecutive data points of equal value in the wave time series

Flag 2: Vertical acceleration between two consecutive data points in the wave time series greater than 0.5g

Flag 3: Difference between the elevation of two consecutive data points in the time series greater than 2.83 H_s

Flag 4: Any wave period longer than 25 seconds

Flag 5: Energy in the spectrum below 0.04 Hz greater than 5% of the total energy

Flag 6: Crest/trough heights greater than five times the standard deviation

Flag 7: The ratio $H_{max}/D > 0.8$

Flag 8: The ratio $H_{max}/H_{1/3} > 2.5$

These flags are applied in the above order. The number of sea states removed by flags1, 2, and 6 were the largest, while the contribution of flags 7 and 8 is very small. Overall, there are 41,120 sea states that passed; only 3.08 % of the data are removed by our additional QC checks.

2.2 Classification of Sea States

Four parameters relating to the selected sea states were examined: viz. steepness, spectral width, H_s/D, and K_pD. For all the sea states, the steepness values, defined as H_s/L_P ($L_P = 2\pi/K_P$), were low, in the range of 0.007 to 0.066, and not particularly significant. Relating to spectral width, however, Karmpadakis et al. (2020) have indicated that this parameter may influence the

distribution of wave heights in a sea-state. More precisely, they observe that the probability of larger wave heights decreases with increasing spectral width. They base this observation on an examination of three sea states corresponding to an Ursell number of 0.005 (considered to be representative of moderate sea states) from one location. We examined this feature using the current data set. We use the following definitions of Ursell number (*Ur*) and spectral bandwidth (ν) :

$$Ur = \frac{H_s}{k_1^2 d^3} \; and \; \nu = \sqrt{\frac{m_0 m_2}{m_1^2} - 1} \qquad (2)$$

where k_1 = wave number based on period T_{m01}.

A total of 1215 sea states with $Ur = 0.005$ were identified; these corresponded to two locations, viz. Wave Hub and St. Mary's Sound. The spectral widths ranged from 0.33-0.62. The exceedance probabilities were calculated for a sample of five sea states (out of 848 available ones at Wave Hub and 367 at St. Mary's Sound), for different spectral widths. These results, shown in Fig. 2, do not reveal any patterns and suggest that the spectral width does not meaningfully influence the distribution of higher wave heights. To expand the investigation to a larger dataset (i.e. all 1215 sea states), we investigated the values of $H_{1/100}$ and $H_{1/300}$ obtained from the time series. They, too, show little dependence on the spectral width, as seen in Fig. 3.

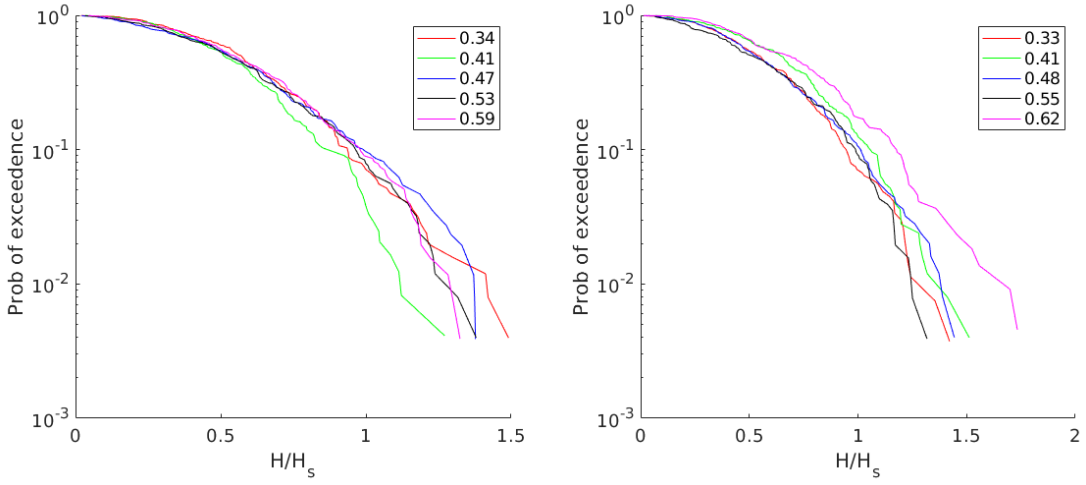

Fig. 2 Normalized wave height distribution (for Ur=0.005) at St. Mary's Sound (left) and Wave Hub (right). Colors indicate spectral bandwidth.

Turning to the other two parameters, Klopman and Stive (1989) interpret H_s/D as the ratio of total local energy density over depth and note, based on measurements, that this is a dominant parameter affecting the distribution of wave heights; it is also a measure of the breaking wave intensity (e.g. Battjes and Janssen 1979). The 41,120 sea states had Hs/D values ranging from 0.05 to 0.63. This would suggest the possibility of breaking being associated with the larger waves. A large portion

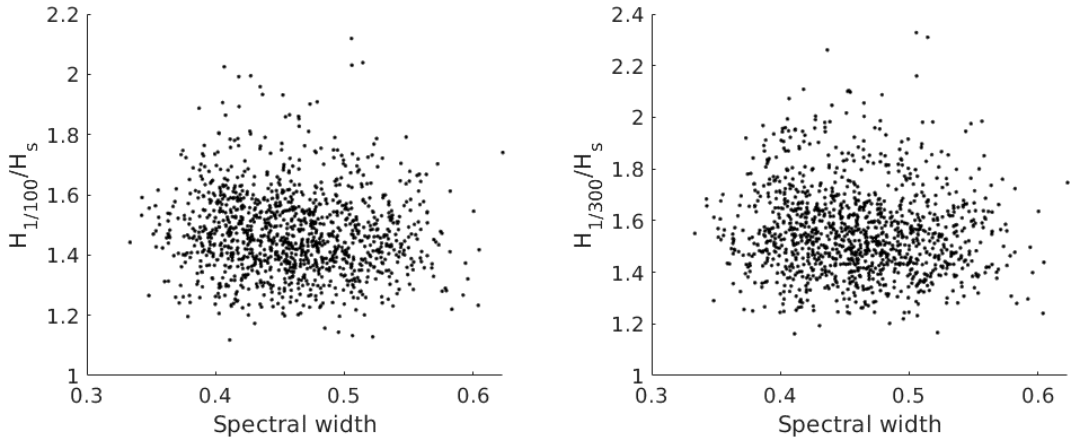

Fig. 3 Variation of normalized wave heights with spectral width for Ur=0.005; $H_{1/100}$ (left) and $H_{1/300}$ (right).

(63%) have H_s/D values between 0.2 and 0.4, while only 24% have H_s/D values below 0.1. This of course is to be expected since we consider only the sea states with $H_s \geq 3$m. We have excluded the lower wave heights, which may be large in number but perhaps of less practical interest, in order to prevent these relatively mild conditions from influencing the statistics relating to model performance. Lastly, the relative depth, K_pD, is commonly used for the classification of deep, intermediate, and shallow water waves. The K_pD range corresponding to the data used here is 0.26 to 5.82. Very few sea states (0.07%) are in shallow water; most of the sea states (97.8%) are in intermediate water depth, and 2.1% are in deep water. A visual description of the data distribution may be obtained from Fig. 4 and based on the pattern, the sea states were divided into six groups (Table 2) for purposes of assessing the wave height distribution formulas given in Section 3. While examining wave periods, though, it will be shown later (Section 4.2) that only the spectral width has some influence on the relationships between characteristic wave periods; hence the comparisons were made without subdivision of the data.

