J. Fluid Mech. (2004), *vol.* 503, *pp.* 273–297.
DOI: 10.1017/S002211200400802X Printed in the United Kingdom

Improved joint probability distribution for ocean wave heights and periods

By PAUL STANSELL[1]†, JULIAN WOLFRAM[2] AND BRIAN LINFOOT[2]

[1]The University of Edinburgh, James Clerk Maxwell Building, The King's Buildings, Mayfield Road Edinburgh, UK

[2]School of the Built Environment, Heriot-Watt University, Riccarton Campus, Edinburgh, EH14 4AS, UK

(Received 21 September 2001 and in revised form 9 September 2003)

Expressions have been derived, from the standpoint of Gaussian random theory, for the joint and marginal distributions of wave envelope amplitude and local wave period when the sampling frequency is equal to the local wave frequency. The new marginal p.d.f. for wave envelope amplitudes shows substantial non-zero probability density at very small amplitudes. The new joint and marginal distributions are found to compare favourably with data obtained from both high frequency measurements of real ocean waves in extreme storms and from measurements of numerical simulations of moderately broadbanded processes with Pierson–Moskowitz spectra. The new p.d.f. of wave envelope amplitudes is found to provide a better approximation to the p.d.f.s of the simulated zero down-crossing wave amplitude than either the traditional Rayleigh p.d.f., applicable for infinitesimal bandwidth; or the narrow bandwidth approximation given by M. S. Longuet-Higgins (*Proc. R. Soc. Lond.* A, 389: 241–258, 1983). The reason for this improvement is that our method takes into account that small waves are likely to have shorter periods than large waves, rather than assuming a constant wave period. There are, however, limitations to the approach adopted which assumes that the individual wave amplitudes can be obtained from the amplitude of the wave envelope. These limitations become more severe as the bandwidth increases. The results obtained apply not only to sea waves, but to any Gaussian linear random process.

1. Introduction

Many results in the theory of random linear waves are derived upon the assumption of a narrow band process. Indeed, when considering the envelope process in the derivation of the Rayleigh distribution for wave amplitudes, an infinitesimally narrow band is assumed. In this paper we examine what happens when the bandwidth becomes finite. Although our discussion is given using the terminology appropriate to surface gravity waves on the ocean, our results will be equally applicable to any Gaussian linear random process.

First, and as illustrated in figure 1, we note that there is a correlation between the wave amplitude, h, and the zero down-crossing wave period, T, measured from real and simulated ocean wave records, $\eta(t)$. Allowing for this when sampling from

† Author to whom correspondence should be addressed: p.stansell@ed.ac.uk.

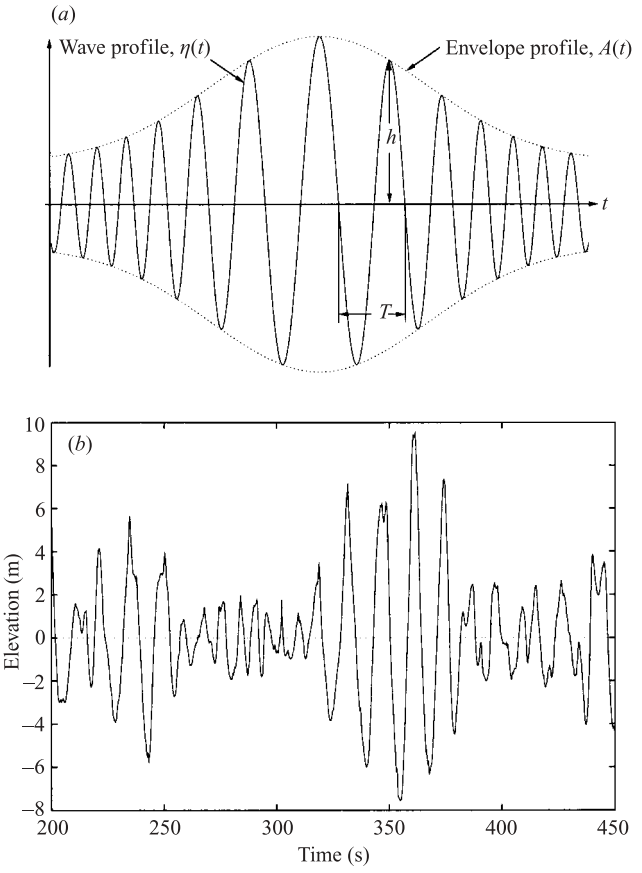

FIGURE 1. (*a*) A schematic illustration of the increase in local wave period with increasing wave envelope amplitude, including notation. (*b*) An example from a wave record measured in the North Sea. This phenomenon is ubiquitous in real broadband ocean wave data.

the wave envelope, $A(t)$, yields distributions of wave amplitudes that contain terms that depend on the bandwidth, ν (defined in (7)). These distributions are, therefore, different from the Rayleigh distribution. In real seas it has been observed (Podgórski *et al.* 2000) that there are frequently more small waves than predicted by the Rayleigh distribution and our new distributions reflect these observations. The associated marginal distributions for wave period and the joint distributions for amplitude and period are also different from those previously obtained.

Second, we find that as the bandwidth becomes finite the wave envelope ceases to be the loci of the amplitudes of the wave surface. As discussed later, this is associated with loops in the phase diagram that do not encircle the origin. The inference from this is that distributions and other characteristics derived from the envelope of a linear random Gaussian process will yield only approximate results when the bandwidth is finite. The inaccuracies become greater as the bandwidth increases, and one measure of this error is obtained. The theoretical results presented here will apply not only to sea waves, but to any narrow (but finite) bandwidth random linear process.

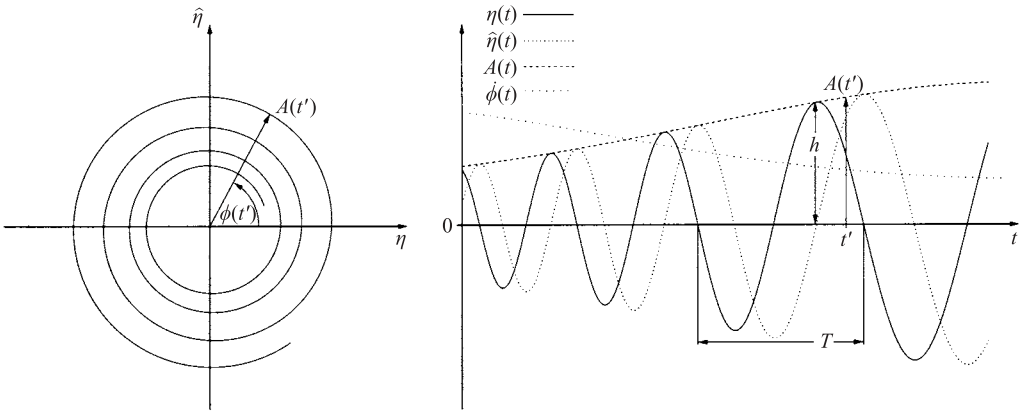

FIGURE 2. Schematic phase plot and time plot for narrowband waves.

Thirdly, we compare the new distributions with existing distributions, and with simulated and real wave data of different bandwidths. We also observe, consistent with our theory, an interesting downward trend in mean amplitude as bandwidth increases.

The general case of the distribution of maxima of a broadbanded process was considered by Rice (1944, 1945) and, in the context of sea waves, by Cartwright & Longuet-Higgins (1956). In the general case there may be more than one surface maximum and more than one surface minimum associated with each zero-crossing wave. This was the case considered by Cavanié *et al.* (1976) and Arhan *et al.* (1976) in deriving their joint p.d.f. for waveheights and periods (for details in English see Ochi 1998). In the practical analysis of sea waves, however, it is usual to associate just one maximum and one minimum, the largest, between sequential up or down-crossings. Any maxima or minima smaller than the largest are ignored. Correspondingly, when defining the wave period it is the interval between zero up or down-crossings that is of interest. For practical purposes it is the joint probability of wave amplitude (or height) and wave period that is of interest to naval architects and ocean engineers estimating wave loadings and the response of dynamically sensitive structures. These distributions should, ideally, be defined with at least the same degree of accuracy as the other well-defined principal elements in such calculations; that is within 1 or 2%. This is particularly important in the tails of the distributions where critical design points are usually located. In practice, however, such accuracy is rarely achieved and many calculations have errors of 10% or more.

