On the distribution of wave height in shallow water

Yanyun Wu[a], David Randell[a], Marios Christou[b], Kevin Ewans[c], Philip Jonathan[a,*]

[a] Shell Research Ltd., Manchester M22 0RR, UK
[b] Imperial College, London SW7 2AZ, UK
[c] Sarawak Shell Bhd., 50450 Kuala Lumpur, Malaysia

ARTICLE INFO

ABSTRACT

The statistical distribution of the height of sea waves in deep water has been modelled using the Rayleigh (Longuet-Higgins, 1952) and Weibull distributions (Forristall, 1978). Depth-induced wave breaking leading to restriction on the ratio of wave height to water depth requires new parameterisations of these or other distributional forms for shallow water. Glukhovskiy (1966) proposed a Weibull parameterisation accommodating depth-limited breaking, modified by van Vledder (1991). Battjes and Groenendijk (2000) suggested a two-part Weibull–Weibull distribution. Here we propose a two-part Weibull-generalised Pareto model for wave height in shallow water, parameterised empirically in terms of sea state parameters (significant wave height, H_S, local wave-number, k_L, and water depth, d), using data from both laboratory and field measurements from 4 offshore locations. We are particularly concerned that the model can be applied usefully in a straightforward manner; given three pre-specified universal parameters, the model further requires values for sea state significant wave height and wave number, and water depth so that it can be applied. The model has continuous probability density, smooth cumulative distribution function, incorporates the Miche upper limit for wave heights (Miche, 1944) and adopts H_S as the transition wave height from Weibull body to generalised Pareto tail forms. Accordingly, the model is effectively a new form for the breaking wave height distribution. The estimated model provides good predictive performance on laboratory and field data.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

There is considerable interest in understanding the characteristics of ocean waves in shallow water. Specifically, from an engineering perspective, a design wave height in shallow water is required in order to determine wave loading on coastal structures, wave run-up and wave overtopping. As discussed by Katsardi and Swan (2011b) in their introduction to modelling of non-breaking unidirectional waves in intermediate and shallow water, the physics of evolving wave fields in shallow water is critically dependent upon water depth. Distributions of wave height in shallow water must therefore be expressed as functions of water depth or related parameters. As summarised in Section 3 below, considerable effort has been devoted to the development and refinement of parametric forms for the statistical distribution of wave height in shallow water based on field and laboratory measurements.

There is a long history of modelling the distribution of wave height in coastal regions; Guedes Soares (2003) provides an

introduction. The LoWiSh Joint Industry Project addresses uncertainties in the specification of the maximum wave height occurring on a continental shelf. One of the objectives of LoWiSh is to review existing distributional forms for wave height in intermediate and shallow water. In the first phase of the project, the distribution of individual wave height from laboratory measurements (Katsardi and Swan, 2011a) was found to be well described by a Weibull form, with parameters expressed in terms of Ursell number. However, the same parameterisation did not hold for the field measurements. The two-part Weibull–Weibull distribution (of Battjes and Groenendijk, 2000) was found to explain the distribution of laboratory wave height well, capturing the discontinuity in slope of the cumulative distribution of wave height in very shallow water. The limiting characteristics of the largest waves in both intermediate and shallow waters were found to be critically dependent upon the effective water depth, $k_L d$, where k_L is a local wave-number based upon a locally measured wave period, and d is water depth. In the recent literature, Mai et al. (2011) report that a modified form of the two-part Weibull–Weibull distribution is appropriate to characterise the distribution of wave height from radar level gauge measurements at three locations in the German North Sea. Katsardi et al. (2013) observe that effective water depth and significant wave height influence the distribution of wave height in shallow water from laboratory measurements, but

that different wave spectral bandwidths and moderate bed slopes (less than 1 : 100) do not. They also observe that the Weibull–Weibull distribution over-predicts largest wave heights.

The objective of the current work is to extend the analysis conducted during the first phase of LoWiSh to establish a universal model for wave height in shallow water, appropriate for all available laboratory, field and numerical model data; and to compare the performance of the new model with competitor models from the literature. The contents of this article are arranged as follows. Laboratory and field data used for model estimation and validation are described in Section 2. Section 3 summarises existing models for the distributions of wave height, and motivates the requirements for the development of the new model, which is explained in detail in Section 4. Section 5 estimates the Weibull-generalised Pareto model for the laboratory and field data, and compares model performance against alternative model forms for the distributions of wave height. Conclusions and recommendations are made in Section 6.

2. Data

Five data sources were used to estimate the Weibull-generalised Pareto (henceforth WGP) model; four correspond to measured data from offshore locations, and the fifth to measured data from a wave tank at Imperial College London. The offshore locations are (a) Ameland Westgat (AWG), at 8 m water depth on the Dutch coast for December 2007; (b) Petten, at 8 m and 20 m water depth on the Dutch coast for different measurement campaigns over the period 2001–2008; (c) the Field Research Facility (FRF) at 9 m water depth on the coast of North Carolina for different measurement campaigns over the period 2003–2007, and (d) North Cormorant (NC), at 160 m water depth in the northern North Sea for the period November 2006 to February 2007.

Wave data at AWG were measured with a Saab Rex WaveRadar sensor. The Saab WaveRadar has been shown to give reliable measurements of the sea surface elevation over the frequency band (0.06, 0.60) Hz by Ewans et al. (2014). The unit on the AWG platform was mounted at 26.5 m above the sea surface, clear from immediate obstructions and in particular without obstruction for waves from between north and north-west, the direction of largest storm waves. The sensor recorded the sea surface elevation continuously at 2 Hz. Waves approaching from the seaward direction of north–north-west traverse small bottom slopes, reaching a maximum slope of 1 : 400 near to the platform. The wave data at the North Cormorant platform were also recorded with a Saab Rex WaveRadar sampling continuously at 2 Hz. The sensor was located on the south-east corner of the platform at an elevation of 28.7 m above mean sea level. Although North Cormorant in no way represents a shallow water location, we include it in the present study to ascertain whether a distribution for wave height can be specified which is applicable generally, not only in shallow water.

Wave data at the AWG and North Cormorant platforms were processed according to the following steps: (a) Air-gap was inverted to estimate surface elevation above a nominal datum, (b) mean water level was calculated for consecutive 10 minute segments, (c) a 2 Hz spline was fitted through each of these 10-minute values to represent a continuously varying mean water level, (d) the continuously varying mean water level time series was subtracted from the surface elevation above the nominal datum time series, and (e) individual waves were identified on zero-crossing basis.

The Petten wave data were recorded with a Directional Waverider buoy located in a nominal water depth of 20 m (the MP1 site), nearly 8 km from the shore, and an Etrometa wave staff located in a nominal water depth of 8 m (the MP3 site). MP3 is behind a bar (approximately 1.9 km seaward) and on the forward face of a second bar. These bars are likely to have introduced local breaking and shoaling, resulting in effects in the MP3 data that are not observed in the deeper water sites at Petten nor the other field locations not affected by bars. The average beach slope between MP1 and MP3 is approximately 1 : 600. The slope is less than 1 : 600 seaward of MP1. The Directional Waverider buoy data at MP1 were recorded at 1.28 Hz, while the wave staff data at MP3 were recorded at 2.56 Hz. Still water level measurements were made at MP3 with a digital level meter. Further details of the measurements can be found in Hordijk (2003,2004).

The Field Research Facility measurements were made with a pressure transducer in 8.5 m water depth. The measurements from the pressure sensor were made at 2 Hz and were converted to surface elevations, using linear wave theory, with a spectral density cap corresponding to an f^{-4} decay rate for frequencies f above 0.25 Hz, to avoid amplification of noise at high frequencies. The beach slope at the pressure transducer location is 1 : 150. Seaward of the location the bottom slope is around 1 : 300. Further details of the measurements can be found in Birkemeier et al. (1997).

