Technical Memo

ECMWF
European Centre for Medium-Range Weather Forecasts

930

Review of Freak Wave Research at ECMWF

Peter A.E.M. Janssen
(Research Department)

August 2025

Series: ECMWF Technical Memoranda

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

Contact: library@ecmwf.int

Abstract

The ECMWF freak wave warning system is based on a random time series analysis using the envelope wave height. The key quantity of interest is the maximum envelope wave height distribution which for (weakly) nonlinear systems depends on envelope skewness and kurtosis. Expressions for skewness and kurtosis for general spectra are presented, but they are very cumbersome. Therefore, a simple parametrization of skewness and kurtosis of the bound waves and the contribution of the free waves to kurtosis is presented. These parametrisations are validated against computations using the theoretical expressions of the statistical parameters. By adjusting a few parameters in the parametrizations a good agreement with the 'exact' computations is found. These parametrizations have been introduced in the operational version of ECMWF's freak wave warning system.

Plain Language Summary

An ocean wave event becomes extreme when a sufficient number of wave components of the sea state are coherent, i.e. have more or less the same phase so that the corresponding contribution to the surface elevation of the different components goes up at the same time and location. The rare event of a coherent sea state can be caused by linear effects such as constructive interference while also nonlinear four-wave interactions, where one wave becomes big because it receives energy from the neighbouring waves, may give rise to large amplitude waves.

However, modern wave prediction systems only predict the energy of the individual waves and not their phase. Attempts to predict the phases of the individual waves fail because after a few hundred periods the sea state becomes chaotic. Therefore, one can only hope that statistical prediction methods may provide information on the likelyhood that extreme events occur.

The envelope of the sea surface elevation is a measure for the local energy of the waves and in first approximation the statistics of the envelope obeys Gaussian statistics (Linear Theory). This approximation has been used sofar to provide estimates of likelyhood of extreme events. Nevertheless, recently it has been realized that nonlinear effects, such as the generation of bound waves and four-wave interactions may give rise to a considerable increase in the likelyhood of extreme events. For particular events, such as the Draupner case or the Andrea wave, the increase is about a factor of 10, while on average the increase is a factor of two to three.

In this Technical Memorandum a description is given of the development of the ECMWF Freak Wave Warning system where, in particular, the focus is on the introduction of a number of nonlinear effects in the calculation of the likelyhood of extreme events.

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 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 present-day understanding of the generation of extreme events follows to a considerable extent from recent developments in (weakly) nonlinear waves. Apart from the field of ocean waves, Freak waves have been reported in, for example, liquid helium, in nonlinear optics, and in microwave cavities. The common denominator between these fields is that they all concern the spatio-temporal dynamics of narrow-band wave packets in a nonlinear dispersive medium which can be modelled by the nonlinear Schrödinger equation.

Here, we will concentrate on ocean waves. 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 narrow-band versions of the Zakharov equation which are closely related to the nonlinear Schrödinger 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).

Nowadays, we judge extreme events by means of the probability distribution function (p.d.f.) of wave height and maximum wave height. Although for linear ocean waves the surface elevation p.d.f. will be close to a Gaussian, finite amplitude ocean waves give rise to deviations from Normality. There are two reasons for this. 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 troughs) and kurtosis, which is a unique function of the square of the skewness: for narrow-band, deep-water waves one finds that kurtosis equals twice the square of the skewness. The resulting deviations from Normality will always occur as long as the waves are sufficiently 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 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. This mechanism provides a plausible explanation for freak wave formation, which is in the present context definitely a rare event. It should be noted that Toffoli *et al.* (2024) have recently in quite some detail analyzed a case in the Southern Ocean which was shown to be exceptional because the ratio of kurtosis and the square of the skewness was much larger than according to well-know statistical relations for the bound waves, suggesting that four-wave interactions were the cause of the generation of this extreme event.

In ocean wave forecasting practice one follows a stochastic approach because the phase of the individual

waves is unknown. And if the phases of the waves were known, the prediction horizon is very much limited because the phase shows a sensitive dependence on errors in the initial condition leading to chaotic behaviour (Annenkov and Shrira, 2001). Here, 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 frequency-direction spectrum $F(\omega,\theta)$ 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 knowledge of the the maximum wave height distribution for given spectral shape and statistical moments.

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. However, modern wave prediction systems do not predict the zero-crossing wave height. These models determine the evolution of the frequency-direction spectrum, and the significant wave height H_S is defined in such a way that it is directly connected to the energy of the waves, i.e. $H_S = 4\sqrt{m_0}$ with m_0 the wave variance which is the zeroth moment of the two-dimensional wave spectrum. The wave energy then equals $\rho g m_0$, with ρ the water density and with g acceleration of gravity. It is customary to omit the factor ρg in the discussions regarding wave forecasting and here this custom is followed so that the terms wave energy and wave variance can be used interchangeably.

Consistent with present-day modelling practice, Janssen (2014) has shown that an alternative technique, which makes use of the envelope of the time series, may have certain advantages. First, the square of the envelope equals the local (linear) energy of the waves, which is a more interesting quantity then the zero-crossing 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 p.d.f. of the envelope height follows the Rayleigh distribution, which only depends on the zeroth moment m_0. However, no such simple functional form for the p.d.f. of the zero-crossing wave height is known. From Monte Carlo simulations for linear waves it is shown that the envelope wave height indeed follows the Rayleigh distribution, while the p.d.f. of zero-crossing wave height depends on spectral shape and only approaches the Rayleigh distribution from below for narrow spectra. Given the p.d.f. for envelope wave height a theoretical expression for the maximum wave height p.d.f. is derived and this expression is validated against results from Monte Carlo simulations.

In this review, a description is given of the work that has been done to use the theoretical approach to estimate extreme events into a practical application which can be used in, for example, the ECMWF freak wave warning system. In §2 a brief overview of the method is given. Starting point for this is Janssen (2014) which develops an analysis of time series based on the envelope ρ. For convenience, and in order to stay close to oceanographic practice, the envelope wave height h is introduced which is defined to be twice the envelope height, i.e., $h = 2\rho$. If the effects of nonlinearity are small, the envelope wave height p.d.f. may be obtained by means of a Taylor expansion of the logarithm of its generating function. This basically gives an expansion around the Rayleigh distribution where the relevant expansion coefficients are the third-order (skewness) and fourth-order (excess kurtosis) cumulants of the random envelope. Following this approach, envelope skewness and excess kurtosis for the bound waves can be obtained from the wave spectrum following a procedure described in Janssen (2009), who applied it to obtain the surface elevations statistics, while the contribution from the free waves is obtained from Janssen (2003), Mori and Janssen (2006), and Janssen and Janssen (2018).

This approach works well as follows from comparisons with p.d.f.'s observed in the laboratory and from comparisons with Monte Carlo simulations. Nevertheless, this statistical theory has a restricted range of validity. This follows from a comparison with maximum wave height data obtained in the field by

Janssen and Bidlot (2009) (see their Fig. 8) where it is evident that the theoretical p.d.f. starts to deviate from the observed one for extreme sea states with $h_{max}/H_S > 2.5$. Since extreme events, such as the Draupner wave, have a maximum envelope wave height that is about three times the significant wave height, it follows that this event is clearly outside the range of validity of the nonlinear approach and an extension of the domain of validity is required. Nowadays there is evidence that for very extreme states the p.d.f. has an exponential tail. Particularly convincing evidence is found from experimental work in nonlinear optics and liquid crystals. The work of Montina *et al.* (2009) suggests that the statistics of the fluctuations can be approximated by a simple empirical form, a stretched exponential distribution. Using this distribution to model the tail of the p.d.f., Janssen (2015b) has shown that this results in a satisfactory agreement with Monte Carlo simulations of a random weakly nonlinear Stokes wave down to values of the simulated p.d.f. of order 10^{-7}.

Given the form of the p.d.f. of envelope wave height, the next step is to establish the maximum envelope wave height probability distribution. In the field of ocean waves, the maximum wave height distribution proposed by Goda (2000) has been used frequently. In Goda's approach an important parameter is the number of events N, of which a choice needs to be made which introduces some arbitrariness. An alternative has been suggested by Naess (1982). His method is based on the reasonable assumption that for envelope time series the number of level upcrossings by the envelope $\rho(t)$ is asymptotically Poisson distributed when the level height increases. In Janssen (2015a) it has been shown that, in particular for the tail of the maximum wave height distribution, Naess's idea yields for linear waves a very good agreement with Monte Carlo simulations of the p.d.f. of maximum envelope wave height. Therefore, it was decided to use the Naess approach in the operational implementation.

Finally, work is presented that is required to develop an operationally feasible warning system. It should be realized that the expressions for skewness and kurtosis introduced by Janssen (2003, 2009) involve the evaluation of four and six dimensional integrals which would result in far too long operational run times. For this reason, simple parametrizations of envelope skewness and kurtosis are introduced, which are based on results obtained from the narrow-band version of the theoretical expression for skewness and kurtosis. These parametrizations are validated against exact computations of the relevant statistical parameters for the Draupner wave and the Andrea Storm, and a reasonable agreement has been obtained.

2 The Envelope method.

Following Janssen (2014) a time series analysis based on the envelope wave height will be used. 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. This method, introduced by Rice (1945), Gabor(1946) and Longuet-Higgins (1983), will be called the envelope method and in Janssen (2014) it is shown that this method gives an accurate estimation of the joint p.d.f. of wave height and period for linear waves and that it is possible to extend the approach into the weakly nonlinear regime. To make things more quantitative let us introduce the local variance of the surface elevation time series η, i.e. $\mathscr{E} = \langle \eta^2 \rangle$ where the brackets denote an ensemble average. In addition, introduce the envelope ρ defined according to $\eta = \rho \cos\phi$. In practice, ρ is obtained from the timeseries of the surface elevation η and its Hilbert transform $\zeta = H(\eta) = \rho \sin\phi$, according to

$$\rho = \sqrt{\eta^2 + \zeta^2}, \tag{1}$$

where for stationary signals surface elevation and its Hilbert transform have the same variance. In terms of the envelope one finds for the local variance

$$\mathscr{E} = \frac{1}{2}\rho^2, \tag{2}$$

while the normalized local variance is given by $E = \mathscr{E}/\sigma^2$ with $\sigma^2 = m_0$ the averaged variance of the surface elevation and m_0 the zeroth moment of the frequency spectrum. The value $E = 1$ then corresponds to the case that the local wave variance equals the average wave variance in the domain of interest. Events such as the Draupner wave event and the Andrea storm have $E \simeq 20$ which illustrates that at the focal point there was a considerable amplication of wave energy, therefore these events are quite extreme. As an alternative measure, introduce the envelope wave height as twice the envelope height ρ and the normalized envelope wave height h becomes

$$h = \frac{2\rho}{4\sigma} \tag{3}$$

and the relation between normalized envelope wave height and local energy is $E = 2h^2$ so that for the Draupner event $h_{max} = 3.1$, while for the Andrea event $h_{max} = 3.26$. For linear waves that obey Gaussian statistics it can be shown that the envelope wave height distribution is the Rayleigh distribution. In terms of the normalised envelope wave height (3) one finds

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

The Rayleigh distribution is the starting point of further developments of the p.d.f. of weakly nonlinear ocean waves. Before we continue the discussion it is time to introduce some important quantities that characterize the sea state as determined by e.g. the angular frequency spectrum $E(\omega)$. One then introduces the spectral moments m_n

$$m_n = \int \mathrm{d}\omega\ \omega^n\, E(\omega), \tag{5}$$

and using these moments one may introduce measures for significant wave height $H_S = 4\sigma$ with the variance σ determined by the zero moment, i.e. $\sigma = \sqrt{m_0}$, while there are two definitions for the mean angular frequency, namely one based on the first moment, i.e. $\omega_1 = \langle\omega\rangle = m_1/m_0$ while an alternative mean frequency may be introduced using the second moment, i.e. $\omega_2 = \langle\omega^2\rangle^{1/2} = \sqrt{m_2/m_0}$. Furthermore, spectral shape can be measured by means of the spectral width parameter ν, introduced by Longuet-Higgins (1983). It is defined as

$$\nu = \sqrt{m_0 m_2/m_1^2 - 1}. \tag{6}$$

and this parameter is connected to ω_1 and ω_2 in the following manner: $\omega_2 = \omega_1(1+\nu^2)^{1/2}$.

2.1 Comparison with the Zero-Crossing method.

Traditionally, a key parameter to express the severity of the sea state has been related to the wave height, which for a single wave is defined as the distance between the crest and the trough of the wave. As typically many waves with different frequency and direction are present at sea, a statistical approach is usually followed. In practice, the wave height distribution is obtained 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 zero-crossing wave height

h_{zc} from the difference of the maximum and the minimum of the surface elevation η in the corresponding time interval. Thus, the zero-crossing wave height is basically determined by sampling with the zero-crossing frequency $(m_2/m_0)^{1/2}$, 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 p.d.f. is a Rayleigh distribution with a variance which depends on spectral width. This follows from Naess (1985) who found an approximate expression for the zero-crossing wave height distribution in the narrow-band limit which will be confirmed by Monte Carlo simulations of linear ocean waves presented here. In particular, for a narrow-band spectrum the zero-crossing wave height p.d.f. is found to be close to the Rayleigh distribution with variance given by $\sigma^2 = m_0$ while for broad-band spectra the effective variance is reduced so that extreme waves are, when compared to the standard Rayleigh distribution, less likely to occur.

In more detail, let t_{max} and $t_{min} = t_{max} + \tau$ be the times dat surface elevation has a maximum and a minimum respectively in the interval between two succesive zero-crossings of interest. The zero-crossing wave height h_{zc} is then given by $h_{zc} = \eta(t_{max}) - \eta(t_{min})$ and it is assumed that the statistics of h_{zc} is stationary so that parameters such as variance, skewness, etc, only depend on the time difference τ. It is then straightforward to obtain the variance $\sigma_{hc}^2 = \langle h_{zc}^2 \rangle$. One finds

$$\sigma_{hc}^2 = \alpha\sigma^2,\ \alpha = \frac{1}{2}\left(1 - r(\tau)\right),\ \tau = T_1/2, \tag{7}$$

where the factor α gives the departure of the effective variance from the narrow-band case, while $r(T_1/2) = R(T_1/2)/R(0) = R(T_1/2)/\sigma^2$ is the normalised autocorrelation function of the surface wave elevation η, and $T_1 = 2\pi/\omega_1$ s the dominant wave period[1]. Note that in the limit of narrow-band waves, the normalised correlation $r(T_1/2)$ approaches -1 so that the variance of zero crossing height becomes identical to the variance of the envelope height shown in Eq. (3).

Introducing the normalised zero-crossing wave height

$$\hat{h} = \frac{h_{zc}}{4\sigma_{hc}} \tag{8}$$

Naess (1985) has shown that the p.d.f. of zero-crossing wave height follows in the narrow-band approximation the Rayleigh distribution

$$p(\hat{h}) = 4\hat{h}e^{-2\hat{h}^2}. \tag{9}$$

However, only in the narrow-band limit, when the normalised correlation $r(T_1/2)$ approaches -1 and $\alpha \to 1$, the zero-crossing wave height p.d.f. equals the distribution for the envelope wave height.

It is emphasized here that the alternative technique, namely the envelope method, shows a different picture as theoretically and from Monte Carlo simulations (see e.g. Janssen, 2014) one finds that for a Gaussian sea state the p.d.f. of envelope wave height always follows the Rayleigh distribution, with variance independent of spectral shape.

We have illustrated these findings on the statistics of zero-crossing wave height and envelope wave height in Fig. 1. The details of the Monte Carlo simulations are given in Janssen (2014). Essentially, linear theory is assumed and the surface elevation for a spectrum of waves is obtained where the amplitudes are drawn from a given frequency 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 $2^{1/2}$

[1]Naess (1985) suggests to use as time scale $T_2 = 2\pi/\omega_2$, but I found a slightly better agreement with the Monte Carlo simulations by choosing T_1.

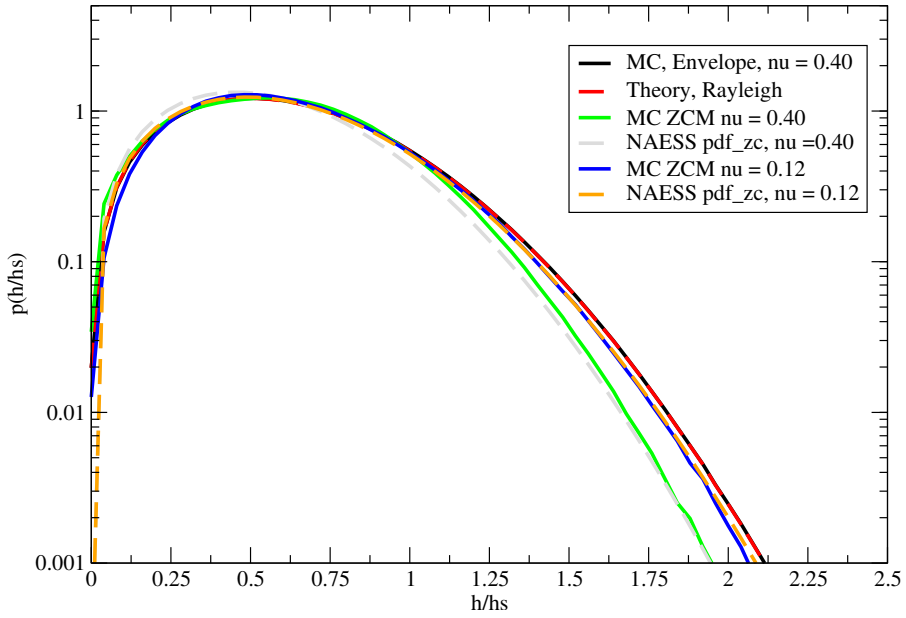

Figure 1: Comparison of p.d.f. of zero-crossing wave height (labelled ZCM) with p.d.f. of envelope wave height for different spectral width parameter ν. The envelope p.d.f. does not depend on ν, whilst the zero-crossing height p.d.f. does. For broad spectra (ν large) probabilities are underestimated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

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. Clearly, it takes many observations to obtain a reliable estimate of the zero-crossing waveheight, so this is definitely a weak point of this approach.

From Fig. 1 it is clear that the p.d.f. of zero-crossing wave height is a sensitive function of spectral width which, compared to the envelope waveheight results, underestimates for broad spectra the probability of extreme events by an order of magnitude. On the other hand, the p.d.f. of envelope wave height is independent of spectral width and is given by the Rayleigh distribution. There is a preference for the envelope method because it measures potential energy of the waves, information which is vital for determining extreme forces on oil rigs and ships. The zero-crossing waveheight would understimate these forces on structures. Therefore, from now onwards we concentrate on the envelope method.

We conclude the discussion by remarking that recently evidence has been presented by Häfner *et* al (2021), that the correlation between crest and trough, i.e. $r(T_1/2)$, is a good predictor for the presence of extreme Rogue waves (defined by the criterion that zero-crossing wave height h_{zc} normalised by significant wave height H_S is larger than 2). In view of the result of Naess (1985) this finding is fairly trivial because zero-crossing wave height will always depend on $r(T_1/2)$ whether there are extreme events or not. In fact, it is straightforward to determine from (8) and (9) the probability that $h_{zc}/H_S > 2$. One finds

$$P(h_{zc} > 2H_S) = e^{-16/(1-r(T_1/2))}. \tag{10}$$

and clearly this probability depends explicitely on $r(T_1/2)$. However, once more, it is argued that zero-crossing wave height is not the most appropriate measure to characterize the impact of an extreme event

on structures such as oil rigs and ships at sea. The envelope wave height seems to be a better indicator. For linear waves the probability that envelope wave height is larger than twice significant wave height is given by $P(h_{env} > 2H_S) = e^{-8}$ and is, of course, independent of the correlation between crest and trough. This result follows from Eq. (10) for $r(T_1/2) = -1$, which corresponds to the limit of narrow-band waves. Thus the envelope provides an upper-bound to the probability of extreme events. This choice follows the old adage that it is better to be safe than sorry.

2.2 Statistics in the weakly nonlinear regime.

In Janssen (2014) the envelope wave height p.d.f. for a weakly nonlinear sea state has been obtained assuming the simple time series description that $\eta = \rho\cos\phi$ where envelope and phase ϕ are obtained from the surface elevation η and its orthogonal complement ζ which follows from the Hilbert transform of η. The envelope wave height distribution is then obtained from the joint p.d.f. of η and ζ.

As a brief intermezzo it is noted that in statistics, a central role is played by the so-called generating or characteristic function $G(\mu)$, which is defined as the Fourier transform of the joint p.d.f. $p(\mathbf{x})$ where $\mathbf{x} = (x_1, x_2, ..., x_N)$ represents N parameters, i.e.

$$G(\mu) = \langle \exp \mathrm{i}\mu \cdot \mathbf{x} \rangle = \int \mathrm{d}\mathbf{x}\, p(\mathbf{x}) \exp(\mathrm{i}\mu \cdot \mathbf{x})$$

and the parameter μ represents the counterpart of $\mathbf{x}$ in Fourier space. The generating function contains all the statistical information on the stochastic process $\mathbf{x}$, e.g. the moments of the p.d.f. are related to derivatives of G with respect to μ at the origin. Hence, the moments are related to the coefficients of the Taylor expansion of G around the origin. This expansion is, however, not very useful because it does not bring out the significance of a special characteristic function, the one corresponding to a Gaussian distribution, which is great importance as linear waves have a normal distribution. Here, using from now onwards the Einstein summation convention, i.e. summation over repeated indices, the Gaussian characteristic function is given by

$$G_0 = e^{-\frac{1}{2}\mu_i\mu_j B_{ij}},$$

and by expanding G around G_0 the so-called cumulants of the distribution function are introduced. These are the coefficients of the Taylor expansion of the logarithm of G, i.e.

$$G = \exp\{-\frac{1}{2}\mu_i\mu_j B_{ij} - \frac{i}{3!}\mu_i\mu_j\mu_k C_{ijk} + \frac{1}{4!}\mu_i\mu_j\mu_k\mu_l D_{ijkl} + ...\},$$

where B_{ij} is the second-order cumulant, C_{ijk} is the third-order cumulant, related to the skewness and, finally, D_{ijkl} is the fourth-order cumulant which is connected to the excess kurtosis. Consistent with the small amplitude expansion for waves with a small steepness ε, there is an ordering in the magnitude of the cumulants, i.e.

$$B_{ij} = \mathcal{O}(\varepsilon^2), C_{ijk} = \mathcal{O}(\varepsilon^4), D_{ijkl} = \mathcal{O}(\varepsilon^6),$$

and the hope is that for sufficiently small ε the expansion converges. In order to obtain the required p.d.f. the Fourier transform of G is required, which for the above expresion for G is not straightforward. Therefore the exponential function is Taylor expanded. Adopting the above ordering one finds to fourth order in ε, thus retaining effects of skewness and kurtosis,

$$G \simeq G_0\left\{1 - \frac{i}{3!}\mu_i\mu_j\mu_k C_{ijk} + \frac{1}{4!}\mu_i\mu_j\mu_k\mu_l D_{ijkl} - \frac{1}{72}\left(\mu_i\mu_j\mu_k C_{ijk}\right)^2\right\}.$$

The corresponding expansion for the pdf $p(\mathbf{x})$ then follows from the Fourier transform of G, i.e.

$$p(\mathbf{x}) = \frac{1}{(2\pi)^N} \int \mathrm{d}\mu G(\mu) \exp(-i\mu \cdot \mathbf{x}).$$

Noting the usual rule for Fourier transformation, namely that each factor μ_i corresponds to $i\partial/\partial x_i$ and denoting the Fourier transform of G_0 by Φ_0 where

$$\Phi_0 = \frac{1}{(2\pi)^{N/2}|B|^{1/2}} \exp(-\frac{1}{2} x_i x_j B_{ij}^{-1})$$

the expansion for the pdf $p(\mathbf{x})$ becomes

$$p(\mathbf{x}) = \left\{ 1 - \frac{C_{ijk}}{3!} \frac{\partial^3}{\partial x_i \partial x_j \partial x_k} + \frac{D_{ijkl}}{4!} \frac{\partial^4}{\partial x_i \partial x_j \partial x_k \partial x_l} + \frac{C_{ijk}^2}{72} \left(\frac{\partial^3}{\partial x_i \partial x_j \partial x_k} \right)^2 \right\} \Phi_0(\mathbf{x}) \tag{11}$$

which is known as the Edgeworth expansion. Compared to the well-known Gram-Charlier expansion, used by e.g. Tayfun and Lo (1990), the difference is the additional term which is proportional to the square of the skewness parameter C_{ijk}. According to the order of magnitude of skewness and kurtosis, the last term of the Edgeworth expansion is, at least in the tail of the distribution on the scale $x_i = \mathcal{O}(1/\varepsilon)$, as important as the term involving the kurtosis D_{ijkl}. Therefore, it should be retained and it gives an important contribution to the tail of the wave height distribution as will be seen in a moment.

Two applications of the Edgeworth expansion are now given. The first one concerns the derivation of the p.d.f. of the surface elevation. This is the most simple case to deal with because only one variable is considered so that $N = 1$. Performing the differentiations for this case one finds as p.d.f. for the normalised surface elevation $x = \eta/\sigma$

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \left(1 + \frac{C_3}{6} H_3(x) + \frac{C_4}{24} H_4(x) + \frac{C_3^2}{72} H_6(x) \right) \tag{12}$$

where the Hermite polynomials are given by

$$H_3(x) = x^3 - 3x,\ H_4(x) = x^4 - 6x^2 + 3,\ \text{and}\ H_6(x) = x^6 - 15x^4 + 45x^2 - 15,$$

while the cumulants are normalised with the variance $\langle \eta^2 \rangle$, i.e. the skewness C_3 and excess kurtosisis C_4 are given by

$$C_3 = \frac{\langle \eta^3 \rangle}{\langle \eta^2 \rangle^{3/2}},\ C_4 = \frac{\langle \eta^4 \rangle}{\langle \eta^2 \rangle^2} - 3.$$

It is clear from Eq. (12) that for the surface elevation p.d.f. the main effect of nonlinearity is given by the skewness of the sea state while effects of kurtosis are relatively minor. In sharp contrast, we will see in a moment that the envelope wave height p.d.f. has a completely different dependence on nonlinear effects as the envelope is symmetric with respect to the mean sea surface so that there is no first order effect of skewness.

The second application concerns the derivation of the envelope wave height p.d.f., which follows from the joint p.d.f. $p(\eta, \zeta)$, so now $N = 2$. Recall that $\eta = \rho \cos\phi$ while $\zeta = \rho \sin\phi$, hence ρ is the envelope while ϕ is the phase of the surface elevation time series. After some straightforward algebra (for some details see Janssen (2014)) and an integration of the joint p.d.f. over ϕ from 0 to 2π (which cancels the term proportional to the skewness C_3), the p.d.f. of the envelope wave height $h = \rho/2\sigma$ becomes:

$$p(h) = 4he^{-2h^2} \left\{ 1 + C_4 \left(2h^4 - 4h^2 + 1 \right) + C_3^2 \left(4h^6 - 18h^4 + 18h^2 - 3 \right) \right\}. \tag{13}$$

where the parameters C_4 and C_3^2 are obtained from knowledge of the two-dimensional wave spectrum and are related to the kurtosis and skewness of the sea state. Here

$$C_4 = \frac{\kappa_4}{8}, \tag{14}$$

with $\kappa_4 = \kappa_{40} + 2\kappa_{22} + \kappa_{04}$ the envelope kurtosis, while the skewness factor becomes

$$C_3^2 = \frac{\kappa_3^2}{72}, \text{ with } \kappa_3^2 = 5(\kappa_{30}^2 + \kappa_{03}^2) + 9(\kappa_{21}^2 + \kappa_{12}^2) + 6(\kappa_{30}\kappa_{12} + \kappa_{03}\kappa_{21}).$$

In Janssen (2015b) it was shown that the skewness factor may be simplified because it was found that for general spectra $\kappa_{21} = \kappa_{03} = 0$ while $\kappa_{12} = \kappa_{30}/3$. As a consequence the skewness factor becomes

$$C_3^2 = \frac{\kappa_{30}^2}{9}. \tag{15}$$

The κ's refer to a number of cumulants of the joint distribution of the surface elevation $\eta = \rho\cos\phi$ and its Hilbert transform $\zeta = \rho\sin\phi$. Introducing the normalized moments

$$\lambda_{mn} = \frac{\langle \eta^m \zeta^n \rangle}{\langle \eta^2 \rangle^{m/2} \langle \zeta^2 \rangle^{n/2}},$$

the normalized skewness elements become

$$\kappa_{mn} = \lambda_{mn},\ m+n = 3$$

while the normalized kurtosis elements become

$$\kappa_{mn} = \lambda_{mn} + (m-1)(n-1)(-1)^{m/2},\ m+n = 4$$

In addition, it is noted that both free and bound waves may contribute to the cumulants. The general expressions for skewness and kurtosis parameters in terms of the wave spectrum are given in Appendix A, which also provides explicit expressions for the case of a narrow-band wave train.