Table 2 Grouping of sea states

Groups	H_s/D	K_pD
A	$H_s/D \leq 0.2$	$K_pD \leq 2$
		$K_pD > 2$
B	$0.2 < H_s/D \leq 0.4$	$K_pD \leq 0.5$
		$0.5 < K_pD \leq 1$
		$K_pD > 1$
C	$H_s/D > 0.4$	$K_pD < 1$

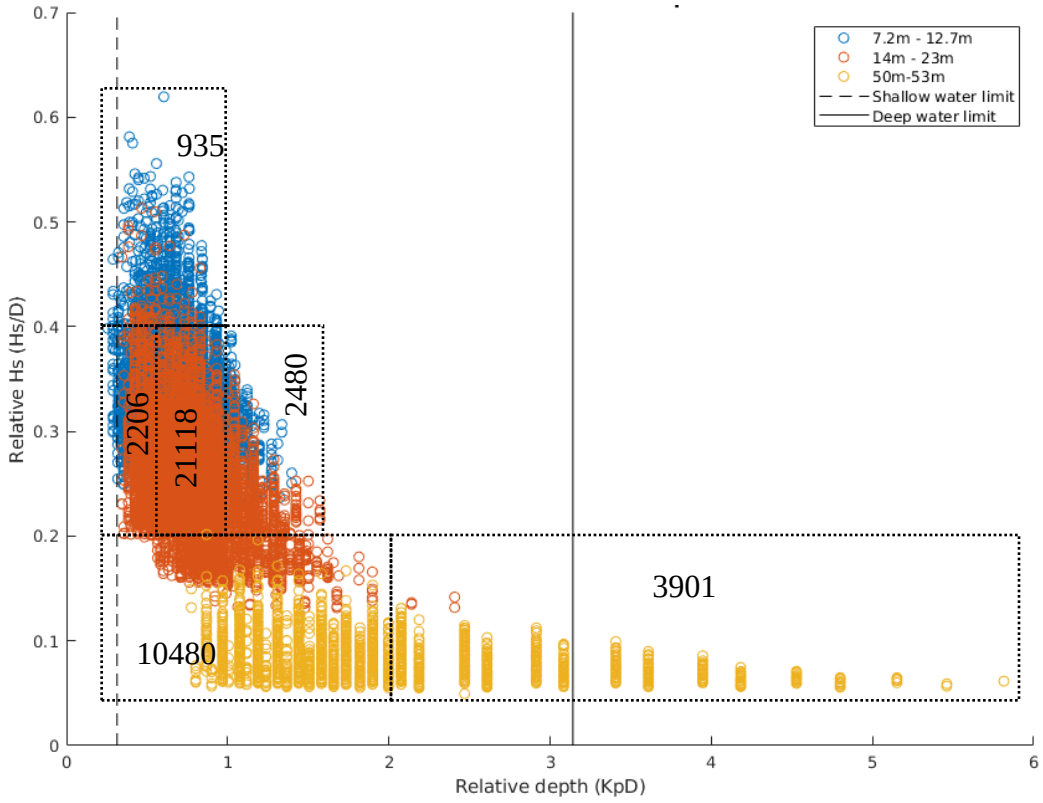

Fig. 4 Sea states that passed QC; colors indicating depth. Numbers indicate number of sea states in the different groups contained within dotted lines.

3. Short-term statistical distributions

3.1 Wave heights

Wave height distribution models examined in recent studies (e.g. Casas-Prat and Holthuijsen 2010, Karmpadakis et al. 2020) can be grouped in to two broad categories, namely deep-water and depth-dependent models. Some depth-dependent models also incorporate the local bottom slope. The exceedance probabilities associated with the seven distribution models considered here are summarized below. Of these, the first five may be regarded as deep-water distributions, and the last two as depth-dependent distributions.

a) Rayleigh distribution (Longuet-Higgins, 1952)

Based on the assumptions of a Gaussian ocean surface and a narrow spectrum, Longuet-Higgins (1952) proposed the Rayleigh distribution for wave heights:

$$Q(H) = \exp\left(-\left(\frac{H}{a}\right)^2\right) \tag{3}$$

where H is the wave height and a = the parameter of the distribution. Maximum likelihood estimation shows that $a = H_{rms}$, the rms wave height. However, the true H_{rms} may not be known *a priori*; for instance, when the time series is not available but only the spectral information is (e.g. from a model output). At such times, an estimate of H_{rms} is obtained from the spectrum as $H_{rms} = \sqrt{8m_0}$ (following Rice (1944, 1945) and assuming a narrow band Guassian process). A distribution using $a = \sqrt{8m_0}$ as the parameter will be referred to as the 'spectral Rayleigh' distribution in this paper.

b) Forristal (1978) distribution

Using Gulf of Mexico hurricane wave records, Forristall (1978) observed that the Raleigh model overpredicts the probabilities of the highest waves. He therefore proposed a two-parameter Weibull distribution which provided a better fit to the observed data:

$$Q(H) = \exp\left(-\frac{1}{\beta}\left(\frac{H}{\sqrt{m_0}}\right)^{\alpha}\right) \tag{4}$$

where $\alpha = 2.126$ and $\beta = 8.42$ were estimated by calibrating the model against the hurricane data.

c) Longuet-Higgins (1980) distribution

Longuet-Higgins (1980) showed that the Rayleigh distribution did in fact fit the Gulf of Mexico hurricane data as well as the two-parameter Weibull distribution if the true H_{rms} is used as the parameter a instead of $\sqrt{8m_0}$ as used by Forristal. He also suggested that the difference between the values of H_{rms} based on water surface elevation record and $\sqrt{8m_0}$ may arise from the 'noise in the spectrum outside dominant peak' and proposed using $H_{rms} = \alpha\sqrt{8m_0}$ to account for the effects of finite spectral bandwidth for those cases where it was necessary to evaluate the parameter a of the distribution from the spectrum (rather than directly as H_{rms}). The scale factor α is given by $\alpha = \sqrt{1 - 0.734\nu^2}$ and the probability of exceedance can be written as

$$Q(H) = \exp\left(-\frac{1}{1 - 0.734\nu^2}\left(\frac{H}{\sqrt{8m_0}}\right)^2\right) \tag{5}$$

In the narrow-banded limit ν = 0, and the distribution is identical to Rayleigh distribution.

d) Naess (1985) distribution

Naess (1985) derived a distribution for the crest to trough wave height in a narrow-band stationary Gaussian wave train using an approach different from that of Longuet-Higgins (1952). He considered both the crest and trough depth to be random variables with a certain correlation ρ (as opposed to assuming $H = 2\eta_{crest}$). However, the resulting expression resembles the Rayleigh distribution with $a = \alpha\sqrt{8m_0}$ as its parameter; but in this case, $\alpha = \sqrt{1/2(1-\rho)}$, where ρ is the correlation between crest heights and trough depths. The autocorrelation of the wave time record may be estimated from the spectrum as $(\tau) = m_0^{-1}\int S(f)\cos(2\pi f\tau)\,df$. While Naess (1985) observed that the precise choice of τ is not crucial, Tayfun and Feddele (2007) suggested taking τ as the abscissa of the first minimum of the normalized autocorrelation function; this treatment is followed in the present study, and the probability of exceedance is given by

$$Q(H) = \exp\left(-\frac{2}{1 - \rho(\tau)}\left(\frac{H}{\sqrt{8m_0}}\right)^2\right) \tag{6}$$

In the narrow-banded limit, $\rho(\tau) = -1$ and the distribution becomes identical to the Rayleigh distribution.

e) Boccotti (1989) distribution

Using the terminology of Casas-Prat and Holthuijsen (2010), the three distributions developed by Boccotti (1989), Vinje (1989), and Tayfun (1990) constitute three "Raleigh-like" models which are intended to better predict the behavior of the larger characteristic wave heights. Casas-Prat and Holthuijsen (2010) found these models to perform equally well. Although Karmpadakis et al. (2020) did not consider the Vinje (1989) and Tayfun (1990) models, they found the Boccotti model to be the most accurate in some categories. Based on these results, we have selected the Boccotti model as representative of this class of models. It, too, incorporates the effects of finite spectral bandwidth, and its functional form is given by:

$$Q(H) = \frac{1 + \rho''(\tau)}{\sqrt{2\rho''(\tau)\big(1 - \rho(\tau)\big)}} \exp\left(-\frac{2}{1 - \rho(\tau)}\left(\frac{H}{\sqrt{8m_0}}\right)^2\right) \tag{7}$$

where $\rho''(\tau) = \frac{\partial^2 \rho}{\partial t^2}$ and τ is the abscissa of the first minimum of the autocorrelation function.

