**Natural Hazards
and Earth
System Sciences**

Dynamical and statistical explanations of observed occurrence rates of rogue waves

J. Gemmrich and C. Garrett

Department of Physics and Astronomy, University of Victoria, Victoria, BC V8W 3P6, Canada

Abstract. Extreme surface waves occur in the tail of the probability distribution. Their occurrence rate can be displayed effectively by plotting $\ln(-\ln P)$, where P is the probability of the wave or crest height exceeding a particular value, against the logarithm of that value. A Weibull distribution of the exceedance probability, as proposed in a standard model, then becomes a straight line. Earlier North Sea data from an oil platform suggest a curved plot, with a higher occurrence rate of extreme wave and crest heights than predicted by the standard model. The curvature is not accounted for by second order corrections, non-stationarity, or Benjamin-Feir instability, though all of these do lead to an increase in the exceedance probability. Simulations for deep water waves suggest that, if the waves are steep, the curvature may be explained by including up to fourth order Stokes corrections. Finally, the use of extreme value theory in fitting exceedance probabilities is shown to be inappropriate, as its application requires that not just N, but also $\ln N$, be large, where N is the number of waves in a data block. This is unlikely to be adequately satisfied.

1 Introduction

The media often present accounts of "rogue" waves, though the meaning of the term is usually left undefined. In some situations, any large surface wave is described as a rogue. More technically, the term is reserved for those waves in the tail of the probability distribution, exceeding the average by some prescribed multiple. Greater scientific interest comes from the question as to whether there are more abnormally large waves in a given sea state than predicted on the basis

of "simple" physics and statistics, and if so, what is the more complicated physics or statistics involved.

The topic has been discussed in numerous papers, with recent reviews and summaries in the proceedings edited by Müller and Henderson (2005), the review paper by Dysthe et al. (2008) and the book by Kharif et al. (2009). Garrett and Gemmrich (2009) is a brief summary. Some of the questions that arise in the analysis of wave data are:

1. What probability distribution functions (pdfs) for crest height or trough to crest wave height should be used for comparison with data?

2. How should these pdfs be used?

3. What physical or statistical effects can lead to a higher occurrence rate of large waves than expected from simple theories?

4. Is generalised extreme value theory (e.g., Coles, 2001) useful in examining the infrequent events of large amplitude that occur in the tail of the pdfs?

The purpose of this short note is to examine these questions, partly motivated by results cited by Dysthe et al. (2008).

2 Theoretical probability distribution functions

2.1 Linear theory

If the sea surface height ζ is made up of a large number of independent sinusoids, then its probability density function $p(\zeta)$ is Gaussian with

$$p(\zeta) = (2\pi\sigma^2)^{-1/2}\exp\left(-\frac{\zeta^2}{2\sigma^2}\right) \tag{1}$$

where σ^2 is the variance $\overline{\zeta^2}$ of the surface elevation.

We are interested here in the height of crests, or the wave height H from trough to crest (just twice the crest height for a plane wave in linear theory). For a sea of random surface waves distributed over a very narrow frequency band, Longuet-Higgins (1952) showed that H has a probability distribution given by the Rayleigh formula

$$p(H) = \frac{4H}{H_s^2} \exp\left(-\frac{2H^2}{H_s^2}\right) \qquad (2)$$

where $H_s = 4\sigma$ is the "significant wave height". The exceedance probability for H/H_s is

$$P(H/H_s > z) = \exp(-2z^2) \qquad (3)$$

The corresponding exceedance probability for crest heights is

$$P(\eta/H_s > z) = \exp(-8z^2) \qquad (4)$$

Naess (1985) has shown that for a finite but still narrow bandwidth the pdf of wave heights is still given by a Rayleigh distribution, but with H_s in Eq. (2) reduced by an amount related to the bandwidth. Thus large wave heights become less likely as the bandwidth increases, though the distribution of crest heights is unaffected. The topic is discussed in detail by Casas-Prat and Holthuijsen (2010).

2.2 Second order effects

In deep water, a single plane wave is distorted at finite amplitude. The second order correction in the Stokes expansion (e.g., Dean and Dalrymple, 1991) gives a surface elevation

$$\zeta = a\cos\theta + \frac{1}{2}ka^2\cos(2\theta) \qquad (5)$$

where a is the linear wave amplitude and θ is the wave phase $kx - \omega t$, with k, ω the wavenumber and frequency, and x, t the space coordinate and time.

Allowing for the second order correction does not affect the trough-to-crest height for a narrow-band spectrum (though it may do so slightly for a finite bandwidth), but does modify the distribution of crest heights. Forristall (2000) reviewed studies by Tayfun (1980) and others that lead to an expected exceedance probability for the crest height η given by

$$P(\eta/H_s > z) = \exp\left\{-\frac{8}{R^2}\left[(1+2Rz)^{1/2} - 1\right]^2\right\} \qquad (6)$$

where R is the measure kH_s of the steepness of the waves, with k the wavenumber. The derivation of Eq. (6) is simple. From Eq. (5), the full crest elevation has

$$\frac{\eta}{H_s} = \frac{\eta_l}{H_s} + \frac{1}{2}R\left(\frac{\eta_l}{H_s}\right)^2 \qquad (7)$$

with η_l the crest height according to linear theory, so that

$$\frac{\eta_l}{H_s} = \frac{(1+2R\eta/H_s)^{1/2} - 1}{R} \qquad (8)$$

leading to Eq. (6) if η_l has the exceedance probability given by Eq. (4); for the nonlinear η/H_s to exceed a particular value, the linear η_l/H_s must exceed the value given by Eq. (8).

Forristall (2000) also cited the formula

$$P(\eta/H_s > z) = \exp\left[-8z^2\left(1 - \frac{1}{2}Rz\right)^2\right] \qquad (9)$$

of Kriebel and Dawson (1993). Expansion in powers of Rz of the exponent in Eq. (9) agrees with that for Eq. (6) in the $(Rz)^0$ and $(Rz)^1$ terms, but the formulae differ significantly for relevant values of R and z.