For each of the field locations, original 20-minute sea state samples were combined into rolling three-hour sea states for consideration here. A small number of the resulting 3-hour sea states were omitted due to missing data or obviously unrealistic values. The resulting number of sea states for analysis per location was 414 (at AWG), 3646 (at NC), 4383 (at FRF) and 676 (at Petten). In addition, we found it useful to consider the Petten samples at 8 m and 20 m separately, henceforth denoted P8 and P20 respectively for brevity.

The experimental tank data were obtained in a specially commissioned wave flume at Imperial College London adopting a 1:100 laboratory length scale and 1:10 laboratory time scale. At laboratory scale, the flume is 60 m long and 0.3 m wide. The waves were generated by an absorbing flap-type wave maker located in deep water (0.7 m at model scale). The waves then propagated up a 1 : 15 slope to a water depth of 0.5 m for three different beach slopes: constant water depth of 0.3 m (flat-bed), 1 : 100 and 1 : 250 gradient. For each test case, 8 separate runs with different random phasing were undertaken. Each run consists of 256 s of complete data, giving a total record length of 2048 s , for each test case. The sampling rate was 128 Hz. Wave trains corresponding to JONSWAP and log-normal spectral shapes with different spectral parameters were measured at up to 8 gauge locations. A total of 175 cases corresponding to different combinations of wave spectrum, bed slope and gauge were recorded (together henceforth referred to as Tank data). The full details of the laboratory set-up, instrumentation, and experimental measurements undertaken at Imperial College London are given by Katsardi et al. (2013). At full scale, each of the 175 cases corresponds to observation of a three-hour sea state.

For each data source and sea state, values of sea state significant wave height (H_S, derived from the zeroth order moment of the wave spectrum) and water depth d were available, together with values of individual wave heights and corresponding individual wave numbers (k_L, determined from individual zero-crossing wave periods using the linear dispersion relationship). We use the median of individual wave numbers as sea state wave number. Further information regarding the data sources and cases used in this work is available on request from the authors.

3. Distributions for wave height

A number of distributional forms have been proposed for wave height. In this section, we begin (in Section 3.1) by summarising some of the more popular distributional forms. With this background, we motivate a new model (in Section 3.2).

3.1. Existing models for wave height

Cumulative distribution function F_H and probability density function f_H for different descriptions of the distributions of wave height $h (> 0)$ as described below, where $H_S (> 0)$ is significant wave height.

3.1.1. Rayleigh distribution *(Longuet-Higgins, 1952)*
Longuet-Higgins (1952) introduced the Rayleigh distribution for wave height, with

$$F_H(h) = 1 - \exp\left[\left(-\frac{h}{\sqrt{8}s}\right)^2\right] \quad , \quad f_H(h) = \frac{2h}{8s^2}\exp\left[\left(-\frac{h}{\sqrt{8}s}\right)^2\right]$$

where s is parameterised in terms of H_S as $s = H_S/4$.

3.1.2. Glukhovskiy distribution *(Glukhovskiy, 1966)*
Glukhovskiy (1966) used a Weibull distribution for wave height in shallow water, with

$$F_H(h) = 1 - \exp\left[-A\left(\frac{h}{H_m}\right)^K\right] \quad , \quad f_H(h) = \frac{AK}{H_m}\left(\frac{h}{H_m}\right)^{K-1}\exp\left[-A\left(\frac{h}{H_m}\right)^K\right]$$

where, following van Vledder (1991)

$$A = \Gamma(1/K + 1)^K \ , \ K = \frac{2}{(1 - H_m/d)} \ , \ H_m = \sqrt{8}\frac{H_S}{4}\frac{\Gamma(1 + 1/k)}{\sqrt{\Gamma(1 + 2/k)}} \ \text{and}$$

$$k = \frac{2}{1 - \sqrt{8}H_S/(4.5636d)}$$

and Γ is the cumulative distribution function of the gamma distribution. Note that all of A, K, H_m and k are assumed > 0.

3.1.3. Forristall distribution in deep water *(Forristall, 1978)*
Forristall proposed the following Weibull distributional form for wave height in deep water, the parameters of which were estimated from measurements of Gulf of Mexico hurricanes, where

$$F_H(h) = 1 - \exp\left[-\frac{1}{\beta}\left(\frac{h}{H_S/4}\right)^\alpha\right],$$

$$f_H(h) = \frac{\alpha}{\beta H_S/4}\left(\frac{h}{H_S/4}\right)^{\alpha-1}\exp\left[-\frac{1}{\beta}\left(\frac{h}{H_S/4}\right)^\alpha\right]$$

where

$$\alpha = 2.125 \text{ and } \beta = 8.42 \ .$$

3.1.4. Tayfun distribution *(Tayfun, 1990)*
Tayfun generalised the asymptotic model of Boccotti (1989) to include the effects of higher order non-linearities, such that

$$F_H(h) = 1 - c_0 f \exp\left[-c_1\left(\frac{h}{H_S/4}\right)^2\right],$$

$$f_H(h) = \left(\frac{2c_1 h}{(H_S/4)^2}f - \frac{1}{H_S/4}df\right)\exp\left[c_1\left(\frac{h}{H_S/4}\right)^2\right]$$

where

$$c_0 = 1 \ , \ c_1 = 1/[4(1 + r_m)] \ , \ f = 1 + \left(\frac{Hs}{4}\right)^2\left(\frac{1 - r_m^2}{4r_m h^2}\right) \text{and}$$

$$df = -\left(\frac{Hs}{4}\right)^2\left(\frac{1 - r_m^2}{2r_m h^3}\right)$$

and parameter r_m is estimated from the wave spectrum as explained in Tayfun and Fedele (2007).

3.1.5. Forristall distribution in shallow water *(Forristall, 2007)*
Forristall developed the following Weibull distributional form for wave height in shallow water, the parameters of which are expressed in terms of Ursell number $U (> 0)$, with

$$F_H(h) = 1 - \exp\left[-\left(\frac{h}{\alpha H_S}\right)^\beta\right] \quad , \quad f_H(h) = \frac{\beta}{\alpha H_S}\left(\frac{h}{\alpha H_S}\right)^{\beta-1}\exp\left[-\left(\frac{h}{\alpha H_S}\right)^\beta\right]$$

where

$$U = H_S/(k^2 d^3) \ , \ \alpha = 0.6917 + 0.1012U \text{ and } \beta = 2.126 + 1.1365U.$$

3.1.6. Battjes-Groenendijk distribution in shallow water *(Battjes and Groenendijk, 2000)*
Battjes and Groenendijk proposed a two-part Weibull–Weibull distributional form for wave height in shallow water, the parameters of which were estimated from laboratory measurements. Above a high threshold $H_{tr} (> 0)$

$$F_H(h) = 1 - \exp\left[-\left(\frac{h}{H_2}\right)^{k_2}\right] \quad , \quad f_H(h) = \frac{k_2}{H_2}\left(\frac{h}{H_2}\right)^{k_2-1}\exp\left[-\left(\frac{h}{H_2}\right)^{k_2}\right]$$

where

$$k_2 = 3.6, H_{tr} = 0.35 + 5.8\tan(\alpha)d \ ,$$

and α is the bed slope. The form of $H_2 (> 0)$ is obtained from equation 7 of Battjes and Groenendijk (2000).