From the wave height p.d.f., Eq. (13), one may then immediately obtain the p.d.f. of wave energy, since $p(h)\mathrm{d}h = p(E)\mathrm{d}E$ with $E = 2h^2$. The result is

$$p(E) = e^{-E}\left[1 + C_4 A(E) + C_3^2 B(E)\right], \tag{16}$$

where

$$A(E) = \frac{1}{2}E^2 - 2E + 1,\ B(E) = \frac{1}{2}E^3 - \frac{9}{2}E^2 + 9E - 3. \tag{17}$$

Finally, for the purpose of estimating the maximum wave height distribution the exceedance probability $P(E > E_c)$ is required. It follows from an integration of the p.d.f. (16) from E to infinity, with the result

$$P(E) = e^{-E}\left[1 + C_4 A(E) + C_3^2 B(E)\right], \tag{18}$$

where

$$A(E) = \frac{1}{2}E(E-2),\ B(E) = \frac{1}{2}E(E^2 - 6E + 6). \tag{19}$$

It should be realized, however, that the expressions for the p.d.f. and c.d.f. have only a restricted range of validity. For example, for given skewness and kurtosis the p.d.f. (13) underestimates the true p.d.f. for large envelope wave heights in the range $h > 2.5$. In other words, the theoretical approach fails just in the wave height range where a number of extreme events, such as the Draupner wave and the Andrea storm, have been reported. The range of validity of the theoretical approach therefore needs to be extended and this will be discussed next.

2.3 Behaviour of the tail of the p.d.f.

Nowadays there is ample evidence that for very extreme (sea) states the envelope wave height p.d.f. has an exponential tail, resulting, compared to the present theory, in much larger probabilities for extreme events. Evidence for an exponential tail follows from numerical simulations (e.g. Montina *et al.* (2009), Walczak *et al.* (2015), Janssen (2014), Janssen (2015b)), and comparison with field data (Janssen and Bidlot, 2009). Also, in nonlinear optics and liquid crystals a considerable amount of experimental evidence is available that suggests that this tail is exponential. It is therefore important to modify the present approach by adding an exponential tail.

For extreme waves in a nonlinear optical cavity Montina *et al.* (2009) have noted that the observed probability distribution function for intensity E can be well approximated by a stretched exponential form, involving a number of fitting parameters. Following this idea, it was realized, after some trial and error, that by using a stretched exponential to approximate the cumulative distribution function only one fitting coefficient was needed. The general form used to extend the range of validity of the theoretical c.d.f. (16) for normalised wave energy reads

$$P(E) = \int_E^\infty \mathrm{d}x\, p(x) = e^{-z},\; z = -\alpha + \sqrt{\alpha^2 + \beta E}, \tag{20}$$

where (see Janssen (2015b)) $\beta = 2(\alpha+1)$ from the condition that the ensemble average wave variance should equal the average in the domain of interest. Therefore, for matching purposes one only needs to determine the parameter α. The idea of using (20) was suggested to me by Residori (2015). Unfortunately, this form is not so easy to justify although the most simple nonlinear system that has a similar c.d.f. for wave energy is one that only has a first and second, bound harmonic (Janssen, 2017).

This simple form has some interesting properties. First of all, the condition $P(E=0) = 1$ is automatically satisfied so that the underlying p.d.f. $p(E)$ is normalized to 1. Second, for small E Taylor expansion of z gives $z = \beta E/(2\alpha)$ hence

$$\lim_{E\to 0} P(E) = e^{-\frac{\beta}{2\alpha}E}, \tag{21}$$

while for large E one finds

$$\lim_{E\to\infty} P(E) = e^{-\sqrt{\beta E}}. \tag{22}$$

Realizing that $E = 2h^2$ this means that in terms of envelope wave height we have for small E a Gaussian distribution while for large E the distribution is exponential.

By differentiation of the c.d.f. (20) it is straightforward to obtain the p.d.f. $p(E)$. By definition

$$p(E) = -\frac{\partial P}{\partial E}$$

so that

$$p(E) = \frac{\beta}{2(z+\alpha)} e^{-z},\; z = -\alpha + \sqrt{\alpha^2 + \beta E}.$$

It is then straightforward to obtain the envelope wave height p.d.f. from the condition $p(h)\mathrm{d}h = p(E)\mathrm{d}E$ and the result is

$$p(h) = \frac{2\beta h}{(z+\alpha)} e^{-z}. \tag{23}$$

In a moment it will be seen that this form of the envelope wave height p.d.f. is an adequate approximation for both the weakly nonlinear regime and for the tail of the distribution.

A relation for α is now obtained by matching the empirical c.d.f. (20) with the theoretical one, given in Eq. (16), which is denoted by P_{th}. The fitting constant α then follows from the condition that at the edge of the range of validity, taken as $E_b = 10$ (corresponding to $h = 2.2$), the empirical c.d.f. equals the theoretical one, i.e. $P(E_b) = P_{th}(E_b)$. This gives for α,

$$\alpha = \frac{f_b^2 - 2E_b}{2(E_b + f_b)}, \; f_b = \log P_{th}(E_b). \tag{24}$$

where f_b is the logarithm of the theoretical c.d.f. at the boundary given by $E = E_b$. In this manner a connection between skewness and kurtosis of the sea state, via f_b, and the fitting parameters of the empirical c.d.f. has been established. For relatively small values of C_4 and C_3 this matching procedure works well, as reported in Janssen (2015b).

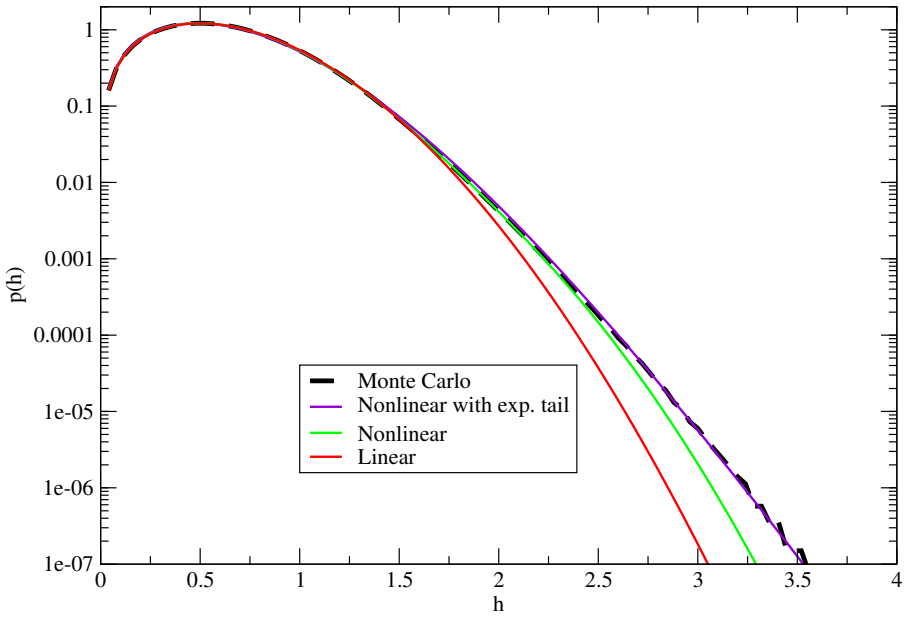

Figure 2: Probability distribution function of envelope wave height for a significant steepness of 0.06 and a dimensionless depth of 1.45, mimicking the Draupner wave event. The Monte Carlo simulation shows clear evidence that the tail of the distribution is exponential. Nonlinear theory combined with the stretched exponential of Residori is in good agreement with the simulation.

It is important to check the validity of the statistical distribution presented in this section by means of a Monte Carlo simulation, and this subject has been more extensively discussed in Janssen (2015b). Since the interest is in extreme events with probabilities of the order 10^{-6} the number of ensemble members N_{ens} needs to be quite large. By trial and error, $N_{ens} = 50,000,000$ is taken, in order to obtain smooth results for the p.d.f.. Assuming that the envelope wave height p.d.f. is determined by the skewness factor C_3 and the kurtosis factor C_4 only, one may use any nonlinear system in the numerical simulation as long as it has the same statistical parameters. For this reason a Monte Carlo simulation is performed with a Stokes wave train, where the amplitude a is drawn from a Rayleigh distribution while the phase is drawn from a uniform distribution. We have studied the statistical properties of the envelope wave height of this nonlinear system. The envelope wave height p.d.f., as obtained from the Monte Carlo Simulation, is shown in Fig. 2. Over a wide range of values, between 10^{-6} and 10^{-2}, the logarithm of

the p.d.f. behaves as a straight line, hence the p.d.f. follows an exponential law. For comparison, also shown are results according to linear theory, and it is clear that this gives a large underestimation of the frequency of extreme events. The nonlinear theory discussed by Janssen (2014) shows good agreement with the simulation up to a wave height of $h = 2.5$, but the probability of extreme events such as the Draupner case with $h = 3.1$ is considerably underestimated. The approach that adds, using the method detailed here, an exponential tail is found to agree very well with the numerical simulation. In fact, the violet line, labeled Nonlinear with exp. tail, corresponds to Eq. (23) with fitting parameter given in (24), therefore the stretched exponential captures both the weakly nonlinear regime with $h < 2.5$ and the tail region with $h > 2.5$. From now onwards the stretched exponential form wil be used for describing the statistical distribution of envelope wave height.

2.4 Maximum envelope wave height distribution: linear theory.

In order to derive the p.d.f. of maximum envelope wave height information is needed on the statistics of the envelope wave height time series and on 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, c.d.f.)

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

and the joint p.d.f. of envelope wave height h and its time derivative $\dot{h}$. Here, we concentrate on linear theory firstly and in the next subsection we sketch how to extend results for weakly nonlinear theory. As according to linear theory the envelope wave height distribution is given by the Rayleigh distribution (see Eq. (12)),

$$p(h) = 4he^{-2h^2}, \tag{25}$$

the cumulative distribution function becomes

$$P(h) = e^{-2h^2}. \tag{26}$$

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

$$p(h,\dot{h}) = \frac{8h}{\sqrt{2\pi}} e^{-2(h^2+\dot{h}^2)}, \tag{27}$$

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 the width of the frequency spectrum. Note that by considering the j.p.d. 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 (27) is valid for a Gaussian sea state and can be obtained in a straightforward manner from the well-known expression of the j.p.d. 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 (27). Alternatively, one may argue that h and $\dot{h}$ are independent and that the joint p.d.f. of h and $\dot{h}$ is simply the product of the marginal distributions of h and $\dot{h}$, i.e. the product of a Rayleigh distribution for h and a Gaussian distribution for $\dot{h}$.

In an earlier version of the ECMWF Freak Wave Warning system Goda's maximum wave height distribution has been used. Goda (2000) obtained the maximum wave height distribution from a surface elevation time series of N independent wave events. The number of waves would be obtained from the ratio of the length T_L of the time series divided by either the mean zero-crossing frequency or the peak frequency. A similar approach was already followed by Davenport (1964) in the problem of estimating the maximum gust. Here we start with Goda's (2000) approach but now applied to time series of the envelope wave height. Clearly, when obtaining the envelope wave height time series from the surface elevation signal the mean or peak oscillation frequency will be removed from our considerations and this will affect the estimate of the number of events. It will now involve the frequency scale $\nu\bar{\omega}$. 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 timeseries indeed the mean oscillation frequency is removed from the signal.

Now, the maximum envelope 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 that there are N possibilities. Therefore

$$p_{max}(h_{max}) = N[1 - P(h_{max})]^{N-1} p(h_{max}), \tag{28}$$

where for linear wavess P is the c.d.f. given in Eq. (25). It is straightforward to implement this expression for the p.d.f. of maximum envelope wave height, but we are also interested in deriving simple, accurate expressions 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 (28). From experience it is known that this approximation works well for $N > 10$.

In the continuum limit, i.e. for large N, the maximum wave height distribution assumes, with $\mathscr{G}(h) = -NP(h)$, the simple form

$$p_{max}(h = h_{max}) = \frac{\mathrm{d}\mathscr{G}}{\mathrm{d}h} \exp(\mathscr{G}) = Np(h_{max}) \exp[-NP(h_{max})], \tag{29}$$

where $p(h)$ is the p.d.f. of envelope wave height and $P(h)$ is the corresponding exceedance probability. Then, the probability that maximum envelope wave height is larger than a given value, say h_c, becomes for large N,

$$P(h_{max} > h_c) \simeq 1 - \exp(\mathscr{G}(h_c)) = 1 - \exp(-NP(h_c)).$$

For linear waves the statistics are Gaussian and the exceedance probability is given in (26) resulting in the well-known double exponential law

$$P(h_{max} > h_c) = 1 - \exp\{-Ne^{-2h_c^2}\}$$

for the maximum envelope wave height probability. Note that in Goda's method the number of events N is independent of the exceedance level h_c.

So the question now is how to choose the number of independent events. It is customary to define an event with respect to a chosen reference level h_c. An event is then a part of a time series of length T_L that starts where the envelope has an upcrossing at level h_c and that finishes at the next upcrossing. The frequency of events is then determined by the upcrossing frequency. The total number of events in a time series of length 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, in order to make Goda's method work, which relies on a constant number of degrees of freedom, it may be

appropriate to introduce an average upcrossing 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

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

where $p(h,\dot{h})$ is the joint p.d.f. of h and $\dot{h}$. For upcrossings $\dot{h}$ is positive so that the integration over $\dot{h}$ is restricted to positive values only. Making use of the joint p.d.f. of h and $\dot{h}$ in Eq. (26) and performing the integrations one immediately finds the simple result

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

and, defining the dimensionless length $T_L^* = \nu \bar{\omega} T_L$, the average number of events becomes

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

This result reflects that the number of degrees of freedom is determined by the number of wave groups. It is emphasized that it has only been made plausible how N depends on the relevant parameters. To some extent the result is uncertain because the choice has been made to connect the number of events with the average upcrossing frequency.

An alternative method to obtain the maximum envelope wave height distribution has been suggested by Naess (1982). His approach is based on the reasonable assumption that for envelope time series the number of level upcrossings by the envelope $\rho(t)$ is asymptotically Poisson distributed when the level height increases. In Janssen (2015a) it has been shown that, in particular for the tail of the maximum wave height distribution, Naess's idea results for linear waves in very good agreement with Monte Carlo simulations of the p.d.f. of maximum envelope wave height.

Naess (1982) 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 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 average frequency with which $h(t)$ 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 (27) 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 c.d.f. P one finds

$$P_H(h) = \exp\{-hN_{slc} e^{-2h^2}\}, \tag{30}$$

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

$$N_{slc} = 2T_L^*/\sqrt{2\pi} \tag{31}$$

is the number of (up)crossings at the significant level $h = 1$ (Here, the subscript 'slc' stands for significant level crossings). This parameter is a measure for the number of degrees of freedom. The maximum envelope height distribution then follows from $p_{max}(h) = dP_H/dh$, or,

$$p_{max} = (4h^2 - 1)\, N_{slc}\, e^{-2h^2} P_H(h), \quad \text{where } h \geq \frac{1}{2}. \tag{32}$$

The restriction $h \geq \frac{1}{2}$ is added in order to prevent the p.d.f. from becoming negative. In addition, note that the approach by Naess is only valid for large level crossings, presumably because 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 (29) N has been assumed a constant. As a consequence, the large h behaviour of (32) differs from (29) because it involves an additional factor h. This different asymptotic

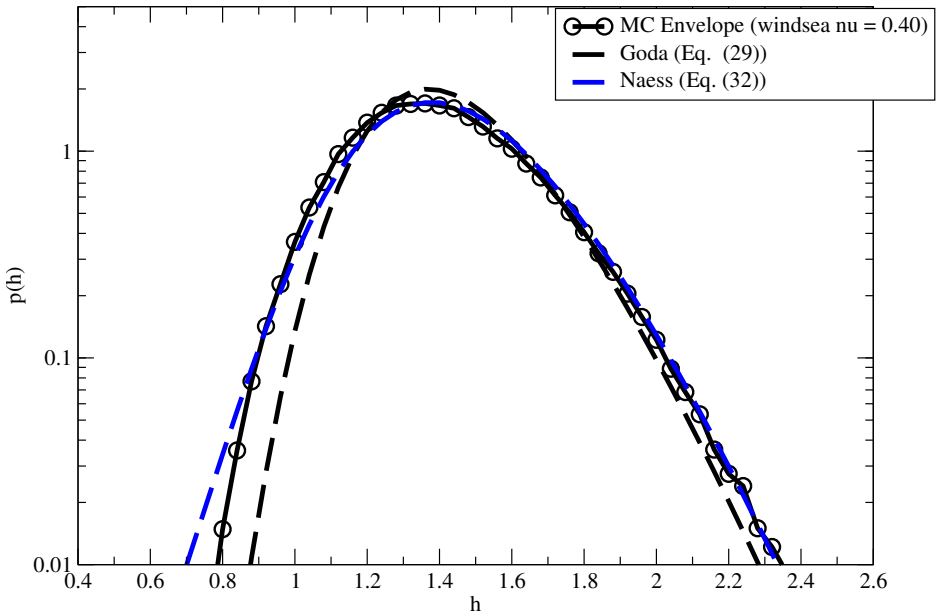

Figure 3: The p.d.f. of maximum envelope wave height for 'old' wind sea ($\gamma = 1$). The duration of the time series is 10 wave periods. Shown is the comparison between results from Monte Carlo simulations, and the theoretical result from Eq. (29) and Naess' (1982) Eq. (32). In the Monte Carlo simulations both amplitude and phase of the waves are regarded as stochastic variables.

behaviour of the two results for maximum wave height distribution is born out by a comparison of (32) and (29) with Monte Carlo simulations of the sea state with a Pierson-Moskovitz spectrum, as shown in Fig. 3. Here, it is important to point out that a Monte Carlo simulation was performed where amplitudes are drawn randomly from a Rayleigh distribution while phases were drawn from a uniform distribution. Fig. 3 shows clearly that the Naess expression for the maximum envelope wave height distribution is in better agreement with the Monte Carlo result, in particular for the tails of the distribution.

2.5 Maximum envelope wave height: nonlinear extension and expectation value.

In the previous sections we have found that for linear waves Naess' expression is adequate in describing the maximum envelope wave height distribution. The extension towards nonlinear waves is not trivial, however, because it requires deriving the joint probability distribution of h and $\dot{h}$ in the weakly nonlinear case. In principle this can be done but it is simply a matter of a lot of work. Therefore, at the moment an

educated guess is being made of the form of a maximum envelope wave height p.d.f., based, of course, on the findings thus far.

The following discussion will be done both in terms of the wave energy $E = 2h^2$, and in terms of the envelope wave height h. Based on Eq. (29) it is posited that the maximum envelope wave height distribution has the form

$$p_{max}(h) = \frac{\mathrm{d}\mathscr{G}}{\mathrm{d}h} \exp(\mathscr{G}), \tag{33}$$

where $\mathscr{G}$ involves the stretched exponential

$$\mathscr{G} = -N \exp(-z),\ z = -\alpha + \sqrt{\alpha^2 + \beta E}. \tag{34}$$

and, in keeping with Naess' approach, the number of degrees of freedom depends on the reference level, i.e.

$$N = N_{slc} \left(\frac{E}{2} \right)^{1/2} \tag{35}$$

where N_{slc} is given by (31). Finally, the parameters α and β follow, see Eqns. (20-24), from matching the stretched exponential distribution with the weakly nonlinear c.d.f. of (16) so that the fitting parameters depend on skewness and kurtosis of the envelope time series.

An important quantity to measure the extremity of an event is the exceedance probability that energy or wave height is larger than a critical value. Using (33) one finds

$$P(h > h_c) = \int_{h_c}^{\infty} \mathrm{d}h\ p_{max}(h) = 1 - \exp(\mathscr{G}). \tag{36}$$

This quantity is very sensitive to the tail of the maximum envelope wave height distribution, but it gives important information on, for example, the probability that waves hit the lower deck of an oil rig.

A more robust indicator of extreme events is the expectation value of maximum envelope wave height. It is here defined using the expectation value of maximum wave energy $\langle E_{max} \rangle$, defined as

$$\langle E_{max} \rangle = \int_0^{\infty} \mathrm{d}E\ E\ p_{max}(E), \tag{37}$$

with

$$p_{max}(E) = \frac{\mathrm{d}\mathscr{G}}{\mathrm{dE}} \exp \mathscr{G}. \tag{38}$$

Now, the expected maximum wave height is defined as

$$\langle h_{max} \rangle = \sqrt{\frac{1}{2} \langle E_{max} \rangle}, \tag{39}$$

and this definition of maximum wave height emphasizes the idea that wave energy is the key parameter in assessing damage to vessels and oil rigs.

Now it should be realized that the maximum energy distribution $p_{max}(E)$ is, as shown below, a narrow distribution. For example the dependence of the number of degrees of freedom $N(E)$ on energy E is fairly weak compared to the stretched exponential function $\exp(-z)$. As a consequence, the function

$N(E)$ of (35) may be approximated by $N(\langle E\rangle)$ and is regarded as a constant in the integration. In that event, for large N the integral may be evaluated almost exactly (with error of $\mathscr{O}(\exp -N)$) with the result

$$\langle E_{max}\rangle = \frac{1}{\beta}\left[G_2 - 2G_1\left(\alpha + \log N\right) + \log N\left(2\alpha + \log N\right)\right], \tag{40}$$

where $N = N_{slc}\sqrt{\langle E_{max}\rangle/2}$, hence Eq. (40) is an implicit relation for $\langle E_{max}\rangle$. The symbols G_n, $n = 1,2$, denote 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

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

It may be shown that $G_1 = \Gamma'(1) = -\gamma$, while $G_2 = \Gamma''(1) = \gamma^2 + \pi^2/6$, where $\gamma = 0.5772$ is Euler's constant.

Although (40) is an implicit equation for the expectation value of maximum wave energy, it is straightforward to solve it by iteration. In practice, this iteration scheme converges very quickly, only 5 iteration are needed at the most. This follows from a comparison with results from a numerical computation of the integral (37) where after 5 iterations agreement up to four digits was achieved.

Furthermore, it is emphasized that the maximum envelope wave height is a random variable and therefore it is important to have knowledge of the width of the maximum envelope wave height distribution. The width $\sigma_{h_{max}}$ is defined as

$$\sigma^2_{h_{max}} = \langle h^2_{max}\rangle - \langle h_{max}\rangle^2$$

It is straightforward, but tedious, to obtain $\langle h_{max}\rangle$ and as a result one finds for sea states that are close to a Gaussian, hence α is large, that to lowest order the square width of the maximum height distribution becomes

$$\sigma^2_{h_{max}} = \frac{\alpha\pi^2}{\beta}\left(\frac{1}{24\log N} + \frac{1}{16\alpha}\right) \tag{41}$$

For typical values of maximum wave height, i.e. $\langle h_{max}\rangle \simeq 2$ and for a length $T_L = 1800s$ and a typical value of N_{slc} of around 400 one finds that $\sigma_{h_{max}}$ is of the order of 18% of significant wave height. Since maximum wave height is about twice significant wave height this means that the 'error' in the expectation of maximum wave height is only 9%.

Finally, it should be emphasized that the expression (40) is very elegant, and, compared to previous work (cf. Janssen, 2015a), it is much simpler. This is, of course, because of the use of the stretched exponential for the 'parent' distribution for envelope wave height. In addition, it should be noted that for a Gaussian sea state, which is achieved by taking the limit of $\alpha \to \infty$ (while realizing that $\beta = 2(\alpha+1)$) the expression for maximum wave energy simply becomes

$$\langle E_{max}\rangle = \gamma + \log N. \tag{42}$$

This is an implicit relation for the average maximum wave energy as the number of degrees of freedom $N = N_{slc}(\langle E_{max}\rangle/2)^{1/2}$ depends on $\langle E_{max}\rangle$, and a definite answer is only found by iteration. In practice, it

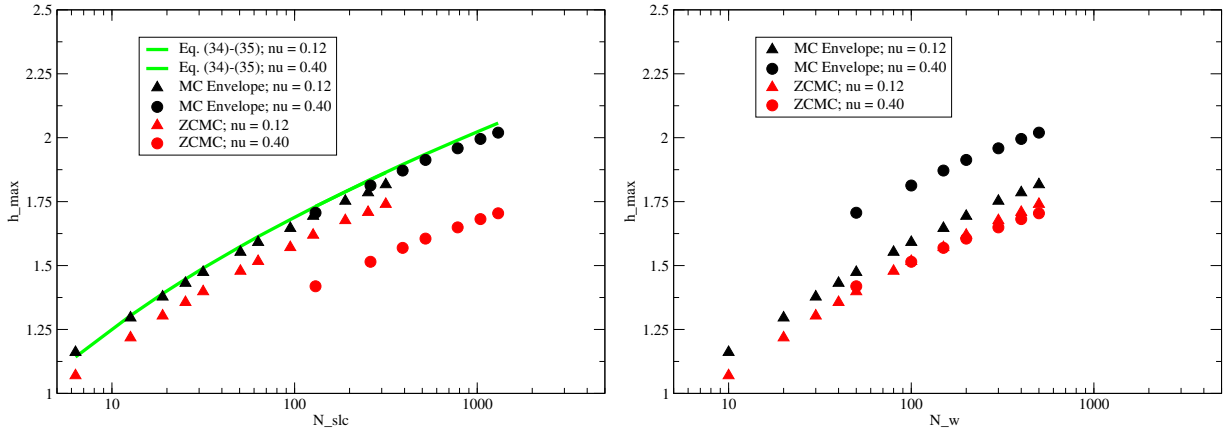

Figure 4: Left Panel: Expected maximum wave height versus number of upcrossings N_{slc} according to the numerical simulations for two spectral widths. The green line shows the theoretical result based on Eq. (39)-(40). For comparison, the maximum wave height (ZCMC) according to the zero crossing method as function of N_{slc} is shown as well. In the right panel the same parameters are shown as in the left panel but now as function of the number of waves $N_w = T_L/T_P$.

is found that this linear result already gives a reasonable approximation (with a systematic error of 20%), so let us explore its properties in a bit more detail. The expectation value of maximum wave height turns out to be a robust measure for extreme seas, being relatively insensitive to how the tail of the pdf is represented by the numerical simulations. Therefore, in order to produce the results displayed in Fig. 4 only 1,000 member ensembles were required. In order to generate more cases the number of waves in the timeseries was varied between 50 and 500 and we took two truncation limits in the wave spectrum, namely a wavenumber cut-off at 64 times the peak wavenumber and a wavenumber cut-off at $\sqrt{2}$ times the peak wavenumber, representing a broad ($\nu = 0.40$) and a narrow ($\nu = 0.12$) spectrum. In Fig 4 we have plotted in the left Panel the expectation value of maximum wave height as a function of N_{slc}. From the comparison between the Monte Carlo results and theory (Eq. (42)) it is clear that the agreement is excellent and it seems that the approach sketched in this paper works for a wide range in the number of degrees of freedom, even, surpisingly for small values of N_{slc}. In addition, although widely differing spectral widths have been chosen, it is clear from the universal behaviour displayed in the left Panel of Fig. 4 that the choice of $\nu\bar{\omega}$ for scaling of the number of degrees of freedom is essentially correct. This is obvious as the procedure to produce the envelope of a timeseries essentially removes information on the peak frequency. An alternative choice, e.g. the number of waves $N_w = T_L/T_P$ (with T_P the peak period), does not give rise to proper scaling behaviour as is plainly clear from the right panel of Fig. 4. On the other hand, from the Figure it is also clear that the maximum zero-crossing wave height shows universal scaling when using the N_w parameter. In addition, especially in case of broader spectra, the popular maximum zero-crossing wave height is considerable smaller than the maximum envelope wave height. For example, for a typical $N_{slc} \simeq 300$ maximum zero-crossing wave height is about 1.6 while maximum envelope wave height is of the order of 1.8. Nonlinear effects due to skewness and kurtosis will increase the expected maximum wave height values by about 10-20 % only. On the other hand, probabilities of extreme events are much more sensitive to nonlinear effects and they may increase results by an order of magnitude in exceptional circumstances, see e.g. in Fig. 2 for the envelope wave height distribution at dimensionless envelope wave height of the value of order 3.

The dependence of the average maximum envelope wave height on N_{slc} and hence on the spectral width ν has consequences for observations of maximum envelope wave height. This implies that the timeseries should be sampled with sufficiently high frequency in order to accurately represent the width ν which

depends on the second moment of the wave frequency spectrum. Considering once more the right Panel of Fig. 4 we see that for fixed N_w the envelope method may underestimate maximum wave height by about 25% when the observations do not represent the high frequency part of the wave spectrum. However, the truncation of the wave spectrum at 1.5 times the peak wavenumber is quite extreme. Some additional experiments were run for a fixed number of waves ($N_w = 200$) for truncations of 4 times, 16 times and 64 times the peak wavenumber resulting in an average normalized maximum envelope wave height of 1.83, 1.90 and 1.91. Therefore, in practice a reduction of maximum envelope wave height of only 4% or less is to be expected due to undersampling.

In order to close this Section it should be mentioned that it is not trivial to perform Monte Carlo integrations for the nonlinear case. In that case this requires to generate results from a large ($N_{ens} = 50,000,000$) ensemble of integrations of the two-dimensional Zakharov equation for a sufficient number of typical periods and the canonical transformation has to be evaluated to third order in steepness in order to capture the effects of the bound harmonics on skewness and kurtosis. An alternative, as discussed in §2.3, is to perform ensembe simulations using a Stokes wave train. As shown in Fig. 2 this works very well for checking the theoretical envelope wave height distribution. Unfortunately, it does not work for checking the theoretical maximum envelope wave height distribution. The reason is that a Stokes wave train produces a periodic signal with period equal to the inverse of the nonlinear frequency. Hence, the number of degrees of freedom equals 1. Referring to Eq. (28), for $N = 1$ the maximum envelope wave height distribution equals the envelope wave height distribution $p(h_{max})$, therefore for typical large numbers of degrees of freedom the theoretical maximum wave height distribution cannot be checked using a Stokes wave train.