The correct wavenumber k to use in the parameter $R = kH_s$ is uncertain. For the narrow band spectrum assumed in the derivation of the underlying Rayleigh distribution, k could perhaps be that at the spectral peak. For a realistic spectrum, Forristall (2000) cites D. L. Kriebel for the suggestion that k be taken to correspond to waves with a period 0.95 times the period of the waves at the spectral peak, so that k is 1.11 times its values at the peak, but this is clearly somewhat arbitrary. Further, it is the local wavenumber that is relevant, and this may be smaller than average for large waves (Gemmrich and Garrett, 2008); i.e., large waves tend to be longer than average. Simulations are clearly required and will be discussed later.

The discussion here is on deep water waves, but Forristall (2000) cites Eq. (9) for the result that, in water of depth d, the steepness parameter R should simply be replaced by $Rf(kd)$, where

$$f(u) = \frac{\cosh u\,(2+\cosh 2u)}{2\sinh^3 u} - \frac{1}{\sinh 2u} \qquad (10)$$

2.3 Presentation

Data on wave height or crest height exceedance may be compared with formulae (3) and (6) or other formulae, but the large waves then correspond to small values of P at the tail end of the distribution. As in Dysthe et al. (2008), it makes for a better presentation of simulations or data to plot $\ln(-\ln P)$ which, for Eq. (4), is

$$\ln[-\ln P(\eta/H_s > z)] = 2\ln z + \ln 8 \qquad (11)$$

and hence a straight line if plotted versus $\ln z$.

Based on simulations for exceedance probabilities down to 10^{-4}, Forristall (2000) suggested the use of the Weibull distribution (we use the notation of Dysthe et al. (2008) here)

$$P(\eta/H_s > z) = \exp\left(-\frac{z^\alpha}{\beta}\right) \qquad (12)$$

and similarly for $P(H/H_s > z)$, with the values of α and β depending on the sea state. This leads to

$$\ln[-\ln P(\eta/H_s > z)] = \alpha\ln z - \ln\beta, \qquad (13)$$

again straight lines if plotted versus $\ln z$. We compare this later with our own simulations for realistic wave conditions, but first apply the transformation $\ln[-\ln P(\eta/H_s > z)]$ to Eq. (6). The curves in Fig. 1 for various values of R are approximated remarkably well by straight lines and we note that this remains the case even if R is increased by the factor in Eq. (10) to allow for finite depth.

We note that the ordinate y of Fig. 1 is simply related to the exceedance probability by the transformation $P(\eta/H_s > z) = \exp(-\exp y)$, with P decreasing as y increases. The vertical distance between two curves in such a plot gives the difference in exceedance probabilities for a specified crest height, whereas the horizontal difference gives the difference in the crest height expected for a specified exceedance probability.

A simple summary of the two effects of finite frequency bandwidth and second-order effects is that the former moves the plot for wave height H above the Rayleigh line but does not affect the plot for η, whereas the latter does not affect the plot for H but moves the plot for η below the Rayleigh line.

The plot of Eq. (9) in Fig. 1, for a reasonable value of R, shows that it may also be approximated well by a straight line for exceedance probabilities P greater than about 10^{-4}, as suggested by Forristall (2000). However, over the range of P from, say, 10^{-2} to 10^{-4}, it departs slightly from the Tayfun formula for the same value of R. More importantly, the plot of Eq. (9) curves down slightly for small values of P, predicting more frequent large waves than Eq. (6). It will be important to check this against numerical simulations but first the results from data will be discussed.

3 Comparison with data

Dysthe et al. (2008) and other authors have compared observations with theoretical expectations, though usually with data sets that are too small to examine waves occurring with a frequency of less than, say, one in 10^5 or so. For example, the data obtained from the Marlin oil platform in the Gulf of Mexico during hurricane Ivan in 2004, and examined by Dysthe et al. (2008), only deal with scaled wave height exceedance probabilities greater than 10^{-4}. Most recently, Casas-Prat and Holthuijsen (2010) examined 10^7 waves recorded by buoys in deep water off the coast of Spain. They found that wave height exceedance probabilities were well accounted for by the Rayleigh distribution modified for finite frequency bandwidth, but the crest height exceedance probabilities were only very slightly higher than predicted by linear theory. Casas-Prat and Holthuijsen (2010) suggested that the absence of more pronounced nonlinear effects was a consequence of the hydrodynamic response of the buoys.

Casas-Prat and Holthuijsen (2010) did find more pronounced nonlinear effects in the crest height exceedance probability obtained from laser altimetric observations of 10^4 waves at a North Sea oil platform. Casas-Prat and Holthui-
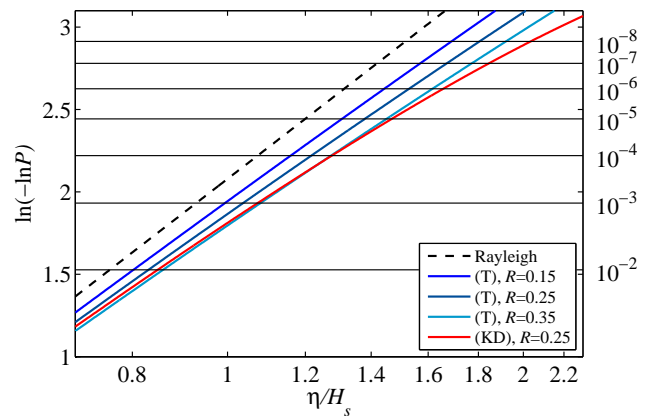

Fig. 1. The function in Eq. (6), demoted by (T), for various values of R and the function in Eq. (9), denoted by (KD), for $R = 0.25$. The dashed line corresponds to the Rayleigh distribution (equivalently $R = 0$). The horizontal lines correspond to different values of the exceedance probability P.

jsen (2010) found that a Rayleigh distribution still fits the data well, i.e. they would still choose $\alpha = 2$ in Eq. (12), but β is increased by 25%. With only 10^4 waves, this is probably reasonably consistent also with the Weibull distribution preferred by Forristall (2000) and shown in Fig. 1 here, using a moderate value for the wave steepness, but we will not pursue further comparison.