f) van Vledder (1991) distribution

For shallow water, wave height distributions will be affected by depth limitations and van Vledder (1991) proposed a method to use a distribution originally developed by Glukhovskiy (1966) and later modified by Klopman and Stive (1989). The functional form of the resulting exceedance probability is given by:

$$Q(H) = exp\left(-A\left(\frac{H}{H_m}\right)^K\right) \tag{8}$$

$K = \frac{2}{1 - \frac{H_m}{D}}$ and $A = \Gamma\left(\frac{1}{K} + 1\right)^K$

Since the distribution is for wave heights normalized by mean wave height (H_m), the mean wave height must be computed from the available spectral information which is usually H_s. H_{rms} can be easily computed from H_s ($H_{rms} = H_s/\sqrt{2}$) and for the distribution given by (6), the following relation exists between H_{rms} and H_m

$$\frac{H_{rms}}{H_m} = \frac{\sqrt{\Gamma\left(\frac{2}{K} + 1\right)}}{\Gamma\left(\frac{1}{K} + 1\right)} \tag{9}$$

van Vledder (1991) proposed an iterative scheme to compute H_m as shown below

$$K^i = \frac{2}{1 - \frac{H_m^i}{D}} \tag{10}$$

$$H_m^{i+1} = H_{rms} \frac{\Gamma\left(\frac{1}{K^i} + 1\right)}{\sqrt{\Gamma\left(\frac{2}{K^i} + 1\right)}}$$

For the first iteration ($i = 1$), H_m^1 is calculated from H_{rms} according to the Rayleigh distribution and the iteration continues till the two succeeding values of the mean wave height is less than a specified accuracy (van Vledder, 1991).

g) Klopman (1996) distribution

Based on the works of Klopman and Stive (1989) discussed above and using the experimental data of Stive (1985, 1986), Klopman (1996) proposed the following modified Glukhovskiy distribution where the wave heights are normalized by H_{rms}:

$$Q(H) = \exp\left(-A\left(\frac{H}{H_{rms}}\right)^K\right) \tag{11}$$

where $K = \frac{2}{1-\frac{\beta H_{rms}}{d}}$ and $A = \left(\Gamma\left(\frac{2}{K} + 1\right)\right)^{\frac{K}{2}}$

and $\beta = 0.7$ was defined using the experimental data. This avoids the computation of the mean wave height as in the van Vledder (1991) distribution.

It may be noted that some other shallow water distributions are also available. These include the two-part Weibull model of Battjes and Groenendijk (2000), the distribution of Mendez et al. (2004) based on wave energy propagation from deep water to shallow water, and the LoWish II distribution (Wu et al. 2016). These models require additional information (such as the seabed slope) which was not readily available for our study locations and were hence excluded from the present investigations.

3.2 Wave periods

The distribution of wave periods is narrower than that of wave heights in wind seas (Goda, 2000) and intrinsically difficult to determine (Rodríguez et al. 2004). Most of the models proposed for the distribution of wave periods were derived as the marginal distribution of the joint probability distribution of wave heights and periods. Rodríguez et al. (2004) investigated the performace of

two prominent period distribution models, viz. those developed by Cavanie et al. (1976) and Longuet-Higgins (1983), in mixed sea states. Although these two models incorporate spectral width information, Rodríguez et al. (2004) observed that their predictions vary considerably from (simulated) data in many sea states. Further, they showed that the effect of the intermodal distance, not incorporated in these models, is significant. Given the inability of these theoretical models in predicting the wave period distribution accurately, employing empirical relations between characteristic wave periods in practice might be justified. Goda (2000) and the US Army Corps of Engineers' Coastal Engineering Manual (2006) relate the characteristic wave heights T_{max} and $T_{1/3}$ to the mean wave period as follows:

$$T_{max} \approx T_{1/3} \approx C\bar{T} \tag{12}$$

where the coefficient C varies between 1.1 and 1.3 and $\bar{T}$ is the mean wave period. We explore the suitability of (12) in Section 4.2.

4. Results and Discussion

4.1 Performance of the wave height distributions

For each sea state, the characteristic wave heights $H_{1/3}$, $H_{1/10}$, $H_{1/100}$, $H_{1/300}$ are estimated employing the distribution models discussed in Section 3 and the spectral information for that sea state. The characteristic wave heights are also computed directly from the water surface elevation time record for that sea state. The % error in the estimation of the characteristic wave heights for different models for each sea state is then computed as

$$\%\ Error\left(H_{1/n}\right) = \frac{H_{1/n\ observed} - H_{1/n\ predicted}}{H_{1/n\ observed}} \tag{13}$$

where $H_{1/n\ observed}$ is computed from the time series data and adjusted using (1). $H_{1/n\ predicted}$ is the value predicted by the model. In estimating these, numerical integration was employed when analytical expressions were not available for $H_{1/n}$. The Root Mean Square (RMS) value of the percentage errors for each model in the six groups is then used to evaluate the model performance in the estimation of characteristic wave heights. These are given in Table 3[1].

[1] Table 3 was also prepared using the unadjusted values of $H_{1/n}$ (not shown). For the most part, the differences in RMS errors were of the order of 1%. Zheng et al. (2006) provided plots of wave height distributions for data based on different sampling rates. For the right side of the distributions, the differences in the curves are minimal. Since the $H_{1/n}$ values used here could be expected to correspond to the right side, the nominal difference in the RMS errors is consistent with the findings of Zheng et al. (2006).

The Rayleigh model with $a = H_{rms}$ from the time series as the parameter is primarily included for reference, since one often does not have access to the time series data. (This is particularly true if an extreme wave analysis is performed using wave models to estimate, say, the 100-year H_s value). It is evident from Table 3 that this distribution consistently outperforms all other models relying on spectral information while predicting $H_{1/3}$ and $H_{1/10}$; however, this superiority relative to other models is lost while predicting the larger characteristic wave heights $H_{1/100}$ and $H_{1/300}$.

Among the distributions that rely on spectral information, the Rayleigh model with $a = \sqrt{8m_0}$ as the parameter is widely used in practice since a may be estimated from H_s. As may be seen in Table 3, this approach underperforms all the other models investigated, for all categories. In fact, the RMS of the percentage error in estimation of $H_{1/n}$ increases with n and is as large as ~25% in some cases.

We now examine the performance of the other distributions in increasing order of H_s/D. First, we consider sea states with $H_s/D < 0.2$ (Group A) which are further divided into two groups based on K_pD. Most of the sea states in this category are from St. Mary's Sound and Wave Hub which, at D = 53 m and 50 m, are the deepest locations in this study. However, the majority of the sea states in this category represent intermediate water depths based on the K_pD values. For these sea states regardless of the relative depth, the Longuet Higgins (1980), Forristal (1978), and Boccotti (1989) distributions perform almost equally well in estimating $H_{1/3}$ with the RMS errors ranging from 6.5%-7.2%. Naess (1985) and the above-mentioned models perform similarly for $H_{1/10}$, with RMS errors ranging from 7.0%-7.8%. This conforms to the findings of Karmpadakis et al. (2020) who report nominal differences in the performance of these models while describing the bulk of the wave height distribution (i.e. the middle part of the distributions, represented here by $H_{1/3}$ and $H_{1/10}$) in comparable conditions. In estimating the larger wave heights (i.e. $H_{1/100}$ and $H_{1/300}$), the Naess (1985) model performs the best with an RMS error of around 9.3% and 12.1% respectively, whereas other deep-water models (except Forristal (1978), for $K_pD>2$) have error values in the range of 9.8%-10.2% for $H_{1/100}$ and 12.8%-13.6% for $H_{1/300}$. It is also evident from Table 3 that the depth-dependent models consistently underperform the deep-water models, especially in the estimation of higher wave heights, for these sea states (i.e. with low relative wave heights). Overall, it can be seen that for $H_s/D < 0.2$, Forristal (1978) model appears to be performing the best for $H_{1/3}$; for higher characteristic wave heights, the Naess (1985) model errors are slightly lower than the

others. Further, the results in Table 3 also suggest that K_pD does not affect the performance of the models.