3 Theory and Parametrization of skewness and kurtosis.

Ocean waves may be regarded most of the time as weakly nonlinear, dispersive waves. Because of this there is a small parameter ε, i.e. the wave steepness, which permits the systematic study of the effect of nonlinearity on wave evolution by means of a perturbation expansion with starting point linear, freely propagating ocean waves. In addition it should be pointed out that the subject of nonlinear ocean waves has conceptually much in common with nonlinear wave phenomena arising in diverse fields in physics. Rapid developments in, say, nonlinear physics, started in the 1960's with the work of K.F. Hasselmann (1960) on resonant four-wave interactions in the context of a stochastic prescription of the sea state (the so-called Hasselmann equation), and with the work of V.E. Zakharov (1968) on a Hamilton description of water waves resulting in an approximate third order deterministic equation for the amplitude of the free-wave action density. The narrow-band limit of this so-called Zakharov equation was shown to result in the nonlinear Schrödinger (NLS) equation, which describes in the frame moving with the group velocity a balance between dispersion and nonlinear focussing. The NLS equation was the common thread between the diverse fields of nonlinear optics, plasma physics and surface gravity waves. In one dimension this nonlinear partial differential equation could be solved exactly, and for initial conditions of compact support it was shown by Zakharov and Shabat (1972) that the solution would evolve towards a series of envelope solitons. These solitons were ideal candidates for freak waves and therefore researchers focussed on (numerical) studies of the NLS equation. A very first attempt to exploit this knowledge in the context of a stochastic description of the sea state was described in Janssen and Komen (1982). As a guess it was assumed that the endstate of the solution of the NLS equation is a modulated wave train where the nonlinear modulation is given by an elliptic function. The resulting probability distribution for the envelope shows considerable deviations from the Rayleigh distribution (the one for linear waves), in particular for large envelope wave height. Additional (numerical) studies

by researchers such as Tanaka (1992), Trulsen & Dysthe (1997), Osborne *et al.* (2000), Onorato *et al.* (2001) followed which concentrated on individual events while Mori & Yasuda ((2002) tried to obtain the p.d.f. of wave height in the case of steep waves suggesting that the kurtosis of the surface elevation may play an important role. Finally, an important advance was reported by Janssen (2003) who started from the Zakharov equation to obtain a kinetic equation for the ensemble mean of the action density spectrum that includes both the resonant four-wave interactions introduced by Hasselmann (1962) *à*nd quasi-resonant four-wave interactions. The quasi-resonant four-wave interactions give rise to a time-dependent contribution to the kurtosis of surface gravity waves which is proportional to the square of the so-called Benjamin-Feir Index (BFI). This contribution will be called dynamic kurtosis from now onwards. Results on kurtosis and the surface elevation p.d.f. for 1-dimensional waves in the narrow-band approximation are further discussed in Mori & Janssen (2006), while Mori *et al* (2011) study for two dimensional propagation the dependence of kurtosis on the Benjamin-Feir Index and directional width. Note that the Benjamin-Feir Index basically measures the importance of wave nonlinearity compared to wave dispersion. For narrow-band waves the effects of the dynamic kurtosis can be quite important as it is basically the ratio of wave steepness ε to the angular width σ_ω of the wave spectrum, i.e. the ratio of two small parameters. Therefore, dynamic kurtosis can in principle give an $\mathcal{O}(1)$ contribution.

The estimation of the expectation value of maximum envelope wave height and of the probabilities for extreme events requires the knowledge of skewness and kurtosis of the bound and free waves. Expressions of these statistical parameters for arbitrary spectra have been obtained for surface elevation statistics by Janssen (2009) and for envelope wave heights by Janssen (2015b). These general formulae involve the evaluation of four- and six-dimensional integrals which are too time consuming in an operational context. Therefore, parametrizations of skewness and kurtosis need to be developed, for deep and shallow water. A shallow water parameterization is highly desirable because most off-shore operations take place in finite depth. It is tried to find these parametrizations by assuming, on the one hand, that the stats obtained from the general formulation for arbitrary spectra is the truth, while the approximate formulae for skewness and kurtosis are obtained by adjusting parameters in the corresponding narrow-band approximations in such a way that a reasonably good agreement with the truth is obtained. For a recent, alternative approach to parametrisation of the skewness and kurtosis see the work of Gramstad and Lian (2024).

Let us first discuss results for skewness and kurtosis using the general formulation for arbitrary spectra, which is then followed by a presentation of the simple parametrisations of the statistics of waves. Simulated WAM spectra were taken from two shallow water cases where freak wave events have been observed, namely the Draupner freak wave event from the 1^{rst} of January 1995, occurring at 15:20, and the Andrea storm which occurred on midnight of the 9^{th} of November 2007. For some details on the Draupner event see Haver (2004) while this event was hindcasted by Cavaleri *et al.* (2016) using a fairly recent version of the ECMWF coupled ocean-wave, weather forecasting system which has a resolution of 10 km in the horizontal and has 137 layers in the vertical. Some time later, L. Bertotti and L. Cavaleri have produced, in addition, a simulation for the Andrea storm with the same coupled system. A detailed description of this storm is given by Magnusson and Donelan (2013). In both cases the water depth at the location where the freak wave events occurred was 70 m and the dimensionless depth k_0D, with k_0 the peak wave number and D depth, was during the extreme event in the intermediate depth regime as the range of dimensionless depth varies between 1.45 and 1.9.

The simulated spectra have, for an operational system, a fairly high resolution of 36 frequencies and 36 directions, where the frequencies are on a logarithmic grid and the directions are on a linear grid. In Fig. 5 a few examples of spectra are shown, one at the time of the Draupner event, one at the time of the Andrea freak wave event and one example of a broad spectrum taken from the Andrea storm at noon on

the 8^{th} of November 2007, illustrating that both narrow and broad-band spectra are present in the time series.

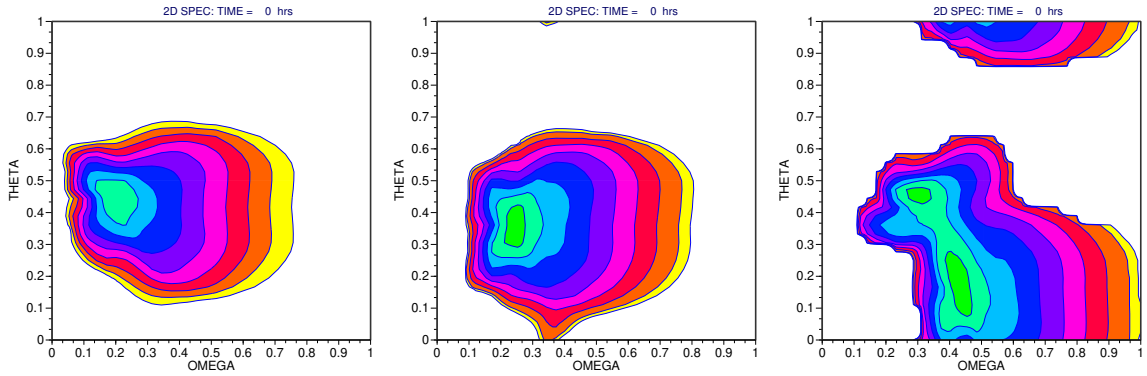

Figure 5: Left panel: Wave spectrum for the Draupner event at 16:00 hrs on 1-1-1995; Middle panel: Andrea main event at midnight of 9-11-2007; Right panel: broad-band spectrum 12 hrs before the Andrea event.

An important point to note is that the wave spectra at the time of the extreme events have a more narrow directional distribution than a typical wind sea in deep water. In order to appreciate this point, in Fig. 6

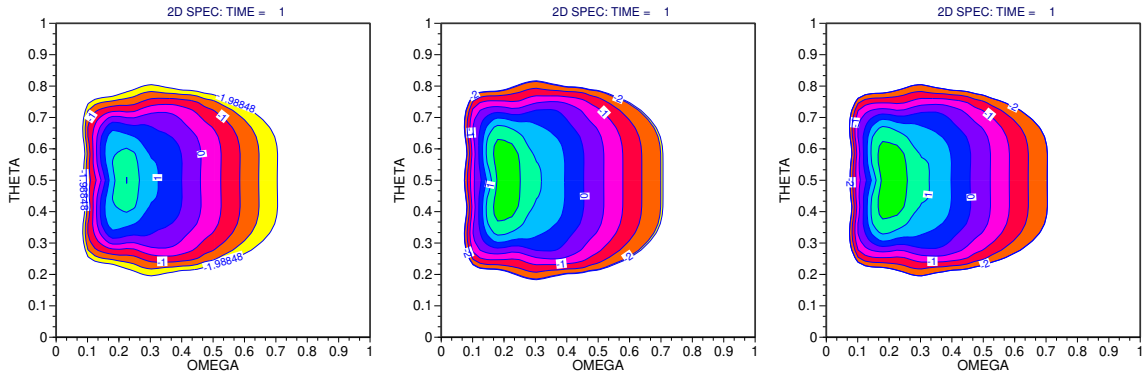

Figure 6: Left panel: Wave spectrum after one day wind forcing in shallow water; Middle panel: Wave spectrum after one day forcing in deep water; Right panel: Shallow water run with bottom friction switched off.

the wave spectrum is shown as obtained from a one-gridpoint model run after 24 hours forcing with a wind speed of 18.45 m/s. The left panel shows the result when depth D = 70 m, which corresponds to $k_0 D = 1.9$, while the middle panel shows the result for deep water. Clearly, the deep water run gives a broader spectrum.

An interesting question to ask is which physical process is responsible for the narrowing of the directional spectrum in shallow water. In the context of a single gridpoint model it is fairly straightforward to investigate this issue. Several candidate source functions were scrutinised, e.g. the nonlinear transfer which contains a depth dependent amplification factor. Switching off the amplification factor did not result, however, in an appreciable change of the width of the directional spectrum. It turned out, as shown in the right panel of Fig. 6, that switching off the bottom friction source function in the shallow water run explains to a large extent the reason why in shallow water the directional spectrum is narrower.

In order to quantify the differences there is a need to introduce a robust measure of directional width near the peak of the spectrum. Usually, the directional width σ_θ is estimated using frequency dependent

moments, i.e.

$$m_0 = \int \mathrm{d}\theta\, F(\omega,\theta),\; m_1 = \int \mathrm{d}\theta \cos(\theta-\overline{\theta})F(\omega,\theta),$$

with $\overline{\theta}$ the mean direction and $F(\omega,\theta)$ the two-dimensional wave spectrum. However, this is not such a robust measure and therefore the frequency dependent moments are integrated over a region around the angular peak frequency ω_p of width $2\Delta\omega$, with $\Delta\omega = \omega_p/2$, i.e.

$$\overline{m}_0 = \int_{\omega_p-\Delta\omega}^{\omega_p+\Delta\omega} \mathrm{d}\omega \int \mathrm{d}\theta\, F(\omega,\theta),\; \overline{m}_1 = \int_{\omega_p-\Delta\omega}^{\omega_p+\Delta\omega} \mathrm{d}\omega \int \mathrm{d}\theta \cos(\theta-\overline{\theta})F(\omega,\theta), \tag{43}$$

and the measure of directional width becomes

$$\sigma_\theta = \sqrt{2\,(1-R_1)},\; R_1 = \overline{m}_1/\overline{m}_0 \tag{44}$$

The above definition of directional width will also be used in a simple parametrisation of dynamic kurtosis. Using the measure in (44) it is found that for the Draupner spectrum in Fig. 5 $\sigma_\theta = 0.32$ while for the deep-water case of Fig. 6 $\sigma_\theta = 0.55$, so there is a considerable difference in width. This difference is quite important for the estimation of the value of the dynamic kurtosis, as it depends on the square of directional width.

3.1 Results for skewness and kurtosis from exact computations.

Using the exact expressions given in the Appendix, i.e. for dynamic kurtosis in (A7), for skewness in (A26) and for bound kurtosis in (A31), time series for skewness and kurtosis have been obtained from the simulated spectra for the Draupner and Andrea case. They are displayed in Fig. 7.

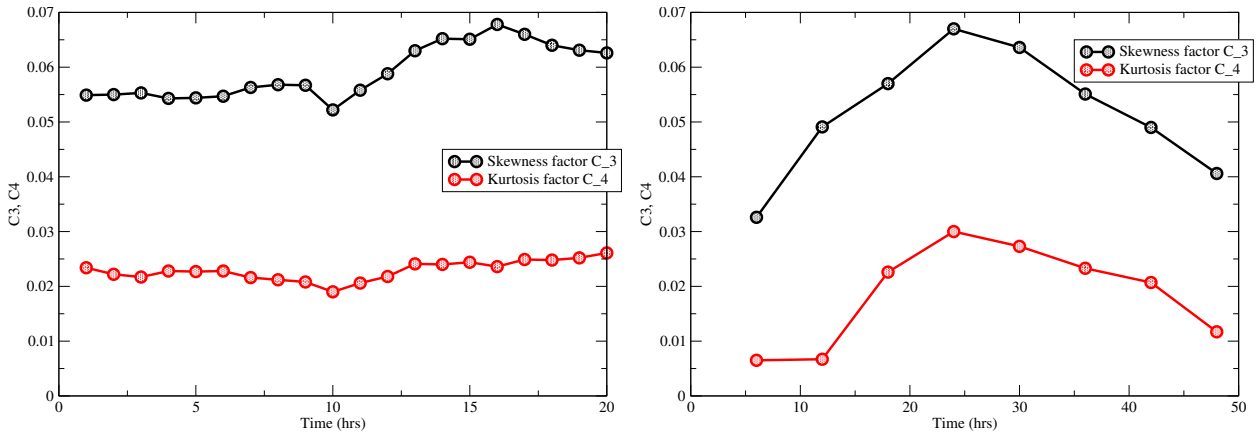

Figure 7: Evolution in time of skewness factor C_3 and kurtosis factor C_4 for the Draupner case (left panel) and the Andrea storm (right panel).

Noting that the Draupner freak wave occurred at 16:00 hrs while the Andrea event occured at 24 hrs, there is a striking difference between the two cases. During the Andrea event there is a clear sign that skewness and kurtosis have a maximum at the time of the freak wave event whereas for the Draupner wave this is only marginally evident for the skewness. In particular, the time series for kurtosis is fairly featureless.

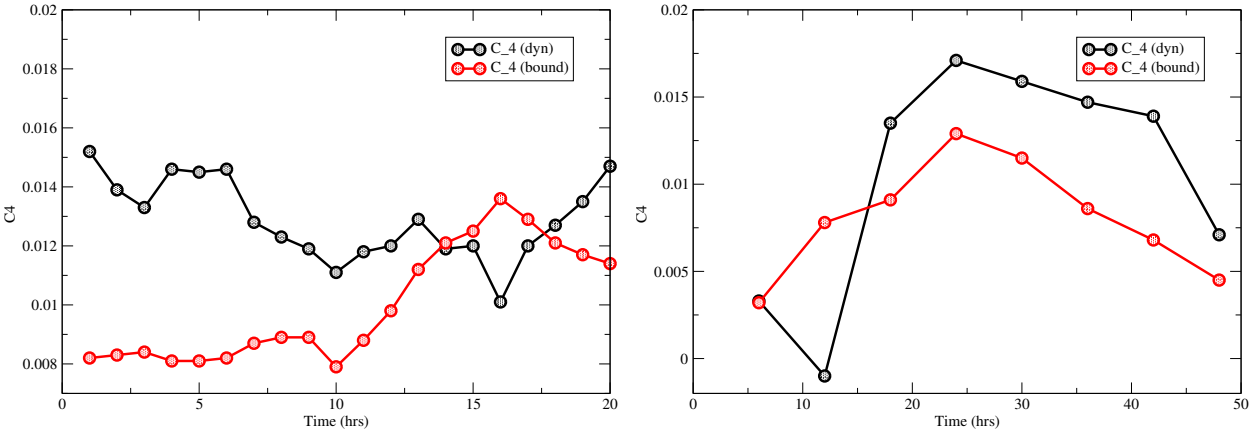

Figure 8: Evolution in time of dynamic and bound-wave part of the kurtosis factor C_4 for the Draupner case (left panel) and the Andrea storm (right panel).

Realizing that kurtosis consists of a dynamic part and a bound-wave part it is therefore also of interest to study their respective timeseries. They are displayed in Fig. 8.

The Figure shows that for both the Draupner event and the Andrea wave the bound-wave part of kurtosis reached an extremum at the time of the occurrence of the freak wave, while the dynamic part of the kurtosis only reached a maximum for the Andrea freak wave event. For the Draupner time series the dynamic kurtosis is fairly flat and even reaches a minimum at the time of the freak wave event. To conclude the discussion of the kurtosis time series it is noted that for the Andrea storm the range of kurtosis is quite large and even becomes slightly negative at 12 hrs. At that time the angular width of the wave spectrum is quite large as shown in the right panel of Fig. 5. Therefore, dynamic kurtosis has a sensitive dependence on angular width, much more so than the bound-wave part.

Finally, as discussed further in the Appendix, there is a marked difference between the statistical information derived from the free waves and the bound waves. For given wave spectrum, bound-wave skewness and kurtosis are independent of time, whereas the dynamic kurtosis evolves in time. This is shown in Fig. A1 (see Appendix A) which displays the evolution in time of dynamic kurtosis for a narrow-band wave train, but it is also very much the case for the Draupner freak wave and the Andrea storm using the expression for the kurtosis valid for arbitrary spectra. In Fig. 9, the evolution of dynamic kurtosis as function of the dimensionless time $\tau = \delta_\omega^2 \omega_p T$ is shown where $\delta_\omega = \sigma_\omega / \omega_p$. This evolution time scale is obtained from the narrow-band considerations in the Appendix. Because realistic spectra are not quite narrow-band a relative width of the spectrum is chosen which is representative for the whole spectrum as defined in Eq. (6), i.e. $\delta_\omega = (m_0 m_2 / m_1^2 - 1)^{1/2}$. Since in practice δ_ω is of the order of 0.3, the time scale of 200 units in the graph corresponds to 300 wave periods. It is seen that on this relatively long time scale the solution is still smooth, but for longer time scales the solution becomes erratic. In agreement with an earlier remark it is seen that the range in kurtosis values for the Draupner event is fairly small since the left panel of Fig. 9 shows a surprisingly universal time behaviour. However, this is much less the case for the Andrea storm, shown in the right panel, hence the range of kurtosis values is much wider.

The question now is how to deal with the time dependency of the dynamic kurtosis, because it would be convenient to have a simple measure for estimating the severity of the sea state. The most straightforward way to deal with this is to take the mean value of the kurtosis obtained from an average over the whole period shown in the Figure. Note that in the calculations the assumption has been made that initially the dynamic kurtosis vanishes, and, therefore, the Figure basically gives the response of the dynamical wave

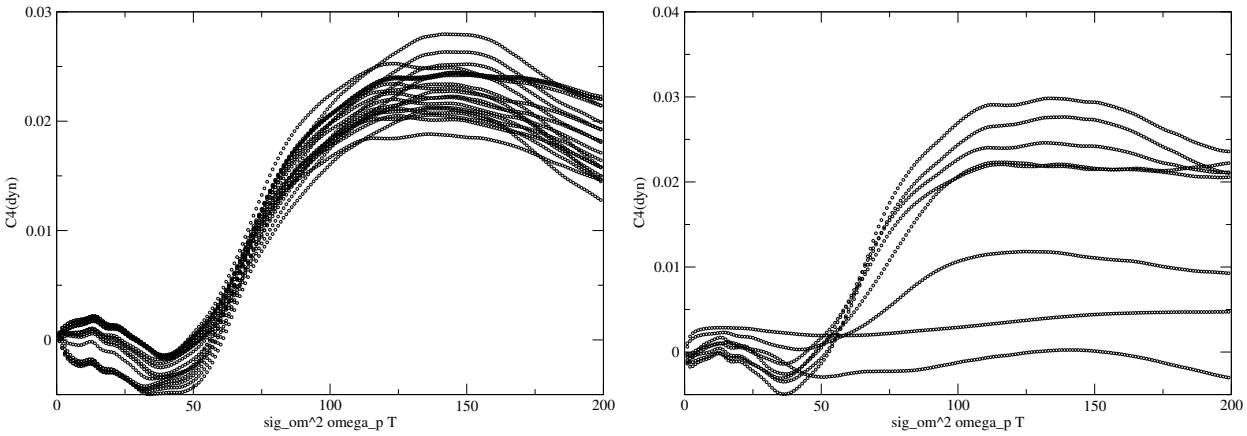

Figure 9: Evolution of dynamic kurtosis $C_4^{dyn} = \kappa_{40}^{dyn}/3$ as function of the dimensionless time $\tau = \sigma_\omega^2 \omega_p T$, where σ_ω is the relative width of the frequency spectrum and ω_p is the angular peak frequency. The different curves correspond to different hours during the coupled simulation. The Left panel shows the Draupner case while the right panel shows the Andrea storm.

system to rapid pertubations caused by e.g. a rapid increase of wind or the passage of a frontal system.

3.2 Probability estimates of extreme events using exact computations.

Let us now try to establish how extreme the Draupner and Andrea freak waves are. For the Andrea storm I had no access to time series so the maximum envelope wave height was estimated by taking twice the maximum crest height, which amounts to 3.26 times the significant wave height (Magnusson and Donelan, 2013). Clearly, we are dealing here with a very extreme event.

For the Draupner case Miguel Onorato provided me with the actual timeseries for the surface elevation $\eta = \rho \cos\theta$. The orthogonal complement $\zeta = \rho \sin\theta$ was obtained from the surface elevation timeseries using the Hilbert transform H, i.e. $\zeta = -H(\eta)$. The envelope ρ then follows from $\rho = \sqrt{\eta^2 + \zeta^2}$. Timeseries of the surface elevation η, its orthogonal complement ζ and its corresponding envelope are shown for the period close to the extreme event in Fig. 10. A number of measurements from the time series can be obtained. First, using the definition that envelope wave height is twice the envelope it was found that the observed maximum envelope wave height was 3 times the significant wave height, or in real terms $h_{max} = 37.02$ m. Secondly, it is straightforward to obtain from the timeseries of η and ζ relevant statistical information as introduced in §2.1 and defined in Appendix C.1.

```
Table 1:
Relevant parameters that characterise the Draupner freak wave event
N  =    2560
DT =       0.46875
XM0_OBS    =   8.878       KAPPA11 =  0.000       KAPPA4  =  1.954
HS         =  11.918       KAPPA02 =  1.000       KAPPA40 =  1.038
KP         =   0.017       KAPPA30 =  0.408       KAPPA22 =  0.244
EPS_W      =   0.052       KAPPA21 = -0.011       KAPPA04 =  0.428
RATIO      =  11.749       KAPPA12 =  0.136       C4_OBS  =  0.244
C3_OBS     =   0.136       KAPPA03 = -0.030       KP*D    =  1.208
```

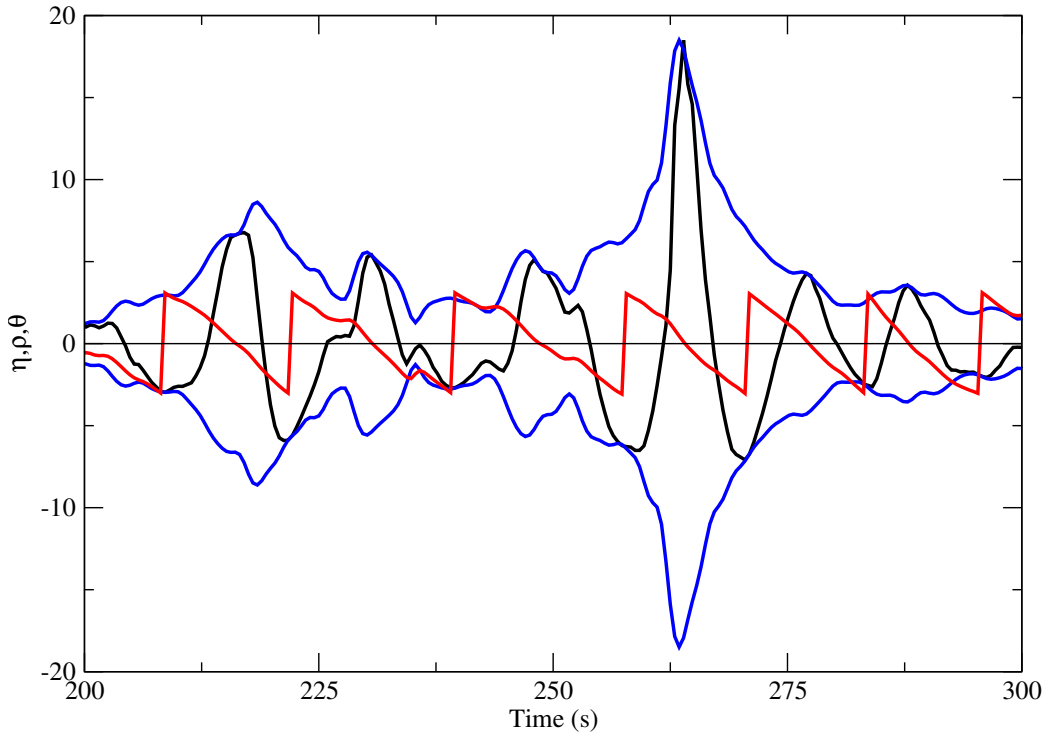

Figure 10: Time series η of the Draupner freak wave event (black). The envelope time series ρ is shown in blue, while the phase of the wave train θ is shown in red. It is clearly evident that the event at time $t = 264$ s is extreme because envelope ρ is maximal while phase θ vanishes. The corresponding local wave energy at the maximum is about 20 times the average wave energy.

For a timeseries length of 1200 s with a sampling period of $dt = 0.46875$ s statistical results are given in Table 1. Both skewness factor $C_3^{obs} = 0.136$ and the kurtosis factor $C_4^{obs} = 0.244$ are, compared to the results of the ensemble mean simulations displayed in Fig. 7, considerably larger. At the time of the Draupner event the ensemble mean simulation gives $C_3^{ens} = 0.0678$ while $C_4^{ens} = 0.0219$ so the difference in skewness is a factor of two while the difference in kurtosis is a factor of ten. In sharp contrast, the significant wave height, based on the second moment $\langle \eta^2 \rangle$, turns out to be a relatively robust parameter as the observed significant wave height is 11.92 m while the modelled wave height is 11.23 m. This suggests that, compared to higher order moments such as skewness and kurtosis, significant wave height is much less sensitive to the phase distribution over a 20 minute period. On the other hand, as shown in Appendix C, modelled wave spectrum differs from reality to some extend. For example, the modelled peak wavenumber equals $k_0 = 0.021$ while the observed peak wavenumber is 15 % lower, i.e. $k_0 = 0.018$. Therefore modelled dimensionless wavenumber $k_0 D$ is about 1.45 while observed dimensionless wavenumber, as given in Table 1, equals 1.208, hence in reality shallow water effects are more pronounced. For a uni-directional spectrum this difference in peak-wavenumber would even signal a dramatic change in the stability of a uniform wavetrain from unstable (model) to stable (observed) modulations, but for finite directional width the differences are much less dramatic and there is no change of stability properties for $k_0 D = 1.383$. For details on this see Appendix B1 and Fig. B1.

In Appendix C we perform a more detailed study of the issue of the representativeness of observations of higher moments of the surface elevation. Simulations with the one-dimensional Zakharov equation show that when as initial condition one takes observed Draupner spectrum and observed phases of the wave components then the kurtosis of the surface elevation is of similar order of magnitude as observed. However, if one performs a Monte Carlo simulations with the observed spectrum and *random* phase then

kurtosis of the sea surface is similar in magnitude as found by the ensemble mean simulations reported in Fig. 7. Therefore it is not meaningful to compare directly freak wave observations with ensemble mean results. This is clearly a limitation of ensemble approach.

In order to understand these discrepancies it should be realized that a wave model gives the ensemble mean of the sea state in a box surrounding the point of interest. This remark is also valid for the probabilities of events, we only know the ensemble average probability distribution function for the envelope wave height, therefore it is not possible to provide estimates of probabilities at a specific location. As a consequence, the statistics (e.g. skewness and kurtosis) derived from a short observed time series are not representative for the statistics of a domain of the size of the chosen resolution. This is readily seen from the estimates of the kurtosis factor C_4 from the Draupner timeseries. Note that such large discrepancies for other sea state parameters such as significant wave height or wave-induced stress are not to be expected, as already suggested above. It follows that freak wave formation and statistical parameters such as kurtosis and to a lesser extent the skewness are sensitive to the distribution of the phases of the individual waves. Only for a coherent sea state nonlinear focussing is to be expected to occur. But when this focussing process occurs at a certain location, this process is not representative for the averaged sea state as produced by a wave model which gives the (ensemble) average over a spatial domain of, in this case, 14×14 km.