For a general time-dependent narrowbanded analytic signal, denoted by $\eta(t)$, figure 2 shows its phasor rotating in the $(\eta, \hat{\eta})$ Cartesian plane, where $\hat{\eta}(t)$ is the Hilbert transform of $\eta(t)$. For sufficiently narrowbanded signals there will be just one maximum and one minimum per zero down-crossing. In this case the signal amplitudes are almost exactly equal to (being very slightly less than) the amplitude of the envelope at the same time. As indicated in the figure, the local phase, $\phi(t)$, is always increasing with time and the rate of change of local phase, or local frequency, $\dot{\phi}(t)$, is always positive. Figure 3 illustrates the characteristics (somewhat exaggerated) of the general broadband case where there may be more than one signal maximum or minimum per zero down-crossing. These additional maxima and minima give rise to loops in the $(\eta, \hat{\eta})$ phase plane that do not encircle the origin. In such cases the signal

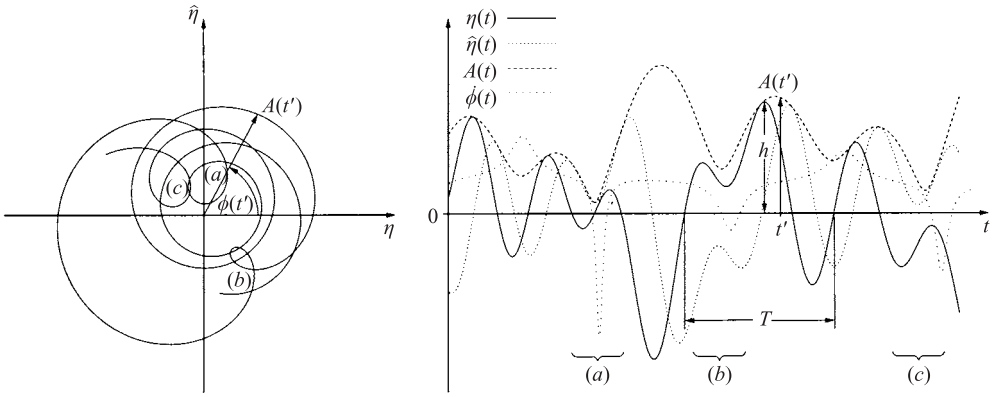

FIGURE 3. Schematic phase plot and time plot for broadband waves. Regions of the phase plot showing phase loops that do not encircle the origin are labelled (*a*), (*b*) and (*c*) and the corresponding regions are labelled similarly in the time plot.

amplitude at its maxima and minima is not always equal to that of the envelope and $\dot{\phi}(t)$ can be negative.

When obtaining his joint distribution for the wave height and period of sea waves Longuet-Higgins (1983) assumed a fixed sampling interval and bandwidth conditions somewhat similar to those shown in figure 2. In § 3 of this paper we consider the same conditions except that we sample the wave envelope, not at a constant rate, but at the rate $\dot{\phi}(t)$ which is not independent of the wave envelope amplitude. (At first sight one might consider a Taylor series expansion of $\dot{\phi}(t)$ in terms of the bandwidth ν in order to estimate a more appropriate sampling interval. As illustrated in figure 3 and discussed later, however, for any finite bandwidth $\dot{\phi}(t)$ varies in a random manner over the sampling interval and so a series expansion at any instant in time would need to be impracticably long.) We also explore the case when the bandwidth increases and small loops appear in the phase plots and $\dot{\phi}(t)$ is occasionally negative. Thus, two types of distribution are obtained: one where we sample at the local frequency, $\dot{\phi}(t)$, and it is assumed always to be positive; and one where we sample at the local frequency and it is allowed to be negative. In both cases we take the amplitude, A, of the wave envelope to be equal to the amplitude, h, of the zero down-crossing wave. In the latter case this approach is inconsistent and becomes progressively less accurate as the bandwidth increases. It is, however, interesting to compare the two approaches and it is shown in § 4 that, although distributions obtained from both approaches give good approximations to both simulated and real broadband data, the better approximation is provided by the model which samples the wave envelope only at times during which the local frequency is positive. This model is shown to provide a clear improvement over that of Longuet-Higgins (1975, 1983).

2. Joint p.d.f. for wave envelope amplitude and local wave frequency

We start from the same basis as did Longuet-Higgins (1975, 1983), however, we relax the restriction he imposed (which followed from the narrowband assumption) that the local wave frequency, $\dot{\phi}$, be always positive . The result is a slightly different joint p.d.f. for the wave envelope amplitude and local period. As may be anticipated, the marginal distribution for the wave envelope amplitude remains Rayleigh. The

marginal distribution for the local wave period, however, is different from that obtained by Longuet-Higgins (1983).

We start with the variables in their usual dimensions. The notation used herein follows the convention that variables having physical units are denoted with an asterisk superscript (*) and those equivalent variables that have been non-dimensionalized are denoted without the asterisk.

Following Longuet-Higgins (1975, 1983) the sea surface elevation is written as

$$\eta^*(t^*) = \sum_{n=1}^{\infty} a_n^* \cos(\omega_n^* t^* + \epsilon_n), \tag{1}$$

where ϵ_n are random phases taken from a uniform distribution. The Hilbert transform of $\eta^*(t^*)$ is

$$\widehat{\eta^*}(t^*) = \sum_{n=1}^{\infty} a_n^* \sin(\omega_n^* t^* + \epsilon_n). \tag{2}$$

The following is the analytic signal through which the real variables A^* and ϕ are defined

$$A^*(t^*)\mathrm{e}^{\mathrm{i}\phi(t^*)} = \eta^*(t^*) + \mathrm{i}\widehat{\eta^*}(t^*). \tag{3}$$

The variables of interest are A^* which is the local wave amplitude or envelope amplitude, and $\mathrm{d}\phi/\mathrm{d}t^* = \dot{\phi}^*$ which is the local wave frequency in radians.

Assuming $\eta^*(t^*)$ is Gaussian distributed, the joint p.d.f. of A^* and $\dot{\phi}^*$can be written as

$$p_{A^*\dot{\phi}^*}(A^*, \dot{\phi}^*) = \frac{A^{*2}}{\sqrt{2\pi m_0^*\left(m_0^* m_2^* - m_1^{*2}\right)}} \exp\left(-\frac{A^{*2}\left(m_2^* - 2m_1^*\dot{\phi}^* + m_0^*\dot{\phi}^{*2}\right)}{2\left(m_0^* m_2^* - m_1^{*2}\right)}\right), \tag{4}$$

where the spectral moments are defined by

$$m_n^* = \int_0^{\infty} \omega^{*n} S^*(\omega^*)\,\mathrm{d}\omega^*, \qquad n = 0, 1, 2, \ldots, \tag{5}$$

and where $S^*(\omega^*)$ is the one-sided frequency spectrum of the surface elevation η^* and ω^* is the angular frequency. (Equation (4) is obtained by assuming that the statistics of the variables η^*, $\widehat{\eta^*}$, $\dot{\eta}^*$ and $\widehat{\dot{\eta}^*}$ are described by a joint Gaussian distribution, applying a change of variables from $(\eta^*, \widehat{\eta^*}, \dot{\eta}^*, \widehat{\dot{\eta}^*})$ to $(A^*, \dot{A}^*, \phi^*, \dot{\phi}^*)$, and integrating out the dependency on $(\dot{A}^*, \phi^*)$.) For the lower order moments it can be shown (see Ochi 1998, p. 104) that

$$\left.\begin{aligned} m_0^* &= E[\eta^{*2}], \\ m_1^* &= E[\eta^*\dot{\eta}^*], \\ m_2^* &= E[\dot{\eta}^{*2}], \end{aligned}\right\} \tag{6}$$

where $E[\cdot]$ stands for the expectation value of $(\cdot)$. The equivalence of (4) and Longuet-Higgins' (1975) equation (A16) is demonstrated in Appendix A.

We will also be using the dimensionless bandwidth parameter (see Longuet-Higgins 1983) defined by

$$\nu^2 = \frac{m_0^* m_2^*}{m_1^{*2}} - 1, \tag{7}$$

and the dimensionless envelope amplitude and local wave angular velocity defined by

$$A = \frac{A^*}{\sqrt{2m_0^*}}, \tag{8}$$

$$\dot{\phi} = \frac{2\pi m_0^* \dot{\phi}^*}{m_1^*}. \tag{9}$$

Equation (4) can be written in term of these dimensionless variables using the transformation

$$p_{A\dot{\phi}}(A, \dot{\phi}) = p_{A^*\dot{\phi}^*}\left(\sqrt{2m_0^*}A, \frac{m_1^*\dot{\phi}}{2\pi m_0^*}\right)\left|\frac{\mathrm{d}A^*}{\mathrm{d}A}\right|\left|\frac{\mathrm{d}\dot{\phi}^*}{\mathrm{d}\dot{\phi}}\right|.$$

Substituting for ν from (7) yields

$$p_{A\dot{\phi}}\left(A, \dot{\phi}\right) = \frac{A^2}{\pi\sqrt{\pi}\nu}\exp\left[-\frac{A^2}{\nu^2}\left(\frac{\dot{\phi}^2}{4\pi^2} - \frac{\dot{\phi}}{\pi} + \nu^2 + 1\right)\right]. \tag{10}$$