We next consider sea states that fall within $0.2 < H_s/D < 0.4$ (Group B), which are divided into three groups based on K_pD values. These sea states are mainly from locations with water depths in the 9.8m -19m range. In the estimation of $H_{1/3}$ and $H_{1/10}$, the relative performance of the deep-water models is similar to that in the case of $H_s/D < 0.2$, especially for the two groups with K_pD >0.5. The RMS errors for the deep-water models are around 5.0%-6.5% for $H_{1/3}$ (excluding Naess (1985)) and 5.4%-6.7% for $H_{1/10}$. However, relative to Group A, the performance of the depth-dependent models is better, and is of the same order as that of the deep-water models; for example, for $H_{1/10}$, the RMS errors for the van Vledder (1991) model are 5.6% and 5.8% for the two K_pD categories; these fall in the aforementioned range for the deep-water models. Further, in the estimation of higher characteristic wave heights, $H_{1/100}$ and $H_{1/300}$, Table 3 shows that for sea states with K_pD >0.5 the depth-dependent van Vledder (1991) model has the lowest RMS error, equal to ~7.7% (based on an average of the two categories) and 9.6% respectively. The deep-water models, show comparable errors for $H_{1/100}$ (7.9-8.7%) and larger errors for $H_{1/300}$ (10.2-12.2%). Thus, the $0.2 < H_s/D < 0.4$ zone appears to be a transition zone where the effect of the water depth seems to impact the wave height distribution. For the smaller characteristic wave heights, the performance of the depth-dependent van Vledder model improved (relative to group A) and was of the same order as the deep-water models; however for the larger wave heights, the improvement is more pronounced, and the model outperforms the deep-water models. This may be attributed to the effects of the water depth influencing the larger characteristic wave heights first (as demonstrated by Karmpadakis et al. 2020).

In the case of Group C (i.e. $H_s/D > 0.4$), most of the sea states are from the shallowest locations with water depths of 7.2 -12.7m range. Since there are only 935 such sea states, they are not further divided based on relative depth; all the sea states in this zone fall in the $0.31 < K_pD < 1$ range. Traditionally $H/D > 0.8$ has been used as the breaking criterion for individual waves, and generally for Rayleigh distributed waves $H_{max} = 2\ H_s$. Hence for these sea states with $H_s/D > 0.4$, depth induced breaking seems probable, especially at higher wave heights. In estimating $H_{1/3}$ and $H_{1/10}$ all models (excluding the first two Raleigh models) perform equally well with RMS errors in the range of 4.9%-6.6%. However, in predicting $H_{1/100}$ and $H_{1/300}$ the van Vledder (1991) and Klopman (1996) distributions perform distinctly better than the deep water-models, with RMS errors of the

order of 7.0 to 8.9%, while the other model errors are much larger (except the Naess model for $H_{1/100}$). In fact, the distinction is more pronounced for $H_{1/300}$, the van Vledder model has an RMS error of 8.1% which is about two-thirds of that of the best performing deep-water model (i.e. 13.0% for the Naess model). It would be reasonable to infer from the results in Table 3 that the van Vledder (1991) model would be a good choice for all the characteristic wave heights for $H_s/D > 0.4$. Karmpadakis et al. (2020) found generally superior performance by the depth-dependent models for comparable sea states at a specific location with water depth = 7.7 m. Since the Group C data correspond to water depths between 7.5 and 12.3 m, our results reinforce their findings.

Overall, within the range of the data considered in this study, it appears that Forristal (1978) model has the lowest error in estimating $H_{1/3}$ and $H_{1/10}$ across all the three groups with few exceptions. The Naess (1985) model may serve as optimal choices for describing the distribution of higher wave heights for sea states with $H_s/D < 0.2$. In this region, the depth-dependent models are mostly less accurate than the others with errors exceeding ~9%. For the sea states that are in $0.2 < H_s/D < 0.4$, a transition appears to occur in the performance of the models (i.e. depth dependent models begin to show improved performance), and the van Vledder (1991) model is generally superior for the larger wave heights. The improvement continues for sea states with $H_s/D > 0.4$ and the van Vledder (1991) model may serve as an optimal choice.

The above results are based on the availability of the spectral information for most of the calculations. If only H_s is available, then the results in Table 3 must be reassessed as it renders the Naess (1985), Boccotti (1989) and Longuet-Higgins (1980) models inapplicable. Hence a comparison of spectral Rayleigh (Longuet Higgins 1952), Forristal (1978), van Vledder (1991) and Klopman (1996) models is necessary. Table 3 shows that the van Vledder model remains the optimal choice for $H_s/D > 0.4$ and for higher wave heights in $0.2 < H_s/D < 0.4$. It may also be seen that the Forristal (1978) model is considerably superior to the other three models for all characteristic wave heights for $H_s/D < 0.2$. It also provides the least errors in estimation of $H_{1/3}$ and $H_{1/10}$ for $0.2 < H_s/D < 0.4$. It is interesting to note that this two-parameter Weibull model, although calibrated on the basis of Gulf of Mexico hurricane data, is as accurate, in several instances, as more complex (spectrum-dependent) models in predicting wave heights around the coast of UK. At other locations, too, Casas-Prat and Holthuijsen (2010) and Kvingedal et al. (2018), observed that this model agrees well with the data. Nevertheless, it must be recognized that maximum likelihood (ML) estimates for the scale (β) and shape (α) parameters, computed using

the measurements, exhibit significant variability. As shown in Fig. 5, the scale parameter ranges from 2.94 to 21.57 (only 36 points were larger than 15) with μ = 7.24 and σ =1.52 whereas the shape parameter ranges from 1.53 to 2.66 with μ = 2.02 and σ = 0.13. Forristal's estimates of 8.42 and 2.126 respectively for the scale and shape parameters are within one standard deviation of the mean of the current ML estimates. Nayak and Panchang (2016) have also observed a wide variation in the ML estimates for the Weibull distribution. The distribution of wave heights in two sea states that provided the minimum (β = 3.87, α= 1.53) and maximum (β= 20.02, α= 2.66) estimates for the parameters are shown in Fig. 6. The figure shows that in both cases, the Weibull distribution itself with the ML estimates fits the data reasonably well, however using Forristal's original parameter values may not always yield the best results.

Having discussed the relative performance of the wave height distribution models, one cannot escape the fact the the RMS percentage errors are usually quite large while estimating the higher characteristic wave heights. In fact, all the models, in all cases except for the depth-dependent models for groups B and C, have RMS errors greater than 10% in estimating $H_{1/300}$. Based on the best fit slopes in Table 3, it can be seen that almost all the distribution models, except for the van Vledder model for $H_s/D > 0.4$, overpredict on average the estimation of $H_{1/300}$. The same could be said about the spectral Rayleigh (1952), Boccotti (1989), and Forristal (1978) models in the estimation of $H_{1/100}$. If an improved estimate is needed, the prediction can be adjusted using the best-fit line slope = $\frac{\text{observed}}{\text{predicted}}$ provided in Table 3. For brevity, an illustration is provided in Fig. 7, which suggest that a factor of 0.958 would improve the Naess (1985) model results for the case shown.

Table 3 RMS of % errors (E) and best fit slope (m) of distributions in different KpD Hs/D regions.

