Assessment of wave height distributions using an extensive field database

I. Karmpadakis*, C. Swan, M. Christou

Department of Civil & Environmental Engineering, Imperial College London, SW7 2AZ, UK

Abstract

The present paper investigates the short-term statistical distribution of wave heights. Specifically, some of the most commonly applied wave height distributions are assessed using field measurements. The latter comprise of wave radar observations from 10 different locations in the North Sea and cover water depths between 7.7 m and 45 m. In total, the field database includes more than 200 million waves, making it one of the largest of its kind in this water depth regime. In using these data, the accuracy of existing wave height distributions has been examined and guidance is provided concerning the best performing models and their domain of applicability. Additionally, insights concerning the influence of key met-ocean parameters are also provided. Taken together, the results in this paper present an overview of the statistical behaviour of wave heights in finite water depths, as observed in the field.

Keywords: wave height, statistics, field data, shallow water, engineering

1. Introduction

A typical design procedure for many offshore and coastal structures involves three main steps. First, characteristic sea-state parameters such as the significant wave height, H_s, and spectral peak period, T_p, are defined using a long-term analysis. This is usually based upon numerical hindcast data, often supplemented/calibrated by *in situ* measurements. Second, using the results of the long-term analysis, sea-states with a specific return-period are defined. The characteristics of these "design" sea-states are then used as input to short-term statistical models; the aim being to derive the distribution of wave heights or crest heights. Targeting wave heights or crest heights with a given probability of exceedance, Q, the third step seeks to define representative "design" waves. These are commonly based upon either regular wave theories or a focused wave approach (Lindgren, 1970; Boccotti, 1983; Tromans et al., 1991); the latter used to define the most probable shape of a large wave in an irregular sea. With the wave form defined, the water particle kinematics and the corresponding wave loading can be defined.

Considering this design methodology, it is clear that the statistical distribution of wave heights represents a key input to the process. This is further demonstrated by the direct use of wave height distributions in a wide range of calculations, particularly in a coastal engineering context. Examples include the design of breakwaters and other coastal structures, as well as the calculation of wave run-up and over-topping; all being generally parametrised on the basis of wave height statistics (De Waal and Van der Meer, 1993; Goda, 2010). While the primary focus in many of these applications lies in extreme sea-states, or the largest wave heights, other applications address operational conditions. Examples include determining the power output from wave energy converters and other renewable energy devices, as well as fatigue calculations.

Given its importance, the distribution of wave heights has received considerable attention. As a result, a significant number of statistical models have been proposed in the literature. One of the main objectives of the present study is to investigate the accuracy of these models (presented in Section 3) under a wide range

*Corresponding author
Email address: i.karmpadakis@imperial.ac.uk (I. Karmpadakis)

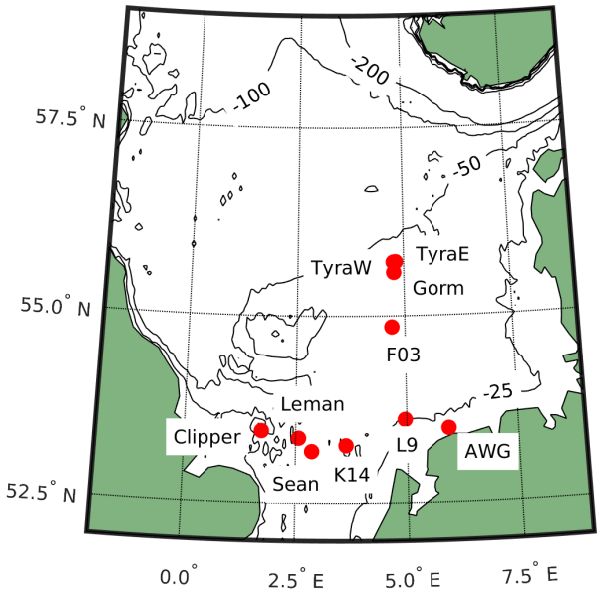


Figure 1: Geographic locations of the offshore platforms in the North Sea, where the field measurements have been recorded. The characteristics of the measurements at each location are provided in Table 1.

of sea-state conditions. To achieve this, an extensive database of field measurements in finite water depths has been employed to compare theoretical predictions with the measured distributions. Taken together, more than 200 million waves arising in intermediate and shallow water depths have been examined. To the authors' knowledge this collection of data constitutes the largest of its kind in the literature. As such, the assessment presented herein can provide insightful guidance for a very wide range of conditions and potential engineering and oceanographic applications. However, it is important to note that similar field data studies can be found elsewhere in the literature; most commonly relating to deep-water conditions. Some examples include Forristall (1978), Longuet-Higgins (1980), Tayfun (1990), Boccotti (2000), Mori and Yasuda (2002), Stansell (2004, 2005) and Casas-Prat and Holthuijsen (2010); the latter providing a detailed review of earlier studies. Indeed, Casas-Prat and Holthuijsen (2010) present one of the largest published datasets (10 million waves) and examine wave height distributions based primarily upon wave buoy measurements. Finally, in analysing a field database of comparable size to the present study, Christou and Ewans (2014) considered 120 million waves (measured by wave radars) to investigate the characteristics of rogue waves, mostly arising in deep water.

The contents of the present study are arranged as follows. An overview of the field measurements and their characteristics is provided in Section 2. Section 3 presents the statistical models utilised in this study. The methodology employed for the assessment of these models is described in Section 4 along with the results arising from its application to the present field database. Physical insights concerning the met-ocean characteristics that drive the statistical distribution of wave heights are presented in Section 5. Finally, concluding remarks are provided in Section 6.

2. Field Data

2.1. Data sources

The field data considered in this study have been obtained from wave radars mounted on the side of fixed offshore platforms. These were provided by Shell International Petroleum Company Ltd. and Mærsk Olie og Gas A/S as part of the LoWiSh Joint Industry Project (Phase III). In total, measurements from 10

Table 1: Sources of field data: water depths (MSL), measuring periods, sampling rate and number of 20-minute sea-states.

Platform name	Water depth [m]	Measuring period	Sampling Rate [Hz]	Number of sea-states
AWG	7.7	12/2007 - 01/2008	2	110,896
		10/2011 - 02/2012	2	
		03/2012 - 02/2017	4	
Clipper	23.3	01/2005 - 04/2008	2	198,323
		02/2011 - 02/2017	2	
L09	25	08/2010 - 11/2011	2	107,488
		12/2011 - 02/2017	4	
K14	27.4	05/2007 - 04/2008	2	147,743
		10/2011 - 02/2012	2	
		03/2012 - 02/2017	4	
Sean	32.3	09/2000 - 05/2001	2	132,087
		11/2007 - 04/2008	2	
		06/2012 - 02/2017	4	
Leman	35.8	02/2006 - 04/2008	2	192,010
		02/2011 - 12/2011	2	
		12/2011 - 02/2017	4	
Gorm	40	2001 - 2013 (not-continuous)	5.12	6,081
F03	42.4	09/2007 - 04/2008	2	14,151
Tyra East	45	2004 - 2013 (not-continuous)	5.12	7,663
Tyra West	45	2001 - 2011 (not-continuous)	5.12	7,667

different platforms in the central and southern North Sea were analysed; Figure 1 showing their geographic locations. The water depths at the measuring locations ranged between 7.7 m and 45 m, covering a large part of the intermediate and shallow water depth regime. The bathymetry surrounding the measuring locations was approximately flat with bed slopes smaller than 1:1,000. The only exception occurs at the shallowest platform (AWG) where the bed slope was 1:500.

For the majority of the platforms considered (AWG, Clipper, L09, K14, Leman and Sean) the available data are continuous and cover the period between 2011-2017. In addition, shorter segments of measurements from earlier years were also available. Recordings from the F03 platform were only available between September 2007 and April 2008; the wave radar being decommissioned. The data from the Tyra (East & West) and Gorm platforms were segmental in the sense that they cover several months every few years; their total duration being approximately 10 years. A summary of the available data for each platform is provided in Table 1.

Along with the sampling periods, Table 1 defines the nominal water depth at each location (measured with respect to the MSL) and the sampling rate of the instrument employed. The sampling rate for the platforms in the Tyra field (Tyra East, Tyra West and Gorm) was 5.12 Hz, making it the highest available. For the rest of the platforms, early measurements were made at 2 Hz, while the majority of the data were sampled at 4 Hz; the transition between the two sampling rates taking place in 2012. The Clipper platform is an exception; the sampling rate of its radar being set at 2 Hz for the whole measuring period.

The instrument used to acquire the surface elevation measurements at all locations was the WaveRadar REX (formerly known as SAAB WaveRadar REX). This wave radar uses a frequency-modulated continuous-wave method to estimate the distance above the water surface. A linearly increasing (9.7 GHz to 10.3 GHz) frequency sweep is followed by a linearly decreasing sweep, resulting in a triangular frequency modulation signal. Since the transmitted signal changes frequency, when it meets with the reflected signal from the sea

surface it creates a beat frequency proportional to the distance between the receiver of the radar and the water surface. This process is repeated continuously during the 10.3 Hz measurement cycle of the instrument and the measurements are averaged to produce the recorded distance. According to the manufacturer's specifications (RS-Aqua, 2014), the error associated with the wave radar is less than ±6 mm for instruments installed at distances less than 50 m above the MSL (as is the case for the present dataset). The radar beam produces a 10° wide cone. This ensures that the area over which the measurements are recorded (the so-called radar footprint) has a diameter of approximately 3.5 m (for the platforms in this study). Due to the design of the electronics modules in the sensor, the WaveRadar Rex is calibrated by the manufacturer at a special facility and does not need any re-calibration during its operation. Moreover, the sensors were placed as far as possible from the platform legs to measure the undisturbed wave field (Forristall, 2005). More information about the operation and accuracy of the specific wave radar can be found in Ewans et al. (2014).

In considering these measurements, some important aspects of the present database are outlined. The first point concerns the adoption of fixed sensors in preferance to wave buoys; the latter being widely utilised in earlier studies. In this respect, Allender et al. (1989) has shown that a buoy can be dragged through or move around the three-dimensional peaks of the largest waves. As such, they do not measure the largest crest elevation. Indeed, James (1986) and Magnusson et al. (1999), amongst others, have demonstrated that due to the hydrodynamic characteristics of buoys the crests of steep waves tend to be linearised. As a result, the surface elevation records from buoys contain more sinusoidal waves than those recorded from a fixed instrument. In considering extreme wave events, Dysthe et al. (2008) note that measurements from fixed instruments give the most reliable information. On this basis, only data recorded from fixed offshore instruments have been included in the present investigation.

The second aspect relates to the sampling rate of each instrument. It is clear that a high rate is required in order to gain insights into the non-linear characteristics of the highest waves. Indeed, the results of Tayfun (1993) and Stansell et al. (2002) show that differences in the estimation of crest heights and wave heights arise even in a linear sea-state, when using a low sampling rate. Specifically, a lower sampling rate will result in an under-estimation of the crest height; the extent of the error depending on the ratio of the sampling time interval over the mean zero-crossing period (T_z). This effect is demonstrated on Figure 2, in which a large wave, generated experimentally, has been down-sampled. The surface elevation was originally recorded at a rate of 12.8 Hz (adjusting for the adopted laboratory scale), and the time-history re-sampled at 4 Hz, 2 Hz and 1 Hz. The maximum deviations in estimating the crest height of this very non-linear wave were then found to be 3.2%, 4.5% and 16%, respectively; the corresponding deviations for wave heights being 2.2%, 3.3% and 11.7%. In Figure 2, the resulting non-dimensional wave profiles are given; the different colours representing different sampling rates. Based upon these effects, it is clear that the measurements used in the present study have notable advantages over other published datasets particularly those using wave buoys; the latter having sampling rates as low as 1.28 Hz.

2.2. Quality Control

While the instruments employed in the present study provide some of the highest quality measurements available from field studies, the harsh operating conditions can adversely affect the collection of data. Indeed, the risk of erroneous measurements, due to a variety of reasons, unavoidably accompanies all field related studies. To this end, considerable attention has been paid into putting together procedures to quality control the recorded data and create datasets that are, as far as possible, free from false measurements.