Transforming to the dimensionless local period, defined by $\tau = 2\pi/\dot{\phi}$ (which can, therefore, take negative values), yields

$$p_{A\tau}(A, \tau) = \frac{2A^2}{\sqrt{\pi}\nu\tau^2}\exp\left[-\frac{A^2}{\nu^2}\left(\frac{1}{\tau^2} - \frac{2}{\tau} + \nu^2 + 1\right)\right]. \tag{11}$$

Equation (11) is very similar to the joint p.d.f. given by (2.16) in Longuet-Higgins (1983), the latter of which can be written

$$p_{\mathrm{LH}}(A, \tau) = N_{\mathrm{LH}}\, p_{A\tau}(A, \tau). \tag{12}$$

The difference lies in the inclusion of a normalization factor

$$\frac{1}{N_{\mathrm{LH}}} = \int_0^\infty \mathrm{d}A \int_0^\infty p_{A\tau}\,(A, \tau)\,\mathrm{d}\tau = \tfrac{1}{2}\left(1 + (1+\nu^2)^{-1/2}\right). \tag{13}$$

This ν-dependent normalization factor arises as a direct consequence of Longuet-Higgins' (1983) exclusion of negative values of $\dot{\phi}$, and therefore, by definition, of τ. Here we impose no such restriction and allow τ to be negative so that

$$\int_0^\infty \mathrm{d}A \int_{-\infty}^\infty p_{A\tau}\,(A, \tau)\,\mathrm{d}\tau = 1.$$

The marginal p.d.f.s are given by

$$p_A(A) = \int_{-\infty}^\infty p_{A\dot{\phi}}(A, \dot{\phi})\,\mathrm{d}\dot{\phi} = 2A\exp(-A^2), \tag{14}$$

$$p_{\dot{\phi}}(\dot{\phi}) = \int_0^\infty p_{A\dot{\phi}}(A, \dot{\phi})\,\mathrm{d}A = \frac{\nu^2}{4\pi}\left(\frac{\dot{\phi}^2}{4\pi^2} - \frac{\dot{\phi}}{\pi} + \nu^2 + 1\right)^{-3/2},$$

$$p_\tau(\tau) = \int_0^\infty p_{A\tau}(A, \tau)\,\mathrm{d}A = \frac{\nu^2}{2\tau^2}\left(\frac{1}{\tau^2} - \frac{2}{\tau} + \nu^2 + 1\right)^{-3/2}, \tag{15}$$

since, for real values of ν and $\dot{\phi}$, it is always the case that

$$\frac{\dot{\phi}^2}{4\pi^2} - \frac{\dot{\phi}}{\pi} + \nu^2 + 1 > 0,$$

and therefore

$$\frac{1}{\tau^2} - \frac{2}{\tau} + \nu^2 + 1 > 0.$$

The expectation value of $\dot{\phi}$ is

$$E\left[\dot{\phi}\right] = \int_{-\infty}^{\infty} \dot{\phi}\, p_{\dot{\phi}}(\dot{\phi})\, \mathrm{d}\dot{\phi} = 2\pi, \tag{16}$$

while the expectation value of τ is

$$E[\tau] = \int_{-\infty}^{\infty} \tau\, p_{\tau}(\tau)\, \mathrm{d}\tau = 1. \tag{17}$$

The probability that $\dot{\phi} < 0$ is

$$\Pr\left(\dot{\phi} < 0\right) = \tfrac{1}{2}\left(1 - \frac{1}{\sqrt{1+\nu^2}}\right) = \frac{\nu^2}{4} + O(\nu^4), \tag{18}$$

and the probability that $\dot{\phi} > 0$ is

$$\Pr\left(\dot{\phi} > 0\right) = \tfrac{1}{2}\left(1 + \frac{1}{\sqrt{1+\nu^2}}\right) = 1 - \frac{\nu^2}{4} + O(\nu^4). \tag{19}$$

From (11), (12) and (15) it can be seen that the marginal p.d.f. for zero-crossing wave period, T, obtained by Longuet-Higgins (1983) is

$$p_{\mathrm{LH}}(T) = \frac{\nu^2 N_{\mathrm{LH}}}{2T^2}\left(\frac{1}{T^2} - \frac{2}{T} + \nu^2 + 1\right)^{-3/2}, \tag{20}$$

where it is assumed that T can be represented by the local wave period, τ, when the latter is non-negative. Also, the marginal p.d.f. for wave amplitude obtained by Longuet-Higgins (1983) is given by

$$p_{\mathrm{LH}}(h) = N_{\mathrm{LH}} \int_0^{\infty} p_{A\dot{\phi}}(h, \dot{\phi})\, \mathrm{d}\dot{\phi} = N_{\mathrm{LH}} h \exp(-h^2)\left[1 + \mathrm{erf}\left(\frac{h}{\nu}\right)\right], \tag{21}$$

where $N_{\mathrm{LH}}(\nu)$ is given by (13).

The cumulative distribution corresponding to (20) is

$$P_{\mathrm{LH}}(T) = \frac{N_{\mathrm{LH}}}{2}\left\{\frac{T-1}{[T^2(1+\nu^2) - 2T + 1]^{1/2}} + 1\right\},$$

and that corresponding to (21) is

$$P_{\mathrm{LH}}(h) = \frac{N_{\mathrm{LH}}}{2}\left\{\frac{\mathrm{erf}(\sqrt{1+1/\nu^2}h)}{\sqrt{1+\nu^2}} - \exp(-h^2)\left[1 + \mathrm{erf}\left(\frac{h}{\nu}\right)\right] + 1\right\}.$$

3. Sampling at the local wave frequency

In the narrowband approach the p.d.f. for individual wave amplitudes, h, is assumed to be the same as the p.d.f. of the wave envelope amplitude, A. As noted above, real sea waves of smaller amplitude tend to have shorter periods, reflected in a higher local wave frequency and a higher zero crossing frequency of the wave surface. Therefore, when the wave envelope amplitude is small one should sample the wave envelope at a higher frequency to obtain the p.d.f. of the individual wave amplitudes.

Here we start by establishing that the mean zero up-crossing frequency, denoted f_z, is simply related to the modulus of local wave frequency. We thus find the mean zero up-crossing frequency and the mean zero down-crossing wave period, denoted T_z. When the local wave frequency is assumed always to be positive we find the result is slightly different.

We also establish a simple relation for the probability of occurrence of zero-crossing waves having negative local wave frequency, $\dot{\phi}$. This relation is a function of bandwidth. It is small waves such as those corresponding phase loop (a) in figure 3 that are ignored by assuming $\dot{\phi}$ is always positive.

We then find two new p.d.f.s for A by integrating the joint p.d.f. of A and $\dot{\phi}$. The second new p.d.f. is obtained by constraining $\dot{\phi}$ to be positive.

3.1. *Mean zero-crossing wave period*

An individual wave is defined as that part of the surface elevation profile which lies between any two consecutive zero down-crossings. It is assumed that the wave heights are equal to twice the local wave envelope amplitude at the times of the zero up-crossing which defines the centre of the wave. Therefore, to find the p.d.f. of wave heights due consideration must be given to the variation of sampling frequency with envelope amplitude.

Consider an ensemble of sea surface elevations each realisation of which is labelled by n. Each sea surface elevation series is represented by an analytic signal like that in (3). In an infinitesimal time interval δt between times t and $t+\delta t$ a zero up-crossing occurs in any one series if and only if $\eta(t)<0$ and $\eta(t+\delta t)>0$. As all values of ϕ are equally likely, the probability that any one realization will have a zero up-crossing between t and $t+\delta t$ is given by the ratio of the arc swept out by the phasor in this time (divided by 2π) to the total number of ensembles. This can be formulated as

$$\Pr\begin{pmatrix}\text{zero upcrossing}\\ \text{between } t \text{ and } t+\delta t\end{pmatrix} = \lim_{N\to\infty}\lim_{\delta t\to 0}\frac{1}{N}\sum_{n=1}^{N}\frac{|\phi_n(t+\delta t)-\phi_n(t)|}{2\pi},$$

$$= \lim_{N\to\infty}\frac{1}{2\pi N}\sum_{n=1}^{N}|\dot{\phi}(t)|\delta t, \tag{22}$$

where Pr (zero upcrossing between t and $t+\delta t$) is the probability of a zero up-crossing in the infinitesimal time interval between t and $t+\delta t$. The absolute value is taken because it is immaterial in which direction, clockwise or anti-clockwise, ϕ is increasing at the time of the measurement – both positive and negative directions can give rise to a zero up-crossing.