Remark:

Finally, returning to Table 1 the parameter *Ratio* should be explained: it gives an indication of the importance of effects of four-wave interactions compared to the effect of the bound waves. For a freak wave warning system in which the statistics of the envelope of the waves plays a central role the Ratio parameter is defined by $R_{env} = \kappa_4/\kappa_{30}^2$, where κ_4 is the envelope kurtosis and κ_{30} is the skewness of the sea surface. For deep-water waves in the narrow-band approximation one finds (see Appendix A) simple expressions for skewness and kurtosis, i.e. $\kappa_{30} = 3\varepsilon$ while $\kappa_4 = 24\varepsilon^2$ with $\varepsilon = k_0 m_0^{1/2}$ the steepness of the narrow-band wave train. In that event the Ratio parameters becomes a constant given by $R_{env} = 8/3$. If, in case of deep-water, the observed R_{env} is much larger than 8/3 then it may be concluded that also four-wave interactions play an important role in the formation of the observed freak wave (see e.g. Toffoli *et al.* (2024) who applied this argument for the surface elevation statistics). Unfortunately, in shallow water the *Ratio* parameter has an additional dependence on dimensionless depth $k_0 D$ and, as detailed in Appendix A, it tends to increase for shallower water if dimensionless depth is larger than 1. From Table 1 it is seen that the dimensionless depth equals 1.20 so that according to the single mode results displayed in Fig. A3 the envelope *Ratio* for bound waves only is about 4.6. As according to Table 1 the observed *Ratio* is 11.8 which is considerable larger than the estimate involving only bound-wave statistics, it is suggested that the contribution of the four-wave interactions to the formation of freak waves for Draupner freak wave event may be considerable. There is a caveat, however, as the theoretical results for the bound wave statistics only hold for an ensemble and not for an individual case. Therefore, the estimation of the *Ratio* parameter may be flawed because theoretical estimates of bound-wave κ_4 and κ_{30} are inaccurate, but it may be argued that the ratio κ_4/κ_{30}^2 might be more accurate.

Now the question is whether and how one can establish the likelyhood of an event, given the spatial resolution of the wave model (in this case 14 km) and given the present knowledge of the physics of freak wave generation. For this reason attention is concentrated on the maximum wave height distribution and the exceedance probability $P_{max}(h_{max} \geq h_{obs})$ is determined that in a domain of 14×14 km over a time span of the integration timestep $\Delta t = 450$ s such an extreme event occurs. This requires that one needs to be able to count the number of events in a spatial domain over a certain time span. This is not a trivial matter and at the moment results are only known for the surface elevation of a linear sea state which has a Gaussian p.d.f. See for this the work of J. Adler (1981) which has been further discussed by Baxevani

and Rychlik (2006) and implemented by Fedele (2012) and Benetazzo *et al.* (2015). In the present case an estimation of the number of events is required for a different problem, namely for the maximum envelope wave height which obeys in the lowest approximation a Rayleigh distribution, but higher order corrections caused by nonlinear effects measured by skewness and kurtosis may be important as well. Such an estimate is not available at the moment. For lack of alternatives the Baxevani and Rychlik estimates have been used as given in Benetazzo *et al.* (2015), and they give for a domain of 14×14 km $\times$ 450 s a surprisingly large number of events, in the order of 300,000 or more.

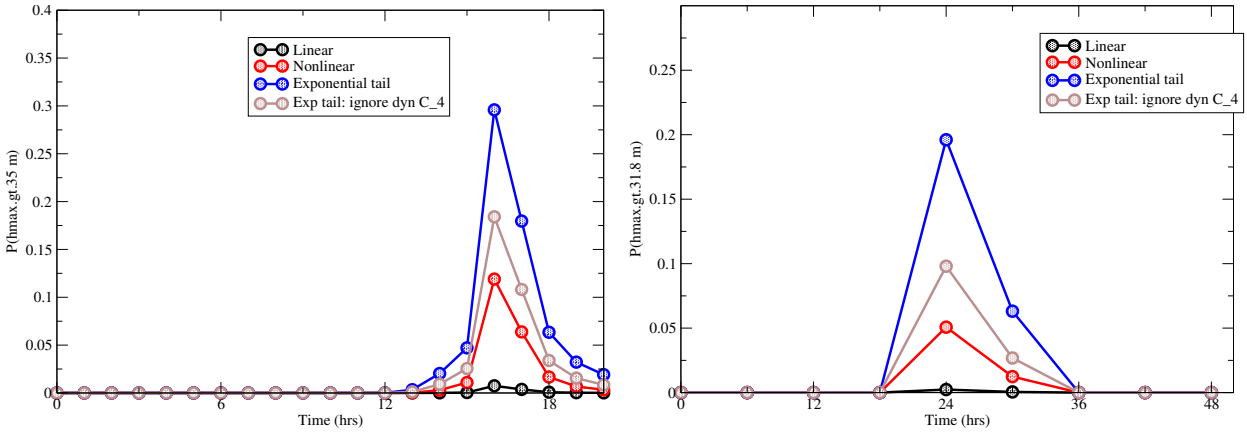

Figure 11: Evolution in time of simulated exceedance probability $P_{max}(h^{obs}_{max})$ for $h^{obs}_{max} = 35$ m (Draupner, left panel) and for $h^{obs}_{max} = 31.8$ (Andrea storm, right panel). For comparison the corresponding results from linear theory are shown, whereas by comparing curves labelled Nonlinear and Exponential tail the impact of the exponential tail is depicted. Finally the impact of dynamic kurtosis is found by comparing curves labelled ignore dyn C_4 and Exponential tail.

In Fig. 11 the timeseries of exceedance probability $P_{max}(h_{max} \geq h^{obs}_{max})$ for the Draupner freak wave (left panel) and the Andrea storm (right panel) are shown. It is clear that when all nonlinear effects are included that these events are rather plausible as the probabilities are already of the order of 20-30 %, while according to linear theory these extreme events are quite improbable. In the Figure is also shown exceedance probabilities when the exponential tail is not used in the calculation. In particular in the case of the Andrea storm this would result in a considerable reduction of probabilities. For the Andrea storm the reduction is a factor of 4 while for the Draupner freak wave the reduction is a factor of three. In addition, it is clear that linear theory would seriously underestimate the occurrence of these events. Finally, the impact of the dynamic contribution to the kurtosis is depicted as well. Omitting this effect gives about a factor of two reduction in the probabilities. All this suggests that it is important to include deviations from Normality related to finite skewness and kurtosis, including the effect of the near-resonant interactions.

3.3 Parametrisation of the bound skewness and kurtosis.

The parametrisations for the bound-wave skewness and kurtosis factors C_3^{bound} and C_4^{bound} have been guided by the narrow band expression (A28) and (A33), presented in Appendix A. This approach seems to work quite well, as will be seen in a moment, because the bound-wave stats are insensitive to the directional width of the wave spectrum. This has been checked numerically for JONSWAP spectra with a variable directional width.

After some trial and error the following parametrisation for the skewness factor C_3 and kurtosis factor C_4

was found:

$$C_3^{bound} = 2.24 m_0^{1/2} (\alpha + 0.9\Delta)\,,\; C_4^{bound} = 7.0 m_0 \left\{ \gamma + \alpha^2 + (\alpha+\Delta)^2 \right\}, \tag{45}$$

where the wavenumber and depth dependent parameters α, γ and Δ are defined in Eq. (A20) and m_0 is the zeroth moment of the frequency spectrum.

It is noted that it is important to choose as wavenumber a characteristic wavenumber that reflects that the wave spectrum is to some extent broad-banded. As a first guess a characteristic wavenumber $\bar{k}$ was chosen that follows from the inverse of the dispersion relation $a\bar{\omega} = \omega(\bar{k}, h)$, where $\bar{\omega} = m_0/m_{-1}$ and a is a tuning parameter. Results for skewness and kurtosis are very sensitive to the choice of this tuning parameter. In the end $a = 0.89$ was chosen which, for deep-water waves, implies that the characteristic wavenumber $\bar{k}$ is larger by about 20% when compared to the peak wavenumber k_0. Thus, in the formulae for α, γ and Δ, for example, the wave number k_0 is replaced by $\bar{k}$, while, to be consistent, this is also done for all wavenumber dependent factors in the dynamic kurtosis parametrisation of the next section.

These parametrisations compare very favourably with the exact computations from the Draupner and Andrea cases, as shown in the scatter plots of Fig. 12.

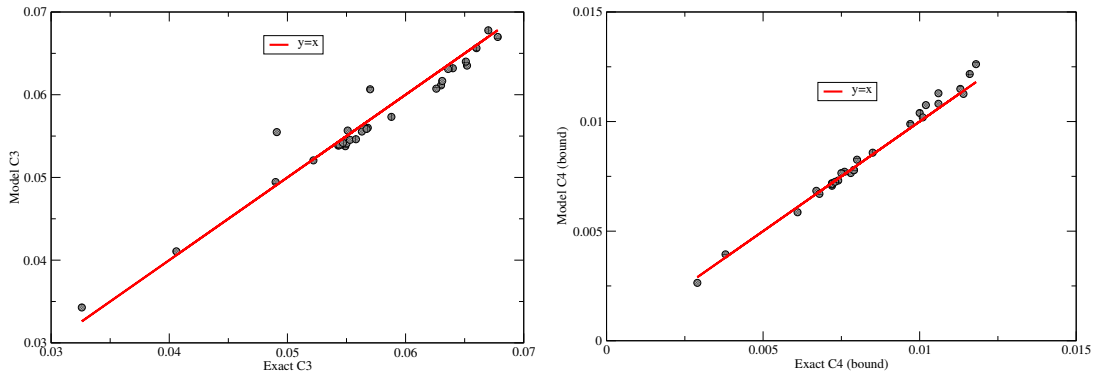

Figure 12: The left panel shows the comparison of parametrized skewness factor C_3^{bound} against exact computation from the Draupner and Andrea events, while the right panel shows the comparison for the kurtosis factor C_4^{bound}.

Quantitatively, the agreement is impressive as follows from a linear fit of the parametrized results against exact computations which gives correlations of more than 96%. Nevertheless, it is noted that the scatter-plot of the skewness factor shows two modest outliers which correspond to the broad angular distribution cases of the Andrea storm.

3.4 Parametrisation of the dynamic kurtosis.

The parametrization of the dynamic part of the kurtosis has been guided by the definition (A15) and the narrow-band result (A17), which is strictly speaking only valid for Gaussian spectra and gives an estimate of the maximum of dynamic kurtosis. Rather then the maximum value to estimate the severity of the sea state, its average value over a time span of about 300 wave periods will be chosen, as already suggested when discussing Fig. 7. In order to obtain a simple parametrization it was attempted to stay as close as possible to the narrow-band result, and in particular attention was paid to the replacement of the definitions of the Benjamin-Feir Index and the ratio of the square of directional width and frequency width to appropriate forms for broad-banded spectra.

After some extensive trial and error the following results were obtained. The dynamic kurtosis factor is given by

$$C_4^{dyn} = J(R)BFI^2, \tag{46}$$

where the form of the the factor $J(R)$ is based on a mathematical analysis given in Janssen & Janssen (2018) regarding the behaviour of the exact narrow-band result around $R = 0$ and $R = 1$. It reads

$$J(R) \approx \frac{\pi}{3\sqrt{3}}\left(1 - \alpha\sqrt{R} + \beta R + \delta R^2\right),\ 0 < R < 1, \tag{47}$$

where $\alpha = 4\sqrt{3}/\pi$ and $\beta = (1/3 + 2\sqrt{3}/\pi)$, and δ is chosen such that $J(R)$ at $R = 1$ satisfies the exact result that $J(R)$ vanishes. This gives the condition $1 - \alpha + \beta + \delta = 0$ from which $\delta = 2\sqrt{3}/\pi - 4/3 = -0.2307$. The expression for $J(R)$ for $R > 1$ follows from the condition $J(R) = -J(1/R)/R$.

In the narrow-band approximation, the evolution of the kurtosis is determined by two dimensionless parameters. The first one is the Benjamin-Feir Index. Introducing the wave steepness $\varepsilon = k_0\sqrt{m_0}$ with m_0 the variance of the surface elevation, while δ_ω is the relative frequency width of the wave spectrum, the deep-water version of the Benjamin-Feir Index BFI is defined as,

$$BFI = \frac{\varepsilon\sqrt{2}}{\delta_\omega} \tag{48}$$

Its shallow water extension is obtained from Eqn. (A13). The relative frequency width is obtained from Goda's peakedness factor Q_p defined as

$$Q_p = \frac{2}{m_0^2}\int_{\mathscr{D}} \mathrm{d}\omega\ \omega E^2(\omega), \tag{49}$$

with $E(\omega)$ the angular frequency spectrum and the integration domain $\mathscr{D}$ consists of all frequencies for which $E(\omega) > E(\omega_p)/4$ (Janssen and Bouws, 1986). The relative angular frequency width δ_ω then follows from

$$\delta_\omega = \frac{1}{Q_p\sqrt{\pi}}. \tag{50}$$

This definition of width emphasizes the peak region of the wave spectrum, which is thought to control the Benjamin-Feir instability.

The second dimensionless parameter measures the importance of directional width δ_θ with respect the frequency width ν_ω. In deep water it is defined as

$$R = \frac{\delta_\theta^2}{2\nu_\omega^2} \tag{51}$$

while the shallow water version is defined in Eqn. (A14). It is important to note that the damping of the Benjamin-Feir instability is caused by the broad-band aspects of the wave spectrum, hence the relative frequency width is given by ν_ω, defined in Eq. (6).

A comparison of the results obtained from the dynamic kurtosis parametrization and the exact computations using spectra from the Draupner and Andrea events is shown in the left panel of Fig. 13. There is a fair agreement but, clearly, compared to for example the bound-wave kurtosis, the agreement is not as good. Fortunately, the probability calculations only involve the total kurtosis, and as shown in the right panel of Fig. 13, this shows a better agreement with correlations of 96%.

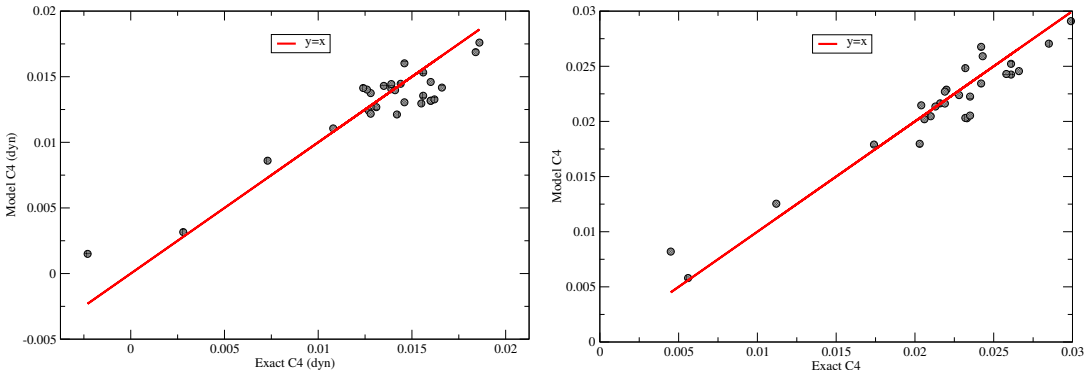

Figure 13: The left panel shows the comparison of parametrized dynamic kurtosis factor C_4^{dyn} against exact computations using spectra from the Draupner and Andrea events, while the right panel shows the comparison for the total kurtosis factor C_4.

4 Examples of Results.

4.1 An example of operational results.

Here, a synoptic example obtained from ECMWF operations is discussed. It concerns an extreme event that was reported by Guidhir *et al.* (2022) in the Irish sea at buoy M4 occuring on the 21^{rst} of February 2022. Datawell and Fugro Wavesense systems recorded, using the zero-crossing method, maximum wave heights of 29.5 m and 28.1 m, respectively. The sample period of the instruments is different ($T_L = 30$ min for Datawell and 17 min for Fugro), and this affects the estimation of maximum wave height as the number of degrees of freedom N_{slc} is proportional to the timeseries length (see Eq. (31)). Operationally at ECMWF, freak wave products are generated assuming a sample period of 20 min. but it is straightforward to calculate from the archived products maximum wave height for a different sample period. In this case the sample period of the Datawell system was taken, i.e. $T_L = 30$ min.

The global maximum wave height map of the 27 hour forecast from the analysis of the 20^{th} of February 2020 at 00 hours is shown in Fig. 14. For the seastate on the 21^{rst} of February at 3 am it is clear that the extreme event at the M4 buoy location (10°W, 55°N) can be regarded as exceptional.

Table 2:

PARAMETERS AT EXTREME WAVE EVENT:	PARAMETERS AT FIELD MAXIMUM:
LAT = 55.0 LONG = 350.0	LAT = 55.5 LONG = 350.0
HMAX = 29.8 +/- 3.0	HMAX = 31.9 +/- 3.4
HS = 14.89	HS = 15.66
FP = 5.99E-002	FP = 5.83E-002
C3 = 5.89E-002	C3 = 5.81E-002
C4 = 3.23E-002	C4 = 7.01E-002
BFI = 0.49	BFI = 0.93
R = 0.35	R = 0.42
U10 = 25.2	U10 = 24.2
AGE = 1.03	AGE = 1.10

In Table 2 values for parameters such as $H_{max}, H_S, F_P, BFI, U_{10}$ and wave age χ are given for the location

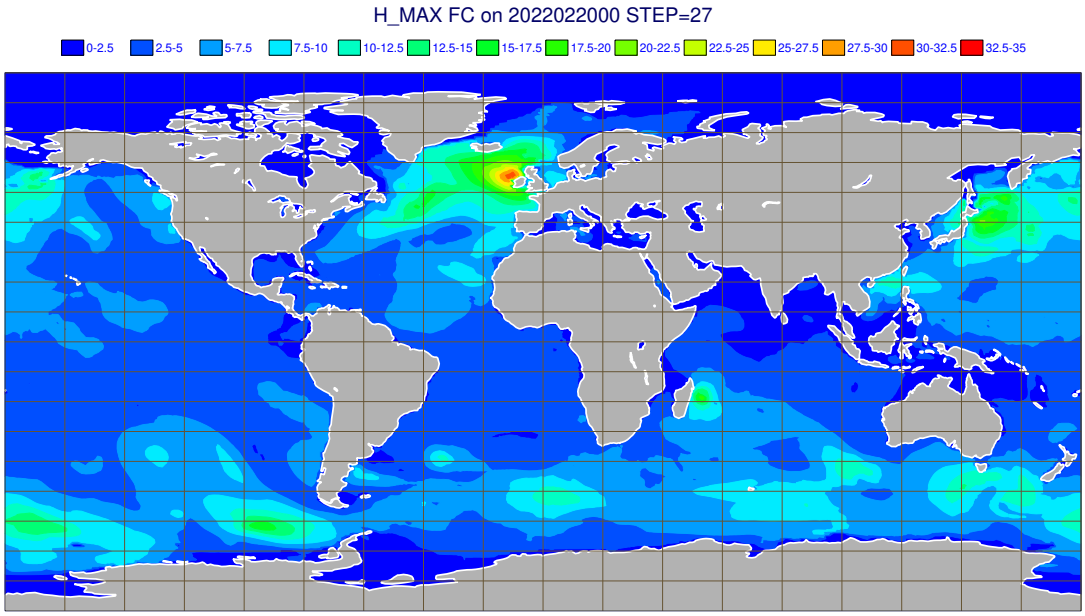

Figure 14: Spatial distribution of maximum wave height H_{max} at the time of an extreme wave height event near wave buoy M4 north west of Ireland.

where according to the observations the extreme wave event occurred. In order to get an impression of how sensitive the parameters of such an extreme event depend on location, also the corresponding parameters at the field maximum for the maximum wave height is shown. This field maximum is just 1/2 degree North of the location of the M4 buoy and comparing parameters such as BFI and C_4 a sensitive dependence of these parameters on location is to be noted resulting in a increase in maximum wave height of 2 m over a distance of around 50 km. Nevertheless, realizing that the observed maximum wave height is obtained with the zero-crossing method and the operational maximum wave height is based on the envelope wave height there seems to be a reasonable agreement between the two.

As a final illustration of the properties of the ECMWF freak wave warning system, some additional global results of a 27 hour forecast of the 20th of February 2022 will be briefly discussed. A number of integral parameters such as significant wave height and several versions of mean period were retrieved which allowed the determination of the moments m_{-1}, m_0, m_1, m_2 needed for the specification of the spectral frequency width parameters. In addition, parameters such as the Benjamin-Feir Index and the angular width were retrieved, enabling the determination of the skewness and kurtosis at every gridpoint and also the determination of the expectation value of maximum envelope wave height. In Fig. 15, results for skewness and kurtosis from a half-degree retrieval are plotted as function of a parameter measuring the stage of development of the sea state, namely the wave age $\chi = c_p/U_{10}$. For young waves, with $\chi < 1$ the envelope skewness C_3 reaches values of 0.1 which corresponds to a surface elevation skewness which is three times larger. The behaviour of the kurtosis parameter C_4 which is the sum of bound-wave kurtosis and dynamic kurtosis is somewhat more complex. Although the bound-wave part of kurtosis is always positive, dynamic kurtosis may become negative, which occurs for spectra with a broad directional distribution. As a consequence, in a number of cases (2% or more) kurtosis turns out to be negative, giving compared to the Gaussian sea state a reduced risk of the occurence of freak waves. On the other hand, for narrow directional spectra, kurtosis is positive, reaching values of up to 0.1, resulting in an enhanced risk for freak waves. Note that these large values of kurtosis are caused by the dynamic part of kurtosis. If one would disregard the dynamic part in the kurtosis calculation than kurtosis would always be positive but would reach at most values of 0.02.

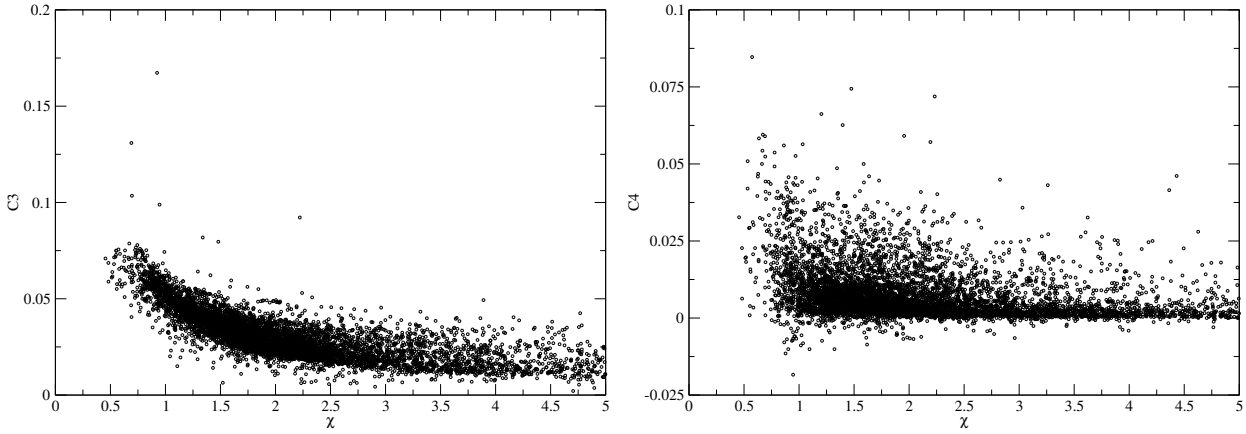

Figure 15: Skewness factor C_3 (left panel) and kurtosis factor C_4 (right panel) versus wave age parameter $\chi = c_p/U_{10}$. Note that there are a number of occasions where the directional wave spectrum is so broad that kurtosis is slightly negative.

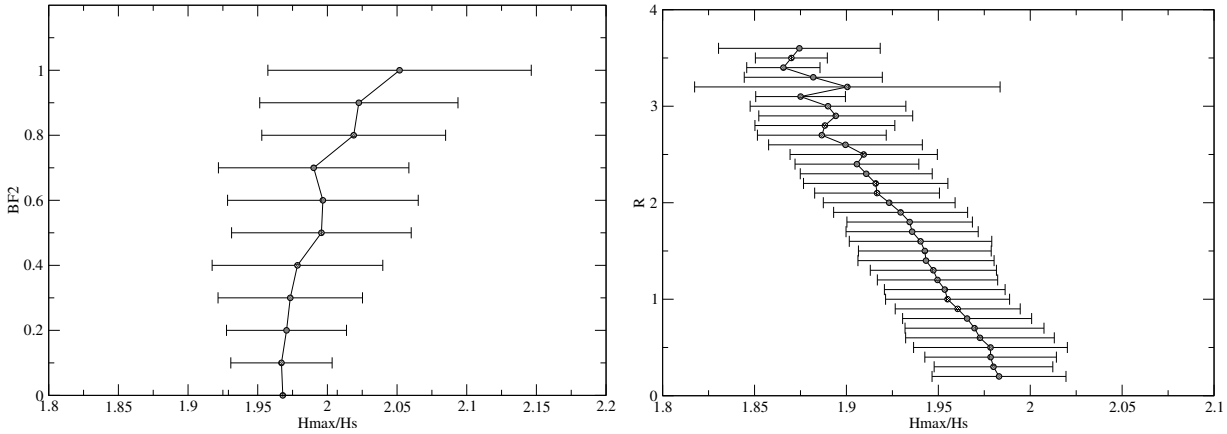

Figure 16: Left Panel: Dependence of expectation value of normalized maximum wave height on the square of the Benjamin-Feir Index (left panel) and on normalized directional width R (right panel). The error bars give the width of the maximum wave height distribution at the BFI^2 or R value of interest

In the next plot, Fig. 16, shows the expectation value of maximum envelope wave height for a 20 min. time series and consider its dependence on parameters such as the square of the Benjamin-Feir Index, BFI^2, and the dimensionless directional width R. These relations are obtained by collecting maximum envelope wave height data as function of a discretized version of the independent parameter and by plotting the bin-average. In agreement with expectations and with results of Burgers *et al.* (2008), based on observations of normalized maximum wave height obtained from AUK platform in the central North Sea, the expectation value of maximum envelope wave height is seen to increase with increasing Benjamin-Feir Index. On the other hand, maximum wave height is seen to decrease, as expected, with increasing dimensionless directional width.

Furthermore, in Janssen and Bidlot (2009) a detailed comparison between observed and modelled distribution of maximum envelope wave height has been made. As argued by these authors, a way to simulate the observed maximum wave height distribution is to start from the theoretical pdf of maximum wave height, the explicit form of which is given in Eqns. 33-35. One then generates from the pdf (33) for given number of waves N_{slc} and given skewness C_3 and kurtosis C_4 a random draw of maximum wave

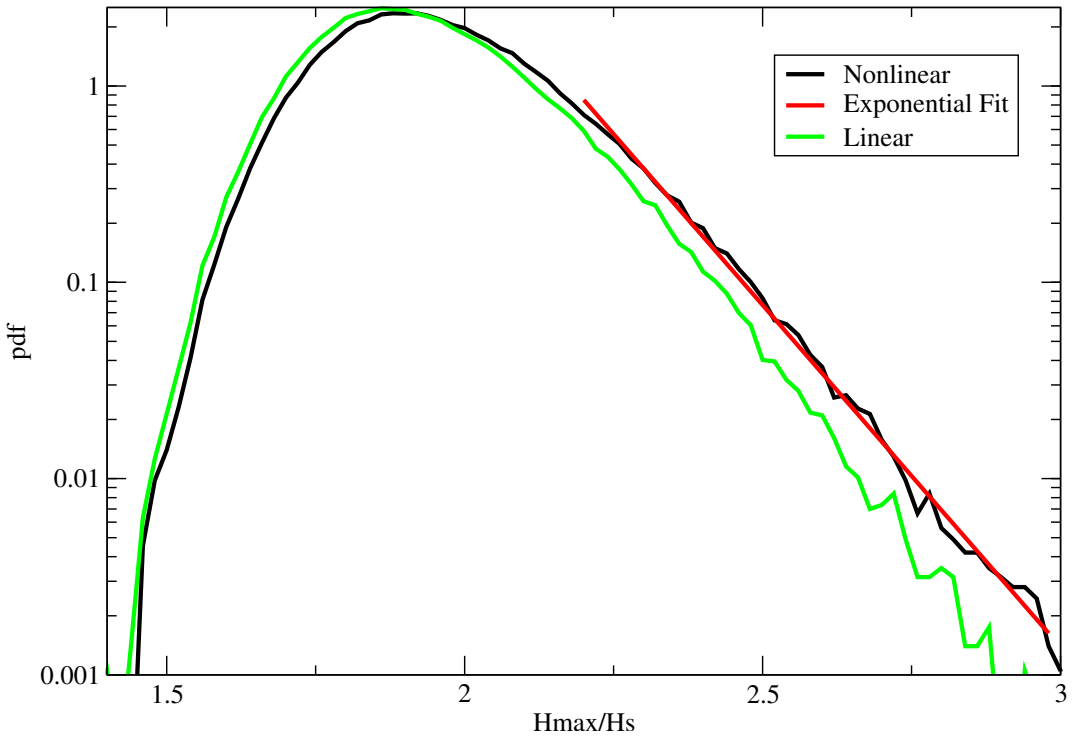

Figure 17: Global maximum wave height distribution obtained by a random draw of H_{max} for given number of waves and given skewness and kurtosis. Including skewness and kurtosis shows that the tail of the distribution is exponential. The distribution according linear theory is shown as well.

height. The usual procedure for this is that one obtains a random draw of maximum envelope wave height from the condition that the cumulative distribution is a random number between 0 and 1. The resulting modelled distribution function is plotted in Fig. 17. It is remarked that the tail of the maximum wave height distribution is basically the number of degrees of freedom N (see (35)) times the envelope wave height p.d.f. (33). Therefore, for large values of maximum envelope wave height the distribution function is an exponential. As can been seen from the present fit, the simulated maximum wave height distribution closely follows the exponential fit, which on a logarithmic plot is a straight line. In order to illustrate effects of nonlinearity, also simulated maximum wave height distribution without nonlinear effects is shown, and clearly, nonlinear effects are seen to play an important role for the extreme events.