At the same time, the quality control (QC) procedures for datasets containing millions of waves, such as the current one, have to be optimally designed to be applied in an automated fashion. In effect, a robust and thorough QC algorithm has to be applied in order to discard all wave records that contain errors. The methodology employed herein is a modification of the "no tolerance" approach presented by Christou and Ewans (2011) and Christou and Ewans (2014). The adopted approach is one of the strictest proposed in the literature and has been shown to be particularly effective in the processing of wave radar measurements. In summary, a number of flags were defined to indicate whether a sample should be accepted or discarded; each one corresponding to a potential source of error. The flags used herein are defined as follows:

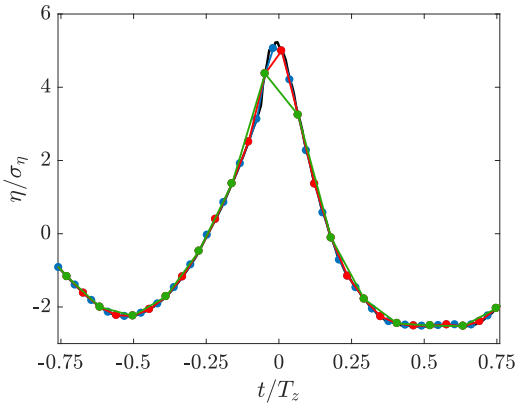


Figure 2: The effect of sampling frequency on the profile of an experimentally generated wave. A laboratory generated wave record, originally sampled at 12.8 Hz [——] has been re-sampled at: 4 Hz [—●—], 2 Hz [—●—] and 1 Hz [—●—]. The x-axis has been normalised by the mean zero-crossing period, T_z, and the y-axis by the standard deviation, σ_η, of the sea-state.

- Flag 1: The occurrences of 10 consecutive points of equal value.
- Flag 2: Zero up-crossing periods longer than 25 s.
- Flag 3: Crest elevations greater than 5 times the standard deviation.
- Flag 4: Number of occurrences of Flag 3.
- Flag 5: Energy in the spectrum below 0.04 Hz greater than 5% of the total.
- Flag 6: Energy in the spectrum above 0.6 Hz greater than 5% of the total.

Individual 20-minute sea-states were discarded when a flag was raised. The only exception to this rule was flag 3, where a maximum of 4 occurrences were allowed (hence the number count given in flag 4). This change was incorporated in order to make sure that the largest waves would not automatically be discarded by the QC procedure. However, the small number of sea-states that had more than 1 but less than 5 occurrences of unexpectedly large crest elevations (flag 3) were visually checked to verify that they did not contain false measurements.

It should be noted that a different steepness flag to the one presented by Christou and Ewans (2014) was used to identify unrealistic spikes in the surface elevation records. This refers to the application of a modified Miche type criterion (Miche, 1944) to identify the steepest waves, followed by visual inspection of the records. Whilst this approach resulted in a much larger number of individual assessments, it was found to be the most robust way to generate a final database of the highest possible quality. The criterion applied, hereafter called flag 7, was defined as

$$H > m\frac{2\pi}{k}\tanh(kd), \tag{1}$$

where $m = 0.12$ and the wavenumber k was calculated using the linear dispersion relation and the corresponding down-crossing wave period (T). In considering this additional flag, it should be clarified that the selection of the value of m only reflects a data driven filter for the steepest waves and not a suggestion for a new wave breaking criterion. In general, the population of suspect samples, identified by this flag, contained a large proportion of small waves; their very short periods leading to the flag being raised. In practice, the erroneous events of most interest were typically characterised by unrealistically large surface elevations (already identified by flags 3 and 4). However, this is an attribute of the present dataset and it does not imply that other datasets will not contain false spikes with reasonable surface elevations.

Time-histories of the water surface elevation containing examples of "failed" or flagged events, identified by the QC algorithm, are shown in Figure 3 (a)-(c). Sub-plot (a) concerns the consecutive occurrence

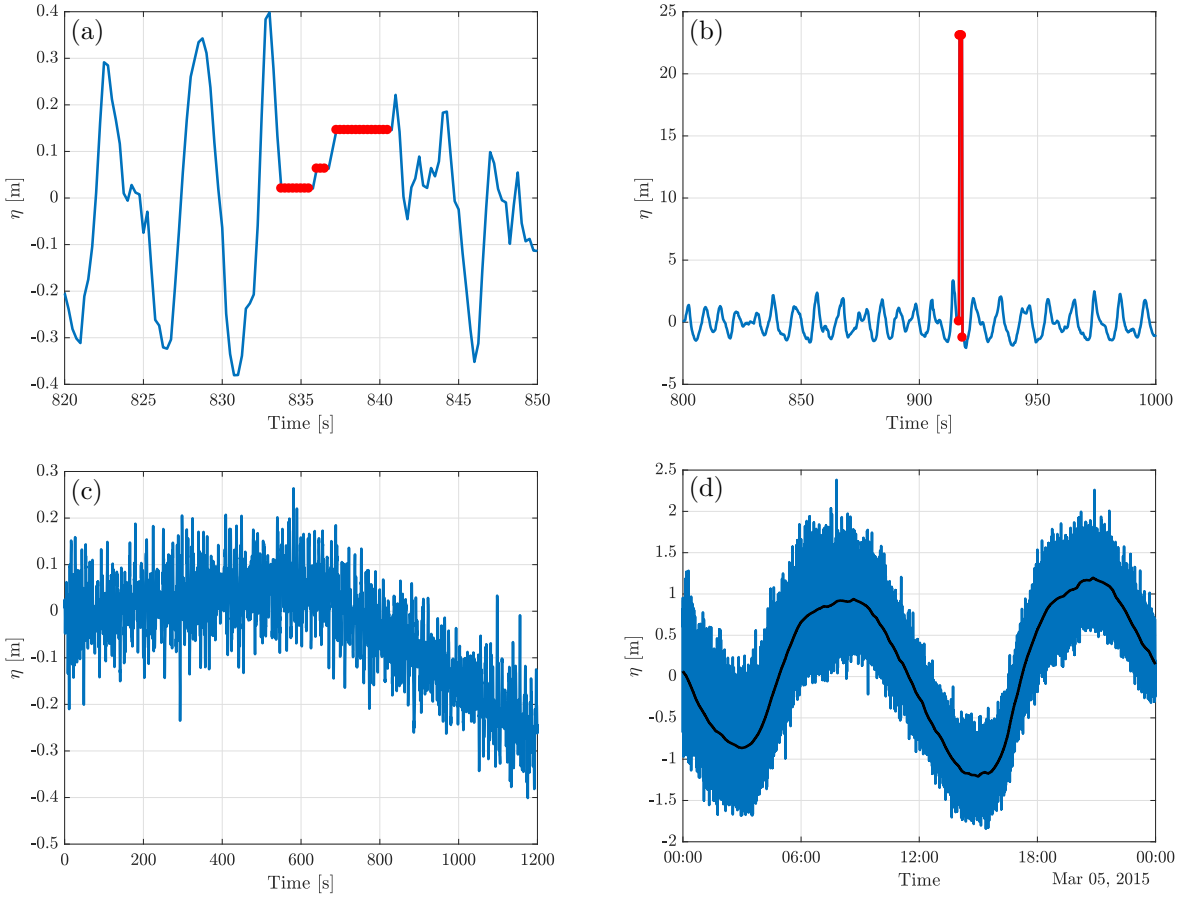


Figure 3: Examples of surface elevation time-histories containing erroneous measurements flagged by the QC algorithm. Sub-plot (a) shows the occurrence of multiple equal values (instrument lock-in), while sub-plot (b) shows a false spike in the record. In sub-plot (c) the sensor exhibits a drift, characterised by unexpectedly long up-crossing periods and sub-plot (d) shows a daily record of surface elevation [——] with the corresponding tidal fluctuations [▬].

of points with equal values (flag 1), also referred to as instrument lock-in; while sub-plot (b) relates to unexpectedly large crest elevations and steepness (flags 3, 4 and 7). Sub-plot (c) shows a recorded drift of the sensor identified by flag 2.

Regarding the processing of the raw surface elevation measurements, one crucial task was performed prior to the application of the QC algorithm. This involved the removal of any tidal fluctuations. These were found to be significant at most locations considered herein. An example of the tidal fluctuations recorded at the AWG platform is provided in Figure 3(d). In this case, it can be seen that the tidal amplitude is approximately 1.5 m and therefore comparable to the H_s of the sea-state. After exploring several methods, the final approach involved fitting a piecewise linear function to consecutive 10-minute segments of the recorded time-histories. The resulting fit was then removed from the signal. In this way, both the tidal elevation and the storm surge, if present, were extracted. In order to verify the accuracy of this method, the results were compared to estimates arising from a harmonic analysis using the *"t_tide"* software package (Pawlowicz et al., 2002). Good agreement was found between these two methods at each location. Having removed the tidal component, the zero-mean surface elevation measurements were divided into 20-minute long sea-states and zero up-crossing and down-crossing analyses performed. The differences between these two zero-crossing methods were found to be insignificant for all locations apart from the shallowest water depth ($d = 7.7$ m). In this case, the largest down-crossing wave heights were found to be larger than the

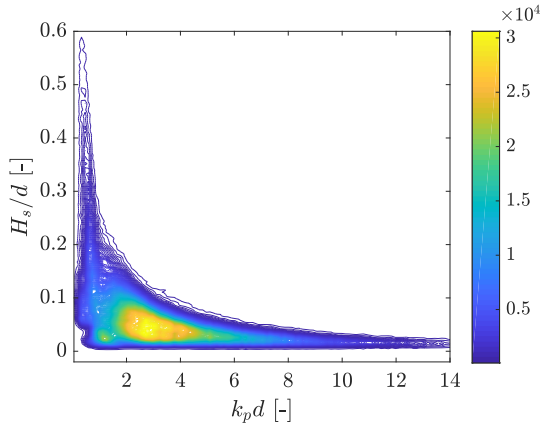


Figure 4: Bivariate density of sea-state severity (H_s/d) and effective water depth (k_pd) for all the quality controlled field observations; the units on the colour-map corresponding to the number of 20-minute sea-state records.

up-crossing wave heights. This was also noted by Guedes Soares et al. (2004). With this in mind, all the data that follow are based upon a down-crossing analysis.

Having completed the quality control procedure, the dataset consisted of approximately 900,000 twenty-minute sea-states; approximately 30% of the original data having been discarded. The number of sea-states that passed the QC at each platform is given in Table 1.

2.3. Available sea-states

The main benefit of long measuring campaigns lies in the variety of wave conditions encountered. This provides greater insights into the behaviour of the sea under widely different circumstances. Of course, for the majority of engineering applications, the most severe sea-states are of primary interest, since this is when the survivability of a marine structure or vessel is most at risk. In order to visually assess the range of environmental conditions, it is common to calculate the bivariate density of the relevant sea-state parameters. Taking all of the available (QC checked) data, this is illustrated in Figure 4. In this case, both the significant wave height, H_s, and the peak wavenumber, k_p, corresponding to the spectral peak period, T_p, have been non-dimensionalised using the water depth, d. As such, the x-axis (k_pd) indicates the effective water depth and the y-axis (H_s/d) the severity of the sea-state. It is worth noting that throughout this paper a spectral definition of the significant wave height has been adopted ($H_s = 4\sigma_\eta = 4\sqrt{m_0}$, where σ_η is the standard deviation of $\eta(t)$ and $\sqrt{m_0}$ the zero$^{\text{th}}$ spectral moment). To obtain this and other spectral estimates, standard spectral analysis techniques were employed for each sea-state; see, for example, Tucker and Pitt (2001).

It can be seen that a large proportion (approximately 40%) of the sea-states have H_s/d values lower than 0.1 and are effectively in deep water ($k_pd > \pi$). This is not surprising, as the weather remains relatively moderate for the majority of the measurement period. However, the data include some severe storms giving rise to very steep shallow water sea-states. These are depicted in the top left side of Figure 4. A value of $H_s/d = 0.5$ means that there will be waves whose elevation above MSL will be one third of the water column. One drawback of the present dataset is that all the measurements for H_s/d values larger than 0.3 come from only one platform (AWG). Nonetheless, there are 3,148 20-minute sea-states above this limit, which is sufficient to provide some variability in the local environmental conditions.

3. Wave height models

Nine wave height models that are widely applied in engineering practice have been selected for comparisons with the measured data. For the purposes of this study, these have been grouped into two broad categories: deep-water [(a)-(d)] and shallow-water models [(e)-(j)]. The criterion for this classification has

been the incorporation of depth-induced wave breaking within the model; the latter being a necessary attribute for the shallow-water category. The probability of exceedance, $Q(H)$, for these wave height models is given below.