3.1.7. Mendez–Losada–Medina distribution *(Mendez et al., 2004)*
Mendez et al. (2004) developed a wave height distribution based on a bore approach for modelling energy dissipation in the inner surf zone, which includes shoaling and breaking, with parameters $H_{rms} (> 0)$ and $H_{max} (> 0)$, such that

$$F_H(h) = 1 - \exp\left[-\phi^2(\kappa)\left(\frac{h}{H_{rms} - \kappa h}\right)^2\right],$$

$$f_H(h) = \frac{2\phi^2(\kappa)H_{rms}h}{(H_{rms} - \kappa h)^3}\exp\left[-\phi^2(\kappa)\left(\frac{h}{H_{rms} - \kappa h}\right)^2\right]$$

for $0 \le h < H_{max}$, where

$$\phi(\kappa) \approx \left(1 - \kappa^{0.944}\right)^{1.187} \text{and } \kappa = \frac{H_{rms}}{H_{max}}.$$

The value of κ is evaluated in terms of Iribarren number I_r as

$$\kappa = (4.7 - 28.8I_r + 26.2I_r^2)\left(\frac{H_{rms}}{d}\right)^{5/2} \text{where } I_r = m\left(\frac{gT_m^2}{2\pi H_{rms}}\right)^{1/2},$$

and T_m is the mean wave period, evaluated here are the mean of the individual wave periods for the sea state.

3.2. Motivating a new model

To account for changing characteristics of the distribution of wave height in shallow water, the notion of a two-part distribution (such as that of Battjes and Groenendijk, 2000), parameterised in terms of sea state and beach characteristics, is attractive and forms the basis of the new model proposed here. It would appear that the Weibull–Weibull distribution of Battjes and Groenendijk (2000) can be improved in a number of important ways. Firstly, in the Weibull–Weibull distribution of Battjes and Groenendijk (2000), the upper Weibull distribution is intended to account for wave limiting processes associated with shallow water, but it has no upper limit (since the Weibull distribution extends to infinity), contradicting observation and theory. Secondly, the two-part distribution has a continuous cumulative distribution function, but a discontinuous density, which is obviously inconsistent with observation. For these reasons, we might seek a new model form that (a) accounts for an upper limit to wave height, and (b) has continuous probability density and hence smooth cumulative distribution.

There is considerable motivation for retaining a Weibull form for the body of the distribution of wave height. For linear waves in a narrow-banded sea state, wave heights are Rayleigh-distributed. It is clear from Section 3 that the Weibull distribution, of which Rayleigh is a special case, has been found to describe distributions of wave height from field data well. Moreover, Glukhovskiy (1966) showed that the Weibull exponent might depend on the ratio of significant wave height to water depth. Extreme value theory (e.g. Coles, 2001) indicates that exceedances $H - u$ of a high threshold u, can be modelled by a generalised Pareto (GP) distribution, with scale parameters $\sigma \, (> 0)$ and shape parameter ξ, for random variables H with any max-stable distribution with tail probability function

$$\Pr(H > h | H > u) = \left(1 + \frac{\xi}{\sigma}(h - u)\right)^{-1/\xi} \text{ for } h > u \,.$$

It would appear therefore that the GP distribution would be a suitable (conditional) model for large wave heights. If the GP shape parameter ξ is negative, the distribution of the excess has an upper bound $h_* = u - \sigma/\xi$. In this context, Miche (1944) suggested, due to steepness and water depth constraints, that the limiting wave height can be described by a function of local wave-number k_L and water depth d. The Miche limit h_{max} satisfies

$$\frac{k_L h_{max}}{2} = \beta \pi \tanh(k_L d) \,.$$

In the original work, the value of β was set at 0.142 and the value of k_L estimated from a non-linear analysis. In the current work we have estimated k_L as the local linear wave-number from the dispersion equation

$$\omega_L^2 = g k_L \tanh(k_L d)$$

where T_L is the local wave period, such that $\omega_L T_L = 2\pi$. We can therefore incorporate a physically-motivated upper end point to the GP tail by setting the GP upper bound equal to the Miche limit. The number of wave heights observed in a sea state, and predicted by theory, varies smoothly with the value of wave height. This requires that the cumulative distribution function should be smooth, and the probability density function at least continuous, for consistency with observation and theory. We can impose density continuity by introducing dependencies between parameters of the new model. Detailed model development is presented in Section 4.

4. The Weibull-generalised Pareto (WGP) model

In this section we describe the Weibull-generalised Pareto (WGP) model for the distribution of the wave height motivated above. The WGP distribution is two-part or piecewise, taking the form of a (truncated) Weibull distribution to the left and a generalised Pareto distribution to the right of some transition threshold u. The model exhibits continuous density and smooth cumulative distribution function on $(0, h_*]$. The model incorporates the Miche formulation for the upper limit, $h_{max} = h_*$.

A full description of maximum likelihood estimation of WGP parameters is given in Section 5. As motivation for the functional forms adopted for some WGP model parameters in terms of known sea state covariates (such as H_S), we also discuss elements of the inferences made here. As detailed in Section 5, during model estimation we found some of the measured data for FRF and P8 to be inconsistent with physical intuition and therefore suspect. Hence, in this section, we choose to motivate WGP model development using AWG, NC, P20 and Tank only, although all data sources are eventually considered in Section 5.

4.1. Model parameterisation

4.1.1. WGP transition threshold

The transition threshold u is chosen to be the extreme value threshold for the GP model. Exceedances of u should therefore be reasonably viewed as being extreme, and consistent with observations from an extreme value model. u therefore needs to be a relatively high quantile of the distribution of wave height. If u is set too low, we violate this extreme value assumption. If u is too high, we are left with insufficient data to estimate the GP model well. The characteristics of the GP distribution (e.g. Scarrott and MacDonald, 2012) are such that, given u is set sufficiently high that the GP distribution is an adequate description of individuals, increasing u will result in no change in ξ or h_*, and a linear trend (with gradient ξ) in σ. We therefore seek a relatively small value for u such that these tail characteristics are present. Fig. 1 illustrates this behaviour for a typical sea state of Tank data. The upper and left lower panels show GP parameter estimates as a function of u. For this particular example, any choice of $u \in (0.06, 0.09)$ m would appear reasonable, for example. A general feature of the distribution of wave height is that the value of sea state H_S lies above the mode of the distribution, on the downward tail of the density. This suggests that a linear form $u = \rho H_S$ in H_S (for some $\rho > 0$) might be a convenient choice for transition threshold u. This is particularly convenient from the perspective of establishing a "designer distribution", the parameters of which are straightforward to specify. In principle, direct estimation of factor ρ might be feasible under the WGP model, but we found this problematic. In the evaluation reported in Section 5, $\rho = 1$ is assumed.
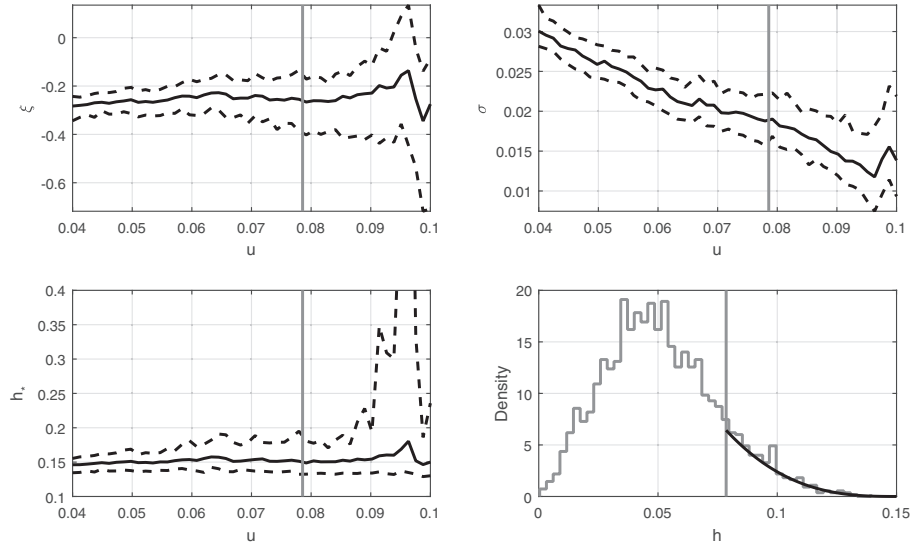
4.1.2. Weibull body