Finally, the present approach has been based on an analysis of time series using the envelope. This differs from the usual method whereby one considers wave height or crest height. In these latter approaches an extreme event is called a freak if wave height is larger than 2.2 times the significant wave height or, alternatively, if the crest height is larger than 1.25 times the significant wave height. In my opinion, the envelope wave height follows closely the crest of the waves and therefore, in the context of the envelope approach, I would call an extreme event a 'freak' if the maximum envelope wave height is 2.5 times the significant wave height $H_S = 4\sqrt{m_0}$ with m_0 the zeroth moment of the wave spectrum. Adopting this definition of a freak wave it is therefore of interest to determine the probability that maximum normalized envelope wave height is larger than 2.5. Based on the Global distribution shown in Fig. 17, this probability is according to linear theory about 4×10^{-3} while according to nonlinear theory, this probability is a factor of 2 larger, namely 9×10^{-3}.

4.2 Verification against buoy observations.

An early version of the results of the envelope method, which was already implemented at ECMWF in the mid 2000's, was validated against Canadian buoys over the two-year period of February 2006 until January 2008. Twenty years ago it was not customary to obtain maximum envelope wave height from buoys timeseries. Janssen and Bidlot (2009) resorted to a procedure that over a given sampling period the maximum of the surface elevation was determined and maximum wave height was then given by twice the maximum of the surface elevation. However, this choice gave rise to some uncertainty and basically the length T_L of the timeseries was used as a tuning parameter to obtain an optimal fit between observed and modelled maximum wave height.

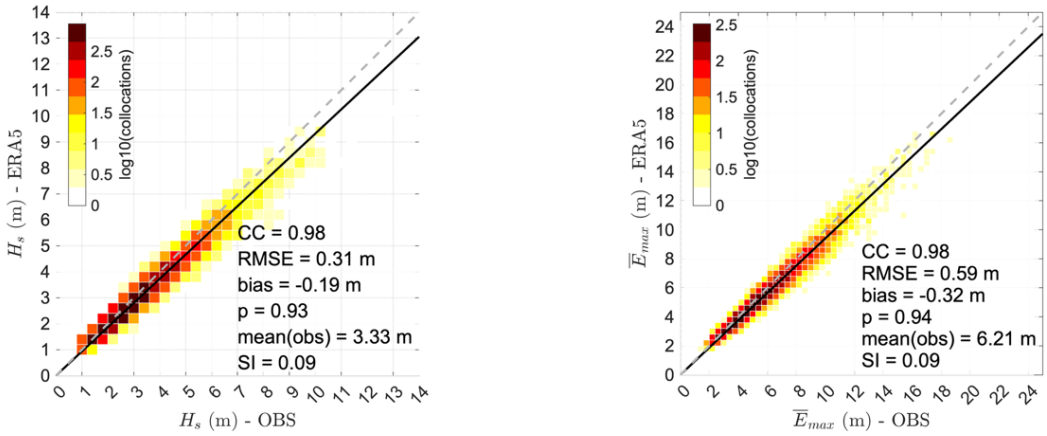

Figure 18: Left Panel: Comparison of ERA5 significant wave height H_S with observed wave height at Ocean Weather Station Papa over a 4 year period. Right Panel: Comparison of ERA5 maximum envelope wave height E_{max} with observed maximum envelope wave height at station Papa.

For the expectation value of maximum wave height over a sampling period of 30 minutes it was found that the bias between modelled and observed was about 10 cm while, with an observed mean value of 3.92 m, the scatter index (defined as the ratio of the standard deviation of error normalised by the mean observed value) was about 16 %, giving a standard deviation of error of 0.63 m. The number of collocations was 30,000. Although tuning was involved this gave a promising agreement as the corresponding scatter index for significant wave height was 13%.

Recently, Barbariol *et al.* (2019) revisited the verification of maximum envelope wave height against buoy observations by considering observed time series of surface elevation η at Ocean Weather Station Papa over the period June 2010 until December 2014. This is a deep water station with a depth of more than 4000 m. The Hilbert transform was applied to these timeseries resulting in an explicit determination of the envelope ρ using Eq. (1). The length of the time series was 20 minutes. These observed maximum wave heights were compared with results from the reanalysis ERA 5, which used regarding the determination of maximum wave height a deep-water version of the statistical model described in this review. Results of the comparison are shown in Fig. 18, and for comparison purposes also the comparison between observed and modelled significant wave height is shown. The bias is about - 0.32 m while the standard deviation of error is 0.59 m, and with a mean observed maximum wave height of 6.21 m the scatter index is 9.5 %. It is remarked that the value of the scatter index is of the same order of magnitude as the estimated 'error' in the expectation value of maximum wave height as obtained from the width of

the p.d.f. of maximum wave height (see discussion around Eq. (41)).

Therefore, also without any tuning there is now a fair agreement between observed and modelled maximum wave height. In closing it is noted that for an accurate estimation of maximum wave height the wave prediction system should produce an accurate value for significant wave height. This is indeed the case as can be inferred from the left panel of Fig. 18, where with a mean observed of 3.33 m and a standard deviation of error of 0.31 m a scatter index of 9.3 % is found. Compared to the routine operational ECMWF wave forecasting verification these are exceptionally accurate estimates of significant wave height as an average scatter index over many northern hemisphere buoys of the wave analysis is typically 13-14 %.

5 Conclusions.

An overview is presented of the theoretical probablistic approach of extreme events, which focusses on the statistical properties of wave energy. Here, wave energy is measured from timeseries of the wave envelope, where the wave envelope is obtained from the surface elevation time series η and the corresponding Hilbert transform ζ. For such a signal, the envelope wave height distribution may be obtained using as starting point the characteristic function of surface elevation and its Hilbert transform. This approach has a restricted range of validity and in order to capture really extreme events an exponential tail has to be added. A procedure to accomplish this has been proposed and results have been validated against Monte Carlo simulations (Janssen, 2015b).

A key quantity to characterize extreme events is the p.d.f. for maximum envelope wave height. Here it is shown that, based on a comparison with Monte Carlo simulations, for envelope wave height the most appropriate approach is the one suggested by Naess (1982). This approach has the added advantage that with the stretched exponential distribution a really simple expression for the expectation value of maximum envelope wave height may be obtained.

In the Appendix a brief overview is given of the analytical formulation of statistical quantities such as skewness and kurtosis in terms of multi-dimensional integrals involving the combination of interaction coefficients and wave spectra. Numerical evaluation of these integrals is in principle possible but requires a considerable amount of computing power. For an operational implementation parametrization of the skewness and kurtosis is highly desirable. In this note I have developed such a parametrization of the wave stats and a satisfactory agreement with the computations using the analytical theory has been obtained. Just recently Gramstad and Lian (2024) have proposed an alternative parametrization of skewness and bound kurtosis using the same exact calculations method for the Jonswap spectrum. No comparison between results has been made so far.

The time series for the exceedance probabilities of maximum wave envelope height suggest that for an area of the size of the resolution of the wave model it is plausible that the Draupner wave and the Andrea event did happen. In addition, knowledge of the deviations from Normality as expressed by skewness and kurtosis play an important part in 'explaining' these events. It is emphasized once more that for a quantity such as envelope wave height it is essential that effects of both bound-wave and dynamic kurtosis need to be included.

In fact, at first it came as a surprise that dynamic kurtosis played such an important role in the freak wave events discussed in this review. First of all, with the work of Janssen and Onorato (2007) in mind, one would expect that the Benjamin-Feir instability plays a relatively minor role for surface gravity waves in intermediate water depth, because the relevant interaction coefficient vanishes for $kD = 1.363$ and

changes sign for even lower dimensionless depth. However, this is only true for one-dimensional modulations. As reminded by Victor Shrira to the author, Hayes (1973) has shown that for two-dimensional modulations there is always instability, though in practice growthrates are vanishingly small for $kD < 0.5$. The consequences of this are discussed in Appendix B, resulting in a parametrization of the nonlinear interaction coefficient that depends explicitely on spectral width in frequency direction space. The upshot is that even for shallow water waves Benjamin-Feir instability may exist and may give a positve contribution to the kurtosis of the sea state.

Secondly, in intermediate water depth wave spectra are much more narrow in direction than in the deep water case. This follows from a plot of the spectra which suggests that the directional distribution of the spectra in waters with dimensionless depth $k_0D = 1.5 - 2$ is much narrower. And experimentation with the single-gridpoint model suggests that the cause for the narrower spectra is connected to the bottom friction term. Perhaps, bottom friction, which mainly affects and removes energy from the long waves, disrupts the normal nonlinear energy fluxes in the spectrum to the the long waves to such an extent that the widening of the angular distribution by the nonlinear transfer is halted. However, more work is required to better understand the role of bottom friction in the formation of the directional distribution.

Before closing, it should be pointed out that, if desired, a very similar approach may be applied to obtain the statistical properties of the surface elevation. However, the details will be different. For example, the maximum surface elevation distribution can be described adequately by means of the method suggested by Goda (2000) but it is still not clear yet how to add an exponential tail to the nonlinear p.d.f. of the surface elevation. Clearly, more research is required to accomplish this.

We conclude with a remark about operational practice. For verification purposes we have concentrated on a time series analysis since most observations that are available are observations of surface elevation in time. Typically, the length of the timeseries is about 20 minutes therefore one deals with around 100 waves. The linear result (41) for maximum wave energy would give a value for normalized maximum wave height of the order of 1.70, whereas using nonlinear effects and the observed values of C_3 and C_4, normalised maximum wave height from the Draupner time series would become about 1.92. Hence, nonlinear effects may have a considerable impact on the expected values for normalised wave height.

However, wave modelling based on the energy balance equation is about predicting the average sea state in a grid box of size $\Delta x \times \Delta y$ by integrating the wave spectrum using a time step Δt. Therefore modern wave prediction can only provide predictions of the sea state in an average grid box. For this reason the most consistent measure one can take is to provide an estimate of the probability of the occurence of extreme events over the 'volume' of the size $\Delta x \times \Delta y \times \Delta t$. As suggested by Fig. 11 there is potential for a sensitive indication of extreme events. However, forecasters will need to adjust their expectations regarding the 'forecast' magnitudes of the normalised maximum wave height. For the Draupner and Andrea cases one typically has normalised maximum wave heights of the order of 3, simply because the number of events is much larger, of the order of 500,000 or more.

Finally, it would be of interest to try to find a simple measure for extreme wave heights given the wave spectrum. In §C.4 I have discussed extensively that for extreme waves the phase information of the individual wavenumber components plays an important role in how extreme the sea state may become by means of constructive interference. Comparing results obtained from the Zakharov equation with random phase and with observed phase of the wave components shows big differences in the kurtosis of the sea state. Therefore, I have tried to introduce an estimate of the severity of the sea state by taking the expression of the surface elevation, estimating the amplitude of the waves using the wave spectrum and giving all wave components the same value of the phase ϕ, i.e, $\cos\phi = 1/3$. Inspecting the increment plot (see Appendix C5) it was a big surprise to see the agreement with the increments to the surface elevation

produced by the observed phase for all wave numbers. Apparently, extreme events are be connected to a coherent sea state. This is most likely related to focussing of wave energy. Of course, focussing may be caused by dispersion giving constructive interference. As is well known, however, the modulational instability which generates wave groups may contribute to focussing as well.

Appendices.

A Skewness and Envelope kurtosis.

A brief summary of the statistical properties of weakly nonlinear, random ocean waves, including shallow water waves is now presented following the work of Janssen (2003), Mori and Janssen (2006), Janssen (2009) and Mori *et al.* (2011). In this approach skewness and kurtosis consists of two contributions, one from the free waves and one from the bound waves. The properties of these contributions differ from each other and therefore they will be discussed separately. For example, while it is straightforward to obtain the narrow-band limit of the bound-wave part of skewness and kurtosis, this is not possible for the free-wave part of the (envelope) kurtosis as for the latter the narrow-band limit does not exist. Therefore, a different approach needs to be followed in order to obtain simplified expressions for the free-wave part of the kurtosis. In contrast to the bound-wave case, it is found that the free-wave contribution to kurtosis has a sensitive dependence on spectral shape, it depends, in particular, on the directional and frequency width.

Following Janssen (2014) a time series analysis is performed in order to evaluate quantities such as the variance of the surface elevation η and its Hilbert transform ζ and a number of skewness terms and the envelope kurtosis. The starting point of the analysis is the Zakharov equation which gives the evolution of the action variable $a(\mathbf{k},t)$ of the free surface gravity waves as caused for four-wave interactions. The form of the Zakharov equation is

$$\frac{\partial a_1}{\partial t} + i\omega_1 a_1 = -i \int \mathrm{d}\mathbf{k}_{2,3,4} T_{1,2,3,4} a_2^* a_3 a_4 \delta_{1+2-3-4}$$

where ω_1 gives the dispersion relation for surface gravity waves and is a positive function of acceleration of gravity g, wave number $|k|$ and depth D. Furthermore, $T_{1,2,3,4}$ is a four-wave interaction coefficient which depends on wave numbers $\mathbf{k}_1$, $\mathbf{k}_2$ etc., the correct form of which was first obtained by Krasitskii (1994). The complete solution, including the bound waves, is obtained by evaluating the canonical transformation. Recall that

$$\eta = \frac{1}{2}(Z + Z^*), \text{ and } \zeta = \frac{1}{2i}(Z - Z^*),$$

where (see Janssen, 2014) the complex function Z is obtained from the canonical transformation by collecting together terms of similar large time behaviour, i.e. Z contains all the terms that vanish for $\Im(t) \to -\infty$. The canonical transformation is only known as an expansion in steepness ε so we write $Z = \varepsilon Z_1 + \varepsilon^2 Z_2 + \varepsilon^3 Z_3$ with

$$Z_1 = 2 \int_{-\infty}^{\infty} \mathrm{d}\mathbf{k}_1 \; f_1 a_1 \; e^{i\theta_1}, \tag{A1}$$

where $f_1 = (\omega_1/2g)^{1/2}$, $\theta_1 = \mathbf{k}_1 \cdot \mathbf{x}$, and

$$Z_2 = 2 \int_{-\infty}^{\infty} \mathrm{d}\mathbf{k}_{1,2,3} \; f_2 f_3 \; e^{i\theta_1} \left\{ \mathscr{A}_{2,3} a_2 a_3 \delta_{1-2-3} + 2\mathscr{B}_{2,3} a_2^* a_3 H_{3-2} \delta_{1+2-3} \right\}, \tag{A2}$$

while

$$Z_3 = 2 \int \mathrm{d}\mathbf{k}_{1,2,3,4} f_2 f_3 f_4 e^{i\theta_1} \left\{ \mathscr{D}_{1,2,3,4} a_2 a_3 a_4 \delta_{1-2-3-4} + \mathscr{C}_{1,2,3,4} a_2 a_3 a_4^* \delta_{1-2-3+4} H_{2+3-4} \right.$$
$$\left. + \mathscr{C}_{-1,2,3,4} a_2^* a_3^* a_4 \delta_{1+2+3-4} H_{4-3-2} \right\}. \tag{A3}$$

The interaction coefficients $\mathscr{A},\mathscr{B},\mathscr{C}$, and $\mathscr{D}$ are given in Janssen (2009). The Heaviside function $H(x)$ is defined in such a way that $H(0) = 1/2$ and in the above formulae the argument of the Heaviside function is a sum of angular frequencies, e.g. $H_{3-2} = H(\omega_3 - \omega_2)$. Note that the term Z_1 represents the contribution of the free waves to the surface elevation, while the terms Z_2 and Z_3 represent the contributions by the bound waves. Therefore, the free-wave contribution to skewness and kurtosis is determined by the statistical properties of Z_1, while the other two terms are required to determine the contributions to bound-wave statistics.

The third and fourth cumulants of surface elevation η and its Hilbert transform ζ are defined as follows: The third cumulants are

$$\kappa_{30} = \frac{\langle \eta^3 \rangle}{\langle \eta^2 \rangle^{3/2}},\ \kappa_{21} = \frac{\langle \eta^2 \zeta \rangle}{\langle \eta^2 \rangle \langle \zeta^2 \rangle^{1/2}},\ \kappa_{12} = \frac{\langle \eta \zeta^2 \rangle}{\langle \eta^2 \rangle^{1/2} \langle \zeta^2 \rangle},\ \kappa_{03} = \frac{\langle \zeta^3 \rangle}{\langle \zeta^2 \rangle^{3/2}}, \tag{A4}$$

while the relevant fourth cumulants are given by

$$\kappa_{40} = \frac{\langle \eta^4 \rangle}{\langle \eta^2 \rangle^2} - 3,\ \kappa_{22} = \frac{\langle \eta^2 \zeta^2 \rangle}{\langle \eta^2 \rangle \langle \zeta^2 \rangle} - 1,\ \kappa_{04} = \frac{\langle \zeta^4 \rangle}{\langle \zeta^2 \rangle^2} - 3, \tag{A5}$$

The procedure to obtain explicit expressions for the cumulants is straightforward but laborious. One substitutes the expressions for η and ζ into the above definitions and evaluates for the bound waves the relevant integrals under the assumption of a homogeneous, Gaussian ocean surface. In order to find a non-trivial result for the free waves one needs to obtain from the Zakharov equation an equation for the free waves contribution of the fourth order cumulants, as detailed in Janssen (2003).

A.1 Free Waves.

The free wave case has been discussed extensively by Mori and Janssen (2006). For the free waves the skewness vanishes while the kurtosis enjoys certain symmetry properties in such a way that $\kappa_{04} = \kappa_{40}$ while $\kappa_{22} = \kappa_{40}/3$. Therefore, using (15) one finds

$$C_4^{free} = \frac{1}{3}\kappa_{40}^{dyn}, \tag{A6}$$

where κ_{40}^{dyn} is explicitely given by Janssen (2003) in terms of the directional angular frequency spectrum $E(\omega,\theta)$, i.e.

$$\kappa_{40}^{dyn} = \frac{12g}{m_0^2} \int \mathrm{d}\theta_{1,2,3} \mathrm{d}\omega_{1,2,3} T_{1,2,3,4} \sqrt{\frac{\omega_4}{\omega_1 \omega_2 \omega_3}} \times G(\Delta\omega, t) E_1 E_2 E_3. \tag{A7}$$

Here, $G(\Delta\omega,t)$ is the time-dependent real part of the resonance function, defined by

$$G(\Delta\omega,t) = \frac{1-\cos(\Delta\omega t)}{\Delta\omega},$$

while the frequency mismatch $\Delta\omega$ is given by $\Delta\omega = \omega_1 + \omega_2 - \omega_3 - \omega_4$, and the fourth wave number follows from the resonance condition in wave number space, i.e. $\mathbf{k}_4 = \mathbf{k}_1 + \mathbf{k}_2 - \mathbf{k}_3$. The frequency ω_4 is obtained by evaluating the dispersion relation at the fourth wavenumber. Eq. (A7) gives the evolution in time of the dynamic part of the kurtosis for given wave spectrum. In general no analytic solution is known so a numerical evaluation of this six-dimensional integral is required. Software has been written to calculate κ_{40}^{dyn} for arbitrary spectra and arbitrary depth D under the restriction that $k_0 D > 1$. The

software has been validated for a number of special cases of which the solution is known, namely for Gaussian spectra in the narrow-band approximation.

It should be realized that Eq. (A7) is quite expensive to evaluate because it involves a six-dimensional integral. With a resolution of 36 frequencies and 36 directions the number of evaluations of the transfer coefficient is $36^6 = 2.18$ Billion, which is quite substantial. For this reason the simulated spectra of the Draupner case were inspected before hand. It turns out that a large part of the frequency-direction space contained small values so that the contribution of these small values could be ignored. This filtering operation reduced the number of evaluations of the transfer coefficient to about 11 million. Therefore, for a limited number of spectra, the filtering operation makes numerical evaluation of the free wave kurtosis feasible.

Another point to mention is that the evaluation of the integral involving a rapidly varying resonance function G is very difficult. For large times the time behaviour becomes erratic but there is, in the case of general spectra, no criterium available which indicates when the numerical solution becomes chaotic. In practice, the solution as a function of time was simply plotted and checked by eye to see whether the solution was smooth or not.

A.1.1 The narrow-band approximation.

Nevertheless, in operational applications the numerical evaluation of the six-dimensional integral is far too expensive, and an efficient approximation is highly desirable. This efficient approximation is found using the narrow-band approximation applied to Eq. (A7). In this approximation most of the wave energy is concentrated around the carrier wave number k_0, so that the wave spectrum has a small relative frequency width $\delta_\omega = \sigma_\omega/\omega_0$ and a small angular width δ_θ. For surface gravity waves on water of finite depth D the dispersion relation for the carrier reads

$$\omega_0 = \sqrt{gk_0T_0},\ T_0 = \tanh x,\ x = k_0D, \tag{A8}$$

while the first and second derivative become

$$v_g = \omega_0' = \frac{1}{2}c_0\left\{1 + \frac{2x}{\sinh 2x}\right\},\ c_0 = \frac{\omega_0}{k_0}, \tag{A9}$$

and

$$\omega_0'' = -\frac{g}{4\omega_0 k_0 T_0} \times \Omega'', \tag{A10}$$

with

$$\Omega'' = \left\{T_0 - x\left(1 - T_0^2\right)\right\}^2 + 4x^2T_0^2\left(1 - T_0^2\right). \tag{A11}$$

Note that for any value of the depth D the second derivative is always negative. Finally, the narrow-band limit of the normalised nonlinear interaction coefficient $X_{nl} = T_{0,0,0,0}/k_0^3$ is given by

$$X_{nl} = \frac{9T_0^4 - 10T_0^2 + 9}{8T_0^3} - \frac{1}{x}\left\{\frac{(2v_g - c_0/2)^2}{c_S^2 - v_g^2} + 1\right\} + \frac{\kappa_1\nu_1}{\mu_1 k_0^3}\frac{\delta_\theta^2}{\delta_\theta^2 + \alpha_\omega\delta_\omega^2} \tag{A12}$$

with κ_1, ν_1 and $\mu_1 = c_S^2$ wavenumber and depth dependent coefficients given in (B6), $c_S = \sqrt{gD}$ the shallow water wave velocity and

$$\alpha_\omega = \frac{c_0^2}{v_g^2}\left(1 - \frac{v_g^2}{c_S^2}\right).$$

Note that the interaction coefficient consists of three terms. The first two terms are well-known (see e.g. Whitham, 1974) as the first term is connected with the nonlinear dispersion relation for surface gravity waves, while the second term is due to effects of wave-induced current caused by one-dimensional, longitudinal modulations. These two terms are of definite sign so they may cancel each other, which, in fact, happens for $x = k_0 D = 1.363$. The third term is new and is connected to the wave-induced current caused by two-dimensional modulations. It also has a definite sign and tends to reduce the dimensionless depth where the nonlinear interaction coefficient vanishes. A complete derivation of this new term and some discussion is presented in Appendix B. Hence, for intermediate water depth waves the nonlinear interactions are, because waves are two-dimensional, still expected to play a relatively important role.

In the narrow-band approximation, the evolution of the kurtosis is determined by two dimensionless parameters. The first one is the shallow water extension of the Benjamin-Feir index. Introducing the wave steepness $\varepsilon = k_0\sqrt{m_0}$, with m_0 the variance of the surface elevation, while δ_ω is the relative frequency width, the Benjamin-Feir Index BFI is defined as

$$BFI^2 = \frac{8\varepsilon^2}{\delta_\omega^2} \times \left(\frac{v_g}{c_0}\right)^2 \times \frac{T_0}{\Omega''} \times X_{nl} \tag{A13}$$

Note that in the deep-water limit one recovers the usual expression for the Benjamin-Feir Index, as $\lim_{k_0 D\to\infty} BFI^2 = 2\varepsilon^2/\delta_\omega^2$.

The second dimensionless parameter measures the importance of directional width δ_θ with respect the frequency width δ_ω. In deep water it is defined as $R = \delta_\theta^2/2\delta_\omega^2$ while in general one has

$$R = 4\frac{\delta_\theta^2}{\delta_\omega^2} \times \left(\frac{v_g}{c_0}\right)^3 \times \frac{T_0^2}{\Omega''} \tag{A14}$$

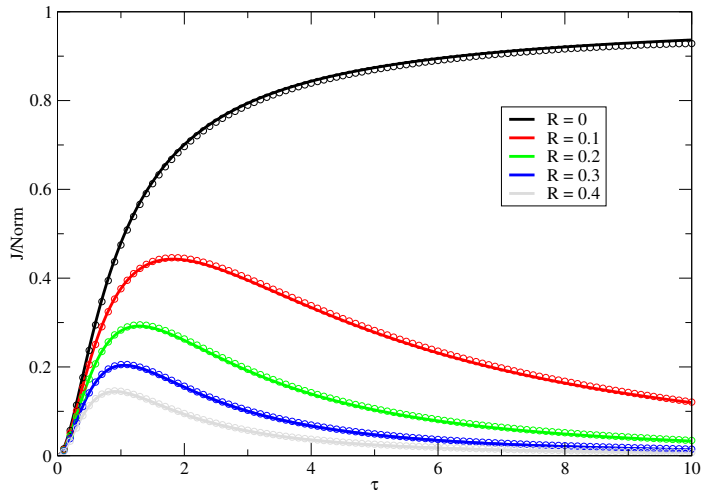

Figure A1: Evolution of normalized kurtosis versus time for different values of R. Results using the numerical solution of (A15) are shown as the full lines, while the circles denote the analytical solution from Fedele (2015).

Applying now the narrow-band approximation to Eq. (A7) the parameter C_4^{free} becomes

$$C_4^{free} = J(R,\tau)BFI^2, \tag{A15}$$

where

$$J(R,\tau) = 2\int \mathrm{d}\nu_{1,2,3}\mathrm{d}\phi_{1,2,3}E_1E_2E_3G(\Delta\omega,\tau), \tag{A16}$$

with resonance function $G(\Delta\omega,\tau) = (1-\cos(\Delta\omega\tau)/\Delta\omega$. Here, $\tau = \delta_\omega^2\omega_0 t$ is dimensionless time, and $\Delta\omega = (\nu_3-\nu_1)(\nu_3-\nu_2) - R(\phi_3-\phi_1)(\phi_3-\phi_2)$. The frequency parameter ν and direction parameter ϕ have been made dimensionless by means of the frequency width and the directional width.

For a Gaussian spectrum

$$E_1 = \frac{1}{2\pi} e^{-\frac{1}{2}(\nu_1^2+\phi_1^2)},$$

the integral in Eq. (A16) may be solved exactly (see e.g. Fedele (2015), and Janssen & Janssen (2018)) and the resulting solution for different values of R as a function of dimensionless time τ has been plotted in Fig. A1. Note that in the plot the parameter J has been normalized with the factor $N(orm) = \pi/3\sqrt{3}$ which is the time-asymptotic value of J for $R = 0$.

From the plot it is clear, that, except for the case of uni-directional propagation (i.e. $R = 0$), the parameter J and therefore the kurtosis parameter C_4^{free} has a maximum at the finite time $\tau_{max} = 1/(3R)^{1/2}$ after which it decays to zero for large times. As an indicator of the severity of the weakly nonlinear sea state one could use the maximum value of J. In Fig. A2 the maximum of J as function of R is plotted as obtained from the exact solution. Janssen & Janssen (2018) obtained an approximate expression for J using the behaviour of the exact narrow-band result around $R = 0$ and $R = 1$. It reads

$$\frac{J}{N} = 1 - \alpha\sqrt{R} + \beta R + \delta R^2,\ N = \frac{\pi}{3\sqrt{3}}, \tag{A17}$$

where $\alpha = 4\sqrt{3}/\pi$ and $\beta = (1/3 + 2\sqrt{3}/\pi)$, and δ is chosen such that $J(R)$ at $R = 1$ satisfies the exact result that $J(R)$ vanishes. This gives the condition $1-\alpha+\beta+\delta = 0$ from which $\delta = 2\sqrt{3}/\pi - 4/3 = -0.2307$. The expression for $J(R)$ for $R > 1$ follows from the condition $J(R) = -J(1/R)/R$. As evident from Fig. A2 the fit (A17) approximates the exact results very well.

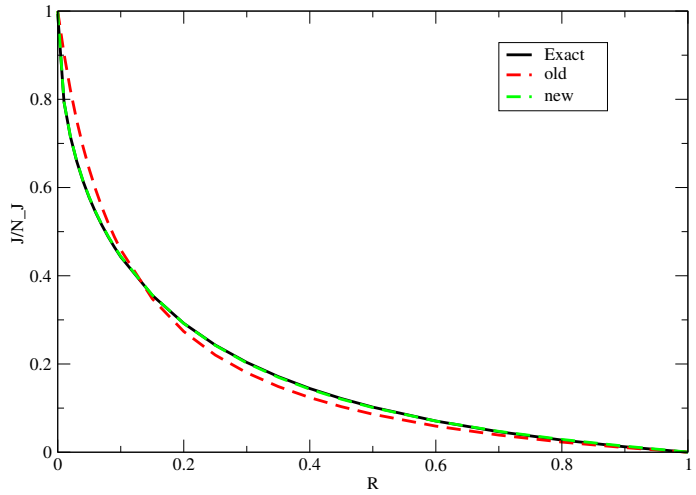

Figure A2: Maximum value of J versus R. "New" refers to the fit (A17) while "old" refers to an earlier attempt.

A.2 Bound Waves.