As the process is stationary and ergodic (by the defining equation (1)) the sum over ensembles can be replaced by an integration over time and the rate, or frequency f_z, of zero up-crossing can be written, from (22), as

$$f_z = \frac{1}{\delta t}\Pr\begin{pmatrix}\text{zero upcrossing}\\ \text{between } t \text{ and } t+\delta t\end{pmatrix},$$

$$= \lim_{T_0\to\infty}\frac{1}{2\pi T_0}\int_0^{T_0}|\dot{\phi}(t)|\,\mathrm{d}t.$$

Alternatively, the time average of $|\dot{\phi}(t)|$ may be written in the expectation notation using the joint p.d.f. $p_{A\dot{\phi}}(A, \dot{\phi})$ given by (10). The result is

$$f_z = \frac{1}{2\pi} E[|\dot{\phi}(t)|] = \frac{1}{2\pi} \int_0^\infty \mathrm{d}A \int_{-\infty}^\infty |\dot{\phi}|\, p_{A\dot{\phi}}(A, \dot{\phi})\, \mathrm{d}\dot{\phi}. \tag{23}$$

(The previous argument could have been bypassed and we could have simple started with $f_z = E[|\dot{\phi}|]/2\pi$ which is fairly intuitive.)

Carrying out the integration over $\dot{\phi}$ in (23) yields

$$f_z = \frac{1}{2\pi} \int_0^\infty 4\pi \exp(-A^2) \left[\frac{\nu}{\sqrt{\pi}} \exp\left(-\frac{A^2}{\nu^2}\right) + A \operatorname{erf}\left(\frac{A}{\nu}\right) \right] \mathrm{d}A \tag{24}$$

Carrying out the integration over A in (24) yields

$$f_z = \sqrt{1+\nu^2} = \frac{\sqrt{m_0^* m_2^*}}{m_1^*}. \tag{25}$$

This is a standard result and is valid generally for a any ergodic broadband Gaussian random process (see, for example, Newland 1984). The total number of zero up-crossings expected in time T_0 is $f_z T_0$, and so the average period between zero up-crossings, which is equal to the average zero-crossing wave period, is

$$\begin{aligned} T_z &= f_z^{-1}, \\ &= \frac{1}{\sqrt{1+\nu^2}}. \end{aligned} \tag{26}$$

3.2. *Proportion of zero-crossing waves with* $\dot{\phi} < 0$

The ratio of the number of zero-crossings with $\dot{\phi} > 0$ to the number with $\dot{\phi} < 0$ can be derived using a similar argument to that used in § 3.1 to find (25).

The frequency of zero up-crossings with $\dot{\phi} > 0$ is given by

$$f_{z|\dot{\phi}>0} = \frac{1}{2\pi} \int_0^\infty \mathrm{d}A \int_0^\infty \dot{\phi}\, p_{A\dot{\phi}}(A, \dot{\phi})\, \mathrm{d}\dot{\phi} = \frac{\sqrt{1+\nu^2}+1}{2}.$$

The frequency of zero up-crossings with $\dot{\phi} < 0$ is given by

$$f_{z|\dot{\phi}<0} = \frac{1}{2\pi} \int_0^\infty \mathrm{d}A \int_{-\infty}^0 -\dot{\phi}\, p_{A\dot{\phi}}(A, \dot{\phi})\, \mathrm{d}\dot{\phi} = \frac{\sqrt{1+\nu^2}-1}{2}.$$

Thus, we see that the frequency, or rate, of zero-crossings with loci in the $(\eta, \hat{\eta})$ phase plane that encircle the origin (for example, all the waves in figure 2) is $(\sqrt{1+\nu^2}+1)/2$, and the frequency of zero-crossings with loci that do not encircle the origin (for example, the small wave giving rise to the phase loop (*a*) in figure 3) is $(\sqrt{1+\nu^2}-1)/2$. The ratio, ρ, of phase loop zero-crossing waves to the total number of zero-crossing waves is

$$\rho = \frac{1}{2}\left(1 - \frac{1}{\sqrt{1+\nu^2}}\right) = \frac{\nu^2}{4} + O(\nu^4). \tag{27}$$

It is interesting to note that for any finite bandwidth there will be at least some zero-crossing waves associated with negative local frequency and that this will increase, in the first instance, with the square of the bandwidth. This result is useful in giving a measure of the level of approximation involved in using the narrowband assumption and the other assumptions made in this paper.

3.3. *Envelope amplitude sampled at the local wave frequency*

In the same way that the p.d.f. for the wave envelope amplitude sampled at regular intervals is given by the Rayleigh p.d.f.

$$R(A) = \int_{-\infty}^{\infty} p_{A\dot{\phi}}(A, \dot{\phi})\, \mathrm{d}\dot{\phi} = 2A\, \exp(-A^2), \tag{28}$$

the p.d.f. of the wave envelope amplitude sampled at the local wave frequency, which we denote using the subscript h in anticipation of its relation to the p.d.f. of waves amplitudes, is given by

$$p_h(A) = N \int_{-\infty}^{\infty} |\dot{\phi}|\, p_{A\dot{\phi}}(A, \dot{\phi})\, \mathrm{d}\dot{\phi}, \tag{29}$$

where N is a ν-dependent normalization factor given by

$$N = \left[\int_0^{\infty} \mathrm{d}A \int_{-\infty}^{\infty} |\dot{\phi}|\, p_{A\dot{\phi}}(A, \dot{\phi})\, \mathrm{d}\dot{\phi} \right]^{-1}.$$

The values of these last two integrals have already been calculated ((23)–(25)) giving, respectively,

$$p_h(A) = \frac{2}{\sqrt{1+\nu^2}} \exp(-A^2) \left[\frac{\nu}{\sqrt{\pi}} \exp\left(-\frac{A^2}{\nu^2} \right) + A \operatorname{erf}\left(\frac{A}{\nu} \right) \right], \tag{30}$$

and

$$N = \frac{1}{2\pi\sqrt{1+\nu^2}}. \tag{31}$$

The cumulative distribution corresponding to (30) is

$$P_h(A) = \operatorname{erf}\left(\sqrt{1 + \frac{1}{\nu^2}} A \right) - \frac{1}{\sqrt{1+\nu^2}} \exp(-A^2) \operatorname{erf}\left(\frac{A}{\nu} \right). \tag{32}$$

We now deal with the case in which we make the approximation that the wave process is sufficiently narrowbanded to allow us to ignore all regions of the envelope for which $\dot{\phi} < 0$. In this case (29) is modified to give

$$p_{h_n}(A) = N_n \int_0^{\infty} \dot{\phi}\, p_{A\dot{\phi}}(A, \dot{\phi})\, \mathrm{d}\dot{\phi}, \tag{33}$$

where N_n is a ν-dependent normalization factor which evaluates to

$$N_n = \frac{1}{\pi(1+\sqrt{1+\nu^2})}. \tag{34}$$

The n subscript on p_{h_n} and N_n stands for the inclusion of non-negative $\dot{\phi}$ only.

The value of the integral in (33) evaluates to

$$p_{h_n}(A) = \frac{2}{1+\sqrt{1+\nu^2}} \exp(-A^2) \left(\frac{\nu}{\sqrt{\pi}} \exp\left(-\frac{A^2}{\nu^2} \right) + A \left(1 + \operatorname{erf}\left(\frac{A}{\nu} \right) \right) \right), \tag{35}$$

and the associated cumulative distribution is

$$P_{h_n}(A) = \frac{1}{1+\sqrt{1+\nu^2}} \left\{ \sqrt{1+\nu^2} \operatorname{erf}\left(\sqrt{1 + \frac{1}{\nu^2}} A \right) - \exp(-A^2) \left[1 + \operatorname{erf}\left(\frac{A}{\nu} \right) \right] + 1 \right\}.$$

Note that neither (30) nor (35) are the same as the p.d.f. for heights of maxima derived by Cartwright & Longuet-Higgins (1956), as the latter is a function of the fourth order spectral moment, whereas the former are not. Taking the narrowband limit of (30) or (35) gives, in both cases, the Rayleigh p.d.f.

One significant difference arising from this new approach is the high density of small amplitudes given by the p.d.f.s in (30) and (35) as compared to the zero density given by (28). In the limit of zero A we have

$$\lim_{A\to 0} p_h(A) = \frac{2\nu}{\sqrt{\pi}\sqrt{1+\nu^2}},$$

$$\lim_{A\to 0} p_{h_n}(A) = \frac{2\nu}{\sqrt{\pi}\left(1+\sqrt{1+\nu^2}\right)},$$

whereas, from (28), we have

$$\lim_{A\to 0} R(A) = 0.$$

The high density of small waves in the new models leads to an expectation value of A which decreases with increasing ν. The expectation value of A calculated from $R(A)$, given by (28), is $E[A] = \sqrt{\pi}/2$, whereas those calculated from the new distributions (30) and (35) are functions of ν given by

$$E[A]_h = \int_0^\infty A p_h(A)\,\mathrm{d}A = \frac{\nu + \arctan(1/\nu)}{\sqrt{\pi}\sqrt{1+\nu^2}}, \tag{36}$$

$$E[A]_{h_n} = \int_0^\infty A p_{h_n}(A)\,\mathrm{d}A = \frac{\pi + 2(\nu + \arctan(1/\nu))}{2\sqrt{\pi}(1+\sqrt{1+\nu^2})}. \tag{37}$$

Note that from the distribution of Longuet-Higgins (1983) given by (21) the expectation value of A is calculated as

$$E[A]_{\mathrm{LH}} = \int_0^\infty A p_{\mathrm{LH}}(A)\,\mathrm{d}A = \frac{N_{\mathrm{LH}}(\pi(1+\nu^2) + 2\arctan(1/\nu)(1+\nu^2) + 2\nu)}{4\sqrt{\pi}(1+\nu^2)}, \tag{38}$$

where N_{LH} is given by (13). This last result for $E[A]$ contrasts with the results from the new distributions as it predicts the expectation value of A to increase with increasing ν. These variations of mean amplitude with ν are compared with real and simulated data in § 4 below.