More provocative and unusual data from a laser altimeter at a North Sea oil platform were presented by Dysthe et al. (2008) and are reproduced here in Fig. 2. The exceedance probability for individual waves was obtained by Dysthe et al. (2008) from the block maxima data by evaluating the probability F that $\eta_{max}/H_s < z$. Then

$$P(\eta/H_s > z) = 1 - F^{1/N} \tag{14}$$

where N is the number waves in a block. For probabilities greater than 10^{-5} or so, the data for the crest height are slightly to the right of the Rayleigh curve, as expected from second order effects for waves with a small steepness and slightly to the left of the Rayleigh line for wave height H, as expected for finite frequency bandwidth. However, the data show a pronounced curvature of the exceedance probability plot for probabilities less than approximately $10^{-4.5}$, particularly for the crest height η. As pointed out by Dysthe et al. (2008), this behaviour is inconsistent with standard models. If the data are reliable, they suggest a dramatic increase in the frequency of occurrence of rare large waves, or in the size of waves occurring with a frequency of less than about one in 30 000 (typically once every 3 days). The water depth at Gorm is 40 m, but even taking the wavenumber k to correspond to waves of 10 second period, this gives $kd = 1.72$ and the function f from Eq. (10) is 1.22, indicating a slight
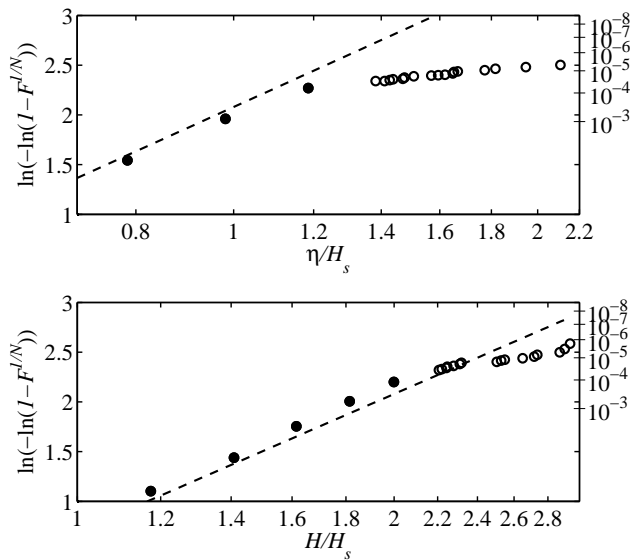
Fig. 2. Data from the Gorm oil field in the central North Sea, re-plotted from Dysthe et al. (2008). As in the figure in Dysthe et al. (2008), the filled circles are representative points, whereas the open circles represent individual records. Here, the dashed lines show the Rayleigh distributions Eqs. (4) and (3) for crest and wave heights respectively. Values of the exceedance probability P are shown on the right hand axis.

shift to the right of the exceedance probability plot, but not a dramatic change in slope at small exceedance probabilities.

Other data obtained using laser altimetry on a North Sea oil platform were reported by Stansell (2004). He examined 354 000 waves observed during stormy periods at the Alwyn North field in a water depth of approximately 130 m and also found a probability of occurrence of rogue waves, with $H \geq 2H_s$, greater than predicted by the Rayleigh formula.

The curvature in the exceedance probability plots shown in Fig. 2 and the associated increased probability of large waves above that predicted by the standard formulae of Forristall (2000) clearly needs further observational confirmation or refutation, but it raises the key question as to what could cause the curvature. We thus proceed to examine the implications of various simulations and statistical considerations.

4 Comparison with simulations

4.1 Second order simulations

As mentioned earlier, Forristall (2000) undertook simulations of typical wave spectra with accurate implementation of second-order interactions. He found that, as expected, crest heights increase over those given by linear theory, slightly

less if the directional spread of the waves is allowed for than if not. Forristall (2000) found that the Kriebel and Dawson (1993) Formula (9) compares well with data and simulations in water that is deep enough to not significantly influence the waves, but gives too great an enhancement of crest heights in shallow water. This agreement of simulations with Eq. (9) is perhaps surprising given that, as shown in Fig. 1, Eq. (9) differs slightly from the supposedly more accurate Eq. (6) for the same value of R.

More importantly, the analysis of Forristall (2000) was for exceedance probabilities no smaller than 10^{-4}, and we see from Fig. 1 that for smaller exceedance probabilities the Kriebel and Dawson (1993) formula departs significantly from the straight line of Eq. (13). It seems worthwhile to check from further simulations whether a plot of $\ln(-\ln P)$ versus $\ln(\eta/H_s)$ is well approximated by a straight line or whether it curves down for small values of P.

We confine our attention to deep water simulations and a simple implementation of the second-order crest height enhancement. This ignores the set-down under wave groups that is included in the accurate simulations of Forristall (2000) but captures the main effect of second order interactions. As in Gemmrich and Garrett (2008, 2010), we have performed simulations using JONSWAP spectra with the peak enhancement factor γ equal to 1, 2, and 3.3, though we focus on $\gamma = 1$ corresponding to the fully-developed seas that are likely to be of the greatest concern. The simulations are based on the Matlab Toolbox "Wave Analysis for Fatigue and Oceanography" (WAFO Group, 2000), which takes random spectral components with sine and cosine terms that are independent and Gaussian. To have reasonably reliable statistics, for each situation we generate 275 time series, each 60 days long, with a 10 s peak period (though this is scaleable so that our results are independent of the choice) and take 10 samples per second. In using Eq. (5), the local wave amplitude a is taken as the crest or trough height from the linear simulations and the wavenumber k is calculated using the linear dispersion relation and a period defined as the time between successive zero up-crossings.

