TECHNICAL MEMORANDUM

755

Notes on the maximum wave height distribution.

Peter A.E.M. Janssen

Research Department

June 2015

European Centre for Medium-Range Weather Forecasts
Europäisches Zentrum für mittelfristige Wettervorhersage
Centre européen pour les prévisions météorologiques à moyen terme

Series: ECMWF Technical Memoranda

A full list of ECMWF Publications can be found on our web site under:
http://www.ecmwf.int/en/research/publications

Contact: library@ecmwf.int

Abstract

This memo discusses a number of updates of the present freak wave warning system (Janssen & Bidlot, 2009) which will be introduced in cycle 41R1 of the IFS. The list of changes is given below and they are discussed in more detail in the remainder of this memo. They are:

1. Finite skewness effects, related to bound waves are introduced, and their impact is verified.
2. In the derivation of the maximum wave height distribution it is assumed that subsequent events are independent. In contrast, here it is assumed that subsequent wave groups are independent, rather than the individual waves. The number of independent events follows by making use of the average length of a group, which depends on the width of the spectrum.

A Monte Carlo simulation for linear waves shows that the above assumption of statistical independence of wave groups is a good choice. In particular, there is a close agreement between the simulated and the theoretical maximum wave height distribution.

1 Introduction.

Recently, there has been considerable progress in the understanding of the occurrence of freak waves, a notion which was first introduced by Draper (1965). Freak waves are waves that are extremely unlikely as judged by the Rayleigh distribution of wave heights (Dean, 1990). In practice this means that when one studies wave records of a finite length (say of 10-20 min), a wave is considered to be a freak wave if the local wave height H (defined as the distance from crest to trough) exceeds the significant wave height H_S by a factor 2. It should be clear that it is hard to collect evidence on such extreme wave phenomena because they occur so rarely. Nevertheless, observational evidence from time series collected over the past decades does suggest that for large surface elevations the probability distribution for the surface elevation may deviate substantially from the one that follows from linear theory with random phase, namely the Gaussian distribution (cf. e.g. Wolfram and Linfoot, 2000). Also, there are now a number of recorded cases which show that the ratio of maximum wave height and significant wave height may be as large as three (Stansell, 2005).

Our increased understanding of the occurrence of extreme events is based on weakly nonlinear theory. Extreme events are judged by means of the probability distribution function (pdf) of wave height and maximum wave height. Although for linear waves the wave height pdf will be close to a Gaussian, finite amplitude ocean waves may give rise to deviations from Normality. There are two reasons for it. First of all, finite amplitude waves generate bound waves such as second and third harmonics which gives finite skewness (connected with sharper crests and wider throughs) and kurtosis. The resulting deviations from Normality will always occur as long as the waves are steep. However, there is another reason why there may be large deviations from Normality but this will only occur when the sea state is coherent, corresponding to a narrow spectrum in frequency and direction. Under those circumstances there is possibly a strong four-wave interaction[1] which may result in quite large amplitude waves corresponding to large values of kurtosis and therefore to considerable deviations from Normality in such a way that extreme events become more likely. Compared to the first reason, this dynamic mechanism is really rare because it only occurs for nonlinear coherent sea states, but when it does occur it may cause large

[1]Extreme events can be simulated by means of the Zakharov equation (Zakharov, 1968, Janssen, 2003), which is the prototype equation for nonlinear four-wave interactions. Yasuda *et al* (1992), Trulsen and Dysthe (1997) and Osborne et al (2000) studied simplified versions of the Zakharov equation and it was found that these extreme waves can be produced by nonlinear self modulation of a slowly varying wave train. An example of nonlinear modulation or focussing is the instability of a uniform narrow-band wave train to side-band perturbations. This instability, known as the Benjamin-Feir (1967) instability, will result in focusing of wave energy in space and/or time as is illustrated by the experiments of Lake et al (1977).

damage. This mechanism provides a plausible explanation for freak wave formation, which is definitely a rare event.

In ocean wave forecasting practice one follows a stochastic approach because the phases of the individual waves are unknown. In this memo we therefore concentrate on a probablistic approach, in particular we try to utilize results from a random time series analysis. It is assumed that the wave spectrum $E(\omega)$ is given and that we know relevant statistical moments such as the variance, the skewness and the kurtosis of the time signal. Noting that maximum wave height is the parameter of first choice to characterize extreme events, the important question then is whether it is possible to obtain a theoretical expression for the maximum wave height distribution for given spectral shape and statistical moments, realizing that a maximum parameter is defined for a given length of the time series. The answer to this question depends on the choice of analysis technique: traditionally in the field of ocean waves one analyzes time series in terms of wave height, defined through the zero-crossing method. In Janssen (2014) it is shown that an alternative technique, namely use of the envelope of the time series, has certain advantages. First, the square of the envelope gives a measure for the local (linear) energy of the waves, which is a more interesting quantity then wave height if one is interested in the impact of waves on a marine structure or a ship. Secondly, in linear theory it is straightforward to show that the probability distribution function (pdf) of the envelope height follows the Rayleigh distribution, while no such simple pdf for the zero-crossing wave height is known. Given the pdf for envelope wave height a theoretical expression of maximum wave height is derived and this expression is validated against results from Monte Carlo simulations.

The content of this memo is as follows. Using Monte Carlo Simulations for linear waves, in §2 a comparison is made between the pdf of wave height using the zero-crossing method and using the envelope method. In agreement with theoretical results of Janssen (2014) it is found that the envelope wave height obeys the Rayleigh distribution while the pdf of zero-crossing wave height depends on spectral shape and only approaches the Rayleigh distribution from below for narrow spectra. We concentrate in §3 on obtaining information about the occurrence of extreme events. To that end we try to derive, using the envelope method, the pdf of maximum wave height starting from the approach suggested by Goda (2000). Given N independent events it is straightforward to obtain the probability that a time series has a given maximum if the pdf of envelope wave height (which includes nonlinear effects as well) is known. The key question is then how to estimate the number of independent events N. In §3 it is argued that successive wave groups may be regarded as independent so that the number of independent events will scale with the 'average' length of a wave group which is proportional to the product of the relative width of the wave spectrum and the mean frequency. The maximum envelope wave height pdf will therefore depend on spectral shape. This is confirmed by means of Monte Carlo simulations. Comparing the numerically obtained pdf with the theoretical one for different spectral shapes a good agreement is obtained, suggesting that with some confidence the thus developed theoretical approach may be used in estimating the probability of extreme events. In §4 we extend the approach of §3 into the weakly nonlinear regime and we show to what extend nonlinear effects may change the pdf of maximum wave height. In §5 of this memo, after a discussion of the consequences for ECMWF's freak wave warning system, a summary of conclusions is presented.

2 Comparison of Envelope method and Zero-Crossing method.

It is common practice to analyze extreme ocean wave events by means of the zero-crossing method. This is a very elegant method, which can be easily used and implemented. One just searches for two consecutive zero-upcrossings in the time series and one determines the wave height from the difference of the maximum and the minimum of the surface elevation η in the corresponding time interval. Thus, wave height is basically determined by sampling with the zero-crossing frequency $(m_2/m_0)^{1/2}$ (where m_n is the n^{th} moment of the spectrum), and to quantify the severity of the sea state one determines the probability distribution function of the zero-crossing wave height. By applying the zero-crossing method to time series of the surface elevation it turns out that the resulting pdf depends on spectral shape. This follows from Monte Carlo simulations. In particular, for a narrow-band spectrum the zero-crossing wave height pdf is found to be close to the Rayleigh distribution while for broad-band spectra extreme waves are, when compared to the Rayleigh distribution, less likely to occur.