The case of bound waves has been extensively discussed by Janssen (2015b). The result for general spectra will be briefly recorded, and also the corresponding results for the case of a single wave train are recorded. It turns out, namely, that, in practice, the narrow-band limit serves as a reasonable approximation to the case of general spectra, so that these simplified results may be used in the operational

implementation of the p.d.f. of extreme events. The wave train is given by the Stokes wave solution up to third order in amplitude that is consistent with the narrow-band approximation for the general case of arbitrary spectra.

The narrow-band limit follows in a straightforward fashion from the complex function Z by using a wavenumber spectrum with a Dirac delta function, i.e. $E(\mathbf{k}) = m_0\delta(\mathbf{k}-\mathbf{k}_0)$ where m_0 is the variance of the sea surface, and $\mathbf{k}_0$ is the peak wave number. In effect, all the interaction coefficients are replaced by their value at the peak wave number. Writing

$$\mathscr{A}_{0,0} = 2\alpha, \mathscr{B}_{0,0} = 2\Delta, \mathscr{C}_{0,0,0,0} = 4\gamma, \text{ and } \mathscr{D}_{0,0,0,0} = 4\beta, \tag{A18}$$

the complex function Z becomes

$$Z = m_0 D + m_0^{1/2} A e^{i\theta} + m_0 B e^{2i\theta} + m_0^{3/2} C e^{3i\theta}, \tag{A19}$$

with $D = \Delta(a^2 - \langle a^2\rangle), A = a(1+\gamma\varepsilon^2 a), B = \alpha a^2, C = \beta a^3$, and $\theta = k_0 x - \omega_0 t + \phi$, with ω_0 the angular peak frequency and ϕ an arbitrary phase. The coefficients α, β, γ and Δ are known functions of peak wavenumber and depth D and they read

$$\Delta = -\frac{k_0}{4}\frac{c_S^2}{c_S^2 - v_g^2}\left[\frac{2(1-T_0^2)}{T_0} + \frac{1}{x}\right] + \frac{k_0^2\kappa_1}{2\omega_0\mu_1}\frac{\delta_\theta^2}{\delta_\theta^2 + \alpha_\omega\delta_\omega^2},$$

$$\alpha = \frac{k_0}{4T_0^3}\left(3 - T_0^2\right), \ \beta = \frac{3k_0^2}{64T_0^6}\left[8 + \left(1-T_0^2\right)^3\right], \ \gamma = -\frac{1}{2}\alpha^2, \tag{A20}$$

where $x = k_0 D$, $T_0 = \tanh x$, $c_S^2 = gD$, $v_g = \partial\omega/\partial k$, and $\omega_0 = (gk_0T_0)^{1/2}$. The second contribution to the mean sea level term Δ is new and is calculated in Appendix B. This additional contribution illustrates nicely that the narrow-band limit of the mean sea level is not unique as it depends on the order in which the limit of vanishing width is taken.

The form (A19) has been used to calculate explicitely all the relevant statistical moments following the method in Janssen (2009, Appendix A.3). Note that in these calculations a certain ordering of the contributions to the surface elevation. has been used. Using the significant steepness $\varepsilon = k_0 m_0^{1/2}$, with typical magnitude in the range 0.01-0.05, one finds that the constant term in (A19) is of order ε^2 while the first, second and third harmonic are of order ε, ε^2 and ε^3 respectively. Calculations of the statistical moments have been performed up to lowest significant order, which means that the calculation is continued up to the first nontrivial contribution of nonlinearity.

A.2.1 Variance.

It is straightforward to express the variances $\langle\eta^2\rangle$ and $\langle\zeta^2\rangle$ in terms of the complex envelope function Z. The result becomes

$$\langle\eta^2\rangle = \frac{1}{2}\left(\langle|Z|^2\rangle + \langle Z^2\rangle\right), \ \langle\zeta^2\rangle = \frac{1}{2}\left(\langle|Z|^2\rangle - \langle Z^2\rangle\right). \tag{A21}$$

For a homogeneous, Gaussian sea state the variances become to lowest significant order

$$\langle\eta^2\rangle = \int \mathrm{d}\mathbf{k}_1 \, E_1 + \int \mathrm{d}\mathbf{k}_{1,2} E_1 E_2 \left\{\mathscr{A}_{1,2}^2 + \mathscr{B}_{1,2}^2 + 2\mathscr{C}_{1,1,2,2}\right\},$$

$$\langle \zeta^2 \rangle = \int \mathrm{d}\mathbf{k}_1 \, E_1 + \int \mathrm{d}\mathbf{k}_{1,2} E_1 E_2 \left\{ \mathscr{A}_{1,2}^2 + \mathscr{B}_{1,2}^2 (H_{1,2} - H_{2,1})^2 + 2\mathscr{C}_{1,1,2,2} \right\}. \tag{A22}$$

The expression for the variance $\langle \eta^2 \rangle$ agrees with an earlier result obtained in Janssen (2009, Eq. (50)). Note that formally the variance of ζ may differ from the variance of η, because of the presence of the additional factor $F_{1,2} = (H_{1,2} - H_{2,1})^2$. The function $F_{1,2}$ equals 1 everywhere except when its arguments are equal. In that event $F_{1,1}$ vanishes. However, if the remainder of the integrand is continuous then removing one point from an integral should not affect the result. Hence, for continuous spectra the variance of η and ζ are the same. The exception is when the remainder of the integrand is singular at the point that is being removed. An example of this is the case of a single wave which has a delta function spectrum. As a consequence, for the single wave case the variances of η and ζ are different. In fact, taking the narrow-band limit of (A22) one finds using Eq. (A18)

$$\langle \eta^2 \rangle = m_0 + 4m_0^2 \left(2\gamma + \alpha^2 + \Delta^2 \right), \; \langle \zeta^2 \rangle = m_0 + 4m_0^2 \left(2\gamma + \alpha^2 \right), \tag{A23}$$

and the variances differ by the amount $4m_0^2\Delta^2$. Expression (A23) is in agreement with the single mode results using Eq. (A19). Note that from the expression of the single mode complex envelope function it is immediately clear that the variance of ζ cannot depend on the parameter Δ, since the Hilbert transform of a constant vanishes.

A.2.2 Skewness.

We need to evaluate skewness terms for the surface elevation η and its Hilbert transform ζ. In terms of the complex function Z one finds

$$\langle \eta^3 \rangle = \frac{1}{8} \left(\left\{ \langle Z^3 \rangle + 3 \langle |Z|^2 Z \rangle \right\} + c.c., \; \langle \eta^2 \zeta \rangle = \frac{1}{8i} \left\{ \langle Z^3 \rangle + \langle |Z|^2 Z \rangle - c.c. \right\},$$

while

$$\langle \eta \zeta^2 \rangle = \frac{1}{8} \left\{ \langle |Z|^2 Z \rangle - \langle Z^3 \rangle \right\} + c.c., \; \langle \zeta^3 \rangle = \frac{i}{8} \left\{ \langle Z^3 \rangle - 3 \langle |Z|^2 Z \rangle - c.c. \right\}.$$

so we only have to evaluate the moments $\langle Z^3 \rangle$ and $\langle |Z|^2 Z \rangle$. To lowest significant order in ε we only need the first two terms of the complex function Z, i.e. (A1)-(A2), Then, by invoking the random phase approximation it is straightforward to establish that

$$\langle Z^3 \rangle = 0 + \mathscr{O}(\varepsilon^5), \; \langle |Z|^2 Z \rangle = 4 \int \mathrm{d}\mathbf{k}_{1,2} \left(\mathscr{A}_{1,2} + \mathscr{B}_{1,2} \right) E_1 E_2 + \mathscr{O}(\varepsilon^5), \tag{A24}$$

where $\mathscr{A}_{1,2}$ and $\mathscr{B}_{1,2}$ are given in Janssen (2009). Hence, the Z-moments either vanish or are real. The direct consequence is that the surface elevation moments involving odd powers of ζ vanish, i.e. $\langle \eta^2 \zeta \rangle = \langle \zeta^3 \rangle = 0$. The remaining moments become

$$\langle \eta^3 \rangle = 3 \int \mathrm{d}\mathbf{k}_{1,2} E_1 E_2 \left(\mathscr{A}_{1,2} + \mathscr{B}_{1,2} \right), \; \langle \eta \zeta^2 \rangle = \frac{1}{3} \langle \eta^3 \rangle. \tag{A25}$$

The eventual result is

$$\kappa_{30} = \frac{3}{m_0^{3/2}} \int \mathrm{d}\mathbf{k}_{1,2} E_1 E_2 \left(\mathscr{A}_{1,2} + \mathscr{B}_{1,2} \right), \; \kappa_{12} = \frac{\kappa_{30}}{3}, \tag{A26}$$

while κ_{21} and κ_{03} vanish. Hence, the skewness terms for general spectra share the same properties as the skewness terms for a single wave train in deep water (Janssen, 2014). Note that in the narrow-band limit the skewness term κ_{30} assumes the simple form

$$\kappa_{30} = 6m_0^{1/2}(\alpha + \Delta), \tag{A27}$$

a result which is in agreement with Janssen (2009).

Finally, in the envelope wave height pdf we have introduced a skewness factor C_3 according to Eq. (15). Making use of the properties of the skewness terms one finds eventually

$$C_3 = \frac{\kappa_{30}}{3}, \tag{A28}$$

and this expression for C_3 has been used to determine the skewness factor in the wave height pdf from the simulated spectra. In Fig A3 we shown the dependence of the skewness factor C_3 as a function of dimensionless depth for the one-dimensional case and for the two-dimensional case which has a pronounced impact on the mean elevation Δ (see Appendix B) and on the skewness factor.

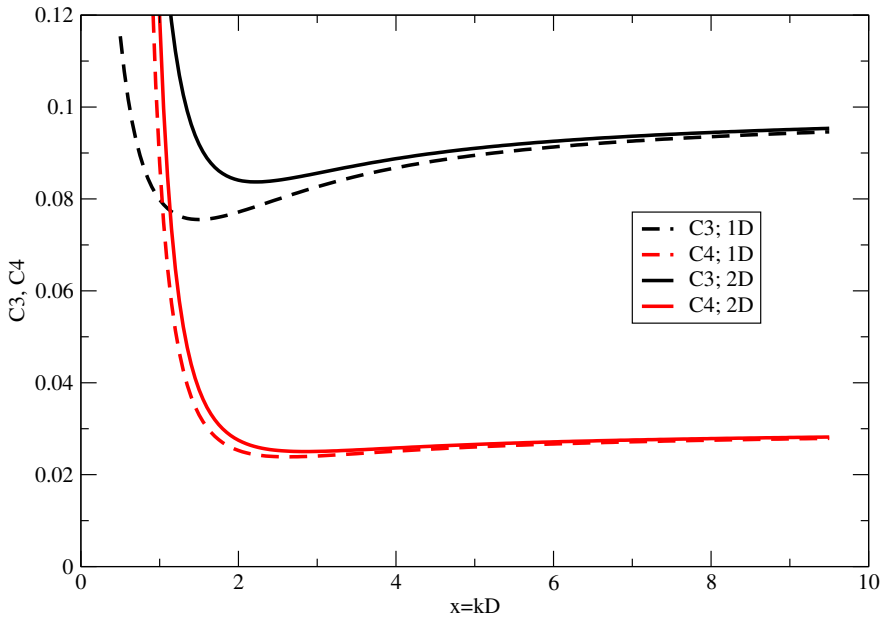

Figure A3: Dependence of skewness factor C_3 and kurtosis factor C_4 on dimensionless depth x according to the single mode results with a significant steepness of 0.1. The difference between 1D and 2D is quite substantial for the skewness factor C_3.

A.2.3 Envelope kurtosis.

The procedure to obtain the fourth cumulants of surface elevation η and its Hilbert transform ζ, i.e. κ_{40}, κ_{22}, and κ_{04} is formally the same as the one used to obtain the third cumulants such as the skewness of the surface elevation. Again, one introduces the complex function Z but it is now more involved because of third-order nonlinearity, i.e. $Z = \varepsilon Z_1 + \varepsilon^2 Z_2 + \varepsilon^3 Z_3$, where Z_1, Z_2 and Z_3 are given by (A1), (A2) and (A3) respectively.

The main purpose is to calculate the envelope kurtosis κ_4. It is defined as

$$\kappa_4 = \kappa_{40} + 2\kappa_{22} + \kappa_{04}, \tag{A29}$$

where the surface elevation kurtosis terms are given by Eq. (A5).

In passing, it is of interest to explain why κ_4 is called the envelope kurtosis. This is most easily illustrated for the case that η and ζ have the same variance, which is true for continuous spectra. From the definition (A29), while using the definitions of κ_{40}, κ_{22}, and κ_{04} it is then straightforward to show that the envelope kurtosis κ_4 is just given, as expected, by the normalized fourth moment of the envelope $\rho = \sqrt{\eta^2+\zeta^2}$, i.e.

$$\kappa_4 = \frac{\langle \rho^4 \rangle}{\langle \eta^2 \rangle^2} - 8 = \frac{\langle |Z|^4 \rangle}{\langle \eta^2 \rangle^2} - 8 \tag{A30}$$

hence, in order to obtain the envelope kurtosis one only needs to determine the fourth moment $\langle |Z|^4 \rangle$.

It is now a straightforward but very laborious task to evaluate the above Z-moment for a homogeneous, Gaussian sea state. This detail is discussed in Janssen (2015b) where also the evaluation of the moments κ_{40}, κ_{22} and κ_{04} is described in some detail. The eventual result for the envelope kurtosis κ_4 becomes

$$\kappa_4 = \frac{32}{m_0^2} \int \mathrm{d}\mathbf{k}_{1,2,3} E_1 E_2 E_3 \left\{ \mathscr{A}_{1,2}\mathscr{A}_{2,3} + \mathscr{A}_{1,2}\mathscr{B}_{2,3} + \frac{1}{2}\mathscr{C}_{1+2-3,1,2,3} H_{1+2-3} + \mathscr{B}_{1,3}\mathscr{B}_{3,2}\mathscr{H}_{1,2,3} \right\} \tag{A31}$$

where, $\mathscr{H}_{1,2,3} = [H_{3-2}H_{3-1} + H_{2-3}H_{1-3}]$ while the coupling coefficients are given in Janssen (2009). Then, the bound-wave part of the kurtosis factor C_4 follows from Eq. (14), hence

$$C_4^{bound} = \frac{\kappa_4}{8}. \tag{A32}$$

Inspecting the expression for the envelope kurtosis it is seen that κ_4 does not depend on the matrix $\mathscr{D}$ which represents the contribution of third harmonics. For equal variance of η and ζ one may give an even more general argument why the envelope kurtosis is independent of the third harmonics. This is related to Eq. (A30) which shows that the envelope kurtosis depends on $\langle |Z|^4 \rangle$ only and it is straightforward to prove that to lowest significant order third harmonics cannot contribute to this fourth moment of Z.

Finally, the general results for the kurtosis were checked by taking the limit of a narrow-band wave train. Using (A18) one finds that the envelope kurtosis becomes

$$\kappa_4 \;=\; 64 m_0 \left(\gamma + \alpha^2 + (\alpha+\Delta)^2\right), \tag{A33}$$

and exactly the same results for the kurtosis parameters are found when one starts from the single mode representation (A19) following the method in Janssen (2009). Again the envelope kurtosis does not depend on the amplitude of the third harmonic, which is given by β. The kurtosis elements κ_{40} and κ_{04} do depend on the third harmonic, however. For completeness, therefore, also the narrow-band results for the other kurtosis elements are given. According to Janssen (2015b) they become

$$\begin{aligned} \kappa_{40} &= 24 m_0 \left(\gamma + \beta + 2(\alpha+\Delta)^2\right), \\ \kappa_{22} &= 8 m_0 \left(\gamma + \alpha^2 + (\alpha+\Delta)^2\right), \\ \kappa_{04} &= 24 m_0 \left(\gamma - \beta + 2\alpha^2\right), \end{aligned} \tag{A34}$$

and, indeed, both κ_{40} and κ_{04} depend on the the contribution by the third harmonic but with opposite signs and therefore, according to Eq. (A29), the envelope kurtosis κ_4 does not depend on the third harmonic.

The single mode results give a reasonable approximation to the statistics of the case of a wind sea spectrum. We use these results to illustrate for a significant steepness of $\varepsilon = k_0 m_0^{1/2} = 0.1$ the dependence of the skewness factor $C_3 = \kappa_{30}/3$ and the kurtosis factor $C_4 = \kappa_4/8$ on dimensionless depth $k_0 D$. This is shown in Fig. A3. Note the sensitive dependence of these statistical factors on dimensionless depth in the range of 1-2.

B Shallow water effects revisited.

B.1 Introduction.

The Benjamin-Feir (B.F.) instability is thought to play a key role in the formation of freak waves, resulting in a considerable amount of research on the properties of this instability. In this Appendix we would like to concentrate on the shallow-water effects. Janssen and Onorato (2007) studied the shallow-water case and found, in agreement with Benjamin (1967), Whitham (1974), Hayes (1973), that for *one-dimensional propagation* the B.F. instability disappears for $kD < 1.363$. This property was also shown to follow from the narrow-band version of the Zakharov equation (Zakharov, 1968) by taking appropriate limits in order to avoid apparent singularities in the nonlinear transfer coefficients. Also, the numerical solution of the Zakharov equation showed that for $kD > 1.363$ the kurtosis was positive, while in the opposite case the sign of kurtosis changed, suggesting that there is a direct connection between the B.F. instability and the generation of freak waves.

In the present version of the ECMWF freak-wave warning system the B.F. instability is taken into account by using the narrow-band approximation of the Zakharov equation. For example, the nonlinear transfer function becomes for 1-D propagation

$$T_{0,0,0,0}/k_0^3 = \frac{9T_0^4 - 10T_0^2 + 9}{8T_0^3} - \frac{1}{x}\left\{\frac{(2v_g - c_0/2)^2}{c_S^2 - v_g^2} + 1\right\} + \text{Extra.} \tag{B1}$$

where $x = k_0 D$, $T_0 = \tanh x$, c_0 is the phase speed ω_0/k_0, $\omega_0 = (gk_0T_0)^{1/2}$, v_g is the group velocity $\partial\omega_0/\partial k_0$, and c_S is the wave speed in shallow water, $c_S^2 = gh$. In the 1-D case the Extra term vanishes.

Numerically, it can be shown that the 1-D version of (B1) vanishes for $k_0 D = 1.363$ (because Extra vanishes) and that for $k_0 D < 1.363$ the nonlinear transfer coefficient becomes negative suggesting that there is no freak wave formation in that case. However, this is not correct for 2-D propagation. Taking into account 2-D modulations there is instability even for $k_0 D < 1.363$ as shown by Hayes (1973) and later by Davey and Stewartson (1974).

Thus, one needs to consider in detail the case of 2-dimensional modulations in shallow water. However, this is not so straightforward when starting from the Zakharov equation because the narrow-band limit of the nonlinear transfer coefficient is awkward. Nevertheless, Onorato and Janssen (2024) have shown that the non-uniqueness of the narrow-band approximation is connected to the initial condition for the current and therefore the narrow-band version of the Zakharov equation is equivalent to the Davey Stewartson equations. Here, the Davey-Stewartson equations are used to obtain an extra contribution to the nonlinear transfer function of the form

$$\text{Extra} = \frac{1}{4T_0}\frac{[2c_0 + v_g(1 - T_0^2)]^2}{c_S^2 - v_g^2}\frac{\delta_\theta^2}{\delta_\theta^2 + \alpha_\omega \delta_\omega^2}. \tag{B2}$$

which is always positive, therefore this tends to enhance the nonlinear transfer. Here, δ_θ and δ_ω are the relative spectral widths in $\theta-$space and $\omega-$space, respectively. Note that the dependence on the width of spectrum is rather peculiar: the limiting value depends on the order in which the limits of δ_θ and δ_ω are taken. Therefore, in the presence of the wave-induced current induced by modulations that have a transverse wavenumber dependence the narrow-band limit is not unique. This is a problematic feature when one is interested in the narrow-band limit of the nonlinear transfer coefficient, but this problem can be resolved by explicitely taking into account the dependence on the modulation wavenumbers in the along and the cross direction with respect the propagation direction.

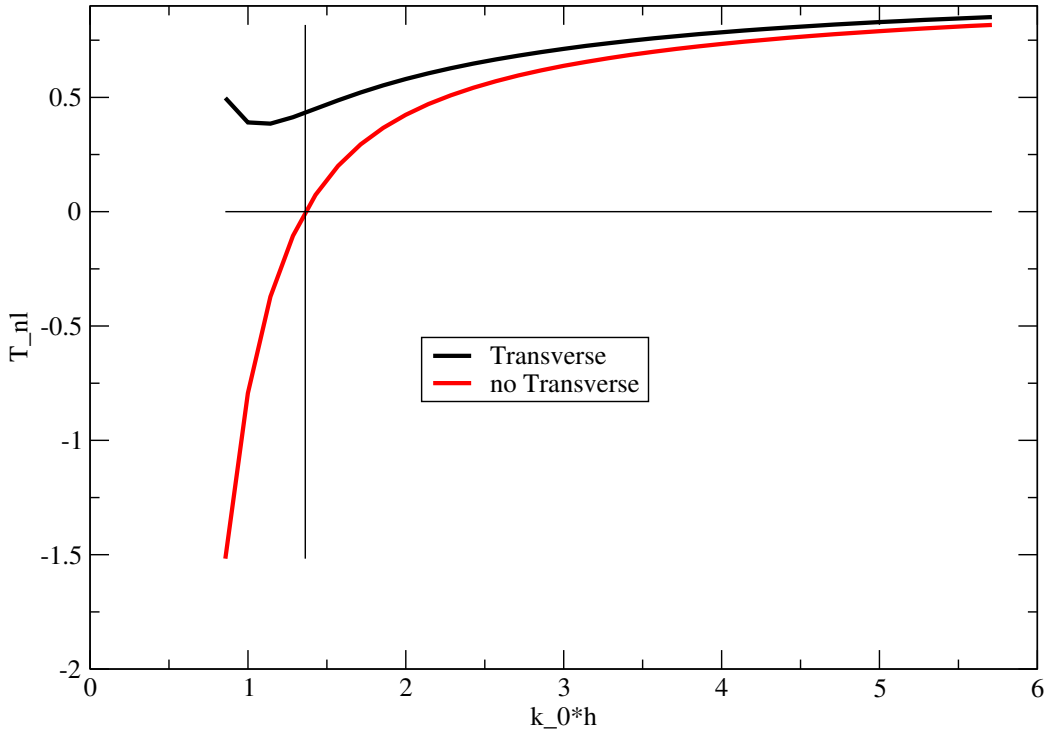

Figure B1: Nonlinear transfer coefficient $T_{0,0,0,0}/k_0^3$ with and without transverse modulations as a function of dimensionless depth k_0D.

The importance of the transverse effect is illustrated in Fig. B1 which shows the nonlinear transfer coefficient $T_{0,0,0,0}$ normalised with the factor k_0^3 as a function of dimensionless depth k_0D for two cases. The first case has equal width in frequency and direction, while the second case is for vanishing directional width. The last case has a vanishing nonlinear transfer for $k_0D = 1.363$ as is indicated by the cross in the figure. It is clear that directional effects in shallow water play an important role, resulting in modulational instability far beyond the one-dimensional threshold value of 1.363.

B.2 Derivation of narrow-band nonlinear transfer coefficient.

The starting point are the Davey-Stewartson equations. There are several forms of this equation available. Here, we take the set which is based on the action variable. Introduce the action variable $A(\mathbf{k})$ defined in such a way that the surface elevation η reads

$$\eta = \int \mathrm{d}\mathbf{k} \left(\frac{\omega}{2g}\right)^{1/2} (A(\mathbf{k}) + A^*(-\mathbf{k})) \exp(i\mathbf{k}\cdot\mathbf{x}). \tag{B3}$$

For a single mode $A(\mathbf{k}) = A\delta(\mathbf{k}-\mathbf{k}_0)$, assuming without loss of generality that the carrier wave with wavenumber k_0 propagates in the x-direction, one then finds

$$\eta = \left(\frac{\omega_0}{2g}\right)^{1/2} \left(Ae^{ik_0x} + A^*e^{-ik_0x}\right). \tag{B4}$$

The Davey-Stewartson equations for the envelope A and wave-induced current Q, caused by transverse modulations, then become

$$i\frac{\partial A}{\partial \tau}+\lambda\frac{\partial^2 A}{\partial x^2}+\mu\frac{\partial^2 A}{\partial y^2}=\nu|A|^2A+\nu_1 QA$$

$$\lambda_1\frac{\partial^2 Q}{\partial x^2}+\mu_1\frac{\partial^2 Q}{\partial y^2}=\kappa_1\frac{\partial^2|A|^2}{\partial y^2} \tag{B5}$$

where

$$\lambda=\frac{1}{2}\omega''(k_0),\ \mu=v_g/2k_0,\ c_0=\omega_0/k_0,$$

$$\nu=k_0^3\left[\frac{9T_0^4-10T_0^2+9}{8T_0^3}-\frac{1}{x}\left\{\frac{(2v_g-c_0/2)^2}{c_S^2-v_g^2}+1\right\}\right],$$

$$\nu_1=\frac{k_0^4}{2\omega_0 v_g}\left\{2c_0+v_g(1-T_0^2)\right\},\ \lambda_1=c_S^2-v_g^2,\ \mu_1=c_S^2,$$

$$\kappa_1=\frac{1}{2}\frac{c_0}{T_0}c_S^2 v_g\frac{2c_0+v_g(1-T_0^2)}{c_S^2-v_g^2}. \tag{B6}$$

The basic solution of Eq. (B5) is the uniform Stokes solution

$$A=a_0e^{[-i(\nu a_0^2+\nu_1 Q_0)\tau]},\ Q=Q_0=\text{constant}, \tag{B7}$$

and to test the stability of this traveling wave one introduces modulations according to

$$A=a_0\{1+\varepsilon a(\mathbf{x},t)\}e^{[-i(\nu a_0^2+\nu_1 Q_0)\tau]},$$

$$Q=Q_0(1+\varepsilon q(\mathbf{x},t)) \tag{B8}$$

with $a=a_+E+a_-E^{-1}$, $q=q_+E+q_-E^{-1}$, where $E=\exp i(lx+my-\Omega\tau)$.

Linearizing Eqns. (B5) while using the form (B8) the resulting dispersion relation for Ω as function of wavenumber l and m becomes

$$\Omega^2=\left(\lambda l^2+\mu m^2\right)\left[2\tilde{\nu}a_0^2+(\lambda l^2+\mu m^2)\right] \tag{B9}$$

where

$$\tilde{\nu}=\nu+\nu_1\kappa_1 m^2/(\lambda_1 l^2+\mu_1 m^2). \tag{B10}$$

From (B9) Hayes (1973) and Davey and Stewartson (1974) showed that the wave train is unstable if

$$\tilde{\nu}(\lambda l^2+\mu m^2)<0. \tag{B11}$$

Note that λ is always negative, μ is positive and ν changes from negative to positive as k_0D increase beyond $k_0D=1.363$ (Hasimoto and Ono, 1972).

Hayes (1973) has shown that it is always possible to choose l and m in such a way that the instability criterion (B11) is satisfied, but instability is practically non-existent for shallow water waves in the range $0\leq k_0D\leq 0.5$.

The severity of the sea state is usually estimated by means of the Benjamin-Feir Index which measures the balance of nonlinearity and dispersion. Here, the measure of importance of nonlinearity is given by the coefficient $\tilde{\nu}$ in the dispersion relation (B10), Note that in the one-dimensional shallow water case one finds $\tilde{\nu} = \nu$ as m then vanishes. This is the expression that has been used so far in the ECMWF freak wave warning system. The 2-D effect is entirely caused by the term $\nu_1 \kappa_1 m^2/(\lambda_1 l^2 + \mu_1 m^2)$. Clearly, the narrow-band limit of this term is not unique as it depends on the order in which the limits $l \to 0$ and $m \to 0$ are taken. Therefore, in this extra contribution spectral width needs to be taken into account explicitely.

In the next step one needs to estimate the size of the modulation wavenumbers l and m. This will be done in terms of the frequency and directional width σ_ω and σ_θ. Writing the original wave numbers k_x and k_y in terms of the polar coordinates k and θ the modulation wave numbers l and m become

$$l = k_x - k_0 = k\cos\theta - k_0,\ m = k_y = k\sin\theta,$$

and expanding around the values $k = k_0$ and $\theta = 0$ one finds in lowest order

$$\delta l = \delta k,\ \delta m = k_0 \delta\theta,$$

therefore

$$\langle \delta l^2 \rangle^{1/2} = \sigma_k,\ \langle \delta m^2 \rangle^{1/2} = k_0 \delta_\theta.$$

Finally, introducing the relative frequency width δ_ω using the relation $\delta_\omega = \sigma_\omega/\omega_0 = v_g/c_0 \times \sigma_k/k_0$ the extra term becomes

$$Extra = \frac{\kappa_1 \nu_1}{\mu_1} \frac{\delta_\theta^2}{\delta_\theta^2 + \alpha_\omega \delta_\omega^2}$$

with $\alpha_\omega = \lambda_1 c_0^2/\mu_1 v_g^2$. This parametrization of the effects of 2D modulations has the desirable property that for a narrow directional distribution with $\delta_\theta \to 0$ the extra contribution vanishes. The front factor $\kappa_1 \nu_1/\mu_1$ and the factor α_ω can be written more explicitely using the expressions for κ_1, ν_1, μ_1 and λ_1. Then α_ω becomes

$$\alpha_\omega = \frac{c_0^2}{v_g^2}\left(1 - \frac{v_g^2}{c_S^2}\right) \tag{B12}$$

and the extra term becomes explicitely

$$Extra = \frac{k_0^3}{4T_0} \frac{\left[2c_0 + v_g(1 - T_0^2)\right]^2}{c_S^2 - v_g^2} \frac{\delta_\theta^2}{\delta_\theta^2 + \alpha_\omega \delta_\omega^2} \tag{B13}$$

In principle, the extra term depends on the ratio $\delta_\theta/\delta_\omega$ which can be eliminated in favour of the ratio R defined in (A14).