Now we examine the large A limit. In the limit of large A the difference between the Rayleigh p.d.f. and those given by (30) and (35) can be expressed in terms of ν as

$$\lim_{A\to\infty} \frac{p_h(A)}{R(A)} = \frac{1}{\sqrt{1+\nu^2}}. \tag{39}$$

$$\lim_{A\to\infty} \frac{p_{h_n}(A)}{R(A)} = \frac{2}{1+\sqrt{1+\nu^2}}. \tag{40}$$

In fact, (39) and (40) are accurate even for high values of ν when A is relatively small, say $A > 1$. The equivalent result for the density derived by Longuet-Higgins (1983) is

$$\lim_{A\to\infty} \frac{p_{\mathrm{LH}}(A)}{R(A)} = N_{\mathrm{LH}}, \tag{41}$$

where N_{LH} is a function of ν given by (13). These last three equations show that in the tail of the density function the prediction of $p_{\mathrm{LH}}(A)$ is greater than that of $R(A)$, whereas those of $p_h(A)$ and $p_{h_n}(A)$ are both less than $R(A)$. Thus, the new

distributions predict a decreasing number of large waves, compared to Rayleigh, as bandwidth increases.

3.4. *Joint p.d.f. for envelope amplitude and period sampled at the local wave frequency*

In this section we derive joint p.d.f.s for envelope amplitude and local zero-crossing wave period when the wave is sampled at the local wave frequency $\dot{\phi}$. The local zero-crossing wave period is denoted by T and constrained to be always positive (unlike $\dot{\phi}$).

We know that $p_h(A)$ can be written as

$$p_h(A) = \int_0^\infty p_{hT}(A, T)\,\mathrm{d}T, \tag{42}$$

for some $p_{hT}(A, T)$ which is defined as the joint p.d.f. of envelope amplitude, and some T which is a non-negative measure of local zero-crossing wave period. We define our non-negative measure of local zero down-crossing wave period by $T \equiv 2\pi/|\dot{\phi}|$. This T is the closest we can get to a surrogate for the non-local measure of zero-crossing wave period which is obtained as the time between consecutive zero down-crossings in real ocean wave time series (see (48) and the text following for support of this statement). The later obviously cannot be negative. The form of $p_{hT}(A, T)$ in (42) can be deduced by writing (29) in the following form for which we integrate only over positive values of $\dot{\phi}$:

$$p_h(A) = N \int_0^\infty |\dot{\phi}|\,[p_{A\dot{\phi}}(A, |\dot{\phi}|) + p_{A\dot{\phi}}(A, -|\dot{\phi}|)]\,\mathrm{d}|\dot{\phi}|,$$

and making a change of variables from $|\dot{\phi}|$ to T we have

$$p_h(A) = -N \int_0^\infty \frac{2\pi}{T}\left[p_{A\dot{\phi}}\left(A, \frac{2\pi}{T}\right) + p_{A\dot{\phi}}\left(A, -\frac{2\pi}{T}\right)\right]\frac{\mathrm{d}|\dot{\phi}|}{\mathrm{d}T}\,\mathrm{d}T. \tag{43}$$

By comparing (42) and (43) the joint p.d.f. of h and T is seen to be

$$\begin{aligned} p_{hT}(A, T) &= \frac{4\pi^2 N}{T^3}\left[p_{A\dot{\phi}}\left(A, \frac{2\pi}{T}\right) + p_{A\dot{\phi}}\left(A, -\frac{2\pi}{T}\right)\right], \qquad T > 0,\ A > 0, \\ &= \frac{4\pi^2 N}{T^3}\frac{h^2}{\pi\sqrt{\pi}\nu}\left\{\exp\left[-\frac{h^2}{\nu^2}\left(\frac{1}{T^2} - \frac{2}{T} + \nu^2 + 1\right)\right]\right. \\ &\quad \left. + \exp\left[-\frac{h^2}{\nu^2}\left(\frac{1}{T^2} + \frac{2}{T} + \nu^2 + 1\right)\right]\right\}, \qquad T > 0,\ A > 0. \end{aligned}$$

where we have used (10). Substituting N from (31), the final result simplifies to

$$p_{hT}(A, T) = \frac{4}{\nu\sqrt{1+\nu^2}\sqrt{\pi}}\frac{A^2}{T^3}\cosh\left(\frac{2A^2}{\nu^2 T}\right)\exp\left[-\frac{A^2}{\nu^2}\left(\frac{1}{T^2} + \nu^2 + 1\right)\right], \qquad T > 0,\ A > 0. \tag{44}$$

Using the p.d.f. of local envelope amplitude and local period, $p_{A\tau}(A, \tau)$ given by (11), (44) can also be written as

$$p_{hT}(A, T) = \frac{1}{\sqrt{1+\nu^2}\,T}\left[1 + \exp\left(-\frac{4A^2}{\nu^2 T}\right)\right]p_{A\tau}(A, T), \qquad T > 0,\ A > 0.$$

We now obtain a slightly different joint p.d.f. by carrying out the same procedure as that given above except we exclude all regions of the wave series for which $\dot{\phi} < 0$.

The left-hand side of (33) can also be written as

$$p_{h_n}(A) = \int_0^\infty p_{h_n T}(A, T)\,\mathrm{d}T, \tag{45}$$

and comparing this to the result after transforming variables in (33) gives

$$p_{h_n T}(A, T) = \frac{4\pi^2 N_n}{T^3}\frac{h^2}{\pi\sqrt{\pi}\nu}\exp\left[-\frac{A^2}{\nu^2}\left(\frac{1}{T^2} - \frac{2}{T} + \nu^2 + 1\right)\right], \qquad T > 0,\ A > 0.$$

Substituting N_n from (34), the final result simplifies to

$$p_{h_n T}(A, T) = \frac{4}{\nu\left(1 + \sqrt{1+\nu^2}\right)\sqrt{\pi}}\frac{A^2}{T^3}\exp\left[-\frac{A^2}{\nu^2}\left(\frac{1}{T^2} - \frac{2}{T} + \nu^2 + 1\right)\right], \qquad T > 0,\ A > 0. \tag{46}$$

3.5. *Marginal p.d.f. for wave periods*

The marginal p.d.f. for wave period is derived from (44) by integrating out the A dependency using

$$p_T(T) = \int_0^\infty p_{hT}(A, T)\,\mathrm{d}A.$$

Evaluating this integral gives

$$p_T(T) = \frac{\nu^2}{2\sqrt{1+\nu^2}}\left\{[(1+\nu^2)T^2 + 2T + 1]^{-3/2} + [(1+\nu^2)T^2 - 2T + 1]^{-3/2}\right\}. \tag{47}$$

The cumulative distribution corresponding to Equation (47) is

$$P_T(T) = \frac{1}{2\sqrt{1+\nu^2}}\left[\frac{(1+\nu^2)T + 1}{\sqrt{(1+\nu^2)T^2 + 2T + 1}} + \frac{(1+\nu^2)T - 1}{\sqrt{(1+\nu^2)T^2 - 2T + 1}}\right].$$

The mean zero-crossing wave period, T_z, can be calculated as the expectation value of the local wave period, T, which is

$$E[T] = \int_0^\infty T p_T(T)\,\mathrm{d}T = \frac{1}{\sqrt{1+\nu^2}}. \tag{48}$$

From (26) and (48) we see that $E[T] = T_z$; thus our local wave period T, defined by $T \equiv 2\pi/|\dot{\phi}|$, is representative of a local zero-crossing wave period since its expectation value is equal to the average zero-crossing wave period T_z.