An alternative technique, based on the use of the envelope of the time series, shows a different picture. I will call this alternative the envelope method. Although this method is not so popular in the field of ocean waves it should be pointed out that in other fields, such as communication theory and nonlinear optics, this approach is found to be very useful. Some aspects of the envelope method have been given in Janssen (2014). The envelope ρ is obtained from the timeseries of the surface elevation η and its Hilbert transform $\zeta = H(\eta)$, according to

$$\rho = \sqrt{\eta^2 + \zeta^2},$$

where for stationary signals surface elevation and its Hilbert transform have the same variance m_0. The local envelope wave height is then defined as $H_{env} = 2\rho$. In agreement with common practice in ocean wave forecasting, wave height will be scaled with $4m_0^{1/2}$ so that we will study the statistical properties of the dimensionless local wave height parameter $h = H_{env}/4m_0^{1/2}$. Theoretically and from Monte Carlo simulations (see e.g. Janssen, 2014) one finds that for a Gaussian sea state the pdf of envelope wave height follows the Rayleigh distribution, independent of spectral shape.

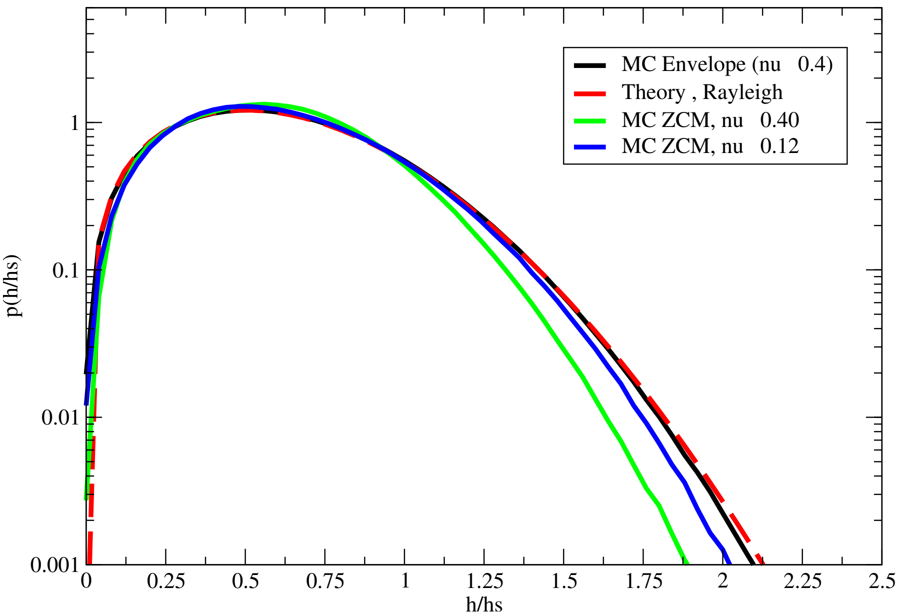

Figure 1: Comparison of pdf of zero-crossing wave height (labelled ZCM) with pdf of envelope wave height for different spectral width parameter ν. The envelope pdf does not depend on ν, whilst the zero-crossing height pdf does. For broad spectra (ν large) probabilities are underestimated by a factor of 10.

We have illustrated the findings on the statistics of zero-crossing wave height and envelope wave height in Fig. 1. In the figure spectral shape is measured by the spectral width parameter ν, introduced by

Longuet-Higgins (1983). In terms of the spectral moments

$$m_n = \int \mathrm{d}\omega\, \omega^n\, E(\omega), \tag{1}$$

with $E(\omega)$ the frequency spectrum, it is defined as

$$\nu = \sqrt{m_0 m_2 / m_1^2 - 1}. \tag{2}$$

The details of the Monte Carlo simulations are given in Janssen (2014). Essentially, we restrict ourselves to linear theory and we write down the surface elevation for a spectrum of waves where the amplitudes are drawn from a given wave number spectrum using a Rayleigh distribution while the phases are drawn randomly. The case of a narrow-band spectrum, with $\nu = 0.12$ corresponds to a wavenumber cut-off of $2^{1/2}$ times the peak wave number while the broad-band case with $\nu = 0.4$ corresponds to a cut-off wavenumber of 64 times the peak wavenumber. As already discussed in Janssen (2014) the envelope method results in an almost continuous representation of wave height while with the zero-crossing method wave height is only sampled with the zero-crossing frequency. Therefore, in order to obtain stable results with the zero-crossing method the number of members of the ensemble was chosen to be 50,000 while the length of the time series of each member was 100 peak periods. From Fig. 1 it is clear that the pdf of zero-crossing wave height is a sensitive function of spectral width underestimating for broad spectra the probability of extreme events by an order of magnitude. On the other hand, the pdf of envelope wave height is independent of spectral width and is given by the Rayleigh distribution. Since for the zero-crossing wave height pdf there is no known analytical expression available and since for the derivation of the maximum wave height statistics knowledge of the wave height pdf is essential we now concentrate on the envelope method.

3 The maximum envelope wave height distribution: linear theory.

In order to derive the pdf of maximum wave height we need information on the statistics of the envelope wave height time series and we need the length T_L of the time series. The relevant statistics are the envelope wave height distribution $p(h)$, the cumulative wave height distribution (or the cumulative distribution function, cdf)

$$P(h) = \int_h^\infty \mathrm{d}h\, p(h), \tag{3}$$

and the joint pdf of envelope wave height h and its time derivative $\dot{h}$. As according to linear theory the envelope wave height distribution is given by the Rayleigh distribution (Janssen (2014)),

$$p(h) = 4he^{-2h^2}, \tag{4}$$

the cumulative distribution function (3) becomes

$$P(h) = e^{-2h^2}. \tag{5}$$

Again assuming linear waves, the joint probability distribution of envelope wave height h and its time derivative $\dot{h}$ is given by

$$p(h,\dot{h})) = \frac{8h}{\sqrt{2\pi}} e^{-2(h^2+\dot{h}^2)}, \tag{6}$$

where $\dot{h}$ has been normalized by means of the parameter $m_0^{1/2}\nu\bar{\omega}$ with $\bar{\omega} = m_1/m_0$ the mean angular frequency while $\nu = (m_0 m_2/m_1^2 - 1)^{1/2}$ is once more the width of the frequency spectrum. Note that by considering the jpd of h and $\dot{h}$ the frequency scale $\nu\bar{\omega}$ is introduced in a natural way which corresponds to the inverse of the relevant timescale of the wave groups. As time has been made dimensionless with this frequency scale the length T_l of the timeseries has to be scaled accordingly, hence the dimensionless length is $T_L^* = T_L \times \nu\bar{\omega}$ The result (6) is valid for a Gaussian sea state and can be obtained in a straightforward manner from the well-known expression of the jpd of envelope ρ and phase ϕ and its time derivatives (see e.g. Janssen (2014)). Integration over phase ϕ and its derivative $\dot{\phi}$ and introduction of the envelope wave height which is twice the envelope then results in (6).

3.1 Goda's Approach.

Goda (2000) obtained the maximum wave height distribution from a time series of N independent wave events assuming a narrow band wave train. The natural sampling frequency would be the mean zero crossing frequency or perhaps the peak frequency as the length T_L of the time series divided by the sampling frequency would give the number of waves. A similar approach was already followed by Davenport (1964) in the problem of estimating the maximum gust. In this note, we follow Goda's (2000) approach but now applied to time series of the envelope height. Clearly, when going from surface elevation to envelope height time series the mean oscillation frequency will be removed from our considerations which will affect the sampling frequency. It will now involve the frequency scale $\nu\bar{\omega}$. Note that this frequency scale can be rewritten as $\nu\bar{\omega} = (m_2/m_0 - \bar{\omega}^2)^{1/2}$ which illustrates that for the envelope time series indeed the mean oscillation frequency is removed from the signal.

Now, the maximum wave height distribution is obtained by writing down the probability that for given number of independent events N the maximum envelope height has a certain chosen value. In other words, the maximum wave height distribution $p_{max}(h_{max})$ is the probability that a certain event attains the maximum value multiplied by the probability that all other events are below the maximum value while realizing there are N possibilities. Therefore

$$p_{max}(h_{max}) = N[1 - P(h_{max})]^{N-1} p(h_{max}), \tag{7}$$

where for linear waves P is the cdf given in Eq. (3). It is straightforward to implement this expression for the pdf of maximum wave height, but we are also interested in deriving a simple, accurate expression for parameters such as the expectation value of maximum wave height. In this case we need an approximate form of $p_{max}(h_{max})$, which is obtained from the large N limit of (7). From experience it is known that this approximation works well for $N > 10 - 20$.