(a) Rayleigh distribution (Longuet-Higgins, 1952)

The Rayleigh distribution for wave heights is given by:

$$Q(H) = \exp\left(-\frac{H^2}{8\sigma_\eta^2}\right), \tag{2}$$

where H denotes the wave heights and σ_η the standard deviation of the free surface elevation. Using the classification defined above, the Rayleigh model falls into the deep-water model category. However, it is widely applied in all water depths.

(b) Forristall distribution (Forristall, 1978)

The probability of exceedance defined by the Forristall model is given by:

$$Q(H) = \exp\left[-\frac{1}{\beta}\left(\frac{H}{\sigma_\eta}\right)^\alpha\right], \tag{3}$$

where $\alpha = 2.126$ and $\beta = 8.42$. The Forristall distribution has been widely adopted in both deep and intermediate water depths. This model is a Weibull distribution calibrated using the field data collected by Hall (1972) and Ward (1974) in the Gulf of Mexico. While it is included in the deep-water models, wave breaking is incorporated in this model through the empirical fit of its parameters to sea-states that include breaking waves.

(c) Naess distribution (Naess, 1985)

This deep-water model provides a modification of the Rayleigh distribution to incorporate finite spectral bandwidth effects. This is achieved by using the autocorrelation function of the water surface elevation, $R(\tau)$, to derive the root-mean-square wave height, H_{rms}, of the sea-state. It is worth noting that the original Rayleigh formulation was also normalised by H_{rms}; the latter being defined as $H_{\mathrm{rms}} = \sqrt{8}\sigma_\eta$ for a deep-water, narrow-banded, linear sea-state. The normalised autocorrelation function, $r(\tau)$, is obtained as

$$r(\tau) = \frac{R(\tau)}{\sigma_\eta^2} = \frac{\int_0^\infty S_{\eta\eta}(\omega)\cos(\omega\tau)d\omega}{\sigma_\eta^2}, \tag{4}$$

where $S_{\eta\eta}(\omega)$ is the energy spectrum and $\omega = 2\pi/T$ is the circular wave frequency. The probability of exceedance for the Naess model can be written as:

$$Q(H) = \exp\left[-\frac{1}{4(1-r(T/2))}\left(\frac{H}{\sigma_\eta}\right)^2\right]. \tag{5}$$

Naess (1985) showed that the model is relatively insensitive to the selection of T, as long as it is bounded by T_m, the mean time interval between maxima, and T_z the mean zero-crossing period. However, a specific value for T was not provided. To address this in the present work, the suggestions of Tayfun and Fedele (2007) have been followed. In this respect, $T/2$ is replaced by T^*, the abscissa of the first minimum of the normalised autocorrelation function. It is worth noting that in the narrow-banded limit ($r(T/2) = r(T^*) = -1$), the Rayleigh and Naess models produce identical results.

Moreover, to apply the model in the prediction of a wave height distribution, spectral information is required. In other words, a spectral shape has to be assumed *a priori*, such that the autocorrelation function can be calculated using Equation (4). Within the present study, this was achieved by considering the JONSWAP spectra (Hasselmann et al., 1973) based upon the given H_s and T_p values. Furthermore, the peak enhancement factor, γ, was determined using the guidelines provided by DNV-RP-C205 (2010). However, any other spectral formulation could be used in a similar way.

(d) Boccotti distribution (Boccotti, 1989)

This is the last of the commonly applied deep-water models and its development follows from the asymptotic behaviour of Gaussian white noise (theory of quasi-determinism). Building upon the work of Rice (1944), Lindgren (1970) and Boccotti (1983), this model provides the distribution of large wave heights and incorporates the effects of finite spectral bandwidth. The functional form of the model is given by

$$Q(H) = \frac{1 + \ddot{r}(T^*)}{\sqrt{2\ddot{r}(T^*)(1 - r(T^*))}} \exp\left[-\frac{1}{4(1 - r(T^*))}\left(\frac{H}{\sigma_\eta}\right)^2\right], \tag{6}$$

where T^* is the abscissa of the first minimum of the autocorrelation function of the surface elevation, $r(\tau)$, and the double over dot denotes the second time derivative: $\ddot{r}(t) = \partial^2 r/\partial t^2$.

Since this model was devised to describe the behaviour of the largest waves (asymptotic model), it is not expected to provide a good estimate for the smallest waves of a sea-state. In fact, the integral of its probability density function (PDF) does not equal unity; the smallest wave heights having probabilities of exceedance larger than 1. Since this model also requires spectral information, the methodology to obtain $r(\tau)$ outlined with respect to the Naess model has again been followed.

(e) Klopman distribution (Klopman, 1996)

This shallow-water model is a modification of the Glukhovskiy (1966) distribution. The original Glukhovskiy distribution accounted for wave breaking using the ratio of the mean wave height over the water depth, H_m/d. Klopman and Stive (1989) corrected an inconsistency in the original model and provided the basis for the Klopman (1996) distribution, given by:

$$Q(H) = \exp\left[-A\left(\frac{H}{H_{\mathrm{rms}}}\right)^K\right], \tag{7}$$

where

$$A = \left[\Gamma\left(\frac{2}{K} + 1\right)\right]^{K/2} \tag{8}$$

and

$$K = \frac{2}{1 - \beta H_{\mathrm{rms}}/d}. \tag{9}$$

In this case, the root-mean-square wave height is obtained as $H_{\mathrm{rms}} = H_s/\sqrt{2}$ based on a narrow-banded Gaussian process. Moreover, Γ is the gamma function defined by $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$ and the factor $\beta = 0.7$ was defined using the experimental data of Stive (1985, 1986).

(f) van Vledder distribution (van Vledder, 1991)

This model is another modification of the original Glukhovskiy distribution. In this case, an iterative method is applied to calculate H_m from H_{rms}. The functional form of the distribution is given by:

$$Q(H) = \exp\left[-A\left(\frac{H}{H_m}\right)^K\right], \tag{10}$$

and the iterative scheme by:

$$K^i = \frac{2}{1 - H_m^i/d}, \tag{11}$$

$$H_m^{i+1} = H_{\mathrm{rms}} \frac{\Gamma\left(\frac{1}{K^i} + 1\right)}{\sqrt{\Gamma\left(\frac{2}{K^i} + 1\right)}}, \tag{12}$$

where i is the counter. For the first iteration, ($i = 1$) van Vledder (1991) uses H_m^1 obtained from the Rayleigh distribution, while the scale parameter A is given in Equation (8).

(g) Battjes-Groenendijk distribution (Battjes and Groenendijk, 2000)

Battjes and Groenendijk (2000) noted that when plotting experimentally recorded wave heights on Rayleigh probability paper, lower and higher waves demonstrate different behaviour. Based on this observation, they used two Weibull models to approximate the two parts of the distribution. In using experimental measurements the Weibull scale and shape parameters were empirically defined and a transition wave height, H_{tr}, between the two parts was calculated. By imposing a continuity constraint in this transition wave height, the functional form of the model was expressed as

$$Q(H) = \begin{cases} \exp\left[-\left(\frac{H}{H_1}\right)^{K_1}\right] & \text{for } H \leq H_{tr} \\ \exp\left[-\left(\frac{H}{H_2}\right)^{K_2}\right] & \text{for } H > H_{tr}. \end{cases} \tag{13}$$

The empirical coefficients are defined by

$$K_1 = 2, \qquad K_2 = 3.6, \tag{14}$$

$$H_{tr} = (0.35 + 5.8 \tan\alpha)d \tag{15}$$

and

$$H_{\text{rms}} = \left(2.69 + 3.24\frac{\sigma_\eta}{d}\right)\sigma_\eta, \tag{16}$$

where $\tan\alpha$ is the (average) bed slope in the direction of dominant wave propagation (shore-normal). Since the model was calibrated using unidirectional laboratory measurements, bathymetric variations in other directions (cross-shore) are not incorporated. The normalising wave heights H_1 and H_2 can be obtained from Table 2 of the Battjes and Groenendijk (2000) paper. Considering the exponents K_1 and K_2, it should be noted that they form a Rayleigh distribution for smaller wave heights and an empirical fit for larger wave heights. A shortcoming of this distribution is that its probability density function becomes discontinuous around the value of the transition wave height, H_{tr}. While this is not physically realistic, it does not affect the integral statistical properties of the model.

(h) Mendez distribution (Mendez et al., 2004)

Adopting a very different approach, this model was derived by estimating the evolution of wave heights as they propagate towards the shoreline. Assuming an initial Rayleigh distribution in deep water, Mendez et al. (2004) apply a wave energy propagation model to define a new wave height distribution in shallow water. In this respect, the effects of wave shoaling (Green's law) and wave breaking (Battjes and Janssen, 1978) are incorporated. The functional form of the model is given by

$$Q(H) = \exp\left[-\phi^2(\kappa)\left(\frac{\xi}{1-\kappa\xi}\right)^2\right], \qquad \text{for} \quad 0 \leq \xi < 1/\kappa, \tag{17}$$

where $\xi = H/H_{\text{rms}}$. The scale parameter $\phi(\kappa)$ was defined by Mendez et al. (2004) as

$$\phi(\kappa) = (1 - \kappa^{0.944})^{1.1877}, \tag{18}$$

while the empirical coefficient κ was obtained by analysing a series of experimental measurements as

$$\kappa = \left(\frac{H_{\text{rms}}}{d}\right)^{2.5}(4.7 - 20.8Ir + 26.2Ir^2), \tag{19}$$

where Ir is the Iribarren number defined by

$$Ir = \frac{\tan a}{\sqrt{H_{\text{rms}}/L_1}}; \tag{20}$$

L_1 being the deep water wavelength corresponding to the spectral mean period, T_1, such that $L_1 = gT_1^2/2\pi$.

(i) LoWiSh II distribution (Wu et al., 2016)

The last shallow-water model was developed during Phase II of the LoWiSh JIP. Following a methodology similar to Battjes and Groenendijk (2000), this is also a two-part model. In this case, a Weibull distribution is used to describe the smaller wave heights and a Generalised Pareto for the larger wave heights. The model was calibrated using a small subset of the field data utilised in the present study, together with experimental data arising from uni-directional laboratory measurements (Katsardi et al., 2013).

In deriving this model, a transition wave height, H_{tr}, between the two parts was selected. Imposing continuity of the probability density functions at H_{tr} and fitting each part of the distribution independently, led to the following functional form:

$$Q(H) = \begin{cases} \exp\left[-\mu\left(\frac{H}{H_s}\right)^k\right] & \text{for } H \leq H_{tr} \\ \exp\left(-\mu\right)\left\{1+\left[1+\frac{\xi}{\sigma}(H-H_{tr})\right]^{\left(-\frac{1}{\xi}\right)}\right\} & \text{for } H > H_{tr}. \end{cases} \tag{21}$$

Considering this parametrization, μ and k are respectively the scale and shape parameters of the Weibull part and ξ is the shape parameter of the Generalised Pareto part; all being defined in Wu et al. (2016). It is worth mentioning that regarding the Generalised Pareto distribution, a negative shape parameter, ξ, implies that there is an upper limit in the distribution. With respect to the LoWiSh II model, the Miche criterion (Miche, 1944) was applied to define the maximum wave height. Unfortunately, the application of this criterion for random waves in intermediate and shallow water depths is not easily justified (Massel, 1996).

4. Assessment of models

Given the variety of available models, the question that immediately arises is which one provides the best description of the observed wave heights, in what parameter space. Considering previous studies in deep-water, Tayfun and Fedele (2007) have found that wave heights measured at a deep-water platform in the North Sea agreed well with the Boccotti distribution. In fact, the deep-water model of Tayfun (1981, 1990) was also included in this study and shown to be in very good agreement with the Boccotti model. A more recent analysis of buoy data, Casas-Prat and Holthuijsen (2010) confirmed the success of the Boccotti distribution. Additionally, a generalized form of the Boccotti distribution has also been found to provide accurate results in deep water (Alkhalidi and Tayfun, 2013).