The Weibull distribution of wave height h below the transition threshold ρH_S is specified in terms of scale parameter $\mu \, (> 0)$ and shape parameter $\kappa \, (> 0)$. The cumulative distribution function takes the form

$$\Pr(H \leq h) = 1 - \exp\left[-\mu \left(\frac{h}{\rho H_S}\right)^\kappa\right] \text{ for } h \leq u \,.$$

Glukhovskiy (1966) suggests that μ is related to $\rho H_S/d$ as

$$\kappa = 2\left(1 - \phi\left(\frac{\rho H_S}{d}\right)\right)^{-1}$$

Fig. 1. Top and lower left hand panels illustrate variation of GP shape ξ, scale σ (m) and upper end point h_* (m) estimated as a function of transition threshold u (m) for a typical Tank sea state. The vertical line corresponds to H_S (m). The bottom right hand panel shows the estimated GP density (black), using sea state $u = H_S$ as transition threshold (vertical line), together with the empirical density of wave heights (grey) as functions of wave height h (m).

where function $\phi(x) = cx$ for some constant c. This yields a Rayleigh distribution as H_S/d tends to zero. We attempted to use a similar form. However, as demonstrated by the estimates in Fig. 2, we found empirically that the a curve of the form $\phi(x) = \lambda x^{1.7}$ for constant λ (> 0) is more appropriate. Note that the value of λ is estimated during maximum likelihood inference below, but that the exponent ($= 1.7$) in the functional form for ϕ is not.

4.1.3. Generalised Pareto scale

H_S is commonly used as scaling factor in distributional forms for individual wave height. With this in mind, we estimated the GP tail form for all data, and found a clear near linear relationship (see Fig. 3) between GP scale estimate and H_S corresponding to $\sigma = \alpha H_S$ with $\alpha \approx 0.22$. We adopt this functional form for σ in all subsequent inference.

4.1.4. Miche limit

Miche (1944) introduced an upper limit $h_{\max}$ for waves described in Section 3.2 with parameter β originally specified as 0.142, but estimated to be 0.12 during the first phase of the LoWiSh Joint Industry Project. When GP shape ξ is negative, the GP distribution has an upper bound $h_* = u - \sigma/\xi$. We incorporate the Miche limit within the WGP model by equating $h_{\max}$ and h_* yielding

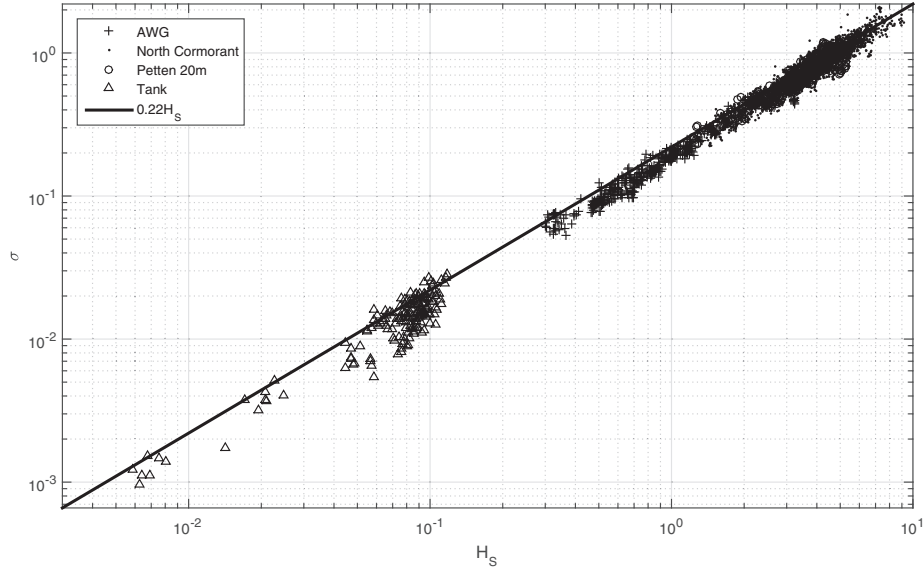$$h_* = u - \frac{\sigma}{\xi} = 2\beta\pi \frac{\tanh(k_L d)}{k_L} = h_{\max}$$

from which

$$\xi = \sigma\left(u - 2\beta\pi \frac{\tanh(k_L d)}{k_L}\right)^{-1} = \alpha\left(1 - 2\beta\pi \frac{\tanh(k_L d)}{k_L \rho H_S}\right)^{-1},$$
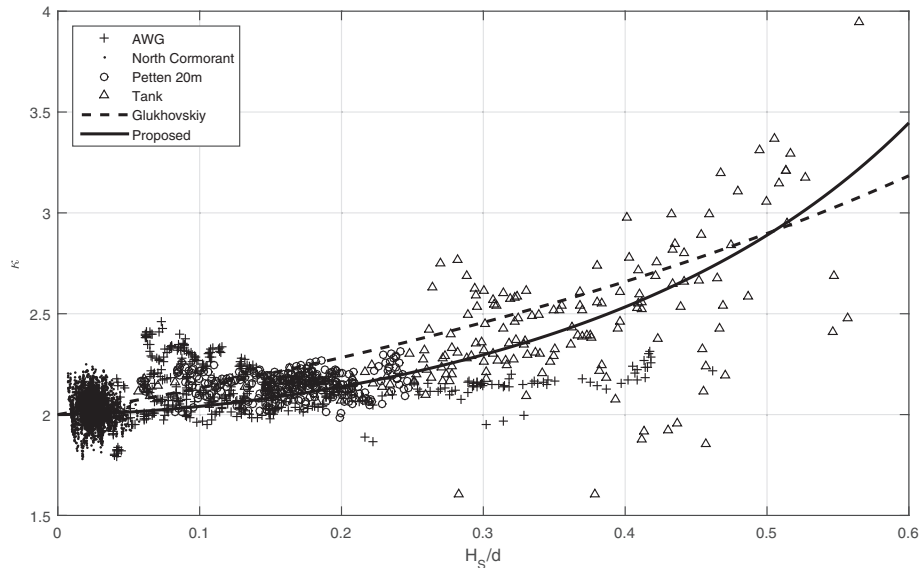
Fig. 2. Parameterisation of Weibull shape parameter in terms of H_S/d. Independent estimates from AWG, North Cormorant, Petten (20 m) and Tank sea states, together with Glukovskij (dashed) and proposed (solid) functional forms.

Fig. 3. Parameterisation of GP scale parameter in terms of H_S. Independent estimates from AWG, North Cormorant (NC), Petten (20 m, P20) and Tank sea states, together with proposed (solid) functional form $\sigma = \alpha * H_S$ with α tentatively set to a value of 0.22.

the right hand expression emerging once the parameterisation $u = \rho H_S$, $\sigma = \alpha \rho H_S$ is adopted. To impose continuity of WGP density at the transition threshold u, the Weibull density $f_W(h)$ and GP density $f_{GP}(h)$ must be equal at $h = u$. Since

$$f_W(h) = \frac{\kappa \mu}{\rho H_S} \left(\frac{h}{\rho H_S}\right)^{\kappa-1} \exp\left[-\mu\left(\frac{h}{\rho H_S}\right)^{\kappa}\right] \text{ and}$$

$$f_{GP}(h) = \frac{1}{\sigma}\left(1 + \frac{\xi}{\sigma}(h - \rho H_s)\right)^{-\frac{1}{\xi}-1} \exp[-\mu],$$

equality at $h = u = \rho H_S$, together with $\sigma = \alpha \rho H_S$ leads to the equation $\mu \alpha \kappa = 1$, enabling the parameterisation of μ in terms of α and λ (the latter via κ).