In the continuum limit, i.e. for large N, the maximum wave height distribution $p_{max}(h_{max})$ then assumes the simple form

$$p_{max}(h_{max}) = Np(h_{max}) \times \exp[-NP(h_{max})]. \tag{8}$$

For fixed N this result can be written in a simple fashion as by definition $dP(h)/dh = -p(h)$ and therefore, introducing the function

$$\mathscr{G}(h) = -NP(h),$$

being equal, apart from a minus sign, to the product of the number of waves N and the cumulative distribution P, we have

$$p_{max}(h = h_{max}) = \frac{\mathrm{d}\mathscr{G}}{\mathrm{d}h} \exp(\mathscr{G}). \tag{9}$$

Close inspection of the result for the maximum wave height distribution shows that this distribution is a double exponential function in general, but for large maximum wave heights (typically of the order of 2 or larger) the pdf simplifies considerable because it becomes

$$p_{max}(h) = N\ p(h).$$

This follows immediately from the basic result (7) and from the continuous limit (8) by realizing that for extreme states the cdf $P(h_{max})$ is small. Using the cdf in (5) one finds that N times the cdf becomes less than a small number, say ε, for the range $h^2 > \frac{1}{2}\log(N/\varepsilon)$. For typical conditions with $N = 100$ and $\varepsilon = 0.01$ one finds a threshold value slightly above 2.1, while for $N = 10$ the threshold value reduces to 1.8.

An important role in this approach is played by the parameter N which is a measure of the number of independent events. It has been suggested (see e.g. Mori and Janssen, 2006) to equate N with the number of waves N and therefore $N = T_L/T_p$ with T_P the peak period and T_L the length of the timeseries. This may be a reasonable choice when the statistics of individual waves is considered but here we study the properties of envelope time series. In the latter case it seems more natural to connect the number of independent events with the number of wave groups (see e.g. Cramer and Leadbetter (1967) and Ewing (1973)). This is the approach followed here and I will validate this choice using numerical evidence. However, only a Gaussian sea state has been considered so far, therefore, for the estimation of the number of degrees of freedom N skewness and kurtosis effects are ignored.

I find the estimation of the number of degrees of freedom not a trivial matter, and the following argument only makes the choice (12) a plausible one. It is customary to define an event with respect to a chosen reference level h_c. An event is then a part of the time series that starts where the envelope has an upcrossing and that finishes at the next downcrossing. The frequency of events is then determined by the upcrossing frequency. The total number of events during the duration T_L^* then determines the number of degrees of freedom N. For a more complete discussion see Elgar *et al.* (1984), where it follows that N and also parameters such as the number of waves in a group depend on the chosen reference level, but it is not clear which level to choose. Therefore, essentially the frequency of events depends on the reference level, so it may be more appropriate to introduce an average frequency. The first measure of frequency that came to mind is basically the average of the rate of change of h with time, $\dot{h}$ normalized with h itself. Hence, the average frequency of events, determined by the average upcrossing frequency, becomes

$$\langle f_{up}\rangle = \langle \dot{h}/h\rangle = \int_0^\infty \mathrm{d}h \int_0^\infty \mathrm{d}\dot{h}\ p(h,\dot{h})\ \dot{h}/h \tag{10}$$

Making use of the joint pdf of h and $\dot{h}$ in Eq. (6) and performing the integrations one immediately finds the simple result

$$\langle f_{up}\rangle = 1 \tag{11}$$

and the average number of events becomes

$$N = \langle f_{up}\rangle T_L^* = \nu\bar{\omega}T_L.$$

This result reflects that the number of degrees of freedom is determined by the number of wave groups. It is emphasized that I have only made plausible how N depends on the relevant parameters. To some extent the result is uncertain because I have chosen to connect the number of events with the average upcrossing frequency. For this reason I introduce a tuning coefficient α, hence

$$N = \alpha\nu\bar{\omega}T_L. \tag{12}$$

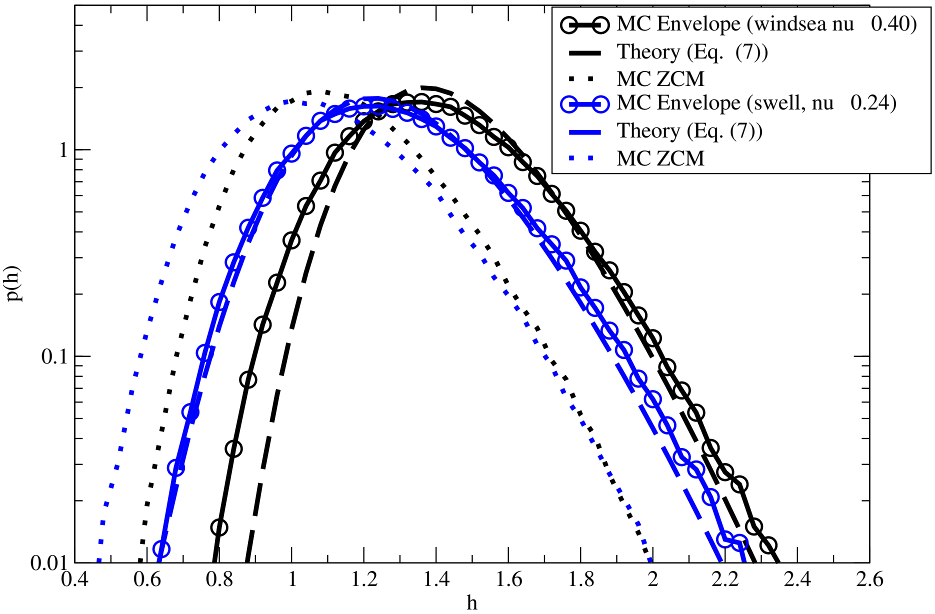

Figure 2: Comparison of pdf of maximum envelope wave height with Monte Carlo simulations for 'old' windsea ($\nu = 0.4$) and for swell ($\nu = 0.24$). For reference, the Monte Carlo result for maximum zero-crossing wave height (ZCM) is shown as well.

which for the following results is chosen to be slightly larger than 1, i.e. $\alpha = \sqrt{4/\pi} = 1.13$.

Fig. 2 shows a comparison between the analytical result (Eqns. (7), (12)) and results from two Monte Carlo simulations, one for 'old windsea' and one for the case of swell, giving different spectral width, namely $\nu = 0.4$ and $\nu = 0.24$ respectively. The number of ensemble members was 150,000 and the wavenumber spectrum was represented on a logarithmic scale with a relative increment of 0.04. Each ensemble member has 10 waves. Both the Monte Carlo simulations and the analytical result suggest that for broader spectra with more degrees of freedom the maximum wave height pdf shifts, as expected, to larger values of normalized wave height. In sharp contrast it should be noted that the maximum wave height pdf according to the zero-crossing method is rather insensitive to a change in ν by a factor of 2. This is surprising since the zero-crossing wave height pdf (the 'parent' pdf) shows a sensitive dependence on spectral width (cf. Fig. 1). Apparently, with increasing ν the number of degrees of freedom increases giving a shift in the pdf towards higher normalized wave height, which is compensated by the reduction of probability according to the 'parent' pdf as shown in Fig. 1.

In summary, it is concluded that the theoretical result (7,12) shows a reasonable agreement with Monte Carlo simulations of the pdf of maximum envelope wave height. Nevertheless, closer inspection of the tail region suggests that theory underpredicts probabilities for extremes in a systematic manner. The higher the normalized maximum wave height, the higher the discrepancy. For this reason, an alternative approach is discussed next.