Note that the additional contribution to the nonlinear transfer coefficient is associated with a finite value of the wave-induced current caused by the two-dimensional modulations, hence a finite value of Q. In terms of the amplitude a of the surface elevation (here action variable $A = (g/2\omega_0)^{1/2} a$) the above result implies that

$$Q = \frac{\kappa_1 g}{2\omega_0 \mu_1} \frac{\delta_\theta^2}{\delta_\theta^2 + \alpha_\omega \delta_\omega^2} |a|^2 \tag{B14}$$

and the additional contribution to the wave-induced current is given by k^2Q/v_g. This will give rise to an additional contribution to the mean sea level. Using Eq. (2.13c) of Davey and Stewartson (1974) one then finds for the mean sea level η_{20} the explicit expression

$$\eta_{20} = -\frac{k_0 c_S^2}{4(c_S^2 - v_g^2)} \left(\frac{2(1-T_0^2)}{T_0} + \frac{1}{x} \right) |a|^2 + \frac{k_0^2 \kappa_1}{2\omega_0 \mu_1} \frac{\delta_\theta^2}{\delta_\theta^2 + \alpha_\omega \delta_\omega^2} |a|^2 \tag{B15}$$

Clearly, also the narrow-band limit of the mean sea level is not unique and therefore one has to take explicit account of the spectral widths.

C The Draupner freak wave and the role of phase information.

C.1 Introduction.

In this Appendix I would like to present some of the details of my analysis of the Draupner Freak Wave. A method is proposed to obtain from the observed timeseries of the surface elevation η the wave envelope, the wave spectrum and statistical parameters such as the second, third and fourth moment in such a way that aliassing is avoided. In addition, also the phases of the wave components are obtained and the important role played by these phases is studied in the evolution of the wave spectrum and in the magnitude of the third and fourth moment. Of course, because of the assumption of homogeneity, the wave phase drops out of the second moment and therefore they play no role in the actual magnitude of the surface wave variance. It is also shown how by using theoretically derived properties of the bound-wave and free-wave kurtosis their respective contributions may be obtained.

C.2 Envelope method to analyse timeseries.

Let us now discuss the application of the envelope method to the analysis of time series, in particular the time series of the sea surface elevation at the time an extreme freak wave event occured. As far as I am aware, this method has been introduced for the first time by Gabor in 1946 in the context of a theory of communication and in the ocean wave context it has been applied by Longuet-Higgins (1983) and Shum and Melville (1984) in the context of the joint probability of wave periods and amplitudes, while Janssen (2014) used it in an attempt to better understand freak waves. A physical interpretation of the envelope of two-dimensional ocean waves as well as a method for wave group analysis was presented by Bitner-Gregerson and Gran (1983).

The envelope method is, in principle, very straightforward to implement. Given a discrete time series produced by sampling the sea surface elevation $\eta(t)$ every Δt seconds over a given period of T_L seconds, so the number of samples is $N = T_L/\Delta t$, one determines the Fourier series of η. Then, the orhogonal complement of η, denoted by ζ, equals to minus the Hilbert transform H of η, i.e. $\zeta = -H(\eta)$. If the Fourier expansion of η is known than it is straighforward to obtain its Hilbert transform because $H(e^{i\omega t}) = -i\mathrm{sgn}(\omega)e^{i\omega t}$. The definition of the envelope ρ is then given by

$$\rho = \sqrt{\eta^2 + \zeta^2} \tag{C1}$$

and the square of the envelope is basically the potential energy of a wave train. The envelope is therefore an attractive quantity to use in the study of freak wave events. In addition, for linear ocean waves the probability distribution function of the envelope ρ is the Rayleigh distribution.

Let us now become more specific. Introduce the complex function

$$Z = \sum_{n=1}^{\hat{N}} a_n e^{-i\omega_n t} \tag{C2}$$

with $\omega_n = 2\pi n/T_L$ while the value $\hat{N}$ will be discussed in a moment, it is found that

$$\eta = \frac{1}{2}(Z + Z^*). \tag{C3}$$

The amplitudes a_n are basically the projection of η on the basis functions, i.e.

$$a_n = \frac{2}{T_L}\int_0^{T_L} \mathrm{dt}\ \eta(t) e^{+i\omega_n t},\ n \le N \tag{C4}$$

The orthogonal complement then immediately follows from the relation $\zeta = -H(\eta)$ and the result is

$$\zeta = -\frac{i}{2}(Z - Z^*). \tag{C5}$$

Inspecting Eqns. (C3) and (C5) it is realized that the role of the pair (η, ζ) is very similar to the role of the canonical variables in the Hamiltonian of surface gravity waves, hence ζ is closely connected to the potential at the sea surface.

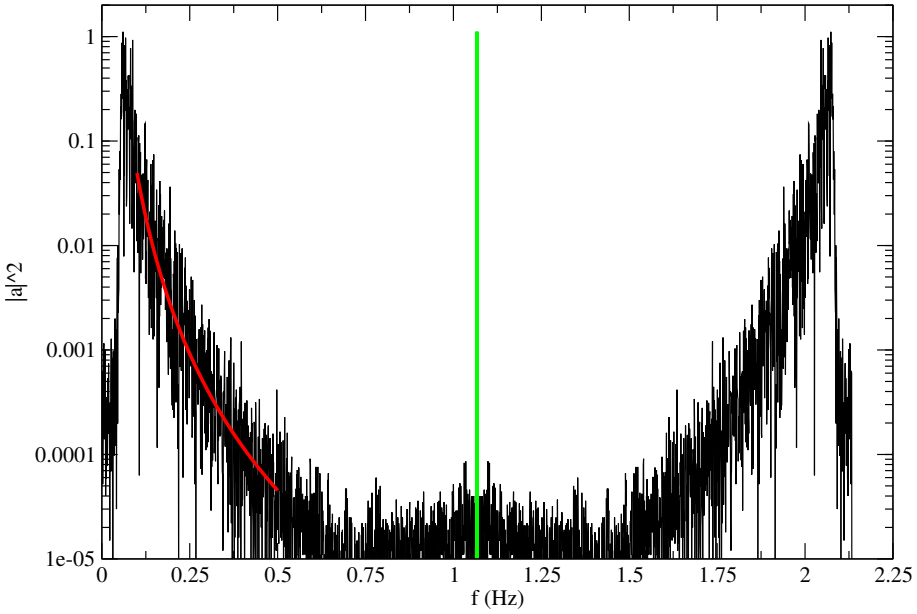

Figure C1: Amplitude spectrum from the Draupner timeseries. Aliassing is clearly evident as the spectrum is symmetric with respect to $f = 1.066$.

It is straightforward to apply this method to a time series. The timeseries from the Daupner freak wave event was obtained from Miguel Onorato. In this case the sampling frequency is 2.1333 Hz and the length of the timeseries is 1200 s so the number of samples $N = 2560$. However, there is, of course an aliassing problem. If one generates from the Fourier coefficients the timeseries of η and ζ using $\hat{N} = N$, then, initially to my great surprise, the so-generated function ζ vanishes! The reason is aliassing which is immediately evident if one plots the amplitude spectrum $|a_n|^2$ as function of frequency as shown in Fig. C1. The usual rule is that, at least for quadratic quantities, two samples per wave period are required

to resolve the aliassing. The two samples per wave period corresponds to the vertical green line in the Figure. Around the green line of symmetry, however, the spectrum is clearly polluted because it is anomously high. Nevertheless. now choosing $\hat{N} = N/2$ we got a realistic function ζ and a realistic looking envelope ρ, although it still seems a noisy. This is shown in the left panel of Fig. C2. In particular, the envelope at the time ≈ 264 of the freak wave event has two peaks which is a sign of

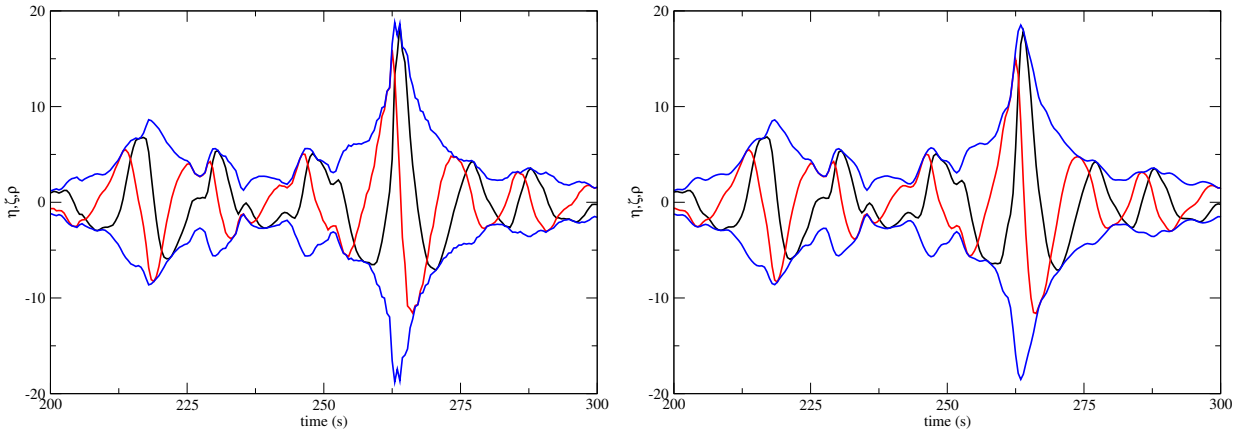

Figure C2: Left Panel: Timeseries of surface elevation η, orthogonal complement ζ and envelope ρ at the time ($t \approx 264$ s) of the Draupner freakwave event. Left panel $\hat{N} = N/2$, Right panel $\hat{N} = N/4$. Black line: η, Red line: ζ, Blue lines: $\pm\rho$.

noisy behaviour. In addition, it should be emphasized that there is not only interest in second moments such as the wave variance but also a reliable estimate of higher order moments such a skewness (a third moment) and kurtosis (a fourth moment) is required. In order to avoid aliasing effects in the fourth moment it is clear that there is a need to truncate even more severily, i.e. one should choose $\hat{N} = N/4$. The consequences of this choice are shown in the right panel of Fig. C2. It is concluded that the more severely truncated option is to be preferred because the envelope is smoother and does not have a double peak at the extreme event. Therefore, from now on $\hat{N} = N/4$ is chosen.

In closing this subsection we briefly discuss the relation between the time series of the surface elevation and the wave spectrum assuming stationary conditions. We start once more with the complex function given by Eq. (C2) and the surface elevation is given by Eq. (C3). The wave spectrum is basically the Fourier transform of the correlation function $R(\tau) = \langle \eta(t+\tau)\eta(t) \rangle$ which, on the scale of the period the wave field, is assumed to be stationary. In terms of the complex function Z the correlation function becomes

$$R(\tau) = \frac{1}{4}\langle (Z(t+\tau) + Z^*(t+\tau))(Z(t) + Z^*(t)) \rangle \tag{C6}$$

Now, the stationary assumption gives the following condition on the complex amplitudes amplitudes, namely

$$\langle a_i a_j \rangle = 0, \langle a_i a_j^* \rangle = |a_i|^2 \delta_{i-j} \tag{C7}$$

As a consequence one finds that $\langle Z(t+\tau)Z(t) \rangle$ and $\langle Z^*(t+\tau)Z^*(t) \rangle$ vanish while

$$\langle Z(t+\tau)Z^*(t) \rangle = \sum_{i=1}^{\hat{N}} |a_i|^2 e^{i\omega_i \tau}, \text{ and } \langle Z^*(t+\tau)Z(t) \rangle = \sum_{i=1}^{\hat{N}} |a_i|^2 e^{-i\omega_i \tau}. \tag{C8}$$

As a result, the correlation function becomes

$$R(\tau) = \frac{1}{4}\langle \sum_{i=1}^{\hat{N}} |a_i|^2 \left(e^{i\omega_i \tau} + e^{-i\omega_1 \tau}\right) \tag{C9}$$

Now, the wave spectrum $E(\omega)$ is defined as the Fourier transform of the auto correlation function, i.e.

$$E(\omega) = \frac{1}{2\pi}\int_0^{T_L} \mathrm{d}\tau e^{i\omega\tau} R(\tau) \tag{C10}$$

Inserting (C9) into (C10 gives for large times T_L the discrete spectrum

$$E(\omega_i) = \frac{1}{2}\frac{|a_i|^2}{\Delta\omega} \tag{C11}$$

with $\Delta\omega = 2\pi/T_L$. Clearly, the wave spectrum $E(\omega)$ is independent of the phase of the individual waves, and conversely, Eq. (C11) determines for given wave spectrum the amplitude of the waves, but, of course, not the phases.

C.3 Method to determine dynamical Kurtosis from timeseries.

We now discuss how one can utilize the knowledge on skewness and kurtosis caused by the bound waves to determine whether free-wave dynamics plays a role in the formation of freak wave events in the field. This idea was suggested and applied to a freak wave event that occurred in the Southern Ocean as reported by Toffoli *et al.* (2024). Let us assume that only bound waves give rise to kurtosis of the sea surface. In those circumstances there is a direct connection between kurtosis and skewness of the sea surface in particular for deep-water conditions. Let us introduce the surface elevation ratio $R_\eta = \kappa_{40}/\kappa_{30}^2$, which is the ratio of surface elevation kurtosis over the square of the skewness. From the narrow-band expressions given in Appendix A one may infer immediately that these ratios are independent of the signficant steepness ε. In fact, in deep water this ratio becomes a constant, i.e. $R_\eta = 2$. Therefore, if an analysis of an observed timeseries showing a freak wave would reveal that R_η is significantly larger than 2, then this is an indication that free-wave dynamics/ four-wave interactions play an important role in the formation of the relevant freak wave event.

Unfortunately, in shallow water things are more complicated as the ratio R_η also depends on dimensionless depth as is illustrated in Fig. C3. The dimensionless depth for the Draupner case is about 1.2 and is close to the maximum of R_η. Now observed $R_\eta \simeq 6.9$ is far above the maximum as found from the narrow-band approximation of the two statistics, which supports the conclusion that in the Draupner freak wave case four-wave interactions are expected to play a significant role.

The objection could be made that this approach is only valid for narrow-band spectra. However, a similar conclusion follows from the exact computations of Gramstad and Lian (2024) using Jonswap spectra with a wide range of values for the overshoot parameter γ and the Phillips parameter α_p. Computations of R_η for the bound-waves gave on average $R_\eta = 2.2$, which for dimensionless depths $k_0 D > 4$ is already close to the values for R_η shown in Fig. C3. Likewise, a similar conclusion follows from the computations presented here using WAM model spectra.

Let us now give a more detailed discussion. In particular, we would like to obtain some general properties of the skewness and kurtosis parameters that allows us to obtain from observed time series information on whether quasi-resonant interactions play a role in freak wave formation. Once more it is noted that skewness is only determined by the bound waves, while the kurtosis κ_4 depends on both quasi-resonant

four-wave interactions and effects by the bound waves. Therefore, we introduce superscripts d and b to denote the dynamical part of the kurtosis and the bound-wave part respectively, hence,

$$\kappa_4 = \kappa_4^d + \kappa_4^b, \tag{C12}$$

where

$$\kappa_4^d = \kappa_{40}^d + 2\kappa_{22}^d + \kappa_{04}^d, \text{ and } \kappa_4^b = \kappa_{40}^b + 2\kappa_{22}^b + \kappa_{04}^b. \tag{C13}$$

Let us first concentrate on the properties of the dynamical parts of the kurtosis. Mori and Janssen (2006) have shown that for general spectra the following relations hold

$$\kappa_{40}^d = \kappa_{04}^d, \text{ while } \kappa_{22}^d = \frac{1}{3}\kappa_{40}^d \tag{C14}$$

hence κ_4^d becomes

$$\kappa_4^d = \frac{8}{3}\kappa_{40}^d. \tag{C15}$$

The proof of this is fairly straightforward, when using the expressions for η and ζ in terms of the complex function Z as given at the beginning of Appendix A, i.e. $\eta = (Z+Z^*)/2$ and $\zeta = (Z-Z^*)/2i$. As a consequence, introducing the complex quantities

$$A = \langle |Z|^4 \rangle, \text{ and } B = \langle Z^3 Z^* + Z^{*3} Z \rangle \tag{C16}$$

while realizing that for a homogeneous system the moment $\langle Z^4 \rangle$ vanishes one finds (see also Janssen (2015b))

$$\kappa_{40} = \frac{3}{8}\left(A + \frac{2}{3}B\right) - 3,\ \kappa_{22} = \frac{1}{8}A - 1,\ \kappa_{04} = \frac{3}{8}\left(A - \frac{2}{3}B\right) - 3. \tag{C17}$$

We see from (C17) that, as expected, the kurtosis elements are depending on only two independent parameters A and B so κ_{40}, κ_{22} and κ_{04} are not independent. In fact, one can show by eliminating A and B from (C17) that the following relation between the fourth cumulants holds,

$$\kappa_{40} + \kappa_{04} = 6\kappa_{22}. \tag{C18}$$

This relation is indeed satisfied by the observed values of kurtosis listed in Table 1 of the main text: with $\kappa_{40} = 1.038$ and $\kappa_{04} = 0.428$ one indeed finds that $(\kappa_{40} + \kappa_{04})/6 = 0.244$ which is the observed value of κ_{22}.

Furthermore, the dynamical part of the kurtosis is determined by near-resonant four-wave interactions. For gravity waves there is only one type of resonant interaction present, namely the one obeying the resonance condition $\omega_1 + \omega_2 = \omega_3 + \omega_4$. This process can only be represented by the parameter A and not by the parameter B. Concentrating therefore on the properties of the quasi-resonant interactions it is immediately clear that for these processes the rules given in Eqn. (C14) hold, which confirms the results obtained by Mori and Janssen (2006). It is emphasized that these relations hold for arbitrary surface elevation and arbitrary envelope probability distributions. For the bound-wave contributions to the kurtosis parameters I have so far not been able to derive general rules for arbitrary spectra and arbitrary p.d.f.'s. As it turns out I need one additional piece of information in order to obtain definite results on the contribution of bound waves and free waves to the fourth order statistics. Several attempts were made but the most straightforward was the approach suggested by Toffoli *et al.* (2024). This

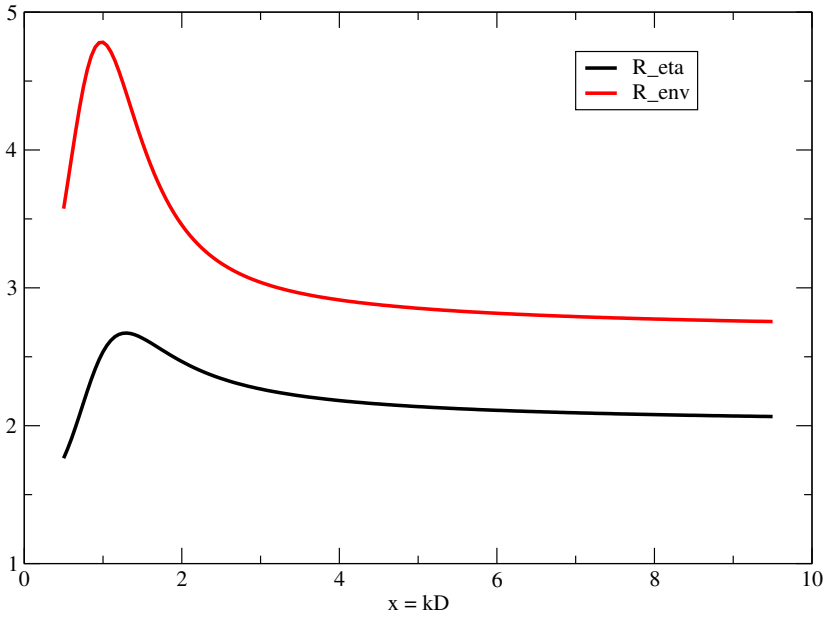

Figure C3: Surface elevation ratio $R_\eta = \kappa_{40}/\kappa_{30}^2$ and Envelope ratio $R_{env} = \kappa_4/\kappa_{30}^2$ as function of dimensionless depth

method utilizes an approximate relation between the skewness κ_{30} and the kurtosis κ_{40} of the bound-waves. Consider as a guideline the case of narrow-band, deep-water waves. From Appendix A one finds, with $\varepsilon = k_0\langle\eta^2\rangle^{1/2}$ the so-called significant slope parameter, that

$$\kappa_{30}^b = 3\varepsilon, \kappa_{40}^b = 18\varepsilon^2 \tag{C19}$$

for the skewness, and kurtosis of the surface elevation, respectively. Therefore, one finds for deep-water waves the familiar relation

$$R_\eta = \frac{\kappa_{40}^b}{(\kappa_{30}^b)^2} = 2, \tag{C20}$$

which is independent of the significant slope. Including finite depth effects it is found that the parameter R_η will depend on the dimensionless depth $x = k_0 D$ with D being the depth. Using the expressions in Appendix A for the narrow-band version of skewness κ_{30} and surface elevation kurtosis κ_{40} one finds for the ratio R_η

$$R_\eta = \frac{2}{3}\frac{\gamma+\beta+2(\alpha+\Delta)^2}{(\alpha+\Delta)^2} \tag{C21}$$

and in Fig. C3 a plot of the proportionality coefficient R_η as a function of $x = k_0 D$ is shown. It is remarked that the parameter Δ which is a measure of the change in mean surface elevation depends explicitely on the spectral parameters δ_θ and δ_ω, as discussed in Appendix B. In Fig. C3 we have therefore chosen for the spectral parameters the values $\delta_\theta = 0.36$, $\delta_\omega = 0.45$ as found for the Draupner event. Clearly, the typical value of R_η is of the order of 2 but in shallow water there is a slight enhancement. For the Draupner case with $x = 1.21$ one finds $R_\eta = 2.66$.

Remark:

It is noted that there is, however, a sensitive dependence on the results for R_η (and R_{env}) on directionality measured by the parameter δ_θ. For the pure one-dimensional case R_η at $x = 1.21$ would increase considerably from 2.66 to 3.78 so the two-dimensional effect discussed in Appendix B plays an important role in shallow waters.

In order to establish now the relative contributions of the quasi-resonant interactions and the bound-waves from observed time series we execute the following steps based on the use of Eq. (C14) for the free wave contribution and relation (C21) for the bound-waves, i.e. $\kappa^b_{40} = R_\eta \kappa^2_{30}$.

$$\begin{aligned} \kappa^d_{40} + \kappa^b_{40} = \kappa_{40} \rightarrow \kappa^d_{40} &= \kappa_{40} - \mathbf{R}_\eta \boldsymbol{\kappa^2_{30}} \\ \kappa^d_{22} + \kappa^b_{22} = \kappa_{22} \rightarrow \kappa^b_{22} &= \kappa_{22} - \boldsymbol{\kappa^d_{40}/3} \\ \kappa^d_{04} + \kappa^b_{04} = \kappa_{04} \rightarrow \kappa^b_{04} &= \kappa_{04} - \boldsymbol{\kappa^d_{40}} \end{aligned} \tag{C22}$$

where in bold we indicate quantities that are either obtained from the observed time series or through the previous operations. The other two kurtosis elements follow from $\kappa^d_{04} = \boldsymbol{\kappa^d_{40}}$ and $\kappa^d_{22} = \boldsymbol{\kappa^d_{40}/3}$.

```
Table 1 Free wave and bound-wave contributions to Kurtosis of Draupner
                Case.
  START PROCESSING
  XN  =    2560.        HS  =    11.919
  XK0 =    0.0175       XK0*D =  1.2075     EPS =XK0*SQRT(XM0) =    0.0521

       OBSERVED:
  XM0_OBS    =   8.879
  C3OBS      =   0.136
  C4_OBS_ENV =   0.244

  INPUT:

  KAPPA30 =    0.408     KAPPA4  =    1.954
  KAPPA21 =   -0.011     KAPPA40 =    1.038
  KAPPA12 =    0.136     KAPPA22 =    0.244
  KAPPA03 =   -0.031     KAPPA04 =    0.428

  RATIO_ETA =    2.663
  KAPPA40/KAPPA30**2 = 6.236

  RESULTS:


  KAPPA40_D =   0.594
  KAPPA40_B =   0.443
  KAPPA22_D =   0.198
  KAPPA22_B =   0.046
  KAPPA04_D =   0.594
  KAPPA04_B =  -0.167

  KAPPA4_D  =   1.586
  KAPPA4_B  =   0.368
```

For given observed $\kappa_{4,0}, \kappa_{2,2}$ and $\kappa_{0,4}$ I have obtained the free wave and the bound-wave contributions to the kurtosis stats for the Draupner time series with a length of 1200 s. The details are presented in Table I. The main result is that the total envelope kurtosis for the Draupner case equals about 1.95 (which is quite a high value) and the dynamic contribution dominates because $\kappa^d_4 = 1.59$ while the bound-wave

contribution is $\kappa_4^b = 0.37$. It is concluded that according to this method it seems that during the Draupner event quasi-resonant four-wave interactions have played a prominent role.

As an alternative method, which emphasizes the envelope wave height as a measure for extreme events, one could also consider for the bound waves a relation between envelope kurtosis κ_4 and the skewness κ_{30}. For narrow-band waves in deep water one finds according to Appendix A that $\kappa_4 = 24\varepsilon^2$ so that the envelope ratio R_{env} becomes

$$R_{env} = \frac{\kappa_4}{\kappa_{30}^2} = 8/3 \tag{C23}$$

which is once more independent of the significant steepness. Including finite depth effects one now finds, using Eqns. (A27) and (A34)

$$R_{env} = \frac{16}{9}\frac{\gamma + \alpha^2 + (\alpha+\Delta)^2}{(\alpha+\Delta)^2} \tag{C24}$$

and this curve is shown in Fig. C3 as well. We will not report the details of this approach, however, because similar results are obtained although the effect of the dynamics is compared to the bound waves somewhat smaller.

Based on the statistical parameters of the envelope wave train we have tried two methods that are able to separate bound-wave and free-wave contributions to the envelope kurtosis. Applied to the case of the Draupner freak event it is seen that the free-wave contribution to the statistics is fairly important, suggesting that for this freak wave event quasi-resonant energy transfer within the wave spectrum plays a role. Therefore, in the next Appendix some simulation results with the Zakharov equation and the role of the phases of the waves will be discussed.

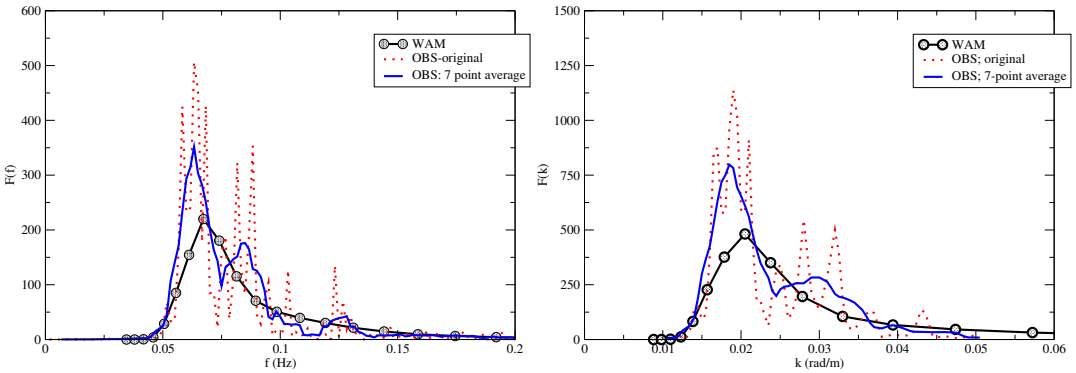

Figure C4: The left Panel shows a comparison of WAM model frequency spectrum at Draupner site with the original, observed spectrum and a 7 point average. The right panel shows the comparison in wave number space interpolated to a linear wave number grid.

C.4 1D simulations for the Draupner event.

During the discussion of the Draupner freak wave event in §3.2 it was pointed out that the observation of such extreme event cannot be represented by a forecasting system that is based on the random phase approximation. Typically, wave forecasting systems provide averaged information related to a domain of, say, 14×14 km over a timestep of the order of 10 minutes. As a consequence the information is an average over more than 100,000 waves, while a freak wave event is obtained from a time series of some 100 waves. In particular, it is expected that higher order moments such as skewness and kurtosis will be

affected by such a large disrepancy in the number of waves and, by means of numerical experimentation, it will be explicitely shown that this is indeed the case.

Here, I will present results of numerical simulations of the 1D Zakharov equation, the numerical details of which have been presented by Janssen (2003). Initial conditions are obtained from the Fourier transformation of the Draupner timeseries with length of 600 s. As shown in Fig. C4 such a spectrum is very spiky and therefore we have done attempts to smooth the Draupner spectrum by averaging 3 and 7 spectral values.

As a first example a comparison is given of the observed frequency spectrum with the modelled WAM spectrum in Fig. C4. Although there is perhaps a reasonable agreement between observed (11.91 m) and modelled (11.23 m) significant wave height, the discrepancy between observed and modelled wave spectrum is considerably larger. This is in particular the case for the unsmoothed original data, but there are also substantial differences between the WAM spectrum and the smoothed observed spectrum using a 7 point average.