Regarding intermediate and shallow water depths, Hameed and Baba (1985), Martin et al. (1988) and Moritz (2002) have reported good agreement between measurements, the Glukhovskiy and Forristall models. Additionally, Mai et al. (2011) compared the Battjes and Groenendijk distribution to radar measurements from the German section of the North Sea. While the original formulation of the model produced relatively poor results, re-calibrating the empirical coefficients using the available data showed better agreement. Cosidering representative wave heights, such as H_{rms} and the maximum wave height H_{max}, Rattanapitikon (2010) has shown small deviations between a wide range of models.

Taking full advantage of the large dataset assembled in the present study, the question of the best performing wave height distribution can be revisited. To address this, the available field data have been compared to the wave height models described in Section 3. The methodology underpinning these comparisons and the results of this process are presented below.

4.1. Methodology

To provide useful comparisons between measurements and model predictions, sea-states with similar characteristics need to be considered. To achieve this, the available data were partitioned using a *data binning* approach. More specifically, the measured sea-states from each offshore platform were grouped into data bins with similar H_s and T_p values; each data bin having a width of 0.5 m in H_s and 1 s in T_p. This bin size was found to be large enough to include a statistically significant number of observations,

but not so large that it introduced significant variability between the sea-state characteristics. The centres of the bins were defined on a fixed grid based on the range of values observed in each dimension of the parameter space. To avoid including results with large statistical uncertainty, data bins that contained less than 1500 waves were excluded from the analysis. In practice, such exclusions are limited to the periphery of the (H_s, T_p) domain. Having completed the data binning, the down-crossing wave heights arising in the sea-states within each data bin were collected and re-ordered, and their respective probability of occurrence determined. Additionally, the corresponding met-ocean parameters (such as H_s, T_p, T_1) of the sea-states included in each bin were averaged and used as inputs for the available models. The predicted wave heights arising from each model were calculated by evaluating their inverse cumulative distribution function.

Based on these distributions, a series of 6 percentiles [50%, 20%, 10%, 2%, 1%, 0.1%] were defined to investigate both the main body and the tail of the distributions. In essence, the divergence between the measured and predicted values was based on the wave heights larger than the six percentiles noted above. To quantify the divergence between predictions and measurements, the following definition of the relative RMS error (ε) was used following Battjes and Groenendijk (2000):

$$\varepsilon = \sqrt{\frac{1}{N}\sum_{i=p}^{N}\left(\frac{H^i_{\text{measured}}}{H^i_{\text{model}}} - 1\right)^2}, \tag{22}$$

where N is the number of zero-crossing wave heights, H_{measured} and H_{model} are the measured and predicted wave heights. In each calculation the summation starts from the p^{th} observation corresponding to the percentile under investigation, and extends towards the smallest probability of exceedance.

Once the error associated with each model was calculated, the data points in each corresponding bin were coloured to reflect the distribution that best matches the data. In this way, maps can be provided to indicate the best fitting distribution given H_s and T_p. In order to combine results recorded at platforms with different water depths into a single figure, the non-dimensional variates H_s/d and k_pd have been used. The main objective in adopting this approach is to provide guidance as to which model should be used under a variety of conditions. Moreover, considering one distribution at a time, the regions where a given model performs best can be clearly demonstrated. Using this approach, the applicability of each model can be defined.

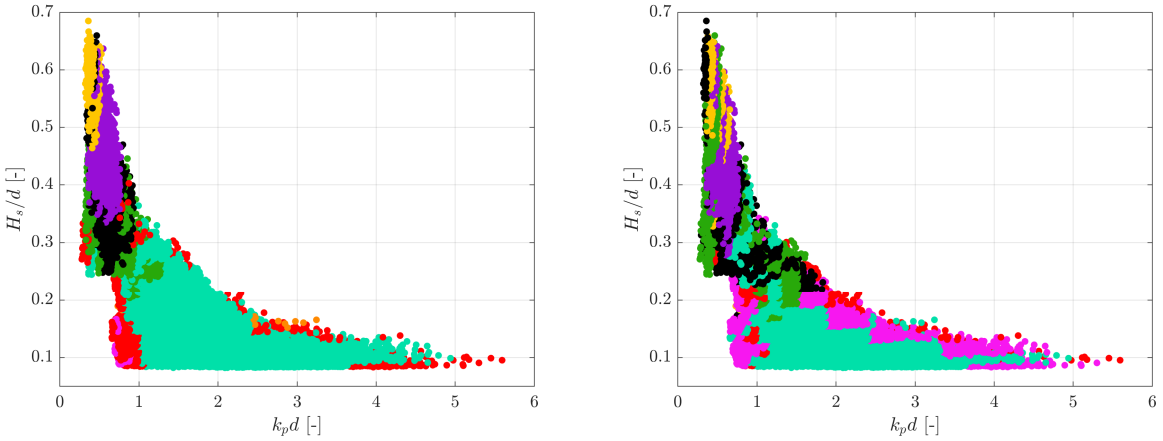
4.2. Results

The maps presented in Figures 5a and 5b were obtained following the methodology outlined above. These concern the amalgamated set of all available field measurements expressed in the non-dimensional parameter space (H_s/d, k_pd). Each sea-state is denoted with a single dot, coloured to represent the best fit model; the models under consideration being listed in the captions. More specifically, Figure 5a concerns an analysis of the data lying above P_{20} percentile; this is taken to be representative of the bulk of the distribution. Similarly, Figure 5b concerns the analysis of data lying above the P_1 percentile; the latter percentile taken to be representative of the behaviour in the tail of the distributions.

Considering these results, it can be observed that no single model provides the best short-term distribution of wave heights across the full range of environmental conditions. In general, the best-fit models appropriate to the bulk of the distribution are consistent across the P_{50}, P_{20} and P_{10} percentiles. This means that the wave heights corresponding to neighbouring data bins are, to a higher degree, best described by the same model. In contrast, there is more scatter in the data relating to the tail of the distributions. This is not surprising as the variability in the wave heights contained in the tail of a distribution is much larger. Indeed, this variability can become progressively more significant when wave breaking occurs, especially in deep water.

Considering the data presented on Figures 5a and 5b, it can be seen that two patterns emerge. First, sea-states with low H_s/d values and $k_pd > 1.2$ are best described by the deep-water models. In contrast, sea-states with larger H_s/d values and $k_pd < 1.2$ are best described by the shallow-water models; the latter incorporating depth-induced wave breaking.

Regarding the first category, the Boccotti model appears to be the best option for the majority of the cases when considering the bulk of the distribution (P_{20} on Figure 5a). This result is in agreement with the



(a) P_{20} (20%) percentile.

(b) P_1 (1%) percentile.

Figure 5: Map indicating the best fitting distribution for all platforms. The distributions are: Rayleigh [●], Forristall [●], Klopman [●], van Vledder [●], Battjes-Groenedijk [●], Mendez [●], LoWiSh II [●], Boccotti [●], Naess [●].

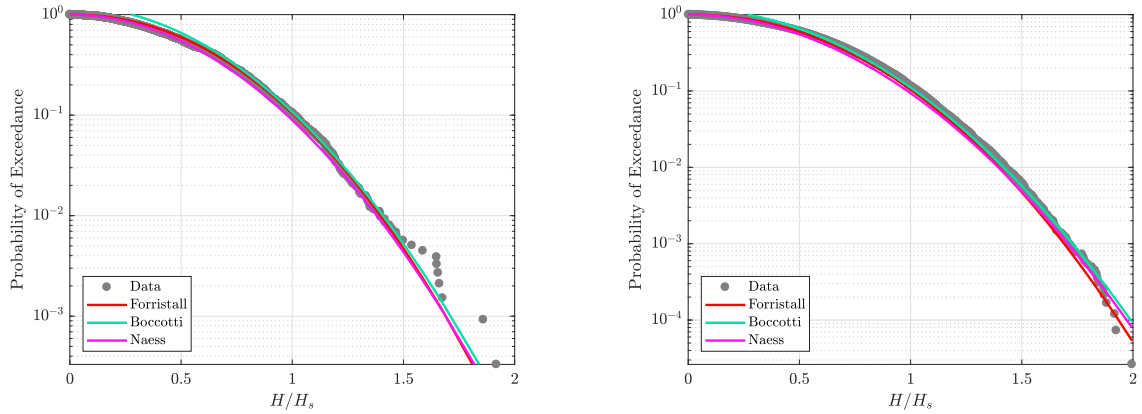
findings of Tayfun and Fedele (2007) concerning field data and Latheef (2013) concerning experimental data. Regarding the tail of the distribution the best fits are again provided by the Boccotti model, but some bins are best described by the Naess model (P_1 on Figure 5b). There are also occurrences where the Forristall model provides the best fit, on both Figures 5a and 5b. Some of these concern relatively benign sea-states, but others relate to steep deep-water sea-states; the latter being closest to the original storm data used to formulate this model.

In considering these results further, it should be noted that the differences in the accuracy of these deep-water models are typically rather small. To highlight this, Figures 6a and 6b serve as an indication of the relative performance of the deep-water models. In these and subsequent figures, the probability of exceedance (Q) is plotted against the normalised wave heights (H/H_s) in semi-logarithmic axes. More specifically, these two figures present comparisons between the predictions of the three deep-water models and the measured data for representative data bins arising from the K14 and L9 platforms. It is clear that the three models (Forristall, Naess and Boccotti) are close to each other; the Boccotti model being closest to the measured data. Moreover, it can be seen that the predicted probability of exceedance for the Boccotti model reaches values larger than 1, indicating that the integral of its PDF will also be larger than 1. Considering that this is an asymptotic model, derived to describe the largest waves, this is not surprising.

It is also important to note that an unavoidable complication emerges when data from all offshore platforms are combined in a single figure (Figures 5a and 5b). This relates to sea-states measured at different locations falling in the same (H_s/d, k_pd) data bin. In such cases, the data points will inevitably overlap and could potentially generate misleading results. Having addressed this point, it was typically found that the same model emerges as the best fit in the majority of overlapping data bins. Indeed, differences only tended to arise when all candidate distributions were in close agreement with the measured data; the examples in Figures 6a and 6b being representative of these cases.

Most importantly, the aforementioned complication can be avoided when data from a single platform are considered at any one time. To this end, Figures 7a to 7f present the best fit maps relating to data from the K14, L9 and AWG platforms, respectively. The selection of these platforms was driven by the fact that their datasets cover a wide range of met-ocean conditions. In each of these figures, the left sub-plot presents results for the P_{20} percentile, while right sub-plot for the P_1 percentile.

The earlier comments concerning the applicability of the deep-water models are easily verified on the basis of Figures 7a to 7d. Indeed, considering the P_{20} percentile for platforms K14 and L9 (Figures 7a and 7c), the Boccotti model provides the best description for the vast majority of sea-states. In a smaller number of cases the Forristall model provides the best description of the data recorded at the K14 platform,



(a) $H_s = 4.75\,\mathrm{m}$ and $T_p = 15\,\mathrm{s}$. (b) $H_s = 5\,\mathrm{m}$ and $T_p = 10\,\mathrm{s}$.

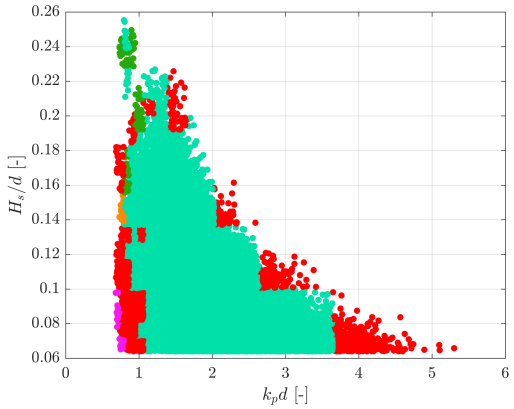
Figure 6: Normalised wave height distributions comparing representative data from the (a) K14 platform ($d = 27.4\,\mathrm{m}$) and (b) L9 platform ($d = 25\,\mathrm{m}$) and the deep-water models.

typically for small H_s/d values. Considering the P_1 percentile for the same platforms, Figures 7b and 7d confirm that the Boccotti model remains the most accurate for the majority of the cases. However, in the extreme tail of the distributions more sea-states are better described by the Naess and Forristall models. More importantly, the larger scatter in the best fitting models in the tail of the distributions is clearly apparent when comparing the results for the P_{20} and P_1 percentiles. As mentioned above, this is primarily caused by the large variability in the tail of the distribution and the close agreement in the predictions of the deep-water models.