3.2 An alternative Approach.

In this context, it should be pointed out that an elegant, alternative way of deriving the maximum envelope height distribution has been presented by Naess (1982). He states, based on Cramér's theorem, that if $\eta(t)$ is a stationary Gaussian process, satisfying certain mild restrictions, then the number of level upcrossings by $\eta(t)$ is asymptotically Poisson distributed when the level height increases. Naess assumes

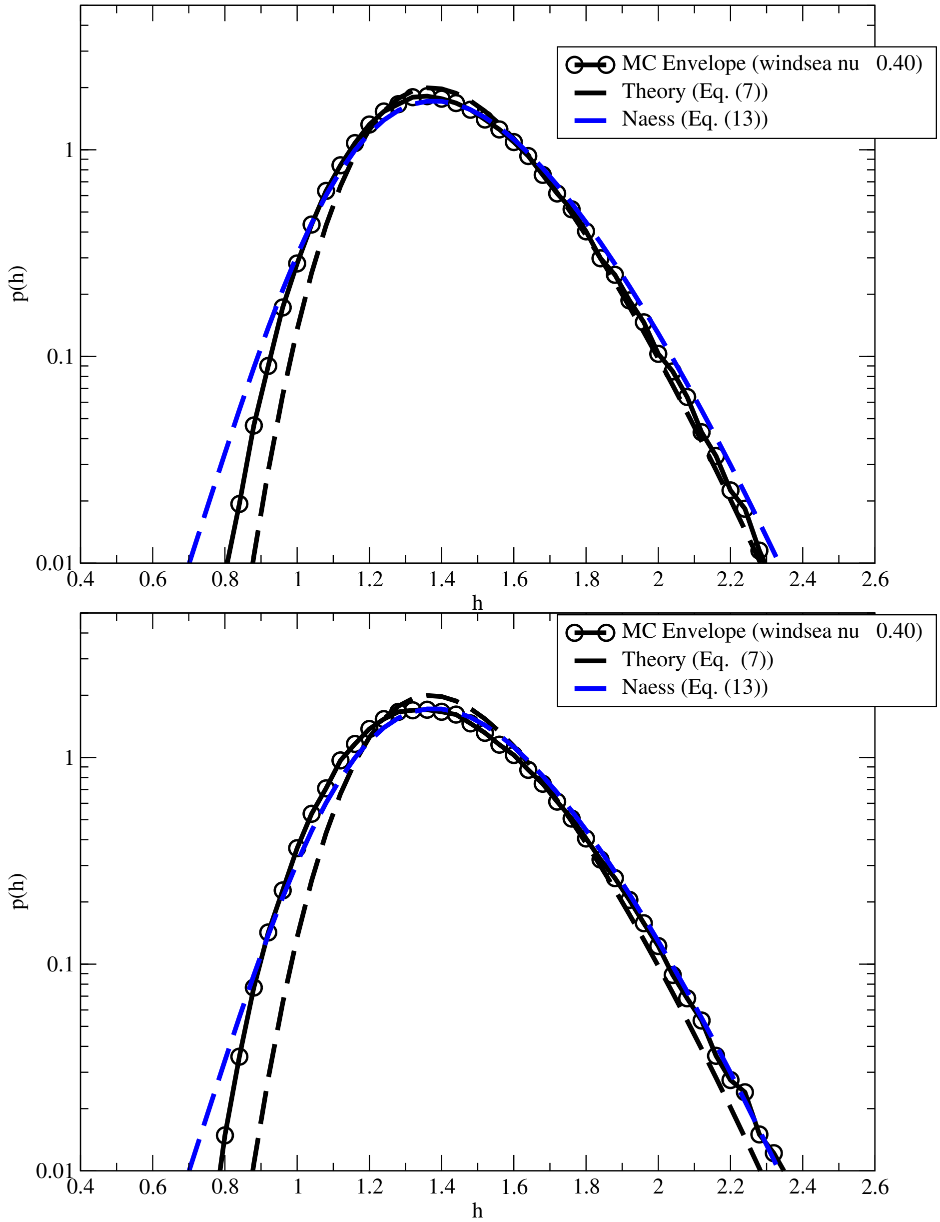

Figure 3: The pdf of maximum envelope wave height for 'old' wind sea ($\gamma = 1$). Shown is the comparison between results from Monte Carlo simulations, and the theoretical result from Eq. (7) and Naess' (1982) Eq. (14). In the top panel the Monte Carlo simulations are based on the Random Phase approximation while in the bottom panel both amplitude and phase of the waves are regarded as stochastic variables.

that this then also holds for the associated envelope process. Let f_{up} be the mean frequency of upcrossings of the level h_c, then for a Poisson process

$$P = \mathrm{Prob}\{h \le h_c; 0 \le t \le T_L^*\} = e^{-f_{up}T_L^*}.$$

The frequency with wich $h(t)$ crosses a reference level h_c with a positive slope, hence positive $\dot{h}$, is then given by

$$f_{up} = \int_0^\infty \mathrm{d}\dot{h}\, \dot{h} p(h_c, \dot{h}),$$

and substitution of (6) and carrying out the integration gives

$$f_{up} = \frac{2h_c}{\sqrt{2\pi}} e^{-2h_c^2}.$$

We now fix T_L^* and denote by $H = \max(h)$ for time $t \in (0, T_L^*)$. Using this in the cdf P one finds

$$P_H(h) = \exp\{-hN_{slc}e^{-2h^2}\},$$

where P_H denotes the probability distribution function of maximum wave height H, and,

$$N_{slc} = 2T_L^*/\sqrt{2\pi} \tag{13}$$

is the number of (up)crossings at the significant level $h = 1$. The maximum envelope height distribution then follows from $p_{max}(h) = dP_H/dh$, or,

$$p_{max} = (4h^2 - 1)\, N_{slc}\, e^{-2h^2} P_H(h), \;\; \text{where } h \geq \frac{1}{2}. \tag{14}$$

The restriction $h \geq \frac{1}{2}$ is added in order to prevent the pdf from becoming negative. In addition, note that the approach by Naess is only valid for large level crossings, presumably because it assumes that for large levels the level upcrossings are statistically independent since these are rare events. An important point to make is that in the present result the dependence of the average number of wave groups N on the reference level has been taken into account, while in the result (8) N has been assumed a constant. As a consequence, the large h behaviour of (14) differs from (8) because it involves an additional factor h. Since we know that Eq. (8) underestimates probability in the tail of the distribution function I expect that the present result might be more correct, in particular for high values of the maximum wave height. This is born out by a comparison of (14) with simulations of the sea state with a Pierson-Moskovitz spectrum, as shown in Fig. 3. However, it should be pointed out that this conclusion depends in a sensitive manner on how the Monte Carlo simulations are being performed. Initially, simulations were done in such a way that the amplitude of the waves is obtained in a deterministic manner from the wavenumber spectrum. If the Monte Carlo simulations are performed in this manner the conclusion is that Goda' approach (7) shows a closer agreement with the numerical pdf than the result of Naess (14). This is shown in the top panel of Fig. 3. But this method is strictly speaking not correct as the amplitudes are random variables as well and, for a Gaussian sea state, should be drawn from a Rayleigh distribution. As a consequence, following the last approach the numerical pdf tends to broaden and now Naess' result is in better agreement with the Monte Carlo result, as seen in in the bottom panel of Fig. 3, at least for the extreme cases.

On the other hand, it should also be pointed out that for low wave height (i.e. $h < 1/2$) Naess' approach is not valid, as already mentioned before, and the pdf becomes negative, which is not really desirable. The Goda approach does not suffer from this problem, and since probabilities for high waves are only underestimated slightly I decided to continue with my original method (7,12).

3.3 Closing Remarks.

Before closing this Section I would like to mention two points. The first one is about an interesting property of the pdf of maximum wave height which allows to estimate the number of degrees of freedom from the observed/simulated maximum wave height pdf directly. The second point concerns the scaling behaviour of the expectation value of maximum wave height with the number of degrees of freedom.