For the case when negative $\dot{\phi}$ are ignored the marginal p.d.f. for wave period is derived from (46) by

$$p_{T_n}(T) = \int_0^\infty p_{h_n T}(A, T)\,\mathrm{d}A.$$

Evaluating this integral gives

$$p_{T_n}(T) = \frac{\nu^2}{(1+\sqrt{1+\nu^2})}[T^2(1+\nu^2) - 2T + 1]^{-3/2}. \tag{49}$$

The cumulative distribution corresponding to (49) is

$$P_{T_n}(T) = \frac{1}{1+\sqrt{1+\nu^2}}\left[\frac{(1+\nu^2)T - 1}{\sqrt{(1+\nu^2)T^2 - 2T + 1}} + 1\right].$$

The expected zero-crossing wave period for the case when $\dot{\phi} < 0$ are excluded can be shown to be the same as that given by (48), that is to say

$$E\,[T_n] = \int_0^\infty T p_{T_n}(T)\,\mathrm{d}T = \frac{1}{\sqrt{1+\nu^2}}.$$

As was the case for wave amplitudes, a significant difference from this new approach is the prediction of a high density of small period waves. In the limit of zero T the p.d.f.s in (47) and (49) give

$$\lim_{T\to 0} p_T(T) = \frac{\nu^2}{\sqrt{1+\nu^2}},$$

$$\lim_{T\to 0} p_{T_n}(T) = \frac{\nu^2}{1+\sqrt{1+\nu^2}},$$

whereas, from (20), we have

$$\lim_{T\to 0} p_{\mathrm{LH}}(T) = 0.$$

Another significant difference is the behaviour in the limit of large T. From (47) and (49) we have

$$\lim_{T\to\infty} p_T(T) \sim \lim_{T\to\infty} p_{T_n}(T) \sim \frac{1}{T^3},$$

whereas, from (20) we have

$$\lim_{T\to\infty} p_{\mathrm{LH}}(T) \sim \frac{1}{T^2}.$$

Thus, as stated by Longuet-Higgins (1983), the mean period cannot be calculated from (20) since

$$\int_0^\infty T p_{\mathrm{LH}}(T)\,\mathrm{d}T \to \infty.$$

4. Comparisons of theory with simulated and measured ocean wave data

Here we present comparisons between the theoretical distributions derived in previous sections and measurements from numerically simulated data and from real ocean waves.

4.1. *Pierson–Moskowitz spectrum*

The simulated data was produced from a Pierson–Moskowitz spectrum using Matlab in conjunction with routines from the WAFO (Brodtkorb *et al.* 2000; WAFO Version 2.0.3 2000) toolbox. Specifically, a set of zero-mean Gaussian wave-trains, $\eta^*(t^*)$, were simulated from a Pierson–Moskowitz spectrum. The spectrum had a peak frequency $T_p^* = 15.754$s and a significant wave height $H_{m_0}^* = 10.358$m. The frequency range of the spectrum was 0 to 4rad s^{-1} divided into 2^{17} individual discrete frequency components. Ten such wave trains were produced in all, each having a duration of 400 min and a sampling frequency of 10 Hz. The total number of simulated zero down-crossing waves analysed was 21 261.

The formula used for the Pierson–Moskowitz spectrum was

$$S_{\mathrm{PM}}^*(\omega^*) = 0.11\, H_{m_0}^{*2}\left(\frac{2\pi}{T_{m_{01}}^*}\right)^4 \omega^{*-5} \exp\left[-0.44\left(\frac{2\pi}{T_{m_{01}}^*\omega^*}\right)^4\right],$$

where

$$T^*_{m_{01}} = 0.7713\, T^*_p,$$

and

$$T^*_{m_{01}} = \frac{2\pi m^*_0}{m^*_1}$$

From the simulated wave-train the zeroth, first and second-order spectral moments were calculated directly from the time series of surface elevation by

$$\left.\begin{aligned} m^*_0 &= \overline{\eta^{*2}}, \\ m^*_1 &= \overline{\eta^* \dot{\eta}^*}, \\ m^*_2 &= \overline{\dot{\eta}^{*2}}. \end{aligned}\right\} \tag{50}$$

From (7) the bandwidth was calculated to be $\nu = 0.4041$.

From $\eta^*(t^*)$ measurements were made to determine the values of all H^*_c, defined as the vertical distance from the mean sea level to the crest maximum, and all H^*_t, defined as the vertical distance from the mean sea level to the trough minimum. These height variables were then non-dimensionalized using the same scaling as that used in (8). Both H_c and H_t are always positive and the non-dimensionalized waveheight is given by their sum $H = H_t + H_c$. Also measured from $\eta^*(t^*)$ were the four quarter periods associated with each wave. These are defined as: T^*_1, the time between the first zero down-crossing and the trough minimum; T^*_2, the time between the trough minimum and zero up-crossing; T^*_3, the time between the zero up-crossing and the crest maximum; T^*_4, the time between the crest maximum and the second zero down-crossing. The quarter periods were then non-dimensionalized using the same scaling as that used in (9). The non-dimensionalized wave period is given by $T = T_1 + T_2 + T_3 + T_4$.

Since the simulated wave train was Gaussian it was symmetric about its zero-mean and so, in order to maximize the data available, all the values of crest heights and trough depths were combined into a single wave amplitude dataset, denoted by $h = \{H_c \bigcup H_t\}$. Also, as the theoretical p.d.f.s for wave period predict, strictly speaking, values for a local wave period, it is expected that the use of quarter periods measured from the simulations will compare better with the theoretical predictions than measurements of whole zero down-crossing periods. The measured values of quarter periods where multiplied by 4 (to scale them to full period size) and combined to give a larger statistical sample. This gives the dataset of measured quarter periods as $T_q = \{4T_1 \bigcup 4T_2 \bigcup 4T_3 \bigcup 4T_4\}$.

Figure 4 shows comparisons between the simulated data so produced and the theories. Shown side-by-side in this figure are predictions and estimates for the various densities (*a*), and the corresponding probability–probability (p–p) plots (*b*). The estimated densities were produced using standard kernel density estimation with Gaussian kernels (Silverman 1986). The uppermost pair of plots show comparisons between the measured wave amplitude, h, and theory. It is clear that for the simulated data the density is non-zero as $h \to 0$. This is predicted by both $p_h(h)$ and $p_{h_n}(h)$ given by (30) and (35). It can also be seen that $p_{\mathrm{LH}}(h)$ given by (21) provides a poorer fit to the data than the basic Rayleigh density $R(h)$, the former being lower for small waves and higher for large waves. The associated p–p plot confirms that $p_{\mathrm{LH}}(h)$ provides a worse fit than $R(h)$, and that $p_h(h)$ and $p_{h_n}(h)$ fit the data better for small wave amplitudes. It appears that the best fit for large h is given by $R(h)$.