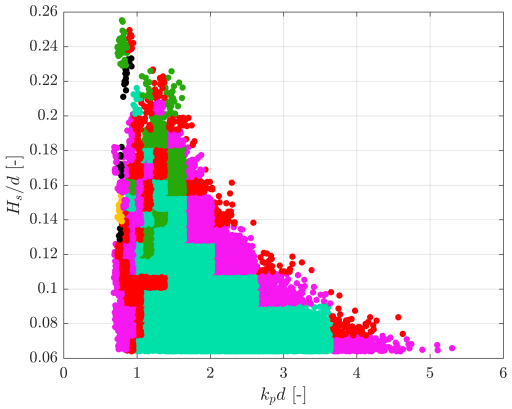
In considering the P_1 percentile, the scatter in the results presented on Figure 7d demonstrates the emergence of a second pattern. This indicates that the tail of the distributions in the most energetic sea-states from the L9 platform (represented by high H_s/d values) are best described by shallow-water models (Klopman and van Vledder). In fact, Bouws (1980) have reported similar findings when comparing storm measurements at a water depth of $d = 30\,\mathrm{m}$ with the Glukhovskiy distribution. This observation marks the transition between the two domains in which the deep-water and shallow-water models are applicable. This arises because depth-induced breaking becomes more important with the increase in H_s/d. As a result, the shallow-water models are more accurate. In this sense, it is not surprising that the transition appears first in the P_1 percentile, as wave breaking first acts to reduce the largest wave heights.

Figures 7e-7f present the results from the AWG platform, the shallowest location in the present dataset of field observations. These results can be used to investigate the behaviour of the shallow-water models in greater detail. In general, the best fits are provided by the Glukhovskiy (Klopman, van Vledder), Mendez and Battjes-Groenendijk distributions. More specifically, the best fit models for the bulk of the wave height distributions, defined by P_{20}, are shown in Figure 7e. In this case, it is clear that for the most moderate sea-states (represented by low H_s/d values), the Glukhovskiy distributions are closer to the measured data. As such, the Klopman and van Vledder models provide the best fit for the majority of the sea-states with $H_s/d < 0.35$. In considering the more energetic range of sea-states lying in the range $0.35 < H_s/d < 0.45$, it is clear that the Mendez model is the most accurate. Finally, the Battjes-Groenendijk model provides the best fit for the most severe sea-states defined by $0.45 < H_s/d < 0.65$.

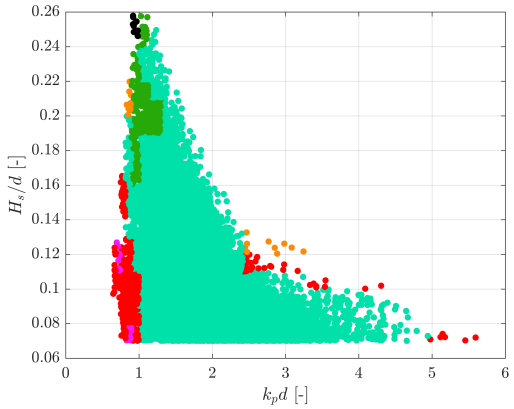
Taken together, these results indicate that as the sea-state severity (represented by H_s/d) increases there is a clear progression in the best performing models. Interestingly, comparisons between Figures 7e and 7f reveal that the pattern of best fit models does not change significantly when considering either the bulk of the data (P_{20}) or the tail of the distributions (P_1). This is different to the results from platforms in deeper water. The explanation for this lies, in part, in the nature of the wave breaking: in deep water this is primarily driven by increased wave steepness, while in shallow water it is depth-induced. It therefore follows that in deeper water the dissipation mechanism is much less uniform than in the shallower water.



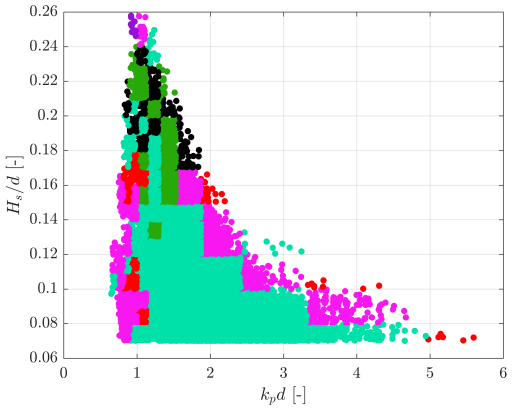
(a) Platform: K14 - P_{20} (20%) percentile



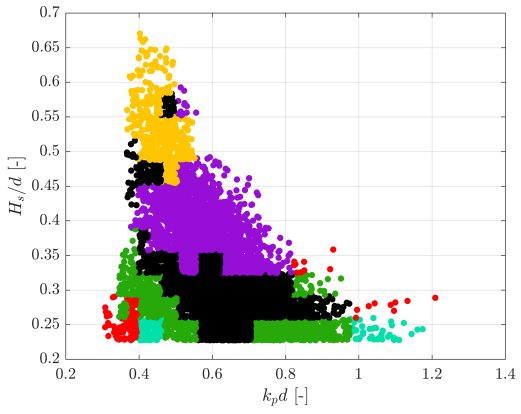
(b) Platform: K14 - P_1 (1%) percentile



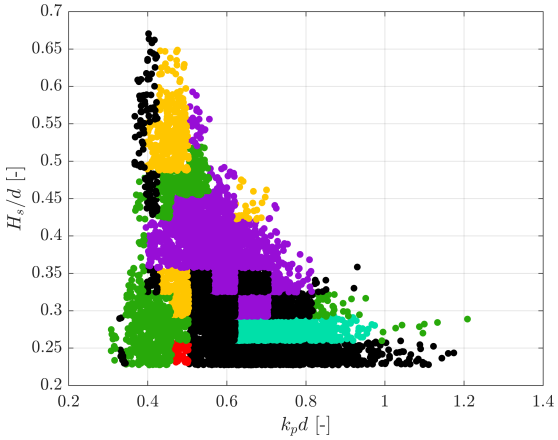
(c) Platform: L9 - P_{20} (20%) percentile



(d) Platform: L9 - P_1 (1%) percentile



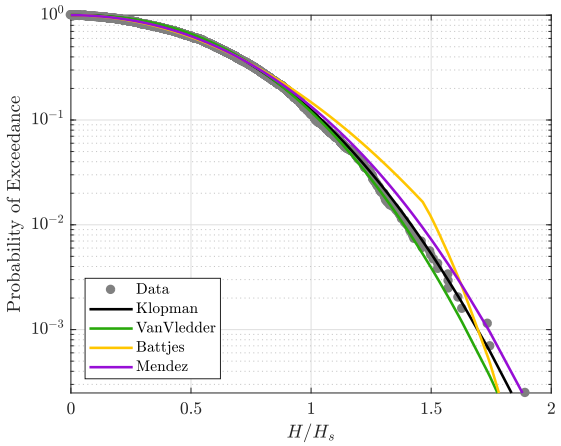
(e) Platform: AWG - P_{20} (20%) percentile



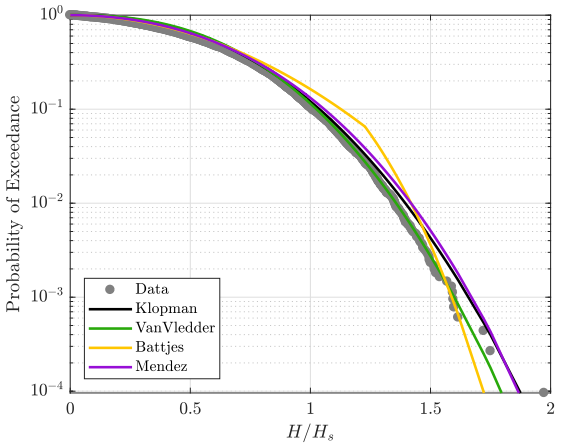
(f) Platform: AWG - P_1 (1%) percentile

Figure 7: Maps indicating the best fitting distribution for the P_{20} and the P_1 percentiles for data recorded at different platforms. The distributions are: Rayleigh [●], Forristall [●], Klopman [●], van Vledder [●], Battjes-Groenedijk [●], Mendez [●], LoWiSh II [●], Boccotti [●], Naess [●].

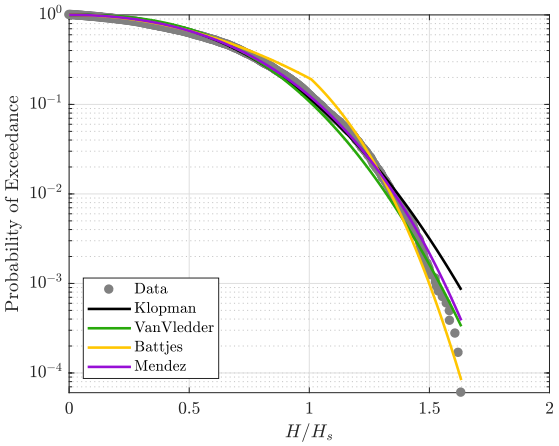
Figures 8a - 8d show comparisons between the measured wave heights and the shallow-water models



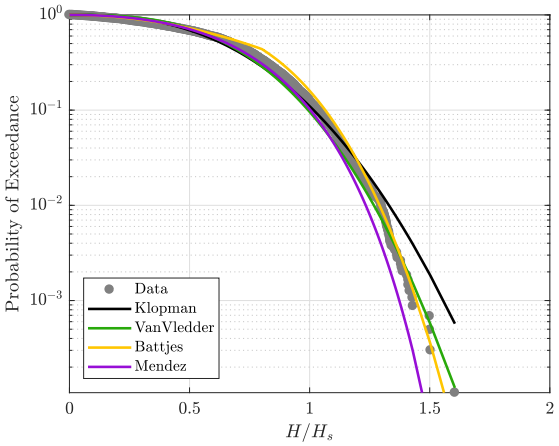
(a) L9: $H_s/d = 0.25$ and $k_p d = 1.06$.



(b) AWG: $H_s/d = 0.32$ and $k_p d = 0.37$.



(c) AWG: $H_s/d = 0.39$ and $k_p d = 0.43$.

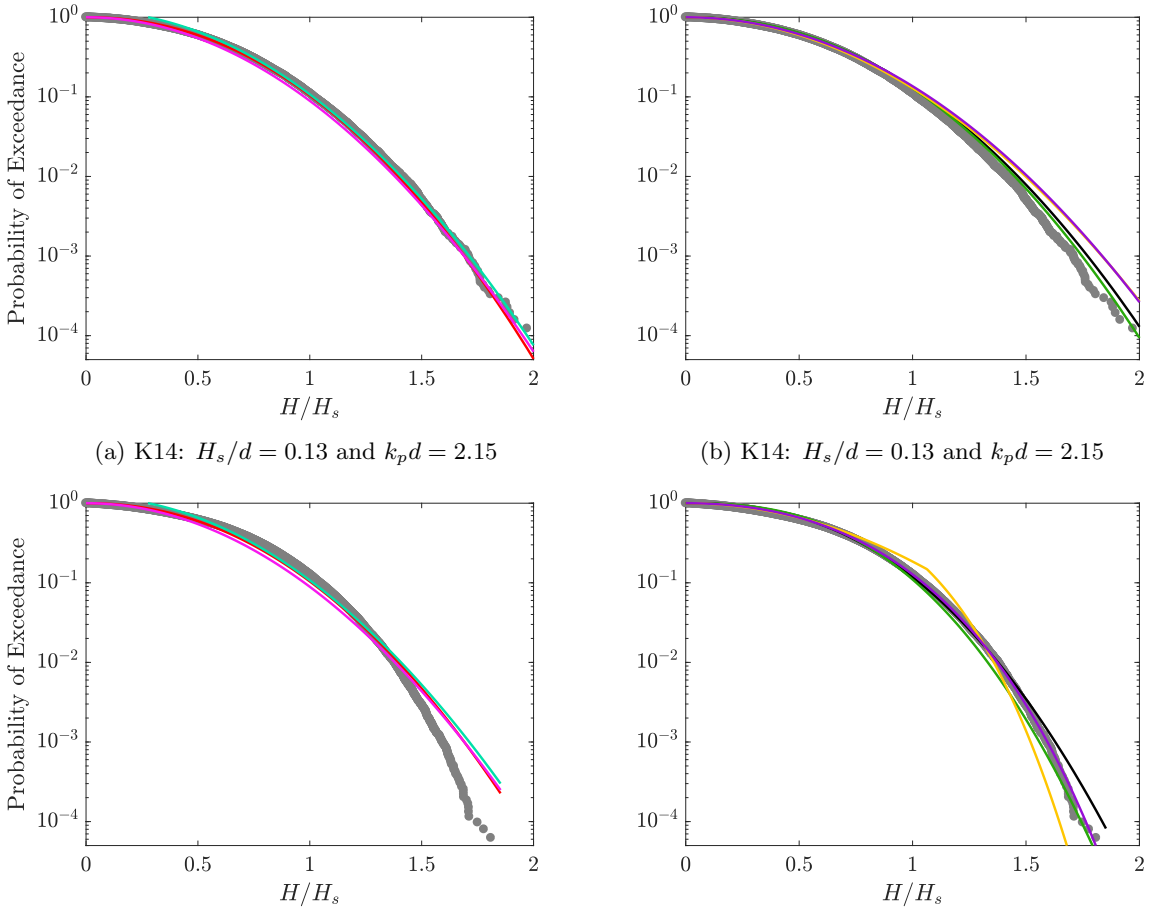


(d) AWG: $H_s/d = 0.55$ and $k_p d = 0.47$.