We have seen that under the extreme circumstances the maximum pdf is just N times the pdf of wave height. Hence, the tail of the maximum wave height pdf just reflects the envelope wave height pdf and, in fact, this provides an opportunity to estimate the number of degrees of freedom from the numerical simulations by determining the ratio of simulated pdf, denoted by $p^{sim}_{max}(h)$ and $p(h)$ as a function of dimensionless maximum wave height. When this quantity is plotted, as done in Fig. 4, then if the above approach is correct one should expect a fairly flat ratio as function of maximum wave height while the average value is then a good indicator for the magnitude of the number of degrees of freedom. In Fig.

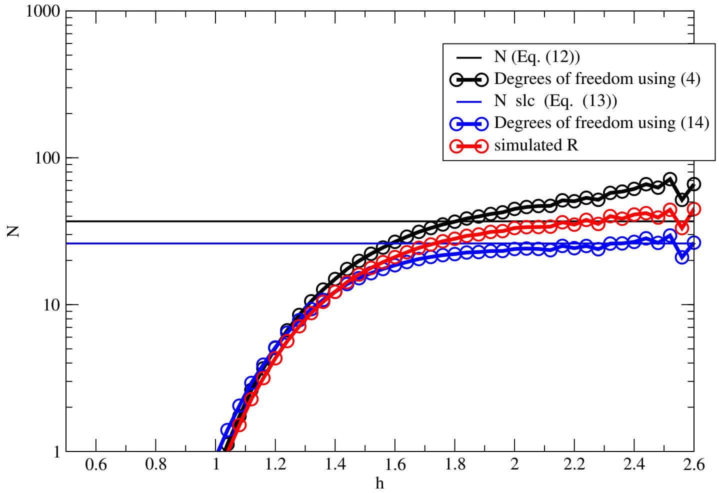

Figure 4: Estimate of the number of degrees of freedom N as function of maximum wave height h as obtained from the ratio of simulated pdf $p^{sim}_{max}(h)$ and $p(h)$. The spectrum is for 'old' windsea with $\gamma = 1$ with a truncation at $2^{1/2}$ time the peak wave number and the length of the time series is 10 peak periods. The number of ensemble members is 1,000,000. Two cases for the normalization are shown, one with $p(h)$ given by (4) while also the result (14) of Naess in the large N-limit is shown. Finally, the ratio $R = p^{sim}_{max}(h)/p^{sim}(h)$, which is entirely based on simulation results, as function of h is shown as the red line.

4 we show estimates of the number of degrees of freedom N as function of maximum wave height h as obtained from the ratio $R = p^{sim}_{max}(h)/p(h)$. In order to get a stable estimate in the tail of the pdf the number of ensemble members was increased to 1,000,000 and, in order to save computation time, the length of the time series was reduced to 10 waves. The spectrum is for 'old' windsea with $\gamma = 1$ with a truncation at 64 times the peak wave number. Two cases for the normalization are shown, one with $p(h)$ given by (4) while also the result using the normalization with the large h-limit of (14) is shown. The thin black and blue lines are theoretical estimates of the degrees of freedom using, respectively, the average number of upcrossings (12) or the number of upcrossings (N_{slc} of 13) corresponding to the significant level $h = 1$. From the Figure it is suggested that N_{slc} is the appropriate choice for the number of degrees of freedom, while the maximum wave height distribution (14) is in close agreement with the numerical simulations since the estimated number of degrees of freedom is a flat function of h. It seems, on the other hand, that the result from Goda's approach (4) fits the numerical simulation less comfortably, in particular for maximum wave heights above 2. Nevertheless, it should be realized that by plotting the ratio R as function of dimensionless maximum wave height h one is really zooming in on details of the tails of the distribution function. If one would, instead, plot the ratio $R = p^{sim}_{max}(h)/p^{sim}(h)$ (hence involving only simulation results) as a function of h, then it seems (see the red line in Fig. 4) that Goda's approach works equally well. Therefore, unfortunately, we cannot draw from this study a definite conclusion on whether the Naess approach or Goda's approach is to be preferred.

Up to now only a few cases have been considered and we studied in detail the shape of the maximum wave height pdf, in particular its tail. Here we explore to what extent there is agreement for a wider

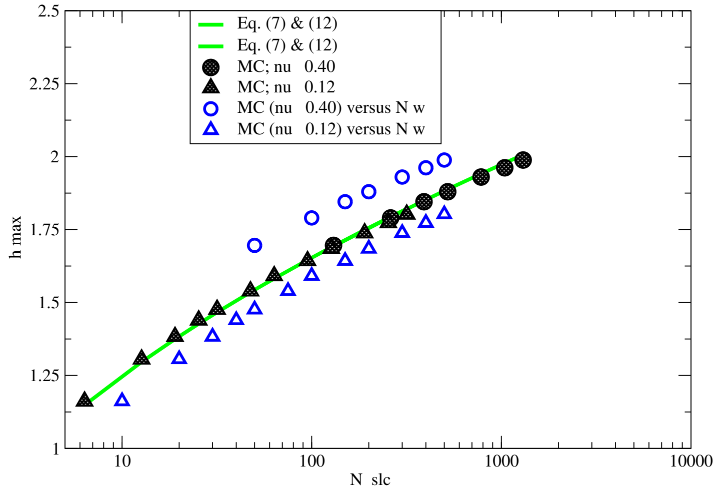

Figure 5: Expected maximum wave height versus number of upcrossings at the significant level (h=1) according to the numerical simulations for two spectral widths. The green line shows the theoretical result based on Eq. (7, 12).However, when results are plotted as function of the number of waves $N_w = T_L/T_P$ (the blue symbols) no scaling behaviour is found.

variety of cases but we restrict the comparison to the expectation value of maximum wave height, i.e. $\langle h \rangle = \int \mathrm{d}h\ p(h)$. The expectation value of maximum wave height turns out to be a robust measure of extreme seas, relatively insensitive to how the tail of the pdf is represented by the numerical simulation. Therefore in order to produce the results displayed in Fig. 5 only 1,000 member ensembles were required. In order to generate more cases we varied the number of waves in the timeseries from 50 to 500 and we took two truncation limits, namely a wavenumber cut-off of 64 times the peak wave number and a wavenumber cut-off of $2^{1/2}$ times the peak wavenumber, representing a broad ($\nu = 0.40$) and a narrow ($\nu = 0.12$) spectrum. In Fig. 5 we have plotted the expectation value of maximum wave height (a key parameter to predict for extremes), as function of the number of upcrossings at the significant level during the time interval T_L (The number of degrees of freedom N_{slc} for short). From the comparison between the Monte Carlo results and theory (Eqns. (7) and (12) it is clear that the agreement is excellent and it seems that the suggested approach works for a wide range in the number of degrees of freedom, even, surpisingly, for small values of N_{slc}. In addition, although we have chosen widely differing spectral widths it is clear from the universal behaviour displayed in Fig. 5 that the choice $\nu\overline{\omega}$ for scaling of the number of degrees of freedom is correct. An alternative choice, e.g. using the number of waves $N = T_L/T_p$ (with T_p the peak period), does not give rise to proper scaling behaviour as is plainly evident from the Figure.

4 Extension to weakly nonlinear waves.

Here, the work is presented to extend the probability for extreme wave height into the weakly nonlinear regime. In particular, we concentrate on the pdf of *maximum* envelope wave height and it will be shown how, given the statistics of envelope wave height (including nonlinear effects) the pdf of the maxima may be obtained. Such an extension into the nonlinear regime is not so easy to obtain for the maximum zero crossing wave height (at least I do not know how to do this), but a similar approach could be followed for the maximum in surface elevation.