Ideally in our simulations we should not start with waves from a JONSWAP spectrum but with a spectrum that only becomes the desired JONSWAP spectrum after the addition of the local second harmonic. However, we have found that the change in spectral shape and level from the addition is small, so we have ignored the problem and not carried out the iteration that would be required to correct it.

The JONSWAP spectrum with $\gamma = 1$ with a 10 second peak period has $H_s = 3.95$ m and $R \equiv kH_s = 0.18$ if, as earlier, we base the wavenumber on the linear dispersion relation and a period of 0.95 times the peak period. To allow for different values of R, we have also conducted simulations with H_s 0.5, 0.8, 1.3, 1.7 times the reference value, giving $R = 0.09, 0.14, 0.23, 0.30$. The resulting exceedance probabilities are shown in Fig. 3.

Nat. Hazards Earth Syst. Sci., 11, 1437–1446, 2011

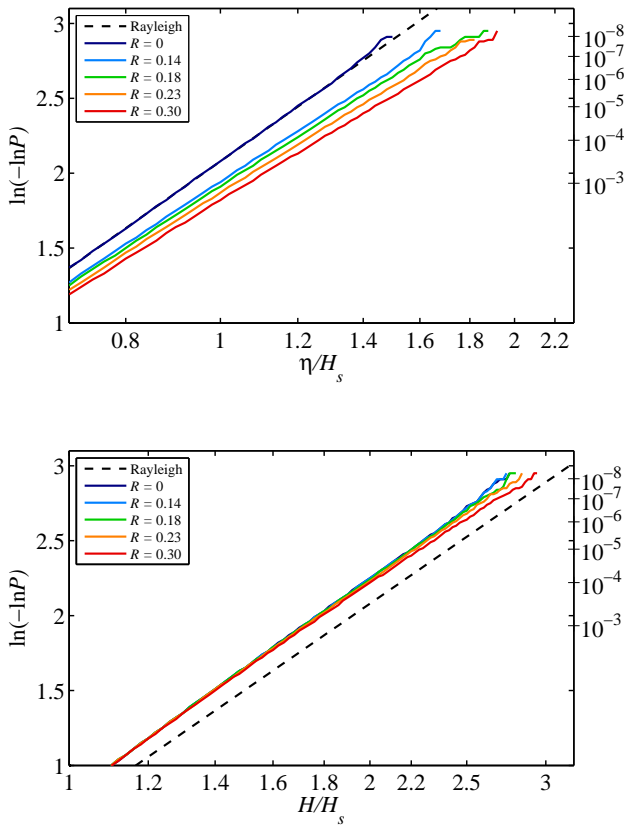www.nat-hazards-earth-syst-sci.net/11/1437/2011/

Fig. 3. The crest height exceedance probability for a simulated JONSWAP sea with $\gamma = 1$ and correction for the second harmonic, with various values of the representative steepness $R = 1.11k_pH_s$. The dashed line corresponds to the Rayleigh distribution, effectively $R = 0$. In this and subsequent figures, values of $\ln(-\ln P)$ are shown on the left hand axis and values of P itself on the right hand axis.

We see that, as for the idealised case discussed earlier, the plots for crest height are close to the straight lines implied by Eq. (13) for probabilities down to 10^{-7}, providing general support for the suggestion by Forristall (2000) but casting doubt on the usefulness of Eq. (9). In particular, it seems that second-order effects cannot account for any downward curvature seen in the exceedance probability plot of data at large values of η/H_s. We also find that the crest height exceedance probability for a purely linear sea (effectively $R = 0$) matches the Rayleigh distribution very well. Thus an observed difference from the Rayleigh distribution must be a consequence of nonlinearities or other effects beyond simple theory.

For the scaled crest height, H/H_s, the exceedance probability displayed in the lower panel of Fig. 3 shows only a very weak dependence on wave steepness, as expected if changes in the crest and trough height very nearly cancel each other. The offset of the plots from the Rayleigh line is the expected effect of finite frequency bandwidth.

4.2 Fourth order simulation

The second-order correction may, as discussed above, be implemented accurately, as by Forristall (2000), or through a local approximation as used here. Higher order corrections to the wave field may also be made using a local description of the wave field, i.e., assuming that a local wave crest is part of a uniform wave train. Dawson (2004) added up to fifth order corrections using this approach. Here we add second, third, and fourth order corrections to the crest height, taking the starting point as the linear simulation. Hence the corrected crest height is taken as

$$\eta = a(1 + \frac{1}{2}ka + \frac{3}{8}k^2a^2 + \frac{25}{24}k^3a^3) \tag{15}$$

where the wave amplitude a and the wavenumber k are taken, as before, from the linear simulations and the linear dispersion relation using the local period. We are thus ignoring the change in the dispersion relation for a regular wavetrain of nonlinear waves but assuming that, within the local wave, the amplification of the wave crest height, even to fourth order, occurs as for a regular wavetrain. Our simulations should thus be regarded as exploratory rather than definitive. The corrections to the displacement of the trough have opposite signs for the second and fourth term in brackets in Eq. (15), so that, at least for a narrow-band spectrum with crest and trough amplitudes nearly equal, the wave height is increased by a factor $(1 + \frac{3}{8}k^2a^2)$.