4.2. Summary of WGP model

The proposed WGP model is piecewise. The transition threshold for the two-part model is ρH_S, where ρ is set to unity for wave height in this work. The left hand side of the model ($h \in (0, \rho H_S]$) takes a Weibull form with density

$$f_W(h) = \frac{\kappa \mu}{\rho H_S}\left(\frac{h}{\rho H_S}\right)^{\kappa-1} \exp\left[-\mu\left(\frac{h}{\rho H_S}\right)^{\kappa}\right]$$

where the shape of the Weibull distribution is parameterised as

$$\kappa = 2\left(1 - \lambda\left(\frac{H_S}{d}\right)^{1.7}\right)^{-1}.$$

The right hand side ($h \in (u, h_{\max}]$) is a (conditional) GP distribution

$$f_{GP|H>H_S}(h|H > \rho H_S) = \frac{1}{\alpha \rho H_S}\left(1 + \frac{\xi}{\alpha}\frac{(h-u)}{\rho H_S}\right)^{-\frac{1}{\xi}-1}.$$

Imposing $\mu \alpha \kappa = 1$, we can parameterise the GP shape parameter ξ as

$$\xi = \alpha\left(1 - 2\beta\pi\frac{\tanh(k_L d)}{k_L \rho H_S}\right)^{-1}.$$

With prior knowledge of sea state water depth d, significant wave height H_S and wave number k_L, the WGP requires just three parameters (scale α, Miche β and Weibull shape adjustment λ) to be estimated.

5. Estimating the WGP model

In this section we report parameter estimates for the WGP model, and evaluate its performance relative to competitor models described in Section 3.

5.1. Maximum likelihood estimation

Maximum likelihood (ML) inference was used to fit the WGP model to individual sea states and sets of sea states. Estimates for WGP parameters α, β and λ for individual sea states per data source are visualised as box-whisker plots in Fig. 4. The figure indicates that estimates for α are in the region of 0.22, for β around 0.15 and for λ around unity. Comparison of observed and fitted densities and distributions for typical sea states of AWG, North Cormorant (NC), Petten 20 m (P20) and Tank are given in Fig. 5. Quantile–quantile plots (not shown) indicate reasonable estimation in general for individual sea states and combinations thereof.

We note from Fig. 4 that estimates for β from FRF and Petter 8 m (P8) are more variable than from other sources, and that very large values of β are estimated for some sea states. This is of particular interest since β locates the corresponding Miche limit. Miche (1944) suggests $\beta = 0.142$. Fig. 4 indicates that for some sea states, at FRF and P8 in particular, the estimated values of β are greater than 4. This appears implausible, as discussed below.

Fig. 4 also indicates that estimation of λ is more problematic for AWG and NC, the latter certainly due to the small values and lack of variability in H_S/d for that source. The figure further shows the corresponding parameter estimates obtained using all sea states
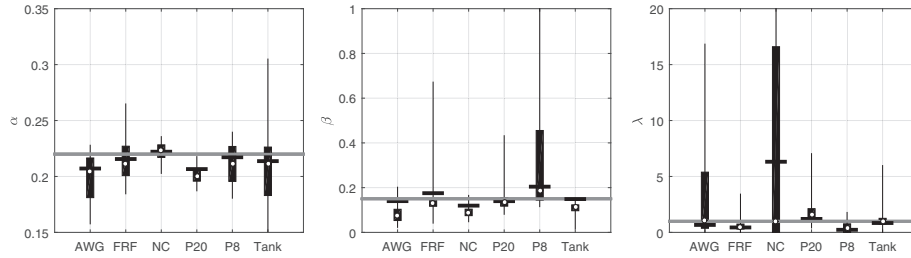
Fig. 4. Box-whisker plots for maximum likelihood estimates of WGP parameters α, σ and λ. White dots indicate the median ML estimate over all individual sea states for the given source, black boxes interquartile range, and whiskers 95% intervals (with respect to individual seas states). Short black horizontal lines indicate the ML estimate using all sea states for the given source and the grey horizontal line the proposed WGP parameter. Note that the 97.5% level for individual sea state β at P8 is truncated at unity (the 97.5% point is at approximately 2) and for individual sea sate λ at NC is truncated at 20 (the 97.5% point is at approximately 80).

(combined) per data source as short horizonal lines. These are consistent with estimates from individual sea states. Maximum likelihood estimates were generated using all sea states for AWG, P20 and Tank (i.e. sources in shallow water), and for AWG, NC, P20 and Tank (i.e. sources with no concerns about data validity). Values for the triplet (α, β, λ) of (0.21, 0.15, 0.92) and (0.22, 0.15, 0.70) were obtained respectively. A number of further estimates restricted to different domains of H_S/d were also obtained for different combinations of sources, showing general consistency with Fig. 4, with estimates for $\alpha \in (0.21, 0.22)$, $\beta \in (0.13, 0.15)$ and $\lambda \in (0.70, 1.16)$. Based on these combined inferences, we propose a parameter set $\alpha = 0.22$, $\beta = 0.15$ and $\lambda = 1.0$ for adoption in applications of the WGP model. Illustrative quality of fit to typical cases from different sources is given in Fig. 5.

Fig. 6 shows individual wave heights h against zero-crossing period for sea states with relatively high H_S/d for each data source in turn. Also shown per source is the expected level of the Miche limit h_* estimated using different assumed values of β (and $\alpha = 0.22$, $\lambda = 1$). For all sources except FRF and P8, values of h do not approach any of the Miche limits (for $\beta = 0.14$ or larger). However, for FRF and P8 there are occurrences which exceed the Miche limit for $\beta = 0.15$. One exceptional value for P8 is particularly suspicious. A similar effect is observed in Fig. 7, which shows the empirical cumulative distribution function of h_{max}/h_* per source assuming $\beta = 0.15$. For FRF and P8, there are clear exceedances of unity, implying that

a value of $\beta > 0.15$ is necessary if these observations are credible. We believe that for some sea states, surface elevations estimated from pressure signals at FRF are inaccurate and give rise to unrealistic parameter estimates. At P8, we believe that surface elevations are measured correctly, but that the sand bar on which the P8 wave staff is mounted causes wave heights to be amplified to values larger than would normally occur over a gently sloping beach, giving rise to the unusually large wave heights observed and the large estimated β.

5.2. Comparison with other models

Using the parameter values $\alpha = 0.22$, $\beta = 0.15$ and $\lambda = 1.0$, we compare the performance of the WGB model with all other models described in Section 2 in describing the empirical distribution of wave height for all individual sea states. As introduction, Fig. 8 shows Rayleigh (Rlg), Glukovskiy (Glk), Battjes–Groenendijk (BtG), Forristall (Frr), Forristall (LoWiSh, FrL), Tayfun (Tfn), Mendez–Losada–Medina (MLM) and WGP distributions for the AWG sea state reported in the top middle panel of Fig. 6. The empirical density is shown in thick grey, and the WGP in thick (dot-dashed) black. In this particular case, all model forms with the exception of MLM appear to perform approximately equally well.