Wave Height	Distribution \ Sea States	(A) $H_s/D \leq 0.2$				(B) $0.2 < H_s/D \leq 0.4$						(C) $H_s/D > 0.4$	
		$K_pD \leq 2$		$K_pD > 2$		$K_pD \leq 0.5$		$0.5 < K_pD \leq 1$		$K_pD > 1$		$K_pD < 1$	
		E	m	E	m	E	m	E	m	E	m	E	m
$H_{1/3}$	Rayleigh	1.85	1.013	1.8	1.013	2.08	1.015	1.89	1.014	1.95	1.016	1.52	1.007
	Spectral Rayleigh	9.45	0.941	9.12	0.953	9.74	0.93	8.28	0.942	7.75	0.953	8.64	0.935
	LH scaled Rayleigh	7.15	1.033	7.08	1.021	9.73	1.085	6.49	1.039	5.71	1.021	5.75	1.026
	Naess	8.19	1.055	8.28	1.052	8.96	1.08	7.75	1.062	7.44	1.054	6.6	1.047
	Forristal	6.52	0.998	6.97	1.011	5.63	0.986	5.03	1	5.51	1.011	4.88	0.993
	Boccotti	7.21	0.979	7.16	0.995	8.22	0.954	6.09	0.977	5.61	0.996	5.91	0.972
	Klopmann	8.84	0.949	8.74	0.959	7.59	0.955	6.43	0.967	6.33	0.974	5.58	0.975
	VanVledder	8.67	0.952	8.63	0.96	6.94	0.965	5.93	0.976	5.97	0.981	4.92	0.992
$H_{1/10}$	Rayleigh	2.93	0.993	2.69	0.995	3.37	0.995	2.82	0.993	2.52	0.996	3.74	0.978
	Spectral Rayleigh	11.35	0.923	10.75	0.935	11.74	0.911	10.38	0.923	9.55	0.934	11.44	0.909
	LH scaled Rayleigh	7.03	1.013	7.25	1.003	8.71	1.064	5.97	1.017	5.75	1.001	5.48	0.997
	Naess	7.58	1.034	7.75	1.033	7.8	1.059	6.7	1.04	6.59	1.034	5.56	1.017
	Forristal	6.96	0.992	7.25	1.006	6.05	0.98	5.44	0.992	5.75	1.005	5.58	0.978
	Boccotti	7.2	0.987	7.38	0.998	6.42	0.979	5.72	0.987	5.87	0.998	6.06	0.972
	Klopmann	9.75	0.942	9.71	0.949	6.83	0.969	6.13	0.978	6.2	0.982	5	0.996
	VanVledder	9.37	0.948	9.44	0.953	5.95	0.99	5.59	0.997	5.82	0.999	5.61	1.031
$H_{1/100}$	Rayleigh	8.34	0.96	7.84	0.964	9.26	0.949	8.27	0.954	7.31	0.963	10.36	0.93
	Spectral Rayleigh	16.29	0.892	15.19	0.907	18.02	0.869	15.81	0.887	13.98	0.903	17.83	0.864
	LH scaled Rayleigh	9.76	0.98	10.17	0.972	8.69	1.015	8.39	0.977	8.66	0.968	9.44	0.949
	Naess	9.32	1	9.33	1.001	8.38	1.01	7.8	0.999	7.77	1	8.48	0.967
	Forristal	9.84	0.975	9.52	0.991	9.99	0.95	8.57	0.969	7.9	0.987	9.62	0.944
	Boccotti	10.03	0.973	9.84	0.982	9.41	0.964	8.63	0.969	8.22	0.979	10.05	0.941
	Klopmann	13.37	0.924	13.26	0.929	9.39	0.962	8.45	0.976	8.11	0.982	6.98	1.004
	VanVledder	12.69	0.933	12.76	0.936	8.05	0.995	7.75	1.007	7.7	1.008	8.48	1.059
$H_{1/300}$	Rayleigh	13.14	0.92	12.06	0.933	14.65	0.901	12.96	0.916	11.14	0.935	16.23	0.881
	Spectral Rayleigh	21.6	0.855	19.49	0.878	24.36	0.826	20.98	0.851	17.93	0.878	24.35	0.819
	LH scaled Rayleigh	13.33	0.939	13.56	0.941	10.72	0.964	11.83	0.938	11.91	0.941	14.52	0.899
	Naess	12.26	0.958	11.96	0.969	11.11	0.959	10.62	0.959	10.17	0.971	12.97	0.917
	Forristal	13.31	0.939	12.18	0.964	14.6	0.907	12.17	0.935	10.42	0.964	14.42	0.9
	Boccotti	13.6	0.937	12.83	0.953	13.56	0.922	12.21	0.935	11.01	0.955	14.94	0.896
	Klopmann	17.88	0.889	17.06	0.902	13.11	0.926	11.41	0.949	10.43	0.965	8.86	0.97
	VanVledder	16.97	0.899	16.42	0.909	10.62	0.961	9.77	0.983	9.42	0.994	8.13	1.03

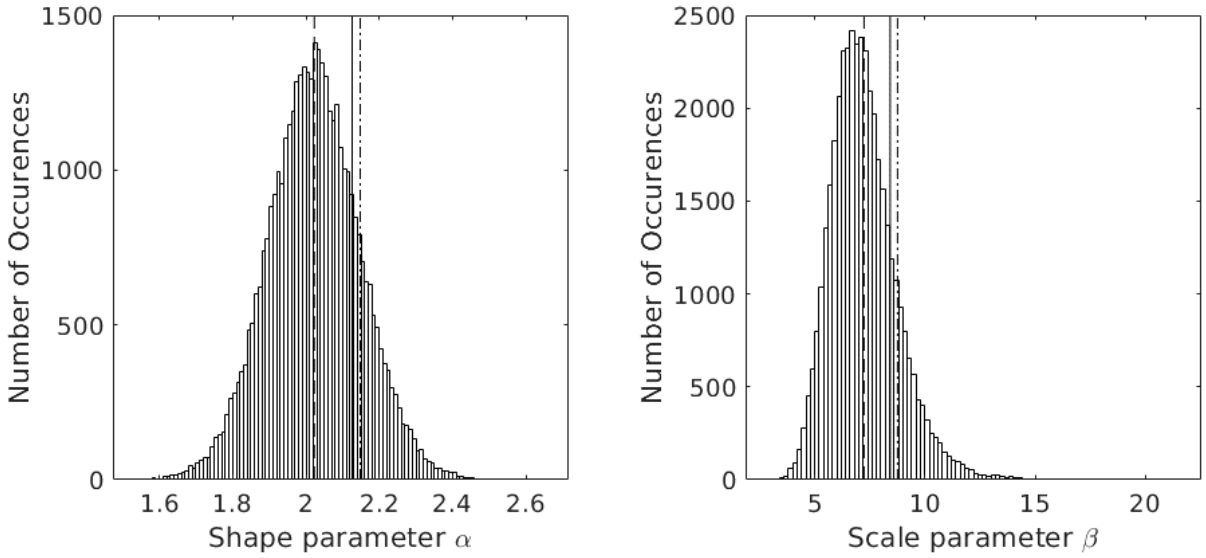

Fig. 5 Frequency distribution of Weibull distribution parameters. Dashed line indicates mean, dash-dot line indicates one std. deviation from mean and solid line indicates Forristal's parameters.

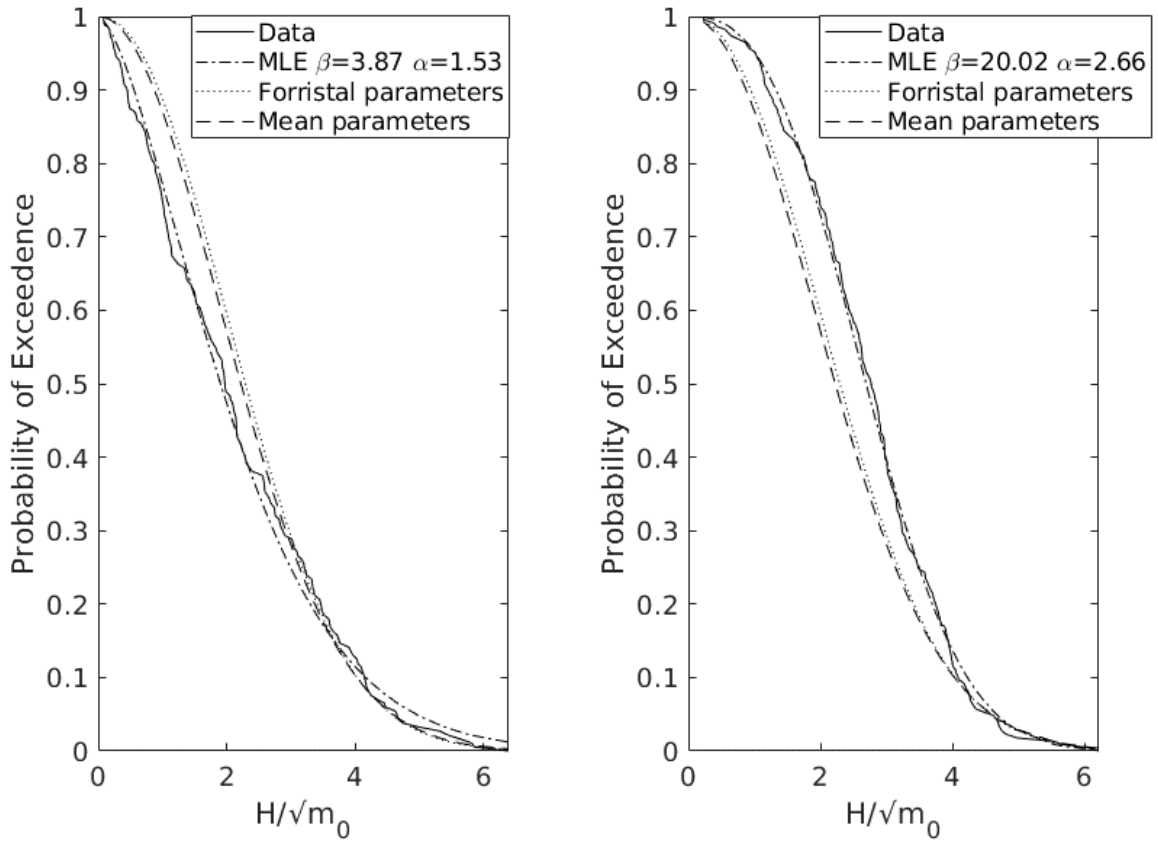

Fig. 6 Normalized wave height distribution for minimum (left) and maximum (right) Weibull parameters (Forristal's parameters β = 8.42, α = 2.126 ; mean parameters , β = 7.24, α = 2.02)