Figure 8: Normalised wave height distribution comparing representative data recorded at the L9 ($d = 25\,\mathrm{m}$) and AWG platforms ($d = 7.7\,\mathrm{m}$) and the shallow-water models.

for representative sea-states recorded at the L9 and AWG platforms. More specifically, Figures 8a - 8b illustrate the quality of the fit in the area where the two Glukhovskiy models are performing best. In contrast, Figures 8c - 8d both concern data from the AWG platform since this is the only platform that contains very high H_s/d values. Figure 8c concerns data with a mean value of $H_s/d \approx 0.4$, and shows that the Mendez distribution provides a good fit, while the performance of the Battjes-Groenendijk model is poor. In contrast, with a higher value of $H_s/d \approx 0.55$, Figure 8d shows that the opposite is true. In this case, the Battjes-Groenendijk model is closer to the data and the Mendez model underestimates the measured wave heights. Interestingly, the effective water depth for these two cases is very similar. This is most probably related to the nature of the experimental data used in the calibration of these models. In both cases the data were generated in uni-directional wave flumes; the numerical studies of Katsardi and Swan (2011) highlighting the nonlinear changes that can arise in uni-directional seas.

The previous behaviour is not observed in the application of the van Vledder model. Indeed, this model appears to be the most robust, even for sea-states in which it is not providing the best fit. This is evident in Figures 8c and 8d, where the van Vledder model, remains in reasonably close agreement with the measured data.



(a) K14: $H_s/d = 0.13$ and $k_p d = 2.15$

(b) K14: $H_s/d = 0.13$ and $k_p d = 2.15$

(c) AWG: $H_s/d = 0.42$ and $k_p d = 0.5$

(d) AWG: $H_s/d = 0.42$ and $k_p d = 0.5$

Figure 9: Normalised wave height distributions recorded at the K14 ($d = 27.4\,\mathrm{m}$) and AWG ($d = 7.7\,\mathrm{m}$) platforms showing comparisons to deep-water [(a),(c)] and shallow-water [(b),(d)] models. The measured data are shown in [●] while the distributions are: Forristall [—], Naess [—], Boccotti [—], Klopman [—], van Vledder [—], Mendez [—], Battjes-Groenendijk [—].

Combining these results provides guidance regarding the best performing wave height models for a wide range of met-ocean conditions. It is also evident that a transition between deep-water and shallow-water models occurs for $H_s/d \approx 0.2$. Below this threshold, the deep-water models provide the best fits; the Boccotti model being the most successful in the majority of the cases. Above this threshold, the shallow-water models provide the best fits. In identifying the best performing shallow-water model, the dependence on H_s/d is clearly defined, as described above.

Having determined the best fit models for a variety of met-ocean parameters, it is instructive to examine how the deep-water and shallow-water models perform for the same sea-states. To this end, two data bins have been considered in Figure 9. Sub-plots (a) and (c) provide comparisons between the measured wave heights and the deep-water models, while sub-plots (b) and (d) show comparisons between the same data and the shallow-water models.

More specifically, Figures 9a-9b present data obtained at the K14 platform with $H_s/d = 0.13$ and $k_p d = 2.15$. Considering this data bin, all three deep-water models (Forristall, Naess, Boccotti) provide reasonable agreement with the measured wave heights. In contrast, the shallow-water models are shown to over-estimate the measured data. Figures 9c-9d present data from the AWG platform with $H_s/d = 0.42$

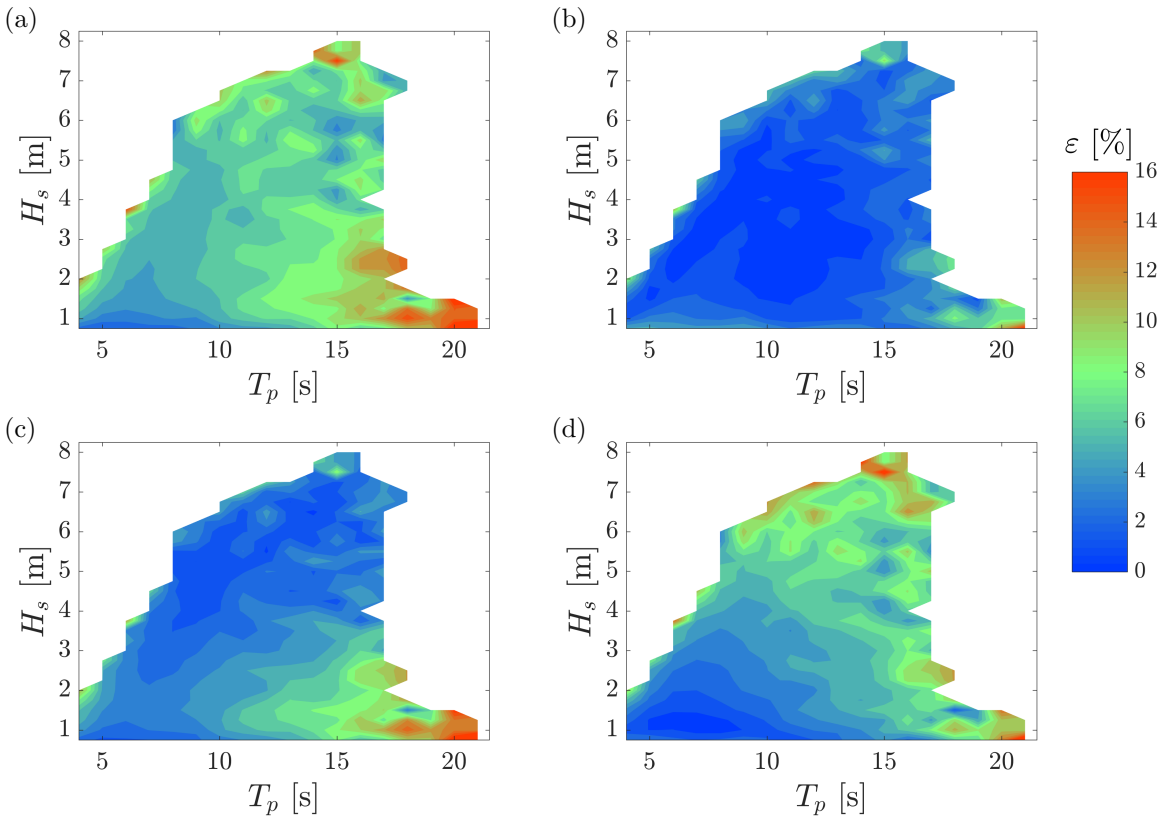


Figure 10: Contour plots indicating the RMS error (ε) in the P_{10} percentile of the wave height distributions for (a) Rayleigh, (b) Boccotti, (c) van Vledder and (d) Battjes-Groenendijk considering data measured at the K14 ($d = 27.4\,\mathrm{m}$) platform. The error levels (expressed as percentages) are mapped onto the (H_s, T_p) parameter space using the *data binning* approach.

and $k_p d = 0.5$. It is clear that under these conditions, the wave height distribution is significantly affected by wave breaking. Evidence of this is provided by the characteristic shift of the distribution towards lower H/H_s values for lower probabilities of exceedance. Given the earlier comments, the deep-water models fail to describe (do not include) the effects of this wave breaking. In contrast, the shallow-water models are much closer to the data; the Mendez and van Vledder distributions showing the closest agreement. These results indicate that the deep-water models cannot describe the shallow-water measurements accurately and vice versa.

Useful insights into the performance of each model can be obtained by investigating the distribution of its error across available sea-states. To this end, Figures 10 and 11 each present contour plots with the relative (RMS) error of four widely applied models; the Rayleigh, Boccotti, van Vledder and Battjes-Groenendijk distributions. All of these data correspond to the P_{10} percentile; Figure 10 concerning data recorded at K14 and Figure 11 AWG. Within these plots, the P_{10} percentile has been selected because it combines information from both the bulk of the data and the tail of the distribution. Additionally, the present results can be used to further validate the earlier findings based upon the P_{20} and P_1 percentiles.

Figure 10 presents the RMS error (ε) in the P_{10} percentile of the wave height distribution mapped onto the (H_s, T_p) parameter space, using the (deeper water) measurements from the K14 platform. Considering the assessment of the Rayleigh distribution (Figure 10a), it can be seen that relatively large errors (>8%) are present over most of the domain. The smallest errors arise in sea-states with very small significant wave heights and spectral peak periods ($H_s \approx 1\,\mathrm{m}$ and $T_p \approx 6\,\mathrm{s}$). At this stage, it should be noted that the main

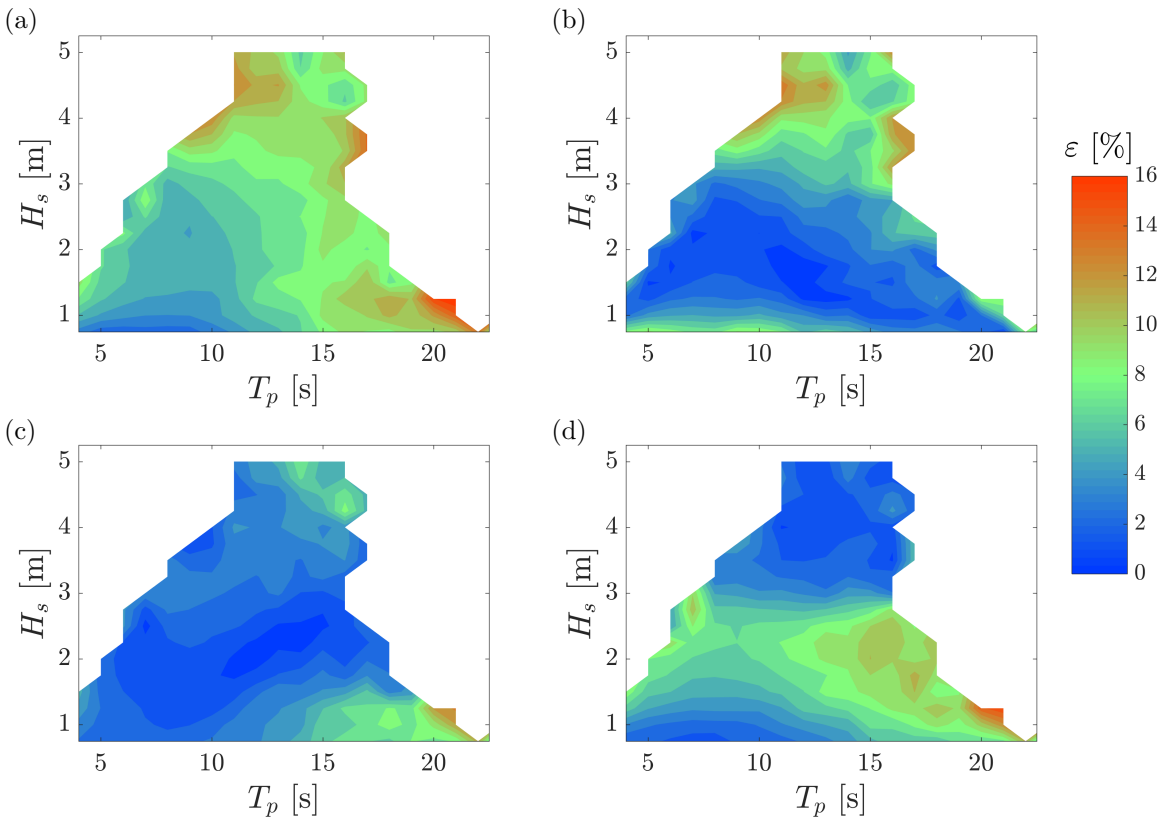


Figure 11: Contour plots indicating the RMS error (ε) in the P_{10} percentile of the wave height distributions for (a) Rayleigh, (b) Boccotti, (c) van Vledder and (d) Battjes-Groenendijk considering data measured at the AWG ($d = 7.7$ m) platform. The error levels (expressed as percentages) are mapped onto the (H_s, T_p) parameter space using the *data binning* approach.

reason for including the Rayleigh distribution is its wide application in engineering practice rather than its expected performance. On the other hand, the Boccotti model (Figure 10b) is characterised by small errors (<3%) over most of the domain. The largest discrepancies appear for very moderate sea-states with long spectral peak periods ($H_s \approx 1$ m and $T_p \approx 20$ s). Minor increases in the error are also observed for sea-states with the largest H_s.