The simulation results are shown in Fig. 4 and correspond to Fig. 3 but with the addition of higher harmonics. We see that this addition of higher harmonics leads to downward curvature of the plots at large values of η/H_s. This suggests that any similar curvature observed in data sets may be a consequence of the higher harmonics. In particular, the Gorm data shown in Fig. 2 have exceedance probability plots for both crest and wave heights that curve over at a point comparable to that expected for a steepness R of approximately 0.25 in the simulation results shown in Fig. 4. This interpretation is weakened, however, by the small offset from the Rayleigh line shown by the Gorm data in the higher probability sections of the plots that are nearly straight lines. A false offset could be produced by the conversion from block maxima having used an inappropriate value of the number N of waves per block, but there is no reason to expect this. In any event, it is important to examine other possible causes of the downward curvature in the plots.

5 Benjamin-Feir instability

Mori and Janssen (2006) have suggested that the enhanced occurrence of large waves is related to the fourth-order cumulant κ_{40} of the surface elevation, related to the kurtosis μ_4 by $\kappa_{40} = \mu_4 - 3$. They show that κ_{40} has only a small contribution from the non-resonant Stokes harmonics that we
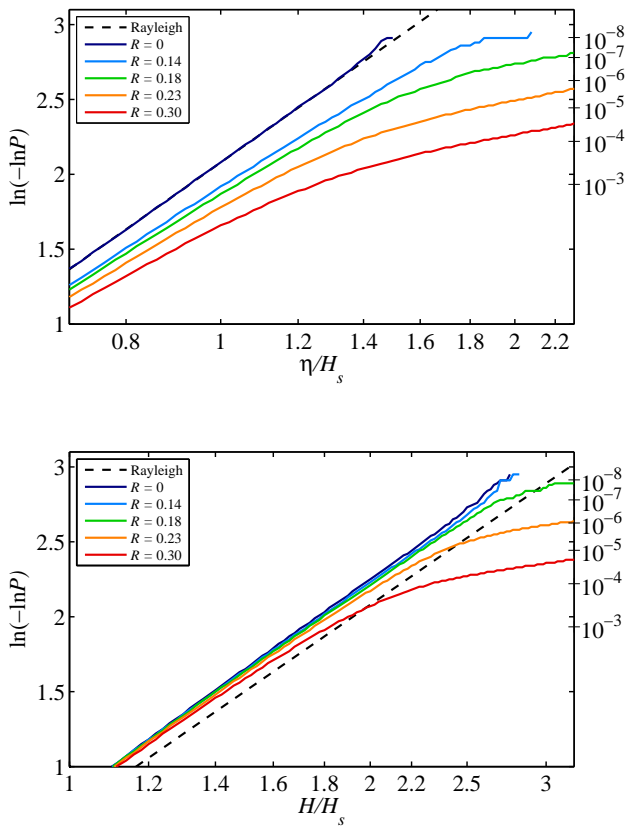
Fig. 4. The crest height exceedance probability for a simulated JONSWAP sea with $\gamma = 1$ and correction for the second, third, and fourth harmonics, for various values of R as in Fig. 3. The dashed line corresponds to the Rayleigh distribution.

have considered so far but may have a significant contribution from the resonant wave-wave interactions that characterise the Benjamin-Feir instability (though, as reviewed by Dysthe et al. (2008), this instability is thought to occur only if the waves are long-crested as well as steep).

The Mori and Janssen (2006) theory applies to waves that are narrow-band and have $H = 2\eta$, so we may scale their Formulae (46) and (47) to give an exceedance probability for crest height given by

$$P(\eta/H_s > z) = \exp(-8z^2)\left[1 + \frac{8}{3}\kappa_{40}z^2\left(4z^2 - 1\right)\right] \quad (16)$$

This shows an increase in the probability of large waves over that given by the Rayleigh distribution and is shown in Fig. 5 for the range of values of κ_{40} that they find appropriate for laboratory and ocean waves.

The notable feature of Fig. 5 is that, while the probability of large waves is enhanced, the plots do not show the downward curvature exhibited by the Gorm data discussed by Dysthe et al. (2008).
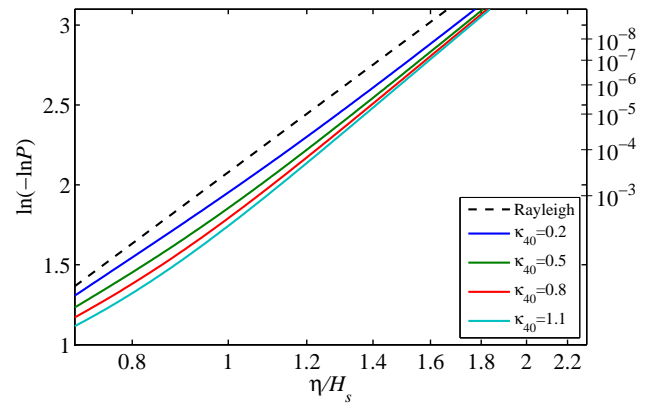

Fig. 5. The crest height exceedance probability according to the theory of Mori and Janssen (2006), for various values of the parameter κ_{40}.

6 Effects of non-stationarity

As remarked by Longuet-Higgins (1952), if a non-stationary time series of waves is treated as if it were stationary, it will show a greater than expected probability of large waves. The point was emphasized by Müller et al. (2005) and by Heller (2006). In the situation studied by Heller (2006), the non-stationarity arises from the passage of a ship through a sea state that is inhomogeneous by virtue of wave interactions with small-scale currents.

We examine the problem analytically and numerically to see how non-stationarity affects the straight line expected in a data presentation such as that of Fig. 1. Analytically, we consider a wave record with significant wave height H_s for the whole record, but H_1, H_2 for the first and second half, with $H_1^2 = H_s^2(1+\epsilon)$ and $H_2^2 = H_s^2(1-\epsilon)$. Then

$$P(\eta/H_s > z) = \frac{1}{2}\left[\exp\left(-\frac{8z^2}{1+\epsilon}\right) + \exp\left(-\frac{8z^2}{1-\epsilon}\right)\right] \quad (17)$$

Figure 6 shows $\ln[-\ln P(\eta/H_s > z)]$ from Eq. (17) for various values of ϵ. (The exceedance probability for the wave height will be similarly affected and does not need to be examined separately.)