Janssen (2014) has obtained a new form for the pdf of wave height starting from the well-known Edgeworth distribution. It reads

$$p(h) = 4he^{-2h^2}\left\{1 + \frac{\kappa_4}{8}\left(2h^4 - 4h^2 + 1\right) + \frac{\kappa_3^2}{72}\left(4h^6 - 18h^4 + 18h^2 - 3\right)\right\}. \tag{15}$$

where

$$\kappa_4 = \kappa_{40} + \kappa_{04} + 2\kappa_{22}.$$

and

$$\kappa_3^2 = 5(k_{30}^2 + \kappa_{03}^2) + 9(\kappa_{21}^2 + \kappa_{12}^2) + 6(\kappa_{30}\kappa_{12} + \kappa_{03}\kappa_{21})$$

Here wave height is normalized with $4m_0^{1/2}$ where $m_n, n = 0, 1, 2, ..$ is the n^{th} moment of the frequency spectrum. Hence, $h = 1$ corresponds to the significant wave height level. The κ's refer to certain third and fourth-order cumulants of the joint pdf of the surface elevation and its Hilbert transform. They are given in Janssen (2014) and in the present treatment they are assumed to be given, so the cumulants κ_3 and κ_4 are known. Assuming that the dynamics of the nonlinear waves is governed by nonlinear four-wave interactions only, the even cumulants (for $n > 0$) consist of two contributions, namely one coming from the bound waves and one coming from the dynamics of the free waves, while the odd cumulants are determined by the presence of the bound waves only. The contribution by the free waves, however, occurs very rarely because for it to be important a coherent sea state is required, i.e. one that has a spectrum that is both narrow in frequency and direction. But when such a coherent sea state occurs large deviations from Normality may result and it is then justified to call these sea states 'freakish'.

The difference with previous work is the additional term proportional to κ_3^2 expressing the effect of skewness on the pdf for (envelope) wave height. This is in particular relevant for the contribution of bound waves to the deviations of statistics from Normality, as bound waves give rise to a considerable skewness. On the other hand, it should be noted that the skewness of free waves is very small and therefore for really extreme events the skewness correction to the wave height pdf is not so important. However, on average, the bound waves will determine the statistics of waves and, therefore, in order to have an accurate desciption of the average conditions the skewness effect needs to be included as well.

In order to derive the pdf of maximum wave height the cumulative wave height distribution

$$P(h) = \int_h^\infty \mathrm{d}h\; p(h),$$

is required. Integration of $p(h)$ from (15) gives

$$P(z) = e^{-2z}\left[1 + \kappa_4 A(z) + \kappa_3^2 B(z)\right], \tag{16}$$

with $z = h^2$, while $A(z) = z(z-1)/4$ and $B(z) = z(2z^2 - 6z + 3)/36$. Because the cumulative distribution function just depends on $z = h^2$ it turns out that the most compact form of the maximum wave height distribution is in terms of the linear wave energy parameter z. Hence, from Eq. (7) one obtains

$$p_{max}(z = h_{max}^2) = \frac{\mathrm{d}\mathscr{G}}{\mathrm{d}z}\exp(\mathscr{G}), \tag{17}$$

and once more $\mathscr{G} = -NP(z)$. The number of events N follows from (12), which for ease of reading is displayed here once more, i.e.

$$N = \alpha\nu\bar{\omega}T_L. \tag{18}$$

Remark that strictly speaking one should estimate the number of degrees of freedom using the statistics including nonlinear effects. This means extending the results displayed in Eqns. (8) and (10). This is straightforward for the probability that the envelope wave height exceeds h_c as the nonlinear extension is given in Eq. (16), however, the nonlinear extension of the joint pdf of wave height h and its time derivative $\dot{h}$ requires a substantial amount of additional work, which has not been done so far.

We have seen that the simplest description of the maxima pdf is in terms of the wave energy. It would therefore be natural to evaluate the expectation value of maximum wave energy, i.e. $\langle z\rangle$. However, since normally extreme seas are characterized by maximum wave height I will evaluate quantities such as the expectation value of h and the width σ of the distribution. It will be found that the width σ of the distribution function (17) is small, so the maximum wave height distribution is quite narrow, which implies that $\langle z\rangle^{1/2}$ and $\langle z^{1/2}\rangle$ may be interchanged without much loss of accuracy.

The expectation value of maximum envelope wave height $\langle h\rangle = \langle z^{1/2}\rangle$ can be found in the limit of large N and small κ_3 and κ_4 in the following manner. By definition

$$\langle h\rangle = \langle z^{1/2}\rangle = \int_0^\infty \mathrm{d}z\, z^{1/2} p_{max}(z).$$

Changing from integration variable z to $x = -\mathscr{G}$ and using the expression for the maximum wave height distribution in (17) gives

$$\langle h\rangle = \int_0^N \mathrm{d}x\, z^{1/2}(x) e^{-x}. \tag{19}$$

and $z = z(x)$ is obtained by solving the relation between x and z, i.e.

$$x = Ne^{-2z}\left(1+\kappa_4 A(z)+\kappa_3^2 B(z)\right)$$

with a perturbation approach. Take the log and rearrange, then

$$z = \frac{1}{2}\log\left(\frac{N}{x}\right) + \frac{1}{2}\log\left(1+\kappa_4 A+\kappa_3^2 B\right)$$

and for small κ_4 and κ_3 one finds in good approximation

$$z = z_0 + \frac{1}{2}\log\left(1+\kappa_4 A(z_0)+\kappa_3^2 B(z_0)\right),\; z_0 = \frac{1}{2}\log\left(\frac{N}{x}\right).$$

Now, utilizing that z_0 is a large quantity, $z^{1/2}$ becomes

$$z^{1/2} = z_0^{1/2} + \frac{1}{4z_0^{1/2}}\log\left(1+\kappa_4 A(z_0)+\kappa_3^2 B(z_0)\right),\; z_0 = \frac{1}{2}\log\left(\frac{N}{x}\right). \tag{20}$$

As a consequence, using (20) in (19) gives

$$\langle h\rangle = \int_0^N \mathrm{d}x\left[z_0^{1/2} + \frac{1}{4z_0^{1/2}}\log\left(1+\kappa_4 A(z_0)+\kappa_3^2 B(z_0)\right)\right]e^{-x} = z_1 + z_2. \tag{21}$$

Consider the first integral

$$z_1 = \int_0^N \mathrm{d}x\, z_0^{1/2} e^{-x} = \frac{1}{\sqrt{2}}\int_0^N \mathrm{d}x\, e^{-x}\sqrt{\log N - \log x}.$$

Because of the exponential function in the integrand contributions to the integral for $x = \mathscr{O}(N)$ are exponentially small and therefore one may expand the square root term in the above integral as if $\log x$ is small. As a result one finds

$$z_1 \simeq \int_0^\infty \mathrm{d}x\, e^{-x} \left\{ y_0 - \frac{\log x}{4y_0} \right\}.$$

with $\hat{z}_0 = \frac{1}{2}\log N$ and $y_0 = \hat{z}_0^{1/2}$, while the upper bound N of the integral is replaced by ∞ as this also only introduces an exponentially small term. According to Gradshteyn and Ryzhik (1965) one has

$$\int_0^\infty \mathrm{d}x\, e^{-x} \log x = -\gamma$$

where $\gamma = 0.5772$ is Euler's constant and therefore

$$z_1 = y_0 + \frac{\gamma}{4y_0}$$

Consider now the second integral in (21). Since this is a small contribution, the factor z_0 may be replaced by $\hat{z}_0$, and therefore

$$z_2 = \frac{1}{4y_0} \int_0^N \mathrm{d}x\, \log\left[1 + \kappa_4 A(\hat{z}_0) + \kappa_3^3 B(\hat{z}_0)\right] e^{-x}$$