Inspecting the unsmoothed data more closely it is immediately evident for someone who is familiar with the Benjamin-Feir instability (also called sometimes the side-band instability) that the observed Draupner spectrum, which shows a main peak and two sidebands, might be a clear illustration of the Benjamin-Feir instability in action. To what extent these sidebands are still changing in time will be studied when the simulation results are discussed. Of course, inspecting the smoothed data shows that the sidebands disappear but there are still clear signs of a second wave system between 0.07 Hz and 0.09 Hz.

It has been suggested (see e.g. Adcock *et al.* 2011) that this second system with peak around 0.08 Hz is really separate from the main system with peak frequency of about 0.06 Hz. Based on a positive anomaly in the mean surface elevation it has been argued that the second system propagates under an angle of 120 $^\circ$ from the mean direction of the main system. Such a double-peaked spectrum is known to be subject to instability even for broad peaks, but the peak frequency of the two peaks should be similar while the angle between the two systems should be large enough, in the order of 90 $^\circ$ (see Onorato *et al.*, 2006). However, according to the timeseries for the Draupner event the peaks have different peak frequencies, while from the WAM simulation there is no evidence of a second system propagating in a different direction from the main wave system as is seen from Fig. 5. Therefore, it is unlikely that the Draupner event is subject to the instability of a double-peaked spectrum. I therefore thought it was sufficient to concentrate on a one-dimensional simulation of the Draupner event following the approach in Janssen (2003).

This approach is based on the Zakharov equation which solves the reduced Hamilton equations for the amplitude of the free waves. The amplitudes are a function of wavenumber and time, and in the software wavenumber space is discretized in a regular linear manner. The observed Draupner spectrum is based on a regular frequency space. Ignoring the effect of the bound waves, which some support by the findings in the previous §C.3, one may obtain a wavenumber spectrum $F(k)$ from the frequency spectrum $F(f)$ using the usual relation

$$F(k)\mathrm{d}k = F(f)\mathrm{d}f \rightarrow F(k) = \frac{v_g}{2\pi}F(f),$$

which gives a wavenumber spectrum on an irregular wavenumber grid since the dispersion relation between (angular) frequency and wavenumber is nonlinear.

As a final step, a wavenumber spectrum on a regular grid, i.e.

$$k_n = 0.0005 \times n, \text{ for } n = 1, \hat{N},$$

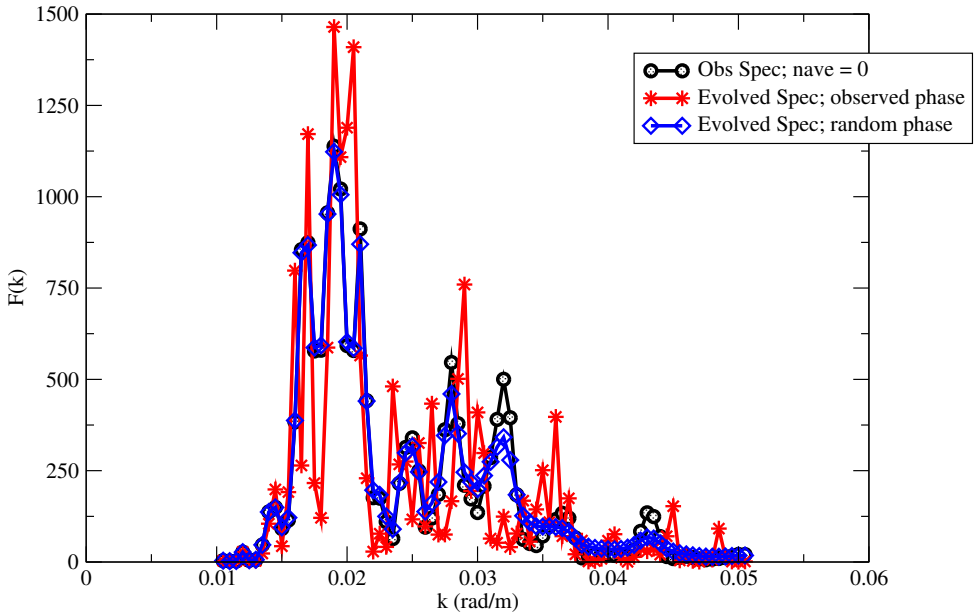

Figure C5: Spectral change of Draupner event due to nonlinear interactions of the original spectra using the observed initial phases and using random phase.

is produced using linear interpolation, and the resulting spectrum is almost identical to the spectrum on the irregular wavenumber grid. The resulting observed and WAM wavenumber spectra are shown in the right Panel of Fig. C4. The differences between observed and modelled wavenumber spectrum are similar to the frequency spectra. An identical set of procedures was performed for the phases of the waves so that wave phase on the above regular wavenumber grid could be provided as an initial condition.

Now, results of numerical simulations with the one-dimensional Zakharov equation are presented using a number of versions of the initial conditions. In agreement with §B.1 the shallow water transverse current effects are mimicked by enhancing the nonlinear transfer coefficient $T_{1,2,3,4}$ in the manner shown in Fig. B1. This assures that there is even Benjamin-Feir instability for shallow water waves.

The first set of experiments is performed with the original observed wavenumber spectra and phases as initial condition. The duration of these runs was chosen to be fairly short, 20 min. i.e. about 75 wave peak periods, because for longer duration other physical effects, such as wave breaking, wind input and bottom dissipation become important as well. The resulting spectrum which is the initial spectrum modified by nonlinear interactions only is denoted by a star in Fig. C5.

In order to show the importance of the initial phase the same experiments were also performed with random phase. The results of these experiments were obtained using an ensemble forecasting system with 500 ensemble members where for each member the spectrum was the same, while the phases of the individual components were obtained from a random draw between 0 and 1 multiplied by the factor 2π. The results for random phase are also shown in Fig. C5 and are denoted by a diamond symbol.

By simply checking out in Fig. C5 the differences between circles and stars for observed phase and comparing these with the differences between circles and diamonds for the random phase case it is clear that phase plays an important role in the changes of the spectrum. While for random phase the low-wavenumber peaks hardly change, in sharp contrast to this, for observed phase there are considerable changes. Above $k \simeq 0.022$ both random phase and observed phase shows differences in the spectrum due to nonlinear interactions, but the observed phase case tends to enhance the secondary peak while the

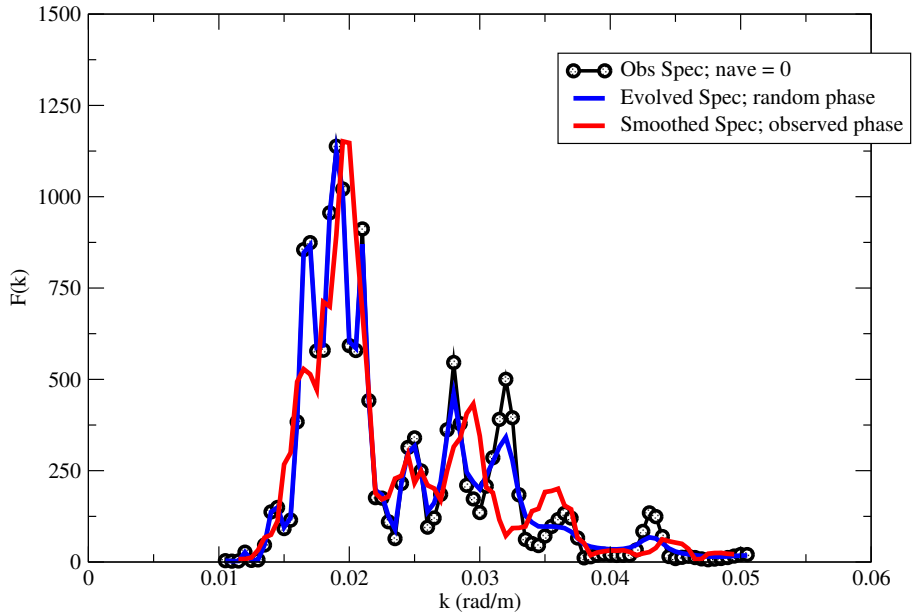

Figure C6: Shown is the 5-point smoothed spectrum in comparison to the initial condition with observed phase.

random phase case tends to smooth the peaks.

It is clear from Fig. C5 that the evolved spectrum using the observed phase becomes fairly erratic, therefore it might be difficult to judge how the nonlinear transfer affects the wave spectrum. For this reason we show in Fig. C6 the 5-point smoothed version of the evolved spectrum from Fig. C5. Clearly, the nonlinear transfer gives rise to an upshift of the main peak, which is quite extra-ordinary since, at least in the random phase case, it is usually expected that the nonlinear transfer gives rise to a downshift of the peak. It is also clear that even the smoothed version of the evolved spectrum starting from observed phases differs considerably from the spectrum obtained from the case of random phase (Janssen, 2003).

The main reason to do these experiments is, of course, to study the dependence of statistics of the evolving nonlinear wave train on the choice of the phases. It turns out that these differences are substantial. Indeed, inspecting Fig. C7 which shows the evolution in time of the spatial average of dynamic kurtosis κ_{40} over a time period of 2400 seconds, it is clear that for random initial phase dynamic kurtosis is small, only of the order of 0.10, in close agreement with the findings obtained from the exact computations for the maximum in kurtosis factor $C_4 = \kappa_{40}/3$ depicted in Fig. 9. However, these results are an order of magnitude smaller than the maximum value of kurtosis obtained from the simulation when observed phase is used as initial condition. The maximum value of dynamic kurtosis κ_{40} is of the $\mathcal{O}(1)$ which is in fair agreement with the observed value given in Table 1. Note, that, as expected, results from the Kinetic equation (Janssen, 2003) are in fair agreement with the random phase simulations.

Finally, it should be noted that Fig. C7 shows some intriguing aspects of the time evolution of κ_{40}. While at initial time the spatial average of dynamic kurtosis is large, owing to dispersion κ_{40} reduces quickly to zero due to phase mixing which is then followed by a number of 'random' solitary peaks. The formation of these solitary peaks is most likely a nonlinear effect as by switching off nonlinearity kurtosis always remains small and frequently less than zero. However, these are the results of a one-dimensional simulation, and two dimensional effects may have to be taken into account as well and may counteract the formation of these solitay peaks. This therefore requires further investigations.

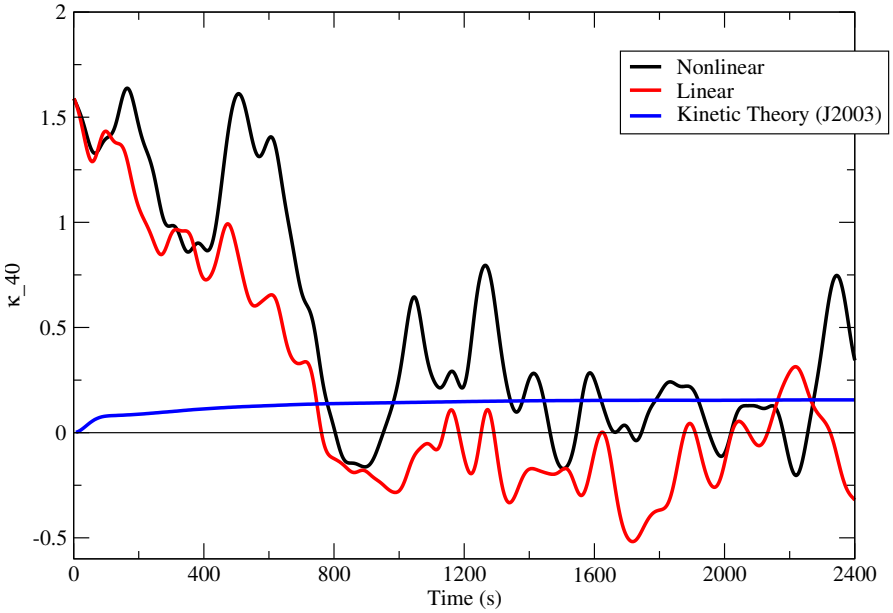

Figure C7: Spatial average of kurtosis κ_{40} as function of time for observed phase, random phase and kinetic theory. The initial condition is the original unsmoothed spectrum with observed phases Switching off nonlinear effects gives a major difference in the evolution of kurtosis κ_{40} .

C.5 Alternative method to detect extreme wave events.

So far we have applied results from the statistics of surface gravity waves to characterize extreme wave events giving statements on e.g. the probability of the exceedance of maximum wave height in a domain of, say, 10 by 10 km during a time span of 20 minutes. It was shown that nonlinear effects related to the bound waves and free waves play an important role in the probability of extreme events and that linear effects only are not enough to explain these extremes.

However, one may wonder whether statistics over a fairly large domain are of any use because the domain is quite large and the user is, of course, interested in whether an extreme wave might occur at the location of an oil rig or a ship. Are there perhaps alternative measures to indicate the severity of the extreme sea state. Still based on a statistical approach, one might contemplate the use and value of the expectation value of maximum wave height which according to the ECMWF freak wave warning system is given by Eq. 40. The advantage of this expression is that it is related to the maximum wave height of a time series of 20 minutes, but the statistics still relates to the domain of the spatial/temporary resolution of the wave prediction system.

Here, it is discussed whether there are alternative methods to characterize extreme surface elevations given only the wave spectrum. The wave spectrum is a very useful quantity because unlike higher-order moments such as skewness and kurtosis, the wave energy and it associated spectrum are under the assumption of stationary statistics independent of the phase distribution. This is illustrated in Fig. C7 where for given wave spectrum one may obtain completely different value of kurtosis for the observed phases of the waves versus random phases.

Let us therefore study once more the single event of the Draupner wave, in particular quantities such as the surface elevation and the envelope of the wave train. From the inverse of Eqn. (C11) we obtain the amplitude as function of angular frequency,

$$|a_n| = \sqrt{2E(\omega_n)\Delta\omega}, \tag{C25}$$

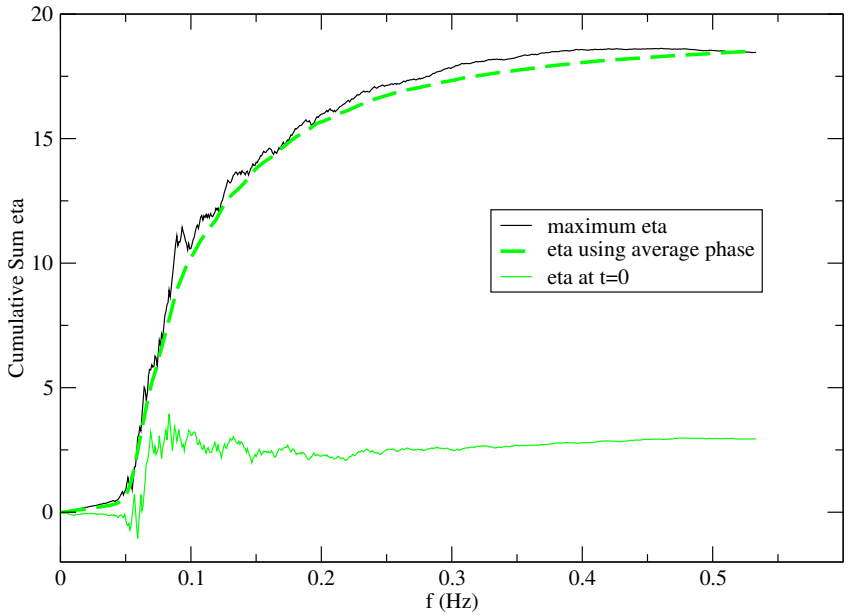

Figure C8: The cumulative sum $\sum_{n=1}^{N}\cos\theta_n$ as function of frequency $f_N = N\Delta\omega/2\pi$ for initial time and for the time of maximum elevation. For the case of maximum elevation also the case of constant phase is shown.The duration of the timeseries T_L is 1200 s.

where $\omega_n = n\Delta\omega$ with $\Delta\omega = 2\pi/T_L$ and T_L the length of the time series. Denoting the phase of the individual waves by ϕ_n one obtains from (C2) and the expressions (C3) and (C5) for surface elevation η and its orthogonal complement ζ the relations

$$\eta = \sum_{n=1}^{\hat{N}} |a_n|\cos(\phi_n - \omega_n t), \quad \zeta = \sum_{n=1}^{\hat{N}} |a_n|\sin(\phi_n - \omega_n t) \tag{C26}$$

It should be clear from the above expression that the observed initial phases ϕ_n and the 'total' phase $\theta_n = \phi_n - \omega_n t$ play a key role in the value of the surface elevation and in the creation of an extreme event.

Let us now concentrate on the role of the phases on the result for surface elevation. Using the observed phases ϕ_n the partial sum $\sum_{n=1}^{N}\cos\theta_n$, with $N < \hat{N}$ as function of N has been plotted in Fig. C8, for initial time and for the time that the envelope reaches a maximum. Inspecting the expression for the surface elevation in Eq. C26 it is clear that for the maximum elevation case the increments shown in Fig. C8 are mainly positive over the whole frequency range which means that the phases θ_n are such that $\cos\theta_n$ is most of the time positive. In order to convince ourselves that this is indeed the case we plot in Fig.C9 the distribution function $p[\cos\theta]$ as function of its argument for the extreme event. It is evidently clear that the distribution is asymetrical with respect to the origin, implying that the average value $\langle\cos(\theta)\rangle$ is positive, its value being about 0.235. In the same Figure we also plot the distribution function $p[\cos\theta]$ for the whole 1200 s time series, which, as expected, is symmetrical with respect to the origin because the average surface elevation vanishes. The extreme event has a distribution of phases that differs to a large extent from the average phase distribution

In passing it is remarked that the phase distribution of the whole time series is found to be uniform. This is shown as follows using the rule $p[\cos\theta]\mathrm{d}\cos\theta = p(\theta)\mathrm{d}\theta$. Assuming that θ is uniformly distributed one finds for the $\cos\theta$-distribution

$$p[\cos\theta] = \frac{1}{\pi\sqrt{1-\cos^2\theta}}. \tag{C27}$$

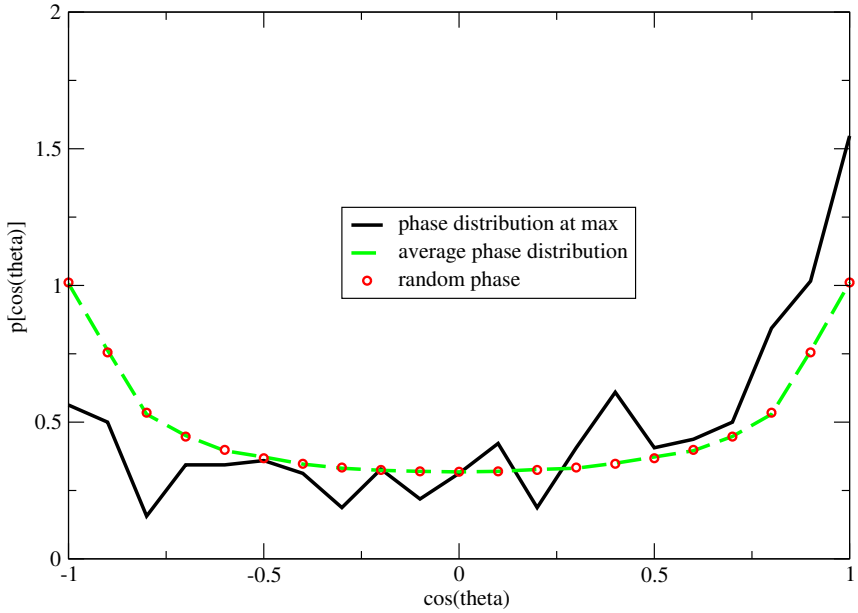

Figure C9: Probability distribution of $\cos\theta$ *at the extreme event which is compared with the average distribution over the whole time series. The latter is in close agreement with the pdf based on the random phase distribution.*

Note that this distribution function has an integrable singularity at the end-points $|\cos\theta| = 1$, therefore only the integral of $p[\cos\theta]$ over a width, in this case, of 0.1 can be directly compared with the numerical results displayed in Fig. C9. There is perfect agreement between results from the time series and Eq. C27 suggesting that the phase distribution over the whole period of 1200 s is indeed uniform.

Of course, the question now is what can one do with this knowledge of the phase distribution in the context of operational wave forecasting where one has no knowledge of the phases of the individual wave components. Just out of curiosity I tried the case that at the extreme event the phases of all the wave components are the same and equal to zero. This leads to a maximum surface elevation of 50 m which is quite extreme since observed maximum elevation is 17.85 m. Hence, I made the choice that each wave component has the same phase with value $\cos\theta = 1/3$. This leads to the result in Fig. C9 with the dashed line. Remarkably, over the whole frequency range, the increment plot with constant phase is in good agreement with the result from the observations.

Why there is such a good agreement is a mystery. Christou and Ewans (2014) would probably argue that this is a case of dispersive focussing. However, the waves in the Draupner event are fairly nonlinear so four-wave interactions are expected to play an important role in the formation of this extreme event. Nonlinearity could lead to phase locking resulting in an almost constant phase on average. More work in understanding why this simple suggestion works is clearly needed.

Finally, let us concentrate on the low-frequency part of the plot in Fig. C9. At the time of the freak wave event even the low-frequencies were phase locked. However, this only happened during a period of 20 seconds around the extreme event. Most of the time, for example at initial time, the low frequencies were not correlated giving a flat increment plot in that range.

References

Adcock, T.A.A., P.H. Taylor, S. Yan, Q.W. Ma and P.A.E.M. Janssen, 2011. Did the Draupner wave occur in a crossing sea? *Proc. Roy. Soc. A:* Mathematical, Physical and Engineering sciences, **467**(2134), 3004-3021.

Adler, R.J., 1981. *The Geometry of Random Fields*, John Wiley, 275 pp.

Annenkov, S.Yu. and V.I. Shrira, 2001. On the predictability of evolution of surface gravity and gravity-capillary waves. *Physica* D, **152-153**, 665-675.

Barbariol, F., J.-R. Bidlot, L. Cavaleri, M. Sclavo, L. Thomson, and A. Benetazzo, 2019. Maximum wave heights from global reanalysis. *Progress in Oceanography* **175**, 139-160.

Baxevani, A. and I. Rychlik, 2006. Maxima for Gaussian Seas. *Ocean Eng.* **33**, 895-911

Benetazzo, A., F. Barbariol, F. Bergamasco, A. Torsello, S. Carniel, and M. Sclavo, 2015. Observation of Extreme Sea Waves in a Space-Time Ensemble. *J. Phys. Oceanogr.* **45**, 2261-2275.

Benjamin, T.B., 1967. Instability of periodic wavetrains in nonlinear dispersive systems. *Proc. Roy. Soc. A* **299**, 59-75.

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

Bitner-Gregerson, E.M. and S. Gran, 1983. Local properties of sea waves derived from a wave record. *Applied Ocean Research* **5**,210-214.

Burgers, G., F. Koek, H. de Vries and M. Stam, 2008. Searching for factors that limit observed extreme maximum wave height distributions in the North Sea. *Extreme Ocean Waves*, E. Pelinovsky and C. Kharif (eds), Springer Science+Business Media B.V., pp. 127-138.

Cavaleri, L., F. Barbariol, A. Benetazzo, L. Bertotti, J.-R. Bidlot, Peter A.E.M. Janssen and Nils Wedi, 2016. The Draupner wave: a fresh look and the emerging view.

Christou, M. & K. Ewans, 2014. Field Measurements of Rogue Water Waves. *J. Phys. Oceanogr.* **44**, 2317-2335.

Davey, A. & K. Stewartson, 1974. On three-dimensional packets of surface waves. *Proc. R. Soc. Lond.***A 338**, 101-110.

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.

Fedele F., 2012. Space-Time Extremes in Short-Crested Storm Seas. *J. Phys. Oceanogr.* **42**, 1601-1615.

Fedele F., 2015. On the kurtosis of deep-water gravity waves. *J. Fluid Mech.* **782**, 25-36.

Gramstad O. and G. Lian, 2024. 2015. Parametrization of sea surface skewness and kurtosis with application to crest distributions. *J. Fluid Mech.* **979**, A4.

Guidhir, M.N., D. Kennedy, A. Berry, B. Christy, C. Clancy, C. Creamer, G. Westbrook and S. Gallagher,

2022. Irish Wave Data- Rogues, Analysis and Continuity, *J. Mar. Sci. Eng.* **10**, 1073.

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

Häfner D., Gemmrich J. and M. Jochum, 2021. Real-world rogue wave probabilities. Sci. Rep., **11**, 10084.

Hasimoto, H, and H. Ono, 1972. Nonlinear modulation of gravity waves. *J. Phys. Soc. Jpn.* **33**, 805-811.

Hasselmann, K., 1962. On the non-linear energy transfer in a gravity-wave spectrum, part 1: general theory. *J. Fluid Mech.* **12**, 481.

Hayes, W.D., 1973. Group velocity and nonlinear dispersive wave propagation. *Proc. R. Soc. Lond.* **A 332**, 199-221.

Haver, S. 2004. A possible Freak Wave Event Measured at the Draupner Jacket January 1 1995, private communication.

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

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

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

Janssen, Peter A.E.M., 2015a. Notes on the maximum wave height distribution. ECMWF Technical Memorandum 755.

Janssen, Peter A.E.M., 2015b. How rare is the Draupner wave event? ECMWF Technical Memorandum 775.

Janssen, Peter A.E.M., 2017, Random note on statistics of a nonlinear system, ECMWF.

Janssen, Peter A.E.M. and Gerbrand J. Komen, 1982. Modification of the Surface Elevation Probability Distribution in Ocean Swell by Nonlinear Spectral Broadening. *J. Geophys. Res.* 87, 4155-4162.

Janssen Peter A.E.M and Miguel Onorato, 2007. The intermediate water depth limit of the Zakharov Equation and consequences for wave prediction, *J. Phys. Oceanogr.* **37**, 2389-2400.

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

Janssen, Peter A.E.M., and Augustus J.E.M. Janssen, 2019. Asymptotics for the long-time evolution of kurtosis of narrow-band ocean waves *J. Fluid Mech.* **859**, 790-818.

Krasitskii, V.P., 1994. On reduced equations in the Hamiltonian theory of weakly nonlinear surface waves. *J. Fluid Mech.* **272**, 1-20.

Lake, B.M., H.C. Yuen, H. Rungaldier and W.E. Ferguson, Jr., (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.

Magnusson, A.K. and M.A. Donelan, 2013. The Andrea wave Characteristics of a Measured North Sea Rogue Wave. Transactions of the ASME Journal of Offshore Mechanics and Arctic Engineering, **135**,

031108-3.

Montina, A., U. Bortolozzo, S. Residori, and F.T. Arecchi, 2009. Non-Gaussian Statistics and Extreme Waves in a Nonlinear Optical Cavity, PRL **103**, 173901.

Mori N. and T. Yasuda, 2002: A weakly non-gaussian model of wave distribution for random wave train. *Ocean Engineering* **29**, 1219-1231.

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

Mori N., M. Onorato, and P.A.E.M. Janssen, 2011. On the estimation of the kurtosis in directional sea states for freak wave forecasting.*J. Phys. Oceanogr.* **41**, 1484.

Naess, A., 1982. Extreme value estimates based on the envelope concept. *Applied Ocean research*, **4**, Issue 3, 181-187.

Naess, A., 1985. On the distribution of crest to trough wave heights. *Ocean Engng* **12**, 221-234.

Onorato M., A.R. Osborne, M. Serio, and S. Bertone, 2001. Freak Waves in Random Oceanic Sea States. *Phys. Rev. Lett.* **86**, 5831-5834.

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.

Onorato, M., A.R. Osborne and M. Serio, 2006. Instability in Crossing Sea States: A mechanism for the formation of Freak Waves. *Phys. Rev. Lett.* **96**, 014503.

Onorato M. and P.A.E.M. Janssen, 2024. On the non-uniqueness of the kernel of the Zakharov equation in intermediate and shallow water: the connection with the Davey-Stewartson equation. *J. Fluid Mech.* **980**, A32.

Stansell, P., 2005. Distributions of extreme wave, crest and trough heights measured in the North Sea. *Ocean Engineering* **32**, 1015-1036.

Residori S., 2015. Private communication and presentation at Cargèse summerschool on Rogue Waves.

Tanaka, M., 1992. The role of the modulational instability in the formation of wave groups. *Breaking Waves*, M.L. Banner and R.H.J. Grimshaw, Eds., Springer Verlag, 237-242.

Tayfun, M. A. and Lo, J.-M., 1990. Nonlinear effects on wave envelope and phase. *ASCE J. Waterway Port Coastal Ocean Engng* **116**, 79-100.

Toffoli, A., A. Alberello, H. Clarke, F. Nelli, A. Benetazzo, F. Bergamasco, B. Ntamba Ntamba, M. Vichi, and M. Onorato, 2024. Observations of rogue seas in the Southern Ocean. *Phys. Rev. Lett.* **132**, 154101

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.

Walczak, Pierre, Stéphane Randoux, and Pierre Suret, 2015. Optical Rogue Waves in integrable turbulence. PRL **114**, 143903.

Whitham, G.B., 1974. *Linear and Nonlinear Waves*. Wiley, New York.

Worsley, K.J., 1996. The geometry of random images. *Chance*, **9**, 27-40.

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.

Zakharov, V.E., and A.B. Shabat, 1972: Exact theory of two-dimensional self-focussing and one-dimensional self-modulating waves in nonlinear media. *Sov. Phys.-JETP (Engl. Transl.)* **34**, 62.