For all cases for all sources, we quantify relative performance of competing distributions in terms of the Kullback–Leibler (KL)
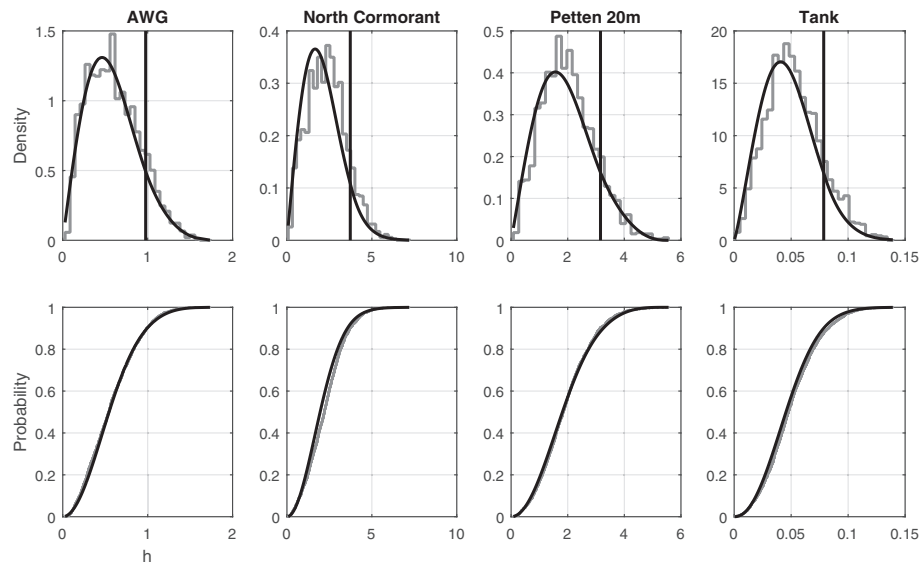
Fig. 5. Illustrative quality of fit to cases from different sources. Empirical (grey) and model (black) estimates for the probability density function (top) and cumulative distribution function (bottom) for each of AWG, NC, P20 and tank.
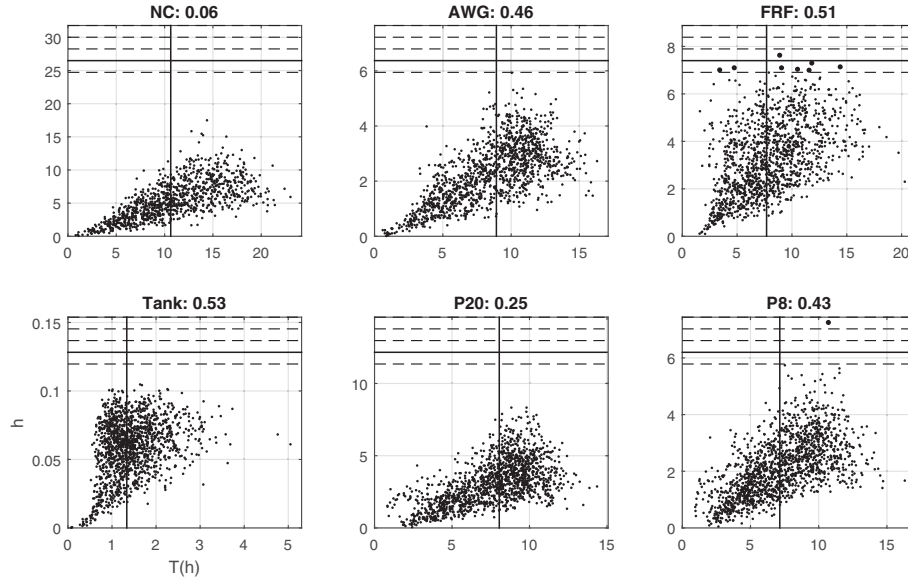
Fig. 6. Scatter plots of individual wave height (h,m) and zero-crossing period ($T(h)$, s) for sea states from each location with high H_S/d ratios (given in panel titles). The solid black vertical line represents the median zero-cross period, and the black horizontal lines corresponding Miche limits for $\beta = 0.14, 0.16, 0.17$ (dashed) and the proposed Miche limit with $\beta = 0.15$.

divergence. KL divergence quantifies the difference between a reference distribution (with cumulative distribution function F_0 and density f_0) and a test distribution (with cumulative distribution function F_1 and density f_1), using the average ratio of density functions

$$D(F_0, F_1) = \int_{-\infty}^{\infty} \log\left(\frac{f_0(h)}{f_1(h)}\right) f_0(h) dh.$$

Comparing a distribution with itself would yield a KL divergence of zero. Relatively small values of KL divergence indicate good agreement between the reference and test distributions. In this work, we use the sample approximation of Perez-Cruz (2008). Using the empirical cumulative distribution per sea state as reference,

Fig. 9 shows pairs of box-whisker plots for KL divergence, comparing whole-distribution (left, black) and tail estimates (right, grey) from different model forms with the empirical estimate per data source. White dots indicate the median KL estimate over all individual sea states, black boxes interquartile range, and whiskers 95% intervals. The tail is defined as exceedances of the 70th percentile of the sample for each sea state.

Inspection of Fig. 9 suggests that several of the model forms do well for various data sources, though emphasis should be given to the performance of the models for the NC, AWG, Tank, and P20 data sets. The Forristall (1978) distribution (Frr) was developed empirically for extreme deep-water sea states in the Gulf of Mexico. Not surprising therefore is its good skill for the deep-water North Cormorant (NC)
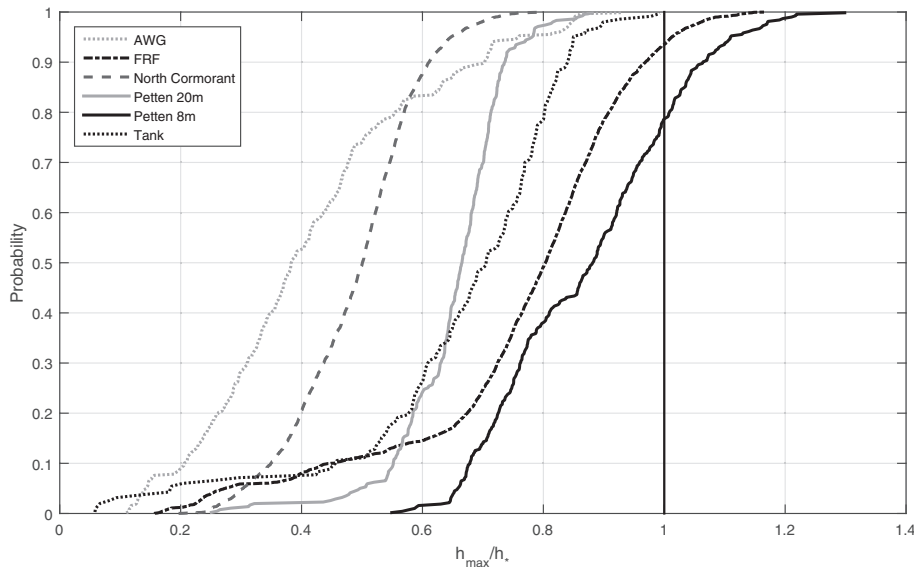
Fig. 7. Distribution of observed maxima h_{max} relative to the proposed Miche limit h_* with $\beta = 0.15$.
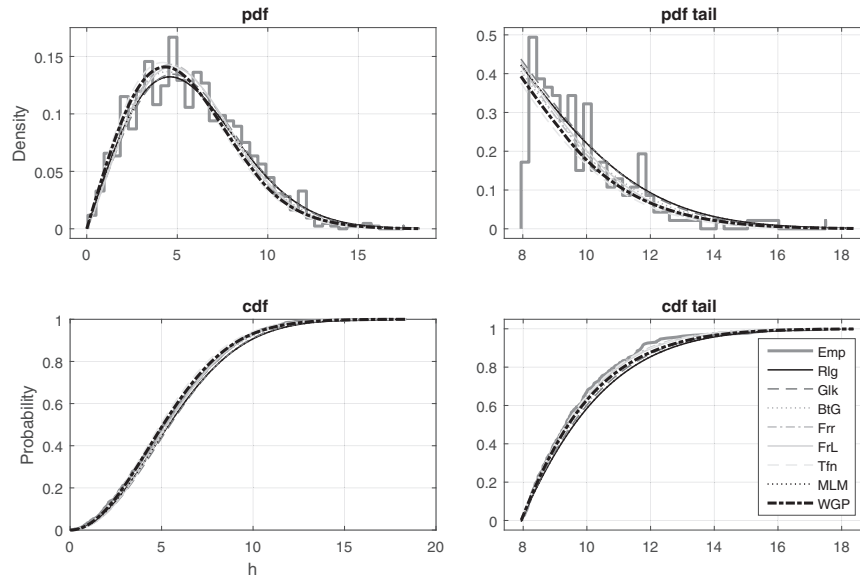
Fig. 8. Illustration of model comparison of Rayleigh (Rlg), Glukovskiy (Glk), Battjes–Groenendijk (BtG), Forristall (Frr), Forristall (LoWiSh, FrL), Tayfun (Tfn), Mendez–Losada–Medina (MLM) and WGP for the AWG sea state illustrated in the top middle panel of Fig. 6. The empirical density is shown in thick grey, and the WGP in thick (dot-dashed) black.