Figure 6 clearly demonstrates an increasing probability of large waves in a non-stationary time series treated as stationary. For large waves, however, the line in Fig. 6 for a particular value of ϵ has a slope equal to that of the line for a stationary time series with $\epsilon = 0$, rather than becoming less steep as occurs with the second-order correction to crest height. The result is not surprising, since the first half of the right hand side of Eq. (17) dominates for large η/H_s.

It is possible that this tendency for $\ln[-\ln P(\eta/H_s > z)]$ to maintain the same slope for non-stationary data is a consequence of assuming an abrupt change in the variance from one value to another. It would be interesting to investigate
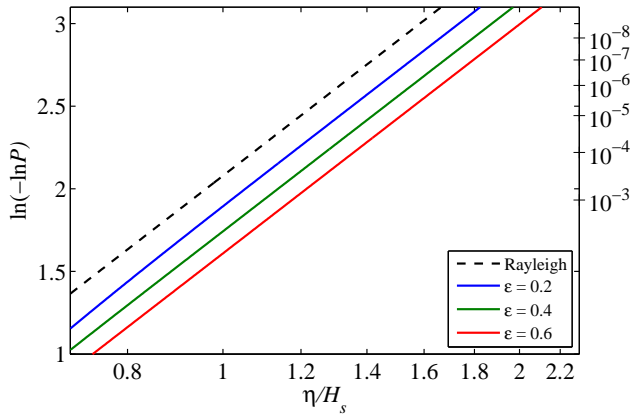
Fig. 6. The effect of non-stationarity, based on Eq. (17), for a time series with a variance that changes halfway through, from $1+\epsilon$ to $1-\epsilon$ times the average, with $\epsilon = 0.2, 0.4.0.6$. The Rayleigh line is equivalent to that for $\epsilon = 0$.
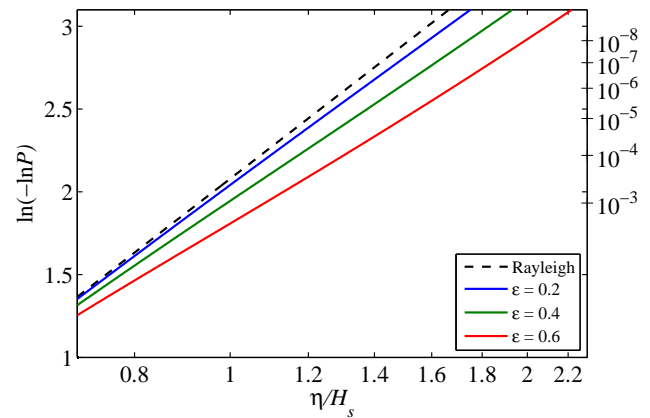

Fig. 7. The effect of non-stationarity, based on Eq. (18), for a time series with a variance that is proportional to $(1+\alpha t)^{-1}$ over the range $-T \leq t \leq T$, with $\epsilon \equiv \alpha T = 0.2, 0.4, 0.6$. The Rayleigh line is equivalent to that for $\epsilon = 0$.

a linearly changing variance, proportional to $1+\alpha t$ over the range $-T \leq t \leq T$ with $0 < \alpha T < 1$, but this does not lead to a simple expression for P. A model that is very close to this, however, takes the variance to be proportional to $(1+\alpha t)^{-1}$. The average variance is no longer equal to the variance at $t = 0$, but, after some algebra, we find that

$$P(\eta/H_{\rm s} > z) = \exp(-Bz^2)\frac{\sinh(B\epsilon z^2)}{B\epsilon z^2} \qquad (18)$$

where $\epsilon = \alpha T$ and

$$B = \frac{4}{\epsilon}\ln\left(\frac{1+\epsilon}{1-\epsilon}\right) \qquad (19)$$

As $\epsilon \to 0$, $B \to 8$ and Eq. (18) reduces to the expected Rayleigh distribution.

We see that the slopes of the plots in Fig. 7 decrease as ϵ increases. For a given ϵ, there is no significant change in slope as $\eta/H_{\rm s}$ increases.

We find similar behavior for the simulated waves from the JONSWAP spectrum with $\gamma = 1$ and no second-order correction, but with $H_{\rm s}$ tapered as for the analytical example discussed above. The exceedance probability plots are very similar to those of Fig. 7 and are not shown.

6.1 Apparent non-stationarity

The flip side of not allowing for non-stationarity can occur if a stationary record is split into blocks. These will have significant wave heights $H_{\rm s}$ that fluctuate around the value for the full record, producing misleading statistics if the crest or wave height is normalized by the block $H_{\rm s}$ rather than the $H_{\rm s}$ from the full record. (For typical wind seas, Donelan and Pierson (1983) find a variation of 12% or so in $H_{\rm s}$ calculated from 17 min blocks of data.) As remarked by Forristall

(2005), a wave that is not really particularly large might appear to be so if it occurs in a block of data with a less than representative significant wave height.

To investigate this, we have taken 20 min blocks from a simulated 40 yr time series of a JONSWAP spectrum with $\gamma = 1$ and a peak period of 10 s. Figure 8 shows the standard presentation of the exceedance probability of crest height.

Surprisingly, the exceedance probability of a given crest height is less if the block values of $H_{\rm s}$, rather than the value from the whole record, are used. This implies that in blocks with smaller than average $H_{\rm s}$ the heights of large crests are even more reduced. However, the exceedance probability plot using the block values of $H_{\rm s}$ is still reasonably well approximated by a straight line; there is no change in behaviour for the large waves.