In contrast, the van Vledder model (Figure 10c) is shown to have smaller errors for the most energetic sea-states (larger H_s). Indeed, in considering this model, the largest errors are recorded for benign sea-states with long spectral periods ($H_s < 2.5$ m and $T_p > 15$ s). Interestingly, the Battjes-Groenendijk model (Figure 10d) exhibits a very similar behaviour to the Rayleigh distribution. Specifically, its error levels are considerable for large values of H_s and T_p. Indeed, it only performs well for low sea-states with small spectral peak periods. This behaviour is not surprising when one considers the formulation of the model. The second (Weibull) part of the distribution is utilised once wave heights exceed a transition wave height. However, this transition wave height has been empirically calibrated for very different wave conditions. As such, the model appears to be utilising only the first (Rayleigh) part of the composite distribution for the majority of the cases. As a result, it predicts very similar behaviour to the Rayleigh model.

Figure 11 presents a similar layout of results for the shallower water measurements recorded at the AWG platform. Again, the Rayleigh distribution (Figure 11a) is shown to perform poorly over the whole domain. Furthermore, the Boccotti distribution is now shown to incorporate larger errors in the steepest sea-states (Figure 11b). This is in agreement with the results presented earlier; the primary reason being the effects of wave breaking in sea-states with large H_s values. In contrast, the van Vledder distribution (Figure 11c) has

(on average) the smallest errors over the whole domain. Interestingly, the smallest errors arise for significant wave heights lying in the range $1.5 < H_s < 3.5$. However, larger errors are observed in the most energetic sea-states ($H_s > 4\,\mathrm{m}$) and for large spectral peak periods ($T_p > 15\,\mathrm{s}$). In contrast, the Battjes-Groenendijk model (Figure 11d) performs best in the most energetic sea-states, while exhibiting larger errors for more moderate sea-states ($H_s < 3\,\mathrm{m}$). Given that the steepest sea-states in this dataset are significantly affected by wave breaking, it is clear that they can be effectively modelled by the Battjes-Groenendijk distribution. Unfortunately, this is the only region where this model demonstrates such good agreement.

Taken together, these results identify the preferred area of application for each of the models considered. In addition, the strengths and weaknesses of each model have been presented. While data from a subset of the available offshore platforms has been presented in detail, the present results are representative of the wide range of met-ocean conditions considered across the available database.

5. Physical insights

Having investigated the applicability of the various models, the available dataset also allows an investigation of the key physical parameters affecting the wave height distribution. Within this section the role of spectral bandwidth and nonlinearity will be addressed. To isolate their effects the measured sea-states were grouped using the spectral bandwidth parameter (ν), the Ursell number (Ur) and sea-state steepness (S_p). These are defined by:

$$\nu = \sqrt{\frac{m_0 m_2}{m_1^2} - 1}, \quad Ur = \frac{H_s}{k_1^2 d^3} \quad \text{and} \quad S_p = \frac{2\pi H_s}{g T_p^2}, \tag{23}$$

where $m_n = \int_0^\infty \omega^n S_{\eta\eta}(\omega)\mathrm{d}\omega$ denotes the n^{th} spectral moment. In defining appropriate combinations of these non-dimensional parameters, 30 sea-states (corresponding to 10 hours of measured data) were selected using a nearest neighbour algorithm to represent distinctive conditions; their wave height distributions being derived as in the *data binning* approach.

Figure 12 presents wave height distributions arising in sea-states with the same nonlinearity (Ur) but different spectral bandwidths (ν). In this context, the Ur number incorporates the effects of both sea-state steepness and effective water depth; a high Ur indicating enhanced nonlinear behaviour. Similarly, large values in the spectral bandwidth parameter ν indicate broad-banded conditions (Tucker and Pitt, 2001). The normalised wave heights in Figure 12 have all been recorded at the K14 platform and correspond to moderate sea-state conditions ($Ur = 0.005$). It can be seen that as the spectral bandwidth increases from $\nu = 0.5$ to $\nu = 0.65$ and $\nu = 0.8$, the wave heights reduce. This is manifested though the shift of the tail of the distribution towards lower values. The reason for this behaviour lies in the progressive de-correlation between wave crests and wave troughs in more broad-banded conditions (Longuet-Higgins, 1980). The results in this figure are representative of a wide variety of moderate sea-state conditions ($Ur < 0.05$) arising at all platforms. Unfortunately, the small variability in spectral bandwidth for steeper sea-states does not allow for a similar investigation at larger Ursell numbers. In this respect, it is worth noting the vast majority of the spectra analysed in this dataset are uni-modal; a small number of multi-modal spectra being observed primarily in benign sea-states.

To address the influence of nonlinearity, sea-states with the same spectral bandwidth ($\nu \approx 0.6$) but different degrees of nonlinearity have been retained. More specifically, Figures 13a and 13b consider sea-states with the same effective water depth ($k_p d$), but different values of sea-state steepness (S_p). In this way, the effects arising in different water depths can be further isolated. In both sub-plots the sea-state steepnesses increase incrementally from $S_p = 0.005$ to $S_p = 0.015$ and $S_p = 0.025$, to capture the effects of increasing nonlinearity as a function of the (depth-independent) sea-state steepness. It can be seen that for the deeper water case ($k_p d = 1.3$ on Figure 13a) the increase in sea-state steepness does not induce any notable changes in the wave height distributions. This is in agreement with the findings of earlier studies (Tayfun and Fedele, 2007) indicating that nonlinear effects do not significantly alter the distribution of wave heights. This is due to the fact that nonlinearities arising at second-order lead to a simultaneous increase in wave troughs and wave crests. As such, the two effects cancel out leaving the wave heights unaffected. However, it is worth noting that the crest height distributions will be affected by these effects (Latheef and

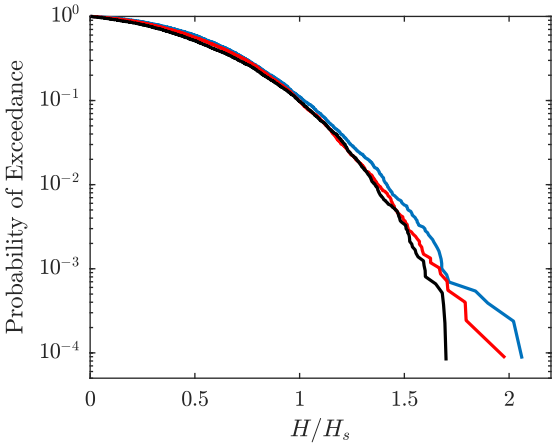
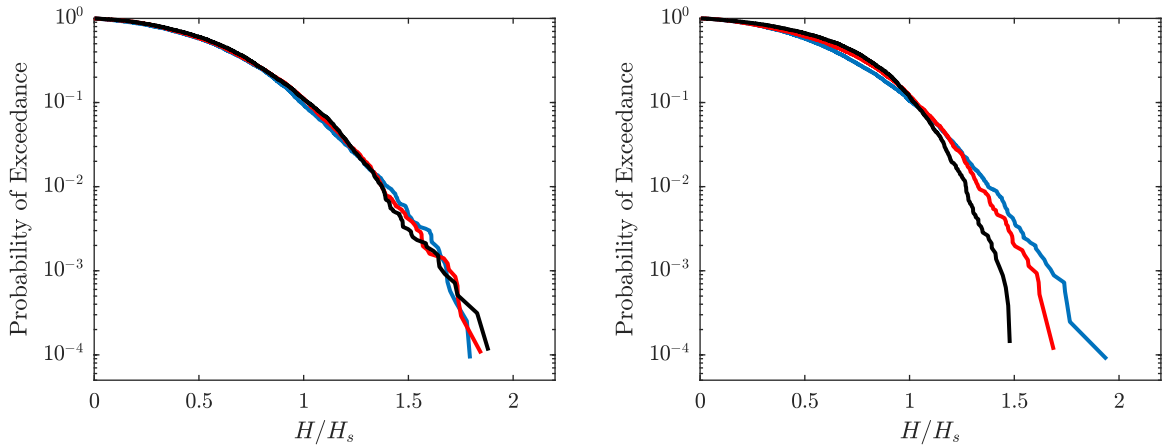


Figure 12: Effect of spectral bandwidth on the normalised wave height distribution (H/H_s). The measurements correspond to moderate sea-state conditions ($Ur = 0.005$) at the K14 platform and $\nu = 0.5$ [——], $\nu = 0.65$ [——] and $\nu = 0.8$ [——].

Swan, 2013). Moreover, in some of the most nonlinear sea-states included in the present dataset, effects higher than second-order are observed. These have been addressed in detail by Karmpadakis et al. (2019), particularly with respect to crest height statistics.

Considering the shallower water cases ($k_p d = 0.55$ on Figure 13b), it is clear that a different situation arises; the increase in S_p now leading to considerable changes in the wave height distribution. Specifically, steeper sea-states show a marked reduction in the largest wave heights indicated by the shift in the tail of the distribution. This shift is accompanied by a small increase in wave heights arising at $Q \approx 4 \times 10^{-1}$. Both these observations are explained by the occurrence of extensive wave breaking. The reductions in the largest wave heights are a direct consequence of the dissipation of wave energy due to wave breaking. At the same time, wave breaking leads to a reduction on the measured H_s of each sea-state compared to what it would have been had no wave breaking occurred. Therefore, the waves that have not undergone wave breaking (typically smaller waves) are now normalised by a lower value. This leads to the observed increase of wave heights for large probabilities of exceedance. In many respects, this situation arises due to the presence of two populations of waves (breaking and non-breaking). This justifies the use of a composite model such as Battjes and Groenendijk (2000) or Wu et al. (2016).

Taken together, these results have indicated the two main parameters that affect the distribution of wave heights in a wide range of sea-state conditions. In relating these to the assessment of the wave height models in Section 4, two important conclusions arise. First, the effect of spectral bandwidth acts to reduce the largest wave heights in moderate, deep water conditions. Second, the effect of nonlinearity does not have a significant impact on the wave height distribution provided that there is no significant wave breaking. However, in the latter case, steeper sea-states exhibit notable reductions of the largest wave heights. As such, the relative success of the deep-water models can be related to the accuracy with which they capture the effects of spectral bandwidth. However, since these models have not been derived to incorporate wave breaking, they cannot approximate steep shallow water sea-states. In contrast, the success of shallow water models in these conditions largely depends on their implementation of wave breaking (typically empirical). While the spectral bandwidth is also incorporated in these models (e.g. empirical calibration of H_{rms}) they cannot capture the differences arising in deeper water depths. As a result, there is no single model that can describe with consistent accuracy the full range of sea-states included in the present dataset. To address this, a new wave height distribution has been derived by the authors using the dataset presented herein together with supplementary experimental measurements. This will be presented in a subsequent publication; the goal being to provide a single model with wide applicability.



(a) K14: $k_p d = 1.3$ (b) AWG: $k_p d = 0.55$

Figure 13: Effect of nonlinearity on the normalised wave height distribution (H/H_s) for sea-states with the same spectral bandwidth ($\nu \approx 0.6$). The sea-state steepnesses correspond to: $S_p = 0.005$ [——], $S_p = 0.015$[——] and $S_p = 0.025$ [——]. Comparisons are made in (a) $k_p d = 1.3$ and (b) $k_p d = 0.55$.