data set; but surprisingly, it also performs well for the intermediate and shallow-water field data recorded at P20 and AWG respectively. Its poor performance for Tank might be explained by the fact that the laboratory measurements are unidirectional, whereas the Forristall model was calibrated using directional field measurements. The Battjes and Groenendijk (2000) distribution (BtG) was developed for laboratory measurements, and as might be expected performs well for the Tank data set. It does not perform so well for the shallow water field data sites P20 and AWG. The good skill of the

BtG model found for the NC data set can be attributed to how well the lower Weibull of BtG describes deep-water waves in general, as the transition height to the upper Weibull was always larger than the maximum in the respective records. Given that the WGP distribution was developed from fits to the NC, AWG, Tank, and P20 data sets, it should not be surprising that it has the best overall skill for all of these data sets, performing well for both deep, intermediate and shallow water field data sets and the shallow-water Tank data set. At the same time, we note that a sample of nearly 5000 sea states
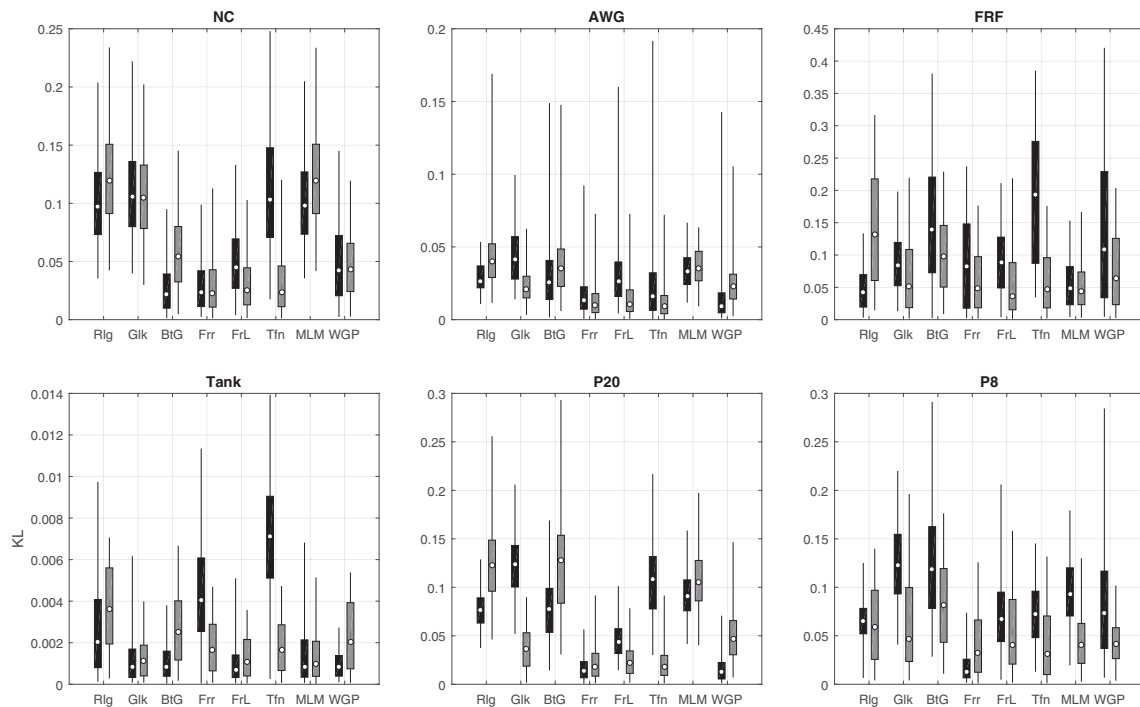
Fig. 9. Pairs of box-whisker plots for Kullback–Leibler divergence (KL) estimates, comparing whole-distribution (left, black) and tail estimates (right, grey) from different model forms with the empirical estimate per data source. White dots indicate the median KL estimate over all individual sea states, black boxes interquartile range, and whiskers 95% intervals (with respect to individual seas states). The tail is defined as exceedances of the 70th percentile of the sample for each sea state.

corresponding to different conditions from four different field and laboratory sources was used to estimate the WGP model, the number of independent observations for model estimation far outweighing the number of model degrees of freedom.

6. Discussion and suggestions for further investigation

The proposed three-parameter Weibull-generalised Pareto (WGP) form for the distribution of wave height in shallow water offers flexibility as a two-part model with continuous density, incorporating both physical and statistical knowledge. The WGP form describes the distributions of wave height in the sample data examined well. These samples include deep, intermediate and shallow water field sources and shallow water tank data. Overall, the proposed WGP performs at least as well as, and typically better than, competitors for these data sets. This is in part not surprising given that relationships between model parameters and sea state covariates, and parameter estimates were inferred using the same data. The estimate for the "Miche β''" parameter is in good agreement with the value originally proposed by Miche (1944).

In this work, we use the median of measured individual wave numbers to represent the sea state wave number k_L in the Miche expression. We have examined the effect of adopting different quantiles of the distribution of individual wave number as alternatives to the median, and find that parameter estimates for the WGP model are relatively insensitive to choices of k_L from the body of this distribution. If, in application, the median of individual wave numbers is not available, we propose it be approximated using the wave number k_P corresponding to peak wave spectral frequency.

Estimates (e.g. for design values) obtained from WGP or any competitor will be uncertain due to the inherent difficulty of modelling unusual events in a sample. The quality of fit of the WGP model to the sample tail (specifically) is generally good. In the exceptional poorly-fitting cases, there are two noteworthy causes of discrepancy. Firstly, the WGP form incorporates the Miche limit explicitly, and therefore can be thought of as corresponding to a breaking wave distribution. However, we have not limited our analysis to sea states showing clear examples of breaking waves due to depth limitation. Such a specification might lead to an improved estimated

of the breaking wave height distribution. Secondly, and possibly related at least for non-saturated sea states, is that tail characteristics (in particular) are influenced by sea state characteristics in addition to H_S, k_L and d. Identifying and including these might improve overall model performance. It is clear that available cases from the laboratory tank experiments are not fully representative; notably, the number of available cases corresponding to truly shallow water, from tank (or indeed from field measurements) is small, of the order of 30 cases). Yet these are exactly the cases for which a more sophisticated parameterisation of the WGP model appears possible, and those which would be most informative about the value of "Miche β''". Further experimental work (particularly tank-based) following a careful statistical experimental design to explore parameter space efficiently and sufficiently, would seem to be justifiable based on the evidence from this study.