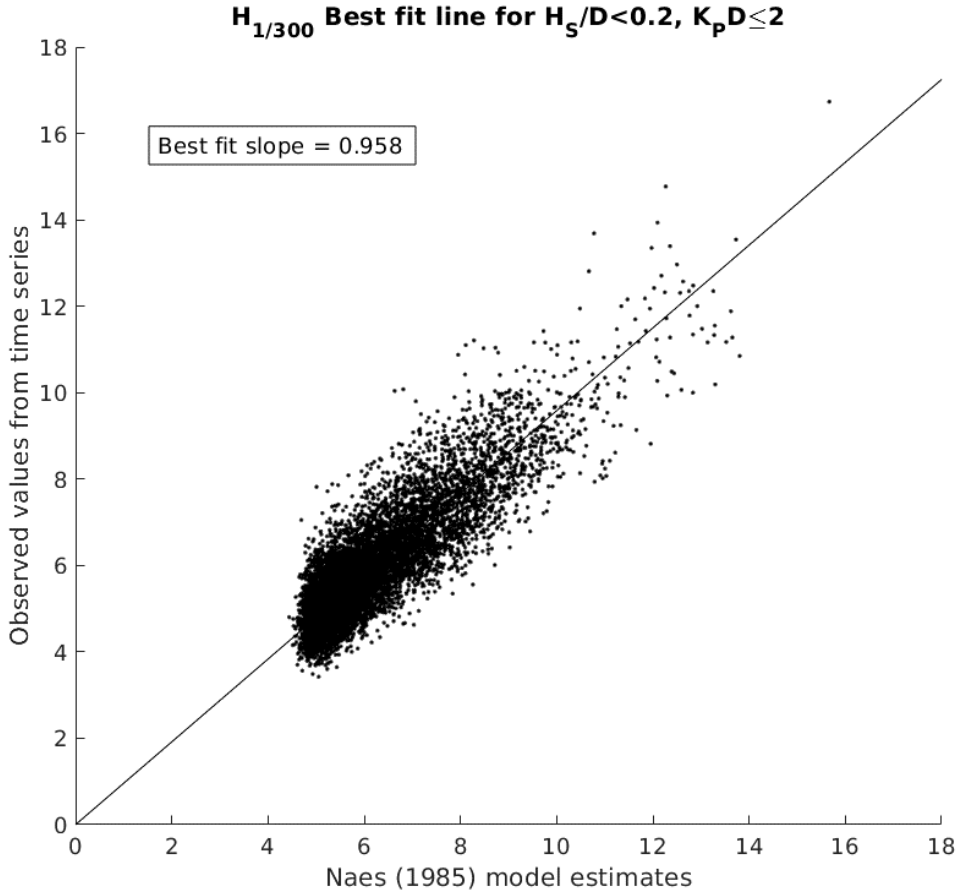

Fig. **7** Best fit line for Naess (1985) model estimates of $H_{1/300}$ (meters) in $H_s/D < 0.2$

4.2 Relationship between characteristic wave periods

We now examine the relationships between characteristic wave periods. For the 41,120 time series considered here, we first examine the relationship of T_{max} with $T_{1/3}$ and T_{mean} based on the time intervals between the zero-upcrossings. Fig. 8 shows that there is considerable variability in these relationships and the best fit lines (with zero intercept) have slopes of 1.38 for T_{max} vs $T_{1/3}$ and 1.97 for T_{max} vs T_{mean} as opposed to (12). Examining the relationship of $T_{1/3}$ and T_{mean} (Fig. 9a) the best fit slope is 1.43 which is slightly outside the range of 1.1-1.3 in (12). However, before proceeding to further investigate the relationship between characteristic wave periods, it should be noted that in practice actual T_{mean} and $T_{1/3}$ are unknowable in the absence of time series information (e.g. when only wave model output is available, as stated earlier). Hence it is necessary to investigate the relationship of the characteristic periods T_{max} and $T_{1/3}$ with mean wave period that can be estimated from the spectrum.

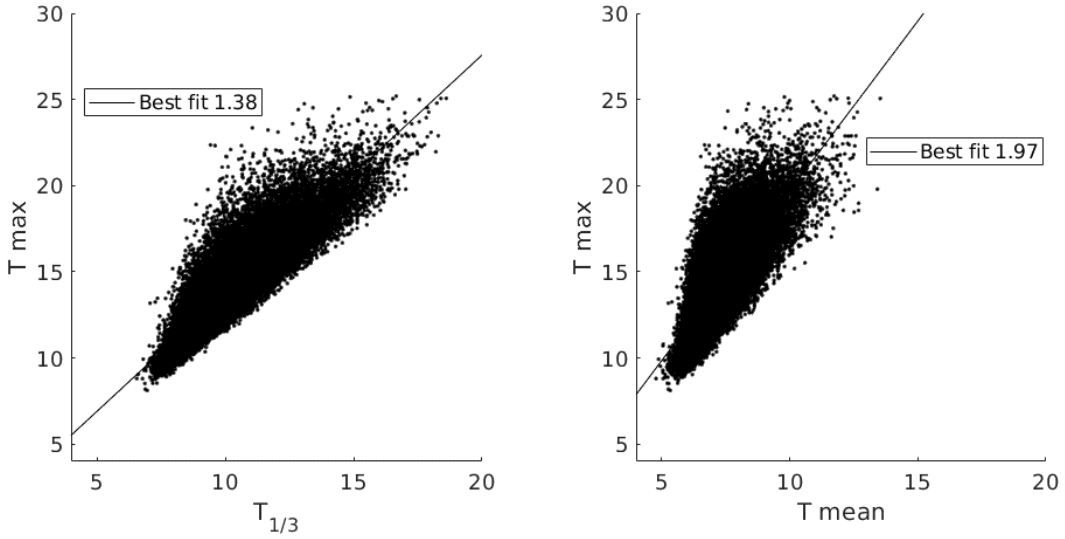

Fig. 8 Relationship of T_{max} to $T_{1/3}$ (left) and T_{mean} (right) in seconds

The mean wave period can be estimated from the spectral moments, either as $T_0 = \sqrt{m_0/m_2}$ or as $T_{m01} = m_0/m_1$ (Holthuijsen, 2007). These quantities, estimated using the 41,120 spectra, are compared with $T_{1/3}$ in Figs. 9 (b, c). It may be seen that the best fit slope while using T_{m01} as the mean is reasonably close to that given by previous recommendations (i.e. eq. 12), whereas the deviation is greater in the case of T_0 as is the scatter, which confirms the expectation of Holthuijsen (2007) that estimates based on the higher sprectral moments could be noisier. In view of this, we will use T_{m01} in this study as the characteristic mean based on the spectrum.