6. Concluding remarks

This study has presented the analysis of an extensive database of field measurements in intermediate and shallow water depths. The focus of this analysis has been the assessment of existing wave height distributions under a wide range of conditions; the aim being to provide guidance as to the most appropriate model to be used in design applications. Comparisons between nine widely applied wave height models and the field data have shown that no single model can describe the measured wave heights across a wide range of incident conditions accurately. Indeed, it was found that for different H_s/d and $k_p d$ values, different models provide the best representation of the available data. Specifically, for $H_s/d < 0.2$ the Boccotti model gives the greatest accuracy, closely followed by the Naess and Forristall models. As H_s/d increases the van Vledder, Mendez and Battjes-Groenendijk models provide the best fits. Additionally, the main effects that drive the formation of the wave height distribution have been outlined. In this respect, the influence of spectral bandwidth and nonlinearity have been shown to have characteristic impacts on the distributions for different effective water depths and sea-state conditions. Although these effects can be approximated in varying ways by the available models, a unified distribution to cover the widest range of conditions is yet to be achieved.

Acknowledgements

The authors are grateful to the following sponsors of the LoWiSh (Limits on Waves in Shallow Water) JIP for funding this research: Shell, Mærsk, BP, Total, ConocoPhillips, Exxon, Woodside and Equinor. The authors would also like to thank Shell and Mærsk for providing access to the field data, whilst confirming that the views expressed herein are entirely their own.

References

Alkhalidi, M.A., Tayfun, M.A., 2013. Generalized Boccotti distribution for nonlinear wave heights. Ocean Engineering 74, 101–106.

Allender, J., Audunson, T., Barstow, S., Bjerken, S., Krogstad, H., Steinbakke, P., Vartdal, L., Borgman, L., Graham, C., 1989. The WADIC project: A comprehensive field evaluation of directional wave instrumentation. Ocean Engineering 16, 505–536.

Battjes, J.A., Groenendijk, H.W., 2000. Wave height distributions on shallow foreshores. Coastal Engineering 40, 161–182.

Battjes, J.A., Janssen, J., 1978. Energy loss and set-up due to breaking of random waves, in: Coastal Engineering 1978, pp. 569–587.

Boccotti, P., 1983. Some new results on statistical properties of wind waves. Applied Ocean Research 5, 134–140.
Boccotti, P., 1989. On mechanics of irregular gravity waves. Atti della ASccad. naz. dei Lincei; A 386. Memorie/Classe di scienze fis., mat. e naturali. Ser. 8; Vol. 19. Sez. 1a,(Matematica, meccanica, astronomica, geodesia e geofisica) .
Boccotti, P., 2000. Wave mechanics for ocean engineering. volume 64. Elsevier.
Bouws, E., 1980. Spectra of extreme wave conditions in the southern north sea considering the influence of water depth, in: Proceedings of the International Conference on Sea Climatology, pp. 51–71.
Casas-Prat, M., Holthuijsen, L.H., 2010. Short-term statistics of waves observed in deep water. Journal of Geophysical Research: Oceans 115, 1–20. doi:10.1029/2009JC005742.
Christou, M., Ewans, K., 2011. Examining a comprehensive dataset containing thousands of freak wave events. Part 1 - Description of the Data and Quality Control Procedure, in: Proceedings of the ASME 30th 2011 International Conference on Ocean, Offshore and Arctic Engineering.
Christou, M., Ewans, K., 2014. Field Measurements of Rogue Water Waves. Journal of Physical Oceanography , 2317–2335doi:10.1175/JPO-D-13-0199.1.
De Waal, J., Van der Meer, J., 1993. Wave runup and overtopping on coastal structures, in: Coastal Engineering 1992, pp. 1758–1771.
DNV-RP-C205, 2010. Environmental Conditions and Environmental Loads. DNV , 9–123.
Dysthe, K., Krogstad, H.E., Müller, P., 2008. Oceanic Rogue Waves. Annual Review of Fluid Mechanics 40, 287–310. doi:10.1146/annurev.fluid.40.111406.102203.
Ewans, K., Feld, G., Jonathan, P., 2014. On wave radar measurement. Ocean Dynamics , 1281–1303doi:10.1007/s10236-014-0742-5.
Forristall, G., 2005. Understanding rogue waves: Are new physics really necessary, in: Proceedings of the 14th Aha Hulikoá Hawaiian Winter Workshop on Rogue Waves, Citeseer.
Forristall, G.Z., 1978. On the statistical distribution of wave heights in a storm. Journal of Geophysical Research , 2353doi:10.1029/JC083iC05p02353.
Glukhovskiy, B.K., 1966. Issledovaniye morskogo vetrovogo volneniya (investigation of sea wind waves), in: Gidrometeoizdat, St. Petersburg, Leningrad.
Goda, Y., 2010. Random seas and design of maritime structures. World Scientific.
Guedes Soares, C., Cherneva, Z., Antao, E.M., 2004. Steepness and asymmetry of the largest waves in storm sea states. Ocean Engineering 31, 1147–1167. doi:10.1016/j.oceaneng.2003.10.014.
Hall, M., 1972. Hurricane-Generated Ocean Currents - Part 1: The Development of a Measurement Program, in: Offshore Technology Conference, pp. 70–73.
Hameed, T.S., Baba, M., 1985. Wave height distribution in shallow water. Ocean Engineering 12, 309–319.
Hasselmann, K., Barnett, T.P., Bouws, E., Carlson, H., Cartwright, D.E., Enke, K., Ewing, J.A., Gienapp, H., Hasselmann, D.E., Kruseman, P., Meerburg, A., Muller, P., Olbers, D.J., Richter, K., Sell, W., Walden, H., 1973. Measurements of Wind-Wave Growth and Swell Decay during the Joint North Sea Wave Project (JONSWAP). Ergnzungsheft zur Deutschen Hydrographischen Zeitschrift Reihe A(8), p.95.
James, I., 1986. A note on the theoretical comparison of wave staffs and wave rider buoys in steep gravity waves. Ocean Engineering 13, 209–214.
Karmpadakis, I., Swan, C., Christou, M., 2019. Laboratory investigation of crest height statistics in intermediate water depths. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences 475, 20190183. URL: https://royalsocietypublishing.org/doi/10.1098/rspa.2019.0183, doi:10.1098/rspa.2019.0183.
Katsardi, V., de Lutio, L., Swan, C., 2013. An experimental study of large waves in intermediate and shallow water depths. Part I: Wave height and crest height statistics. Coastal Engineering , 43–57doi:10.1016/j.coastaleng.2012.09.007.
Katsardi, V., Swan, C., 2011. The evolution of large non-breaking waves in intermediate and shallow water. I. Numerical calculations of uni-directional seas. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences 467, 778–805. doi:10.1098/rspa.2010.0280.
Klopman, G., 1996. Extreme wave heights in shallow water. Technical Report.
Klopman, G., Stive, M., 1989. Extreme waves and wave loading in shallow water, in: Wave and Current Kinematics and Loading: E&P Forum Workshop, Paris, France, 25-26 October 1989, Delft Hydraulics.
Latheef, M., 2013. Surface wave statistics in directionally spread seas. Ph.D. thesis. Imperial College London.
Latheef, M., Swan, C., 2013. A laboratory study of wave crest statistics and the role of directional spreading. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences 469, 20120696.
Lindgren, G., 1970. Some properties of a normal process near a local maximum. The Annals of Mathematical Statistics 41, 1870–1883. doi:10.1214/aoms/1177733256.
Longuet-Higgins, M.S., 1952. On the statistical distribution of sea waves. Journal of Marine Research 11, 245–266.
Longuet-Higgins, M.S., 1980. On the distribution of the heights of sea waves: Some effects of nonlinearity and finite band width. Journal of Geophysical Research , 1519doi:10.1029/JC085iC03p01519.
Magnusson, A., Donelan, M., Drennan, W., 1999. On estimating extremes in an evolving wave field. Coastal Engineering 36, 147–163.
Mai, S., Wilhelmi, J., Barjenbruch, U., 2011. Wave height distribution in shallow waters, in: Coastal Engineering Proceedings, p. 63. doi:10.1016/0029-8018(85)90070-8.
Martin, C., Grant, C., Sproson, R., et al., 1988. Wave height distributions in shallow water, in: Oceanology'88: Proceedings of an international conference, Society of Underwater Technology.
Massel, S., 1996. On the largest wave height in water of constant depth. Ocean Engineering 23, 553–573.
Mendez, F.J., Losada, I.J., Medina, R., 2004. Transformation model of wave height distribution on planar beaches. Coastal

Engineering 50, 97–115.
Miche, A., 1944. Mouvements ondulatoires de la mer en profondeur croissante ou decroissante. Annales Des Ponts et Chaussées 114, 42–78.
Mori, N., Yasuda, T., 2002. A weakly non-gaussian model of wave height distribution for random wave train. Ocean Engineering 29, 1219–1231. doi:10.1016/S0029-8018(01)00075-0.
Moritz, H.R., 2002. Observing large waves using bottom-mounted pressure and current meters, in: Ocean Wave Measurement and Analysis (2001), pp. 44–53.
Naess, A., 1985. On the distribution of crest to trough wave heights. Ocean Engineering 12, 221–234. doi:10.1016/0029-8018(85)90014-9.
Pawlowicz, R., Beardsley, B., Lentz, S., 2002. Classical tidal harmonic analysis including error estimates in matlab using t_tide. Computers & Geosciences 28, 929–937.
Rattanapitikon, W., 2010. Verification of conversion formulas for computing representative wave heights. Ocean Engineering , 1554–1563doi:10.1016/j.oceaneng.2010.09.013.
Rice, S.O., 1944. Mathematical analysis of random noise. Bell Labs Technical Journal 23, 282–332.
RS-Aqua, 2014. WaveRadar REX Manual. Technical Report K.
Stansell, P., 2004. Distributions of freak wave heights measured in the North Sea. Applied Ocean Research 26, 35–48. doi:10.1016/j.apor.2004.01.004.
Stansell, P., 2005. Distributions of extreme wave, crest and trough heights measured in the North Sea. Ocean Engineering 32, 1015–1036. doi:10.1016/j.oceaneng.2004.10.016.
Stansell, P., Wolfram, J., Linfoot, B., 2002. Effect of sampling rate on wave height statistics. Ocean Engineering 29, 1023–1047. doi:10.1016/S0029-8018(01)00066-X.
Stive, M., 1985. A scale comparison of waves breaking on a beach. Coastal Engineering 9, 151–158.
Stive, M., 1986. Extreme shallow water wave conditions. H533, Delft Hydraulics, The Netherlands .
Tayfun, M.A., 1981. Distribution of crest-to-trough wave heights. Journal of Waterway, Port, Coastal, and Ocean Engineering 107, 149–158.
Tayfun, M.A., 1990. Distribution of large wave heights. Journal of Waterway, Port, Coastal, and Ocean Engineering 116, 686–707.
Tayfun, M.A., 1993. Sampling-rate errors in statistics of wave heights and periods. Journal of Waterway, Port, Coastal, and Ocean Engineering 119, 172–192.
Tayfun, M.A., Fedele, F., 2007. Wave-height distributions and nonlinear effects. Ocean Engineering 34, 1631–1649. doi:10.1016/j.oceaneng.2006.11.006.
Tromans, P.S., Anaturk, A.R., Hagemeijer, P., et al., 1991. A new model for the kinematics of large ocean waves-application as a design wave, in: The First International Offshore and Polar Engineering Conference, International Society of Offshore and Polar Engineers.
Tucker, M.J., Pitt, E.G., 2001. Waves in Ocean Engineering. doi:10.1016/0141-0296(92)90048-U.
van Vledder, G., 1991. Modification of the Glukhovskiy distribution. Delft Hydraulics Report H1203 .
Ward, E.G., 1974. Ocean Data Gathering Program - An overview, in: Proceedings of the Fourth Annual Offshore Technology Conference, pp. 771–775.
Wu, Y., Randell, D., Christou, M., Ewans, K., Jonathan, P., 2016. On the distribution of wave height in shallow water. Coastal Engineering 111, 39–49.