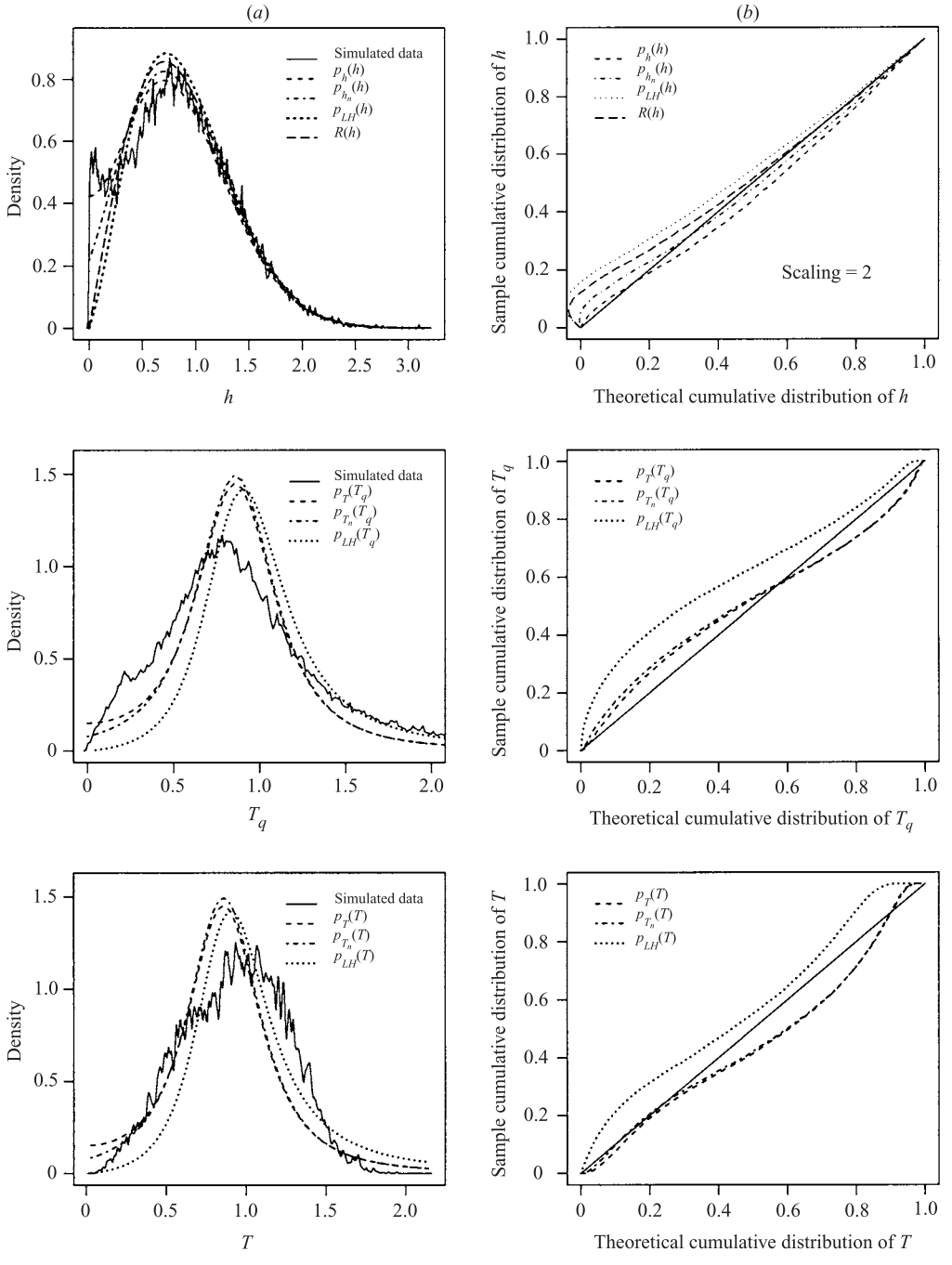

FIGURE 4. (*a*) comparisons between theoretical predictions (with $\nu = 0.4041$) and measurements of the probability densities of wave amplitude, quarter period and whole period made from simulated wave data. (*b*) corresponding probability–probability plots. (Note that the scaling in the top right plot is set to 2 which means that the horizontal distance between the line of equality and data has been doubled in improve clarity of this figure.)

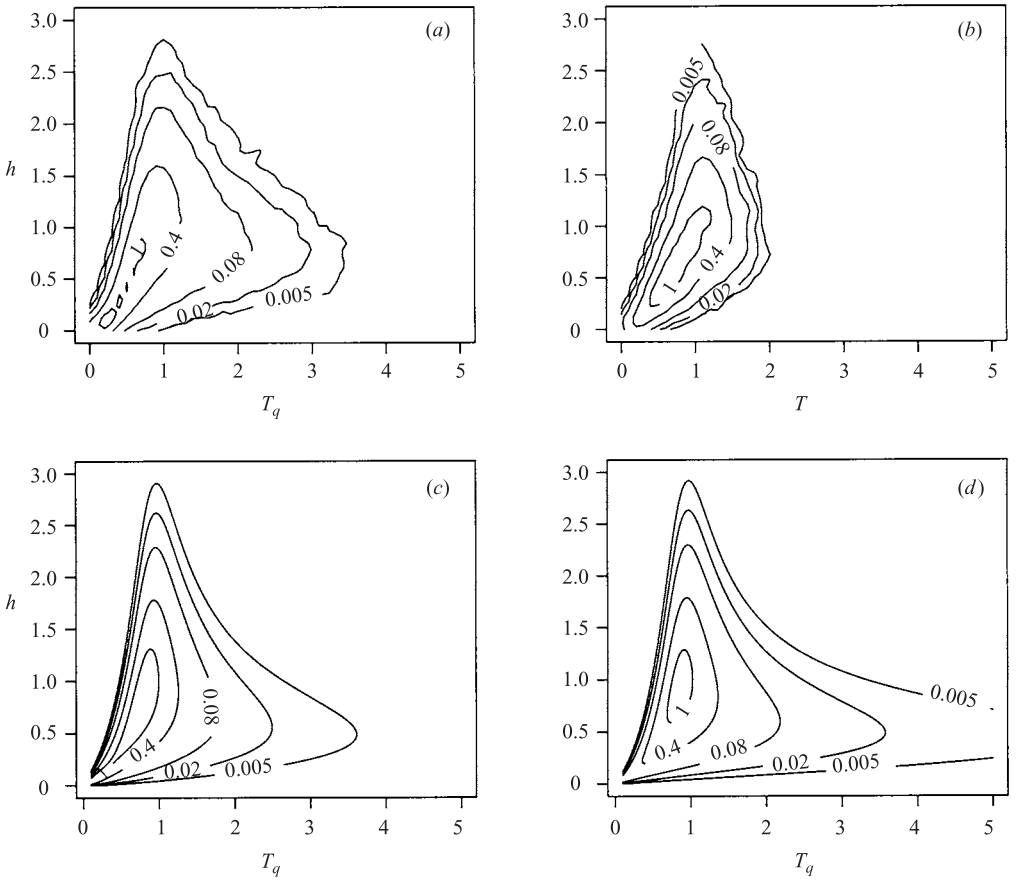

FIGURE 5. (*a*) joint density for wave amplitude and quarter period measured from simulated data (specifically: joint density of combined dataset consisting to the following data pairs $(4T_1, H_t)$, $(4T_2, H_t)$, $(4T_3, H_c)$ and $(4T_4, H_c)$). (*b*) joint density for wave amplitude and whole period measured from simulated data (specifically: joint density of dataset consisting to the following data pairs $(T, (H_t + H_c)/2)$). (*c*) Corresponding prediction of $p_{h_n}(h, T_q)$ from (46) with $\nu = 0.4041$. (*d*) Corresponding prediction of $p_{\mathrm{LH}}(h, T_q)$ from (12) with $\nu = 0.4041$.

The middle and lower pairs of plots in figure 4 show the results for quarter periods, T_q, and whole periods, T, respectively. None of the theories make particularly good predictions of the simulation results for period or quarter period, although, from the p–p plots for quarter periods it seems that the predictions of (47) and (49) are better than that of (20).

Figure 5 shows the joint p.d.f.s for wave period and amplitude. For all four plots the contour lines are set at the same levels. Figure 5(*a*) shows the joint p.d.f. of data pairs $(h, T_q) = \{(H_t, 4T_1) \bigcup (H_t, 4T_2) \bigcup (H_c, 4T_3) \bigcup (H_c, 4T_4)\}$ obtained from the simulated data. Figure 5(*b*) shows the same as (*a*) but for whole periods instead of scaled quarter periods, that is to say, $(h, T) = \{(H_t, T) \bigcup (H_c, T)\}$. Figure 5(*c*) shows the joint density $p_{h_n}(h, T_q)$, predicted by (46) with $\nu = 0.4041$, and figure 5(*d*) shows $p_{\mathrm{LH}}(h, T_q)$, predicted by (12) with $\nu = 0.4041$. With the contour levels used here, the plots of $p_h(h, T_q)$ and $p_{h_n}(h, T_q)$ are found to be almost indistinguishable. However, $p_{h_n}(h, T_q)$ is presented in figure 5 (and later in figure 7) as it provides a slightly better fit to the data as it extends slightly less into the higher wave-period region.

Storm ID	Number of 20 minute records	Total number of waves in unfiltered data	$\overline{H^*}$ (m)	max (H^*) (m)	$\overline{T^*}$ (s)	max(T^*) (s)
23	177	25 632	5.437	21.94	10.61	21.65
25	111	17 348	4.796	15.88	9.89	17.91
124	173	21 942	5.847	21.14	10.85	28.71
195	201	30 084	4.857	18.72	10.28	20.80
Combined	662	95 006	5.266	21.94	10.45	28.71

TABLE 1. Summary of real storm data.

It is clear from these four plots that $p_{h_n}(h, T_q)$ shows fairly good agreement with the scaled quarter periods, whereas $p_{\mathrm{LH}}(h, T_q)$ suffers from a higher density of large wave periods, but neither make particularly good predictions when whole periods are considered.

4.2. *Real ocean wave data*

The real wave data used in this study were collected continuously over the durations of four separate severe storms in the northern North Sea. The data, totalling nearly 221 hours, are divided into 662 surface elevation time series each of 20-min duration. The four storms are of varying bandwidth but all are essentially uni-modal wind-driven seas without significant swell. The mean of the bandwidth was $\nu = 0.68$, with a standard deviation of 0.16. The instrumentation used for data collection were laser altimeters sampling at 5Hz and mounted on the underside of the deck platform of the North Alwyn fixed steel jacket oil and gas platform, operated by TotalFinaElf.

Note that the wave records are wholly unfiltered, not being smoothed by any means other than that arising from the finite sampling rate of the measurement instruments. All zero-downcrossing waves were identified in these records and measured from each were: crest amplitude, trough amplitude, quarter periods and whole periods. A more detailed summary of the data from these storms, identified as Storms 23, 25, 124 and 195, is given in Table 1.

The real data is not statistically stationary and so it was not sensible to use any single values for m_0^* and m_1^* to non-dimensionalize the whole dataset. Instead, the data were divided into sections of 20-min duration and sets of values of H_c^*, H_t^*, and T_i^*, $i = 1, 2, 3, 4$, were measured from each. Local values of m_0^* or m_1^* where also calculated from each 20-min section and these where then used to non-dimensionalize the appropriate amplitudes and periods according to (8) and (9).