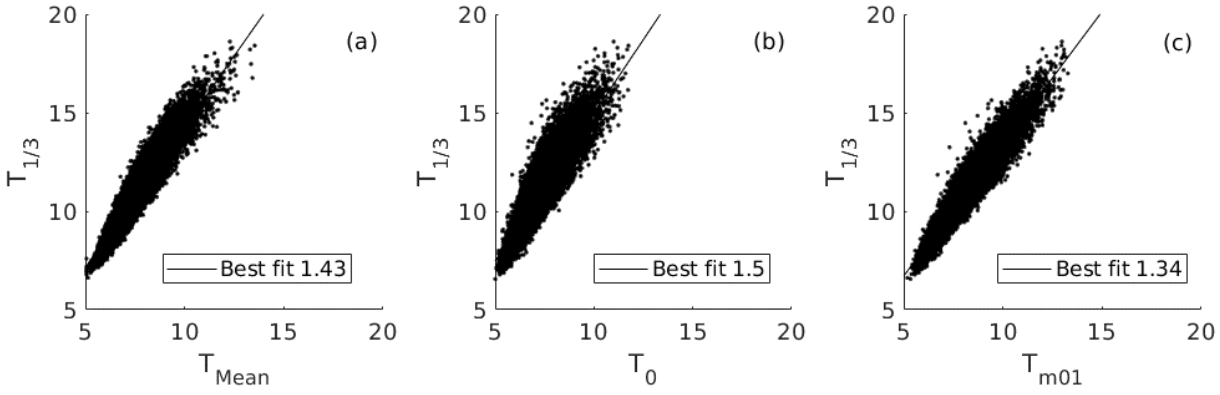

Fig. 9 Relationship of $T_{1/3}$ to estimates of the mean period (a) T_{mean}, *(b)*T_0 and (c) T_{m01} (in seconds).

The relationship between characteristic periods $T_{1/3}$ and T_{max} computed from the time series and T_{m01} estimated from the spectrum are shown in Fig. 10. The best fit line for $T_{1/3}$ vs T_{m01} with zero intercept has a slope of 1.34, which is slightly outside the empirical range of 1.1-1.3 associated with (12). However, the best fit slope of 1.85 for T_{max} is well outside this range. Owing to this, we tried to see if any of the other parameters stated earlier (H_s/D, K_pD, Ursell number, and the spectral width) had a bearing on the comparisons. No significant influence of H_s/D, K_pD, and Ursell number was observed. However, the spectral width (ν) does appear to play a role. As seen in Fig. 10, larger periods are associated with higher spectral widths. Hence we attempted to find an improved empirical relation for $T_{1/3}$ and T_{max} as functions of T_{m01} and ν. Applying linear regression to 80% of the data and then testing it on the remaining 20%, the following relationships were established:

$$T_{1/3} = -1.12 + 1.21\,T_{m01} + 4.42\,\nu \quad (14)$$

$$T_{max} = -2.1 + 1.25\,T_{m01} + 13.94\,\nu \quad (15)$$

In the case of $T_{1/3}$, the linear relation with T_{m01} employing best fit slope gives a good estimate with an R2 score of 0.92 and the inclusion of the spectral width only slightly improves the R2 score to 0.947 and 0.946 for the training and testing, respectively. However, in the case of T_{max}, the best fit line shown in Fig. 10 does not explain the variability in T_{max} well and the R2 score for the best fit line is 0.6758. Including an intercept showed no positive benefit. However, addition of the spectral width as another variable increases the R2 score to 0.771 for the training data set (and 0.772 for the test data set), which shows that incorporating spectral width improves the reliability of the empirical relationships between characteristic wave periods.

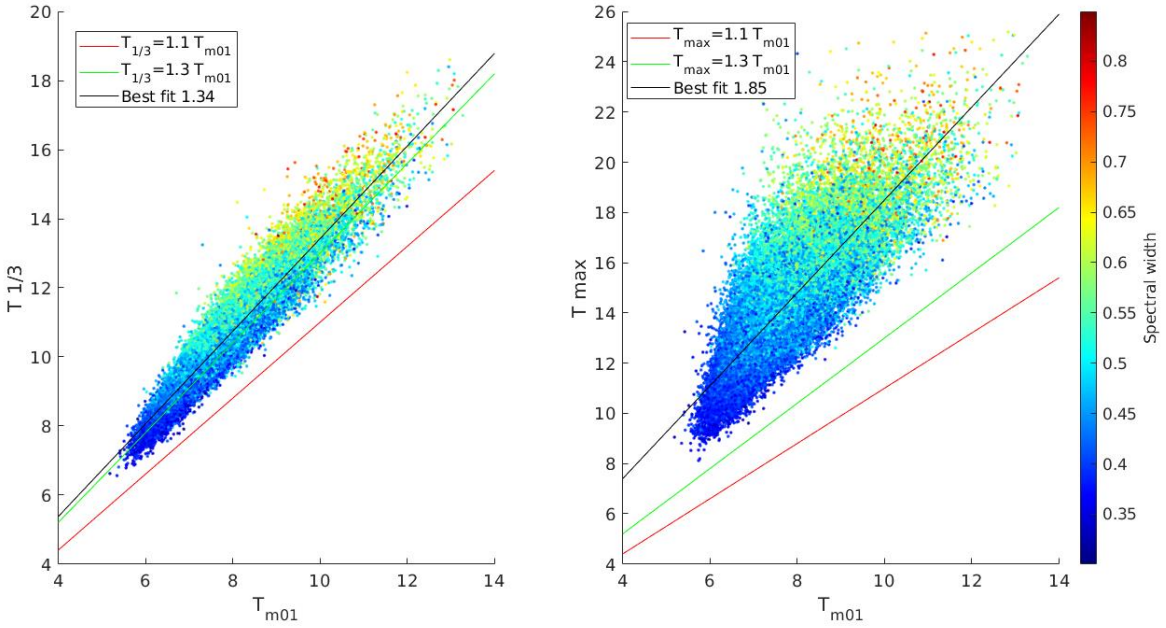

Fig. 10 Relation between $T_{1/3}$ and T_{m01} (left) and $T_{1/3}$ and T_{m01} (right), in seconds; color indicates spectral width

5. Conclusions

We analyzed the performance of seven wave height distributions, using a large set of wave data from intermediate and shallow water depths. A total of 41,120 sea states with Hs ≥ 3 m around the UK coast were used. It is observed that the widely used Rayleigh distribution with $a = \sqrt{8m_0}$ as its parameter has the highest percentage error in estimating characteristic wave heights, $H_{1/3}$, $H_{1/10}$, $H_{1/100}$, $H_{1/300}$. While the Rayleigh distribution with parameter $a = H_{rms}$ computed from the time series (which is usually not possible in practice), has significantly lower errors than other models in estimation of $H_{1/3}$ and $H_{1/10}$, this advantage is lost in the estimation of higher characteristic wave heights $H_{1/100}$ and $H_{1/300}$. For estimating the higher characteristic wave heights, the results in Section 4 indicate that deep-water models in general and the Naess model in particular perform well for $H_s/D < 0.2$, and depth-dependent models in general and the van Vledder model in particular perform well for $H_s/D > 0.4$. For sea states with $0.2 < H_s/D < 0.4$, a transition is visible when progressing from the lower wave heights that seem to be well predicted by deep-water models to higher wave heights by the van Vledder model. These transitions affirm the findings of

Karmpadakis et al. (2020). For estimating $H_{1/3}$ and $H_{1/10}$ (i.e. the lower characteristic wave heights), Table 3 shows that the Forristal (1978) model performs well.

The higher accuracy of the spectrum-dependent Naess (1985) model in some cases bolsters the argument for the modeling agencies to provide the spectrum by way of output, in addition to the usual bulk parameters; this would enable model output users to then use such a distribution to estimate the desired $H_{1/n}$. But in case only H_s is available, the original Weibull model of Forristal (1978) provides better estimates than the often-used Rayleigh distribution. While Casas Pratt et al. (2010) and Kvingedal et al. (2018) have also found it to be reliable, we find that it has low errors (of the order of 7% or less) only for $H_{1/3}$ and $H_{1/10}$. For larger characteristic wave heights, the errors are much larger. In fact, for all the models, all categories (except for the depth-dependent models for groups B and C) have RMS errors much greater than 10% for $H_{1/300}$ estimates. To some extent, improvements may perhaps be attained by resorting to the best-fit slopes provided in Table 3. Further, in case of the Forristal model, analysis of the current data provides grounds for additional caution: there is wide variation in the parameter estimates for the two-parameter Weibull distribution and the mean values obtained here are somewhat lower than Forristal's estimates.