7 Generalised extreme value theory

While comparison with proposed exceedance probabilities is possible, it is also sometimes suggested that extreme values for wave or crest heights should be fitted with the canonical functions that emerge from extreme value theory (e.g., Coles, 2001) . This is a general theory that leads to a family of asymptotic formulae for the exceedance probability for the maximum in blocks of data containing a large number of individual events. A particular member of this family is the Gumbel distribution and we can see how this distribution arises in a situation in which the exceedance probability for individual events is given by Eq. (12).

We start with

$$F \equiv P(\eta_{\max} < z) = \left[1 - \exp\left(-\frac{z^\alpha}{\beta}\right)\right]^N \qquad (20)$$
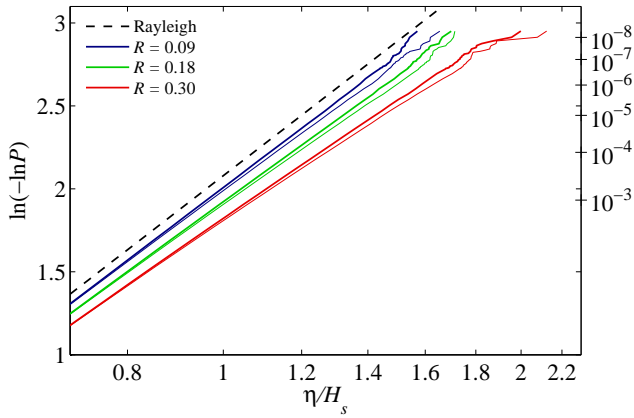
Fig. 8. The effect of apparent non-stationarity, based on using H_s from short data blocks (thick lines) rather than from the whole record (thin lines).

Fig. 9. The function $\ln[-\ln(1 - F^{1/N})]$, with F from Eq. (27) taking $\alpha = 2$ and $\beta = 1/8$, plotted against $\ln z$ and compared with the straight line representing $2\ln z + \ln 8$.

where $\eta_{\max}$ is the maximum of η/H_s in a block of N waves. Hence, following Nerzic and Prevosto (1997) but changing their notation,

$$P\left(\frac{\eta_{\max} - a_N}{b_N} < z\right) = \left\{1 - \exp\left[-\frac{(b_N z + a_N)^\alpha}{\beta}\right]\right\}^N \quad (21)$$

We now define

$$a_N = (\beta \ln N)^{1/\alpha}, \qquad b_N = \frac{a_N}{\alpha \ln N} \quad (22)$$

so that

$$\exp\left[-\frac{(b_N z + a_N)^\alpha}{\beta}\right] = \exp\left[-\ln N \left(1 + \frac{z}{\alpha \ln N}\right)^\alpha\right] \quad (23)$$

$$\simeq \exp(-\ln N - z) \quad \text{if} \quad z \ll \alpha \ln N \quad (24)$$

$$= N^{-1} e^{-z} \quad (25)$$

Hence

$$P\left(\frac{\eta_{\max} - a_N}{b_N} < z\right) \simeq \left(1 - N^{-1} e^{-z}\right)^N \to \exp\left(e^{-z}\right)$$

$$\text{as} \quad N \to \infty \quad (26)$$

so that

$$F \equiv P(\eta_{\max} < z) \simeq \exp\left[-\exp\left(-\frac{z - a_N}{b_N}\right)\right] \quad (27)$$

This is the Gumbel distribution. We may compare it with a plot of (13) by noting that

$$P(\eta/H_s > z) \equiv 1 - F^{1/N} \quad (28)$$

and evaluating the function $\ln[-\ln(1 - F^{1/N})]$ with F from (27). Figure 9 shows this for the Rayleigh distribution with
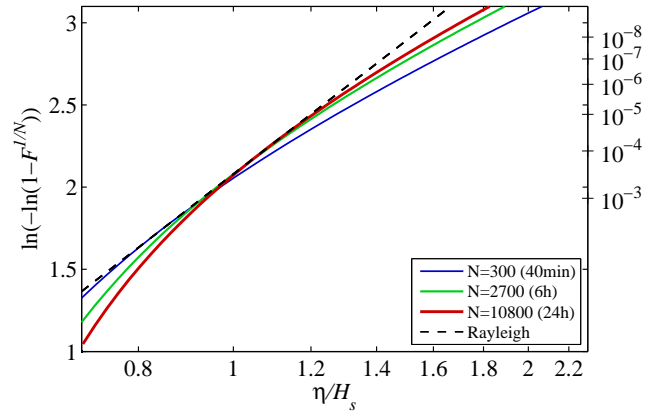
$\alpha = 2$ and $\beta = 1/8$, for various values of N. The Gumbel distribution leads to a curve that can be shown to be tangent, for large N, to the Rayleigh line at a point corresponding to an exceedance probability $P = 1/N$. (This is a general result; the approximating Gumbel distribution is tangent at $P = 1/N$ to the Weibull line for any values of α and β.)

Importantly, the Gumbel curve clearly does not provide a reliable way of extrapolating the distribution to large values of the block maximum $\eta_{\max}$, corresponding to rare events, with an exceedance probability smaller than $1/N$. The main reason for this is that the theory requires not just N but also $\ln N$ to be large, as seen from Eq. (24). This is unlikely in practical situations.

Thus the superficially attractive option of fitting generalised extreme value (GEV) distributions to block maxima obtained from data, without any a priori assumption about the pdf of the individual crest heights, does not seem appropriate.

8   Discussion and conclusions

We have argued that the appropriate way to present the exceedance probability P for wave size is that used by Dysthe et al. (2008), namely to plot $\ln(-\ln P)$ against the logarithm of the wave parameter (scaled height or crest height). In such a plot, the Weibull distributions proposed by Forristall (2000) lead to straight lines. We have shown a straight line provides a good approximation for the Tayfun (1980) distribution of crest heights and also for simulated time series based on the JONSWAP spectrum with a local second-order correction. This conclusion extends down to exceedance probabilities of the order of 10^{-7}, supporting the conclusions of Forristall (2000) that were limited to exceedance probabilities greater than 10^{-4}. The change in the slope of such a plot for low

probability large waves, as discussed by Dysthe et al. (2008) for waves from the Gorm oil field in the North Sea, thus has no simple explanation in terms of the second-order correction. Further examination of the quality of the Gorm data, and confirmation or refutation from other long data sets, is clearly required. However, if the result holds up, we have shown that it might be a consequence of third and fourth order Stokes corrections to the wave height and crest height. Our simulations of these higher order corrections may be oversimplified as we treat each wave as if it were part of a uniform wave train, so further theoretical work is also required.