Utilizing once more the assumption that κ_3^2 and κ_4 are small the logarithm is expanded. Elimination of z_0 and rearrangement then gives

$$z_2 = \frac{1}{4y_0}\left(\alpha\kappa_4 + \beta\kappa_3^2\right)$$

where

$$\alpha = \frac{1}{8}\left\{ 2\hat{z}_0(\hat{z}_0 - 1) + (1 - 2\hat{z}_0)G_1 + \frac{1}{2}G_2 \right\}$$

and

$$\beta = \frac{1}{72}\left\{ \hat{z}_0(4\hat{z}_0^2 - 12\hat{z}_0 + 6) - (6\hat{z}_0^2 - 12\hat{z}_0 + 3)G_1 + 3(\hat{z}_0 - 1)G_2 - \frac{1}{2}G_3 \right\},$$

where in the integrals the upper bound N has once more be replaced by ∞. The symbol G_n denotes integrals involving exponentials and logarithms. These are related to the Gamma function $\Gamma(1+z)$ and its derivatives,

$$\Gamma(1+z) = \int_0^\infty t^z e^{-t} \mathrm{d}t = \int_0^\infty e^{z\log t} e^{-t} \mathrm{d}t,$$

and therefore

$$G_n = \left.\frac{\mathrm{d}^n}{\mathrm{d}z^n}\Gamma\right|_{z=0} = \int_0^\infty \log^n t\, e^{-t} \mathrm{d}t, n = 1, 2, 3, \ldots$$

It may be shown that $G_1 = \Gamma'(1) = -\gamma$, $G_2 = \Gamma''(1) = \gamma^2 + \frac{\pi^2}{6}$, and $G_3 = \Gamma'''(1) = -2\zeta(3) - \gamma^3 - \gamma\pi^2/2$. Here, $\zeta(3) = 1.20206$ is the Riemann zeta function with argument 3. Now, returning to the logarithmic form and combining the results for z_1 and z_2 one finds for $\langle h \rangle$

$$\langle h \rangle = y_0 + \frac{1}{4y_0}\left\{ \gamma + \log\left[1 + \alpha(\hat{z}_0)\kappa_4 + \beta(\hat{z}_0)\kappa_3^2\right] \right\}, \tag{22}$$

The analytical result has been compared with results of numerical computations of $\langle h\rangle$ and the agreement is astonishingly good. Notice that the assumption has been made that skewness and kurtosis are small, in agreement with the assumptions on weakly nonlinear waves. Therefore, in the operational model, when $\langle h\rangle$ is computed, the kurtosis is restricted to the range $-0.33 < \frac{1}{8}\kappa_4 < 1$.

Next, a sketch is given of how the width σ of the maximum wave height distribution has been obtained. By definition

$$\langle z\rangle = \sigma^2 + \langle h\rangle^2,$$

therefore the expectation value of z denoted by $\langle z\rangle$ and defined as

$$\langle z\rangle = \int_0^N \mathrm{d}x\, ze^{-x} = \int_0^N \mathrm{d}x\, e^{-x}\left\{z_0 + \frac{1}{2}\log\left[1+\kappa_4 A(z_0) + \kappa_3^2 B(z_0)\right]\right\}$$

is needed. The integrations can be performed in a similar fashion as before, and for linear waves (i.e. $\kappa_3 = \kappa_4 = 0$) the relative width $\sigma/\langle h\rangle$ becomes

$$\frac{\sigma}{\langle h\rangle} \simeq \frac{\pi}{2\sqrt{6}\left(\log N + \frac{1}{2}\gamma\right)}.$$

For a typical choice of the number of waves, $N = 100$, the relative width is found to be around 13%. The difference between $\langle z\rangle^{1/2}$ and $\langle z^{1/2}\rangle$ then turns out to be less than 1%. In this sense the maximum wave height distribution is narrow, allowing the interchange of $\langle z\rangle^{1/2}$ and $\langle z^{1/2}\rangle$ without much loss of accuracy.

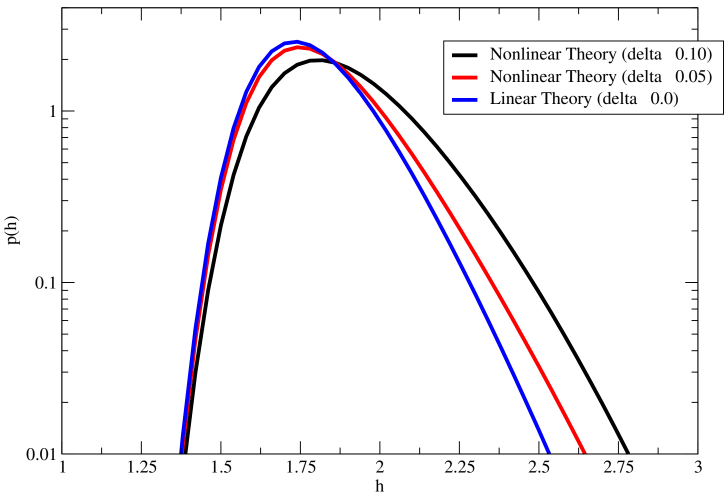

Figure 6: Effect of nonlinearity on maximum wave height pdf. Comparison of the nonlinear case with significant steepness $\Delta = 0.1$, and $\Delta = 0.05$ with the linear case.The length of the time series is 100 waves.

The approach that has been followed here is valid for arbitrary values of skewness and kurtosis, albeit that these statistical moments should be sufficiently small in order to agree with the assumption of weakly nonlinear waves. Let us now study the impact of nonlinearity on the shape of the maximum wave height distribution. Effects of nonlinearity are expressed by the skewness parameter κ_3 and the kurtosis parameter κ_4 shown in Eq. (15). We only study, as in Janssen (2014), the most simple case, which is the one for a narrow band wave train, and we only consider the effect of the bound waves. Introducing the significant slope parameter

$$\Delta = k_p\sqrt{m_0}, \tag{23}$$

one obtains for kurtosis and the square of the skewness

$$\kappa_3^2 = 72\Delta^2, \ \kappa_4 = 24\Delta^2. \tag{24}$$

The resulting modifications on the maximum wave height distribution are shown in Fig. 6. Apart from linear theory two cases of effects of finite steepness are shown. The case $\Delta = 0.1$ is quite extreme and corresponds to a very young windsea with a Phillips parameter $\alpha_p = 0.04$, while the second case corresponds to old windsea with a Phillips parameter $\alpha_p = 0.01$. Surprisingly, even for the frequently occuring old windsea case one notices an increase in the tail of the pdf (by a factor of three at $h = 2.5$). For the extreme young windsea case there is a dramatic impact on the tail of the the maximum pdf, showing an increase of probability by a factor of 10 for $h = 2.5$. On the other hand, by comparison, the expectation value of maximum wave height is fairly insensitive to increases in slope Δ as $\langle h \rangle$ only increases from 1.8 until 1.9 for an increase in Δ from 0 to 0.1. Therefore, the expectation value of h is clearly not given the whole story but its usefulness derives from the fact that, while realizing that extremes are rare, one only needs a few samples to obtain an estimate of the 'correct' value.

Although the theoretical wave height pdf $p(h)$ (15), including nonlinear effects, has been validated against Monte Carlo simulations by Janssen (2014), so far this has not been done for the maximum wave height distribution $p_{max}(h)$. Realizing, however, that in the tail of the distribution $p_{max}(h) = Np(h)$ we can have some confidence in the size of the impact of nonlinearity on the tail of the maximum wave height distribution. In addition, it is noted that the expression for skewness and kurtosis given in (24) considerable underestimates the corresponding values for a broad band spectrum (cf. Janssen (2009)). Finally, it is noted that for a coherent sea state (although their occurence is rare) non-resonant four-wave interactions may have a dramatic impact on the maximum wave height distribution as this may result in kurtosis values of $\mathscr{O}(1)$ which are usually much larger than the contribution due to the bound waves. All this suggests that nonlinear effects play an important role in understanding the formation of extreme sea states.

5 Operational implementation and Conclusions.

In this memo I have sketched how to obtain from the envelope wave height statistics the maximum wave height distribution function. The starting point is the approach of Goda which essentially gives the maximum wave height pdf for a given number of independent extreme events or the number of degrees of freedom N. In order to obtain N one basically performs an extreme value analysis on the envelope time series by identifying extreme events making use of the number of upcrossings at a conveniently chosen envelope wave height level h_c. It turns out that the number of degrees of freedom N is closely related to the number of wave groups in a wave train, and, hence N scales with the scale $\nu\bar{\omega}_1$, where ν is a measure of the width of the angular frequency spectrum while $\bar{\omega}_1$ is the mean angular frequency. The resulting maximum wave height pdf compares favourable with Monte Carlo simulations for narrow and broad wave spectra. In particular, the simulations do suggest that the number of degrees of freedom scales with the number of wave groups as shown in Fig. 5 but not with the number of waves.