It is interesting to examine sea states where there might be an appreciable degree of breaking. Mendez et al. (2004) provide an expression for the fraction N_{br} of breaking waves in a sea-state in terms of the dimensionless ratio κ of root-mean-square wave height H_{rms} to maximum wave height H_{max} introduced in Section 3.1

$$\log(N_{br}) = \left(1.033 + 0.297\kappa - 3.816\kappa^2 + 2.517\kappa^3\right)\log(\kappa).$$

Using this expression we estimated the fraction of breaking waves in all sea states for all sources. We found that not one of the sea states corresponding to field data from any field location had $N_{br} > 0.35$. For the Tank data, 8 sea states were identified with $N_{br} > 0.35$. For these sea states only, the performance of all models was evaluated and is shown in Fig. 10. From the figure, we observe that the models of Battjes–Groenendijk(BtG) and WGP are best performing for the whole distribution, and that Glukovskiy (Glk) and Forristall (LoWiSh, FrL) models are best-performing for the tail. We caution however that the number of sea states used in this comparison is small, that the extent of breaking in these sea states is also relatively small, and that all of Glk, BtG, FrL, Mendez–Losada–Medina(MLM) and WGP provide reasonable fits to the whole distribution and the tail.
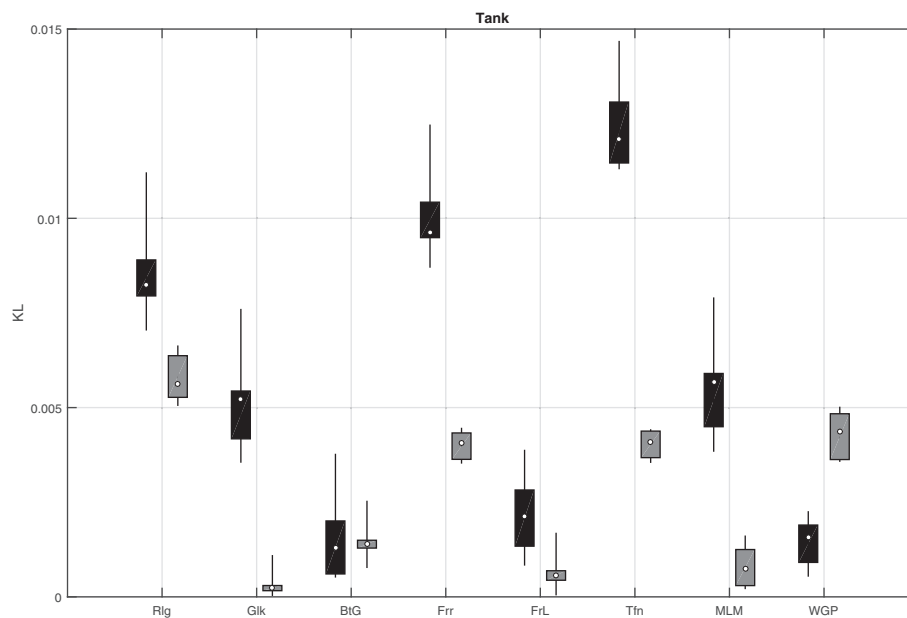
Fig. 10. Pairs of box-whisker plots for Kullback–Leibler divergence (KL) estimates, comparing whole-distribution (left, black) and tail estimates (right, grey) from different model forms with the empirical estimate for Tank data with $N_{br} > 0.35$. White dots indicate the median KL estimate over all individual sea states, black boxes interquartile range, and whiskers 95% intervals (with respect to individual seas states). The tail is defined as exceedances of the 70th percentile of the sample for each sea state.

Controlled laboratory conditions are expected to produce cleaner data than possible in from field conditions. However, there is uncertainty as to just how well laboratory conditions represent sea states in the real ocean. For example, attempts to reproduce directional spreading and wind forcing in laboratory basins are fairly recent additions to experimental programmes; our tank data from Imperial College London were long-crested and without wind. The advantages of adding directional spreading and even wind to the parameter space for future laboratory experiments are likely to outweigh the added cost and complexity in the experimental programme and provide valuable enhancements to the description of the distribution of ocean wave height.

A number of more sophisticated model forms are available in the literature, which might be appropriate enhancements to the current WGP model, especially as larger more informative samples from the field and laboratory become available. Examples include the work of Frigessi et al. (2002), Behrens et al. (2004), MacDonald et al. (2011) and Randell et al. (2015). At the current time, however, given sample quality and the need for a simple "designer" distribution for straightforward application, we judge the proposed WGP model fit for purpose.

7. Acknowledgements

References

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

Behrens, C.N., Lopes, H.F., Gamerman, D., 2004. Bayesian analysis of extreme events with threshold estimation. Stat. Model. 4, 227–244.

Birkemeier, W.A., Long, C.E., Hathaway, K.K., 1997. DELILAH, DUCK94 and SandyDuck: Three nearshore experiments. ASCE Proc. 25th International Conference on Coastal Engineering.

Boccotti, P., 1989. On mechanics of irregular gravity waves. Atti Acc. Naz. Lincei, Mem. VIII 19, 111–170.

Coles, S., 2001. An Introduction to Statistical Modeling of Extreme Values. Springer ISBN 1852334592

Ewans, K., Feld, G., Jonathan, P., 2014. On wave radar measurement. Ocean Dyn. 64, 1281–1303.

Forristall, G.Z., 1978. On the statistical distribution of wave heights in a storm. J. Geophys. Res. 83, 2353–2358.

Forristall, G.Z., 2007. The distribution for wave height in shallow water estimated from laboratory measurements. Unpublished (developed and reported as part of LoWiSh Phase 1).

Frigessi, A., Haug, O., Rue, H., 2002. A dynamic mixture model for unsupervised tail estimation without threshold selection. Extremes 5, 219–235.

Glukhovskiy, B.K.h., 1966. Investigation of sea wind waves (in Russian), Leningrad. Proc. Sea Climatology Conf. 51–71.

Guedes Soares, C., 2003. Probabilistics Models of waves in the coastal zone. In: Lakhan, V.C. (Ed.), Advanced in Coastal Modelling. Elsevier,

Hordijk, D., 2003. Report on Field Measurements Petten Sea Defense: Storm Season 2002–2003. Rijkswaterstaat Report RIKZ/OS/2003.135X,

Hordijk, D., 2004. Report on Field Measurements Petten Sea Defense: Storm Season 2002–2003. Rijkswaterstaat Report RIKZ/OS/2004.123W,

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. Coast. Eng. 73, 43–57.

Katsardi, V., Swan, C., 2011. Omae2011-49957: an experimental study of shallow water wave statistics on mild bed slopes. Proc. 30st Int. Conf. of the OMAE, Rotterdam, NL.

Katsardi, V., Swan, C., 2011. The evolution of large non-breaking waves in intermediate and shallow water. I. Numerical calculations of uni-directional seas. Proc. R. Soc. A 467, 778–805.

Longuet-Higgins, M.S., 1952. On the statistical distribution of the height of sea waves. J. Mar. Res. 11, 245–266.

LoWiSh Joint Industry Project, Limits on waves in shallow water, http://www.ogp.org/metocean/JIPs#461233-lowishlimits-on-waves-in-shallow-water.

MacDonald, A., Scarrott, C.J., Lee, D., Darlow, B., Reale, M., Russell, G., 2011. A flexible extreme value mixture model. Comput. Stat. Data Anal. 55, 2137–2157.

Mai, S., Wilhelmi, J., Barjenbruch, U., 2011. Wave height distributions in shallow waters. Coast. Eng. Proc. 1, 63

Mendez, F.J., Losada, I.J., Medina, R., 2004. Transformation model of wave height distribution on planar beaches. Coast. Eng. 50, 97–115.

Miche, R., 1944. Mouvements ondulatoires de la mer en profondeur constante ou decroissante forme limite de la houle lors de son deferlement. Ann. Ponts Chaussees 285–319.

Perez-Cruz, F., 2008. Kullback–Leibler divergence estimation of continuous distributions. Information Theory, 2008. ISIT 2008. IEEE International Symposium on. pp. 1666–1670.

Randell, D., Turnbull, K., Ewans, K., Jonathan, P., 2015. Bayesian inference for non-stationary marginal extremes. (Submitted to Environmetrics August 2015, draft at www.lancs.ac.uk/~jonathan),

Scarrott, C., MacDonald, A., 2012. A review of extreme value threshold estimation and uncertainty quantification. Revstat 10, 33–60.

Tayfun, M.A., 1990. Distribution of large wave heights. J. Waterw. Port Coast. Ocean Eng. 116, 686–707.

Tayfun, M.A., Fedele, F., 2007. Wave-height distributions and nonlinear effects. Ocean Eng. 34, 1631–1649.

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