It has seemed worthwhile to seek other potential causes of the change in slope. We have found that the Benjamin-Feir instability, as discussed by Mori and Janssen (2006), gives a higher probability of large waves (like the second- order correction) but not the change in slope of the plot at low probabilities. This slope change also seems to be unexplained by the effects of non-stationarity, though this does increase the likelihood of large waves in a record treated as stationary. The slope change is also unaccounted for by statistical fluctuations in the value of H_s for short blocks from a longer, stationary, record.

It could be argued that using the exceedance probability based on a Weibull distribution of crest or wave heights prejudges the situation and that it is better to appeal to extreme value theory for block maxima and fit observations to the formulae predicted by that theory. We have shown, however, that, while the Rayleigh distribution of crest heights predicts an asymptotic Gumbel distribution for the exceedance probability as the number N of waves in a block becomes large, this is typically not a good approximation to the true exceedance probability. The reason is simply that the requirement for not only N, but also $\ln N$, to be large, is not likely to be met.

We conclude that crest height exceedance probabilities less than, say, a few times 10^{-5}, may be considerably greater than obtained from the standard Weibull formulae of Forristall (2000) and this should be allowed for in predictions. Simple Stokes corrections, beyond second order, seem to be the only process that can give the observed shape of the probability curve, but continued investigation of low probability large waves is required. This will be difficult, of course, as 10^5 waves only occur every ten days or so, so that statistically reliable results require very long data sets as well as a valid measurement technique, but it is surely these very rare, very large, waves that are of the greatest scientific and practical interest. In particular, we remark that even if an extreme wave with a scaled crest or wave height is predicted to occur only once every 30 yr (corresponding to an exceedance probability of approximately 10^{-8}) at a fixed location, a fleet of 100 ships would experience it every few months (though less frequently, of course, if attention is limited to high sea states).

References

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

Casas-Prat, M. and Holthuijsen, L. H.: Short-term statistics of waves observed in deep water, J. Geophys. Res., 115, C09024, doi:10.1029/2009JC005742, 2010.

Dawson, T. H.: Stokes correction for nonlinearity of wave crests in heavy seas, J. Waterway, Port, Coastal Ocean Eng., 130, 39–44, 2004.

Dean, R. G. and Dalrymple, R. A.: Water Wave Mechanics for Engineers and Scientists, World Scientific, 1991.

Donelan, M. and Pierson, W. J.: The sampling variability of estimates of spectra of wind-generated gravity waves, J. Geophys. Res., 88, 4381–4392, 1983.

Dysthe, K., Krogstad, H. E., and Müller, P.: Oceanic rogue waves, Annu. Rev. Fluid Mech., 40, 287–310, 2008.

Forristall, G. Z.: Wave crest distributions: Observations and second-order theory, J. Phys. Oceanogr., 30, 1931–1943, 2000.

Forristall, G. Z.: Understanding rogue waves: Are new physics really necessary?, in: Rogue Waves, edited by: Müller, P. and Henderson, D., Proceedings of the 14th "Aha Huliko"a Hawaiian Winter Workshop, 29–35, 2005.

Garrett, C. and Gemmrich, J.: Rogue waves, Phys. Today, June, 62–63, 2009.

Gemmrich, J. and Garrett, C.: Unexpected waves, J. Phys. Oceanogr., 38, 2330–2336, 2008.

Gemmrich, J. and Garrett, C.: Unexpected waves: Intermediate depth simulations and comparison with observations, Ocean Eng., 37, 262–267, 2010.

Heller, E. J.: Freak ocean waves and refraction of Gaussian seas, in: Extreme Events in Nature and Society, edited by: Albeverio, S., Jentsch, V., and Kantz, H., Springer, 189–210, 2006.

Kharif, C., Pelinovsky, E., and Slunyaev, A.: Rogue Waves in the Ocean, Springer, 2009.

Kriebel, D. L. and Dawson, T. H.: Nonlinear effects on water waves in random seas, J. Offshore Mech. Arctic Eng., 113, 142–147, 1993.

Longuet-Higgins, M. S.: On the statistical distribution of the heights of sea waves, J. Mar. Res., 11, 245–65, 1952.

Mori, N. and Janssen, P. A. E. M.: On kurtosis and occurrence probability of freak waves, J. Phys. Oceanogr., 36, 1471–1483, 2006.

Müller, P. and Henderson, D. (Eds.): Rogue Waves, Proceedings of the 14th "Aha Huliko"a Hawaiian Winter Workshop, available at: http://www.soest.hawaii.edu/PubServices/2005pdfs/TOC2005.html, 2005.

Müller, P., Garrett, C., and Osborne, A.: Rogue waves, Oceanography, 18, 66–75, 2005.

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

Nerzic, R. and Prevosto, M.: A Weibull-Stokes model for the distribution of maximum wave and crest heights, Proc. Seventh International Offshore and Polar Engineering Conference, 367–377, 1997.

Stansell, P.: Distributions of freak wave heights measured in the North Sea, Appl. Oc. Res., 26, 35–48, 2004.

Tayfun, M. A.: Narrow-band nonlinear sea waves, J. Geophys. Res., 85, 1548–1552, 1980.

WAFO Group: WAFO: A Matlab toolbox for analysis of random waves and loads, Center for Mathematical Sciences, version 2.0.02, 124 pp., 2000.