It is important to note that we have obtained results for the envelope wave height distribution as the envelope wave height is thought to be physically relevant when one is interested in forces/drag on oil riggs and ships. The usual analysis of extreme waves is performed by means of a zero crossing method for wave height. The zero crossing method underestimates, see Fig. 1, the probability of extreme forces on a structure and as a consequence the expectation value of the maximum wave height pdf shifts to much lower values (by 0.4 units for 10 waves) of maximum wave height (see Fig. 2) compared to the envelope

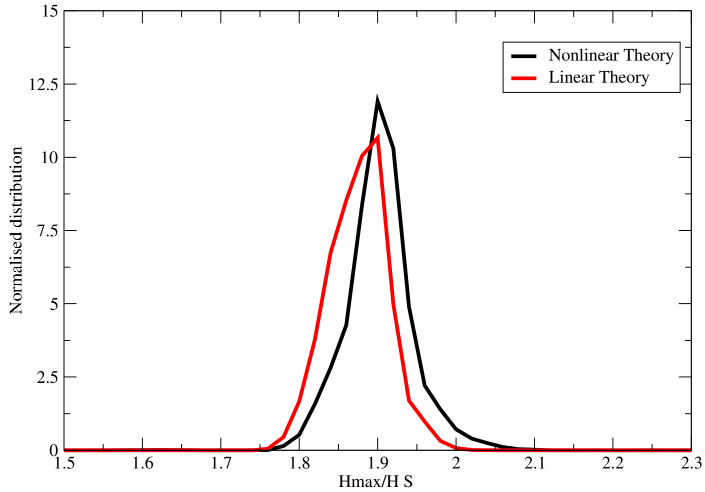

Figure 7: Geophysical distribution of normalised maximum waveheight obtained from one global spectral field with spatial resolution of 1°*. The effect of nonlinearity on the distribution is shown as well. The length of the time series is 20 minutes.*

distribution. Therefore, one should be careful when comparing observed maximum wave heights with this theoretical model using the envelope wave height. For a fair comparison, also observed maximum wave heights should be obtained using the same time series analysis. Although the latter method is perhaps not routinely used in observational practice, it should be pointed out that it has been tried before with considerable success on actual observed time series by Shum and Melville (1984).

In a certain number of aspects the present results are uncertain. I have done numerical experiments to try to decide whether Goda's approach or Naess' approach is more adequate. At present, as discussed in the concluding part of §3, there is no conclusive evidence in favour of one or the other approach. In addition, in order to estimate the number of degrees of freedom in the weakly nonlinear case the appropriate joint pdf is required. In principle this can be obtained but it requires considerable effort. Nevertheless, for the purpose of estimating the expectation value of maximum envelope wave height the chosen approach has been shown to be adequate over a wide range of the number of degrees of freedom as follows from Fig. (5).

The findings in this memo are the basis of an update to the present ECMWF freak wave warning system (Janssen & Bidlot, 2009). These updates have been introduced in Cy40R3, and are used routinely since the operational introduction of CY41R1. I have simply introduced the Eqns. (7) and (12) (with $\alpha = \sqrt{4/\pi}$) while the expectation value of maximum wave height is given by (22). The length of the time series was chosen to be 20 minutes, in agreement with the length of a typical buoy time series. In order to appreciate the effect of nonlinearities I have plotted for an arbitrary date the geophysical distribution of the normalized maximum wave height in Fig. 7. Effects of nonlinearity are quite considerable realizing that the 'critical' normalized wave height equals 2. The expectation value of maximum wave height is the key indicatior for extreme events in our freak wave warning system. However, it may also be of interest to generate the actual maximum wave height pdf at the location of interest. Using (7) this can be done as all relevant parameters such as the first three moments of the spectrum and skewness and kurtosis are archived at ECMWF.

Acknowledgement. Discussions with Jean-Raymond Bidlot are much appreciated.

References

Benjamin, T.B., and J.E. Feir, 1967. The disintegration of wavetrains on deep water. Part 1. Theory. *J. Fluid Mech.* **27**, 417-430.

Cramer, H. and M.R. Leadbetter, *Stationary and Related Stochastic Processes*, John Wiley, New York, 1967.

Davenport, A.G., 1964. Note on the distribution of the largest value of a random function with application to gust loading. *Proc. Inst. Civil Engrs.* **28**, 187-196.

Dean, R.G., 1990. Freak waves: A possible explanation. In A. Torum & O.T. Gudmestad (Eds.), *Water Wave Kinematics* (pp. 609-612), Kluwer.

Draper, L., 1965. 'Freak' ocean waves. *Marine Observer* **35**, 193-195.

Elgar, S., R.T. Guza and R.J. Seymour, 1984. Groups of Waves in Shallow Water. *J. Geophys. Res.* 89, 3623-3634.

Ewing, J.A., 1973. Mean length or runs of high waves. *J. Geophys. Res.* **78**, 1933-1936.

Goda, Y., 2000. *Random seas and Design of Maritime Structures*. 2nd ed. World Scientific, 464 pp.

Gradshteyn, I.S. and I.M. Ryzhik, 1965. Tables of integrals, series, and products, Academic Press Inc, New York and London.

Janssen, P.A.E.M., 2003. Nonlinear Four-Wave Interactions and Freak Waves. *J. Phys. Oceanogr.* **33**, 863-884.

Janssen, P.A.E.M., 2009. Some consequences of the canonical transformation in the Hamiltonian theory of water waves. *J. Fluid Mech.***637**, 1-44.

Janssen, P.A.E.M., 2014. On a random time series analysis valid for arbitrary spectral shape, *J. Fluid Mech.* **759**, 236-256.

Janssen, P.A.E.M., and J.-R. Bidlot, 2009. On the extension of the freak wave warning system and its verification. ECMWF Technical Memorandum 588.

Lake B.M., H.C Yuen, H. Rungaldier, and W.E. Ferguson, 1977. Nonlinear deep-water waves: Theory and experiment. Part 2. Evolution of a continuous wave train, *J. Fluid Mech.* **83**, 49-74.

Longuet-Higgins, M.S., 1983. On the joint distribution of wave periods and amplitudes in a random wave field. *Proc. Roy. Soc. London* **A389**, 241-258.

Naess, A., 1982. Extreme value estimates based on the envelope concept. *Applied Ocean research*, **1982**, 181-187.

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

Osborne, A.R., M. Onorato, and M. Serio, 2000. The nonlinear dynamics of rogue waves and holes in deep water gravity wave trains. *Phys. Lett. A* **275**, 386-393.

Shum, K.T. and W.K. Melville, 1984. Estimates of the Joint Statistics of Amplitudes and Periods of Ocean Waves Using an Integral Transform Technique. *J. Geophys. Res.* **89**, 6467-6476.

Stansell, P., 2005. Distributions of extreme wave, crest and trough heights measured in the North Sea.

Ocean Engineering **32**, 1015-1036.

Trulsen K., and K. Dysthe, 1997. Freak Waves-A Three-dimensional Wave Simulation. in *Proceedings of the 21st Symposium on naval Hydrodynamics*(National Academy Press), pp 550-558.

Wolfram J., and B. Linfoot, 2000. Some experiences in estimating long and short term statistics for extreme waves in the North Sea. Abstract for the *Rogue waves 2000 workshop*, Ifremer, Brest.

Yasuda, T., N. Mori, and K. Ito, 1992. Freak waves in a unidirectional wave train and their kinematics. *Proc. 23rd Int. Conf. on Coastal Engineering*, Vol. 1, Venice, Italy, American Society of Civil Engineers, 751-764.

Zakharov, V.E., 1968. Stability of periodic waves of finite amplitude on the surface of a deep fluid. *J. Appl. Mech. Techn. Phys.* **9**, 190-194.