In terms of wave periods, the best-fit slopes frequently deviate from empirical relations (eq. 12). However, estimates of $T_{1/3}$ and T_{max} from the spectum based mean T_{m01} can be enhanced by incorporating spectral width information. This may be accomplished through equations (14) and (15).

Data Availability Statement

Data are available at the Channel Coastal Observatory website (https://coastalmonitoring.org).

References

Amrutha, M. M. and Kumar, V. S. (2015). Short-term statistics of waves measured off Ratnagiri, eastern Arabian Sea. *Applied Ocean Res.*, 53, 218-227.

Battjes, J. A., and Janssen, J. P. F. M. (1979). Energy loss and set-up due to breaking of random waves. In *Proc.Coastal Engineering Conference* (1, 569–587). ASCE, New York, NY.

Battjes, Jurjen A, and Groenendijk, H. W. (2000). Wave height distributions on shallow foreshores. *Coastal Engg*, 40, 3, 161-182.

Boccotti, P. (1989). On mechanics of irregular gravity waves. *Atti Della Accademia Nazionale Dei Lincei, Memorie, 19,* 110–170.

Cartwright, D. E. and Longuet-Higgins, M. S. (1956). The statistical distribution of the maxima of a random function. *Proc. Royal Soc. London. Ser. A. Math. and Phys. Sci., 237*(1209), 212–232.

Casas-Prat, M. and Holthuijsen, L. H. (2010). Short-term statistics of waves observed in deep water. *J. Geophys. Res., 115*(C9).

Chakrabarti, S. K., and Cooley, R. P. (1977). Statistical distribution of periods and heights of ocean waves. *J. Geophys. Res., 82,* 9, 20, 1363–1368.

Dattatri, J., Raman, H., and Shankar, N. J. (1979). Height and period distributions for waves off Mangalor harbour (India) - west coast. *J. Geophys. Res., 84*(C7), 3767–3772.

Dhoop, T. and Mason, T. (2018). Spatial characteristics and duration of extreme wave events around the English coastline. *J. Marine Sci. Eng, 6*(1).

Earle, M. D. (1975). Extreme wave conditions during Hurricane Camille. *J. Geophys. Res, 80*(3), 377–379.

Forristall, G. Z. (1978). On the statistical distribution of wave heights in a storm. *J. Geophys. Res., 83*(C5), 2353.

Goda, Y. (2000). *Random seas and design of maritime structures*. *Advanced Series on Ocean Engineering* (Vol. 33). World Scientific. Singapore.

Glukhovskiy, B. Kh. "Issledovaniye morskogo vetrovogo volneniya (investigation of sea wind waves)." *Gidrometeoizdat, St. Petersburg, Leningrad* (1966).

Holthuijsen, L. H. (2007). Waves in oceanic and coastal waters. Cambridge University Press.

James, I. D. (1986). A note on the theoretical comparison of wave staffs and wave rider buoys in steep gravity waves, *Ocean Eng., 13(2),* 209–214.

Karmpadakis, I., Swan, C., and Christou, M. (2020). Assessment of wave height distributions using an extensive field database. *Coastal Eng, 157*, 103630.

Kvingedal, B., Bruserud, and E. Nygaard (2018). Individual wave height and wave crest distributions based on field measurements from the northern North Sea. *Ocean Dynamics.* 68, 1727–1738.

Klopman, G. (1996). Extreme wave heights in shallow water. *Delft Hydraulics Report H2486, WL*.

Klopman, G., and Stive, M. (1989). Extreme waves and wave loading in shallow water. *Delft Hydraulics, The Netherlands, 660*.

Krogstad, H. E. (1985). Height and period distributions of extreme waves. *Applied Ocean Res.*, *7*(3), 158–165.

Larsen, L. H., Larsen, and H., L. (1981). The Influence of bandwidth on the distribution of heights of sea waves. *J. Geophys. Res*, *86*(C5), 4299–4301.

Longuet-Higgins, M. S. (1980). On the distribution of the heights of sea waves: Some effects of nonlinearity and finite band width. *J. Geophys. Res*, *85*(C3), 1519.

Longuet Higgins, M. S. (1952). On the Statistical Distribution of the Heights of Sea waves. *J. Marine Res.*, *11*(3), 245–266.

Magnusson, A. K., M. A. Donelan, and W. M. Drennan (1999), On estimating extremes in an evolving wave field, *Coastal Eng*., 36, 147–163.

Mason, T. and Dhoop, T. *Quality Assurance and Quality Control of Wave Data*; Channel Coastal Observatory: Southampton, UK, 2017. Available online: http://www.channelcoast.org/ccoresources/dataqualitycontrol/

Mendez, F. J., Losada, I. J., and Medina, R. (2004). Transformation model of wave height distribution on planar beaches. *Coastal Engg*, *50*(3), 97–115.

Mori, N. and Yasuda, T. (2002). A weakly non-gaussian model of wave height distribution for random wave train. *Ocean Engg*, *29*(10), 1219–1231.

Naess, A. (1985). On the distribution of crest to trough wave heights. *Ocean Eng*, *12*(3), 221–234.

Nayak, S. and Panchang, V. (2016). Coastal wave-height statistics during Hurricane Ike. *J. Waterway, Port, Coastal and Ocean Engg*, *142*(3), 1–11.

Nolte, K. G., and Hsu, F. H. (1979). Statistics of Larger Waves in a Sea State. *J Waterway Port Coastal Ocean Div Proc ASCE*, *105*(4), 389–404.

Panchang, V. and Li, D. (2006). Large waves in the Gulf of Mexico caused by Hurricane Ivan. *Bulletin of the American Meteorological Soc*, *87*(4), 481-490.

Panchang, V., Kwon Jeong, C. and Demirbilek, Z. (2013). Analyses of extreme wave heights in the Gulf of Mexico for offshore engineering applications. *J.Offshore Mechanics and Arctic Engg*, *135*(3).

Rice, S. O. (1945). The mathematical analysis of random noise. *Bell Syst. Tech. J.*, 24, 46-156.

Rodríguez, G., Soares, C. G., and Pacheco, M. (2004). Wave period distribution in mixed sea-states. *J. Offshore Mechanics and Arctic Engg*, *126*(1), 105–112.

Santos, V.M., Haigh, I.D. and Wahl, T., 2017. Spatial and temporal clustering analysis of extreme

wave events around the UK coastline. *J. of Marine Sci. Engg*, *5*(3), p.28.

Singhal, G., Panchang, V.G. and Lillibridge, J. L. (2010). Reliability assessment for operational wave forecasting system in Prince William Sound, Alaska. *J. Waterway, Port, Coastal, and Ocean Engg*, *136*(6), pp.337-349

Stive, M. J. F. (1985). A scale comparison of waves breaking on a beach. *Coastal Engg*, *9*(2), 151–158.

Stive, M. J. F. (1986). Extreme shallow water wave conditions. *H533, Delft Hydraulics, The Netherlands*, *660*.

Tayfun, M. A. (1983). Effects of spectrum band width on the distribution of wave heights and periods, *Ocean Engg*., 10(2), 107–118.

Tayfun, M. A. (1990). Distribution of Large Wave Heights. *J. Waterway, Port, Coastal, and Ocean Engg*, *116*(6), 686–707.

Tayfun, M. A., and Fedele, F. (2007). Wave-height distributions and nonlinear effects. *Ocean Engg*, *34*(11–12), 1631–1649.

Coastal Engineering Manual (2008). U.S. Army Corps of Engineers, Washington, D.C.

van Vledder, G.P. (1991) Modification of the Glukhovskiy distribution. *Technical report, Delft Hydraulics Report H1203*.

Vinje, T. (1989). The statistical distribution of wave heights in a random seaway. *Applied Ocean Res.*, *11*(3), 143–152.

Wu, Y., Randell, D., Christou, M., Ewans, K., and Jonathan, P. (2016). On the distribution of wave height in shallow water. *Coastal Engineering*, *111*, 39–49.

Zheng, G., Cong, P. and Pei, Y. (2006). On the improvements to the wave statistics of narrowbanded waves when applied to broadbanded waves. *J. Geophysical Res. Oceans*, *111*(C11).