In the real data, due to its highly nonlinear nature, the distributions of wave crest height and trough depth differ markedly. We therefore use $h = (H_t + H_c)/2$ to representative wave amplitude measured from the real data. Figure 6 shows comparisons of the non-dimensionalized real wave data with the theories. The mean value of $\nu = 0.68$ was used in the theoretical predictions. Shown side-by-side are predictions for the various densities (a) and the corresponding p–p plots (b). The uppermost pair show comparisons of measured wave amplitudes with theory. For the real data it is again clear that the density is non-zero as $h \to 0$ and, as with the simulated data, this is predicted by both $p_h(h)$ and $p_{h_n}(h)$ but not by $p_{\mathrm{LH}}(h)$ or $R(h)$. Again, the prediction of $p_{\mathrm{LH}}(h)$ fits least well and is worse than that of $R(h)$. The associated p–p plots suggest that $p_{\mathrm{LH}}(h)$ gives the poorest fit and $p_{h_n}(h)$ gives the best.

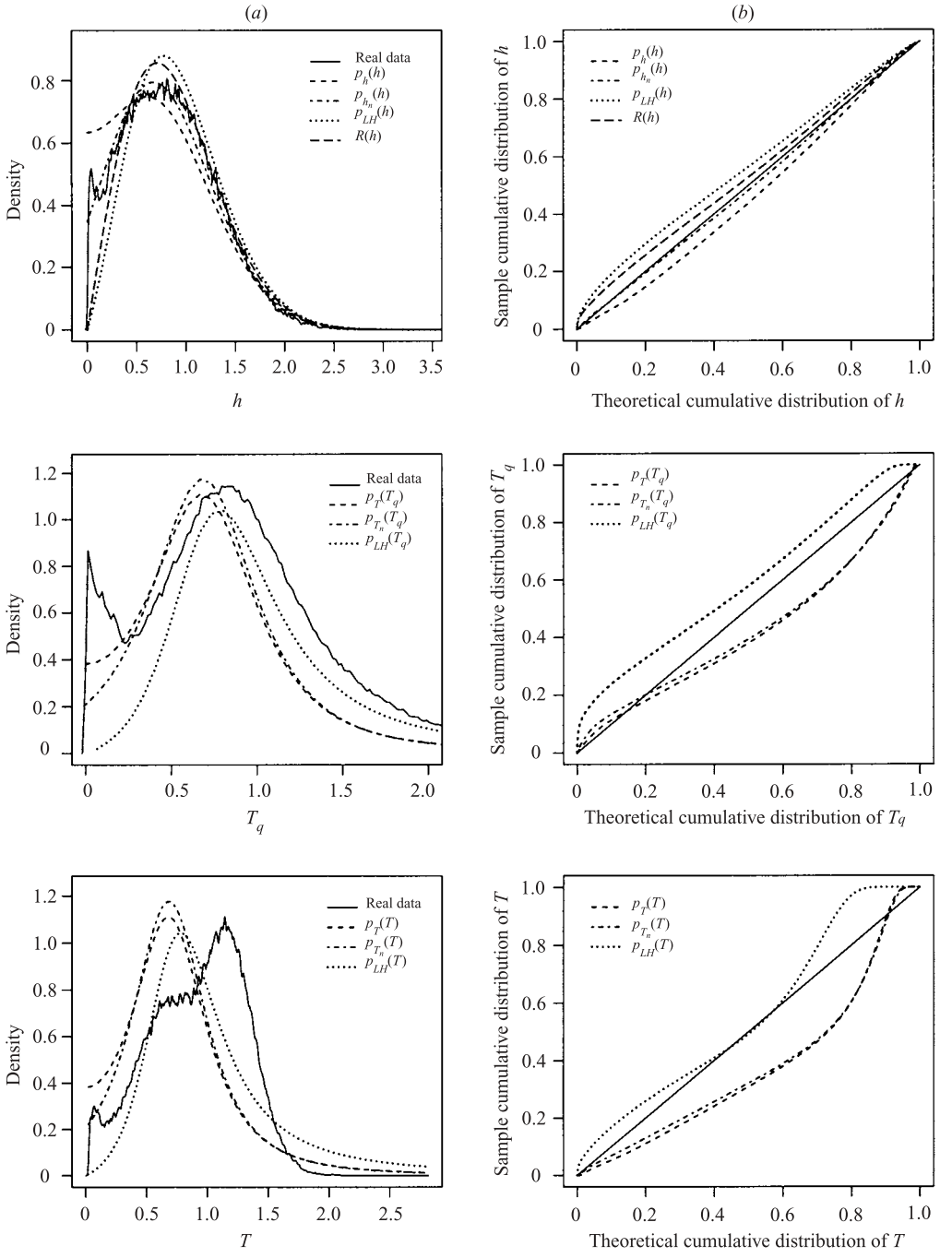

FIGURE 6. (*a*) comparisons between theoretical predictions (with $\nu = 0.63$) and measurements of the probability densities of wave amplitude, quarter period and whole period made from real ocean wave data. (*b*) corresponding probability–probability plots.

The middle and lower pairs of plots in figure 6 show the real data results for quarter periods and whole periods respectively. The measured quarter periods were calculated in the same way as they were for the simulation data above. As with the

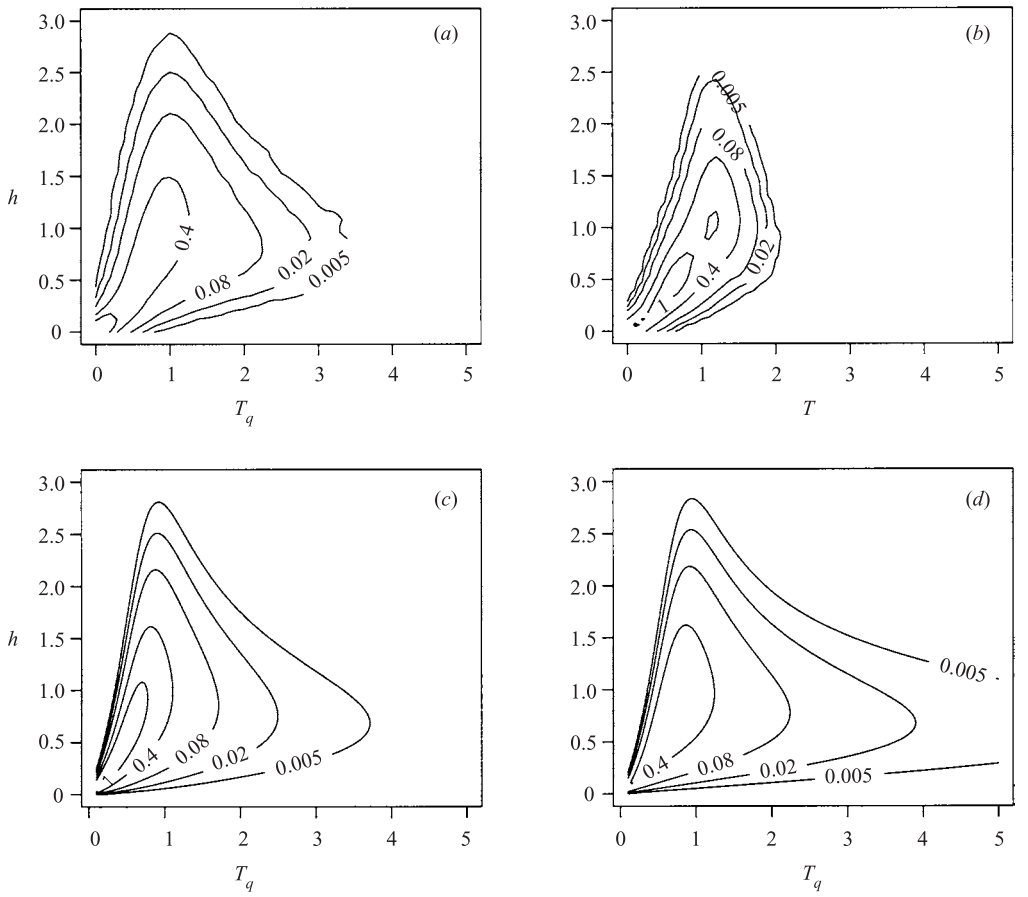

FIGURE 7. (*a*) joint density for wave amplitude and quarter period measured from real ocean wave data (specifically: joint density of combined dataset consisting to the following data pairs $(4T_1, H_t)$, $(4T_2, H_t)$, $(4T_3, H_c)$ and $(4T_4, H_c)$). (*b*) joint density for wave amplitude and whole period measured from real ocean wave data (specifically: joint density of dataset consisting to the following data pairs $(T, (H_t + H_c)/2)$). (*c*) Corresponding prediction of $p_{h_n}(h, T_q)$ from (46) with $\nu = 0.63$. (*d*) Corresponding prediction of $p_{\mathrm{LH}}(h, T_q)$ from (12) with $\nu = 0.63$.

simulated data, none of the theories makes particularly good predictions for periods or quarter periods.

Figure 7 shows the joint p.d.f.s for wave periods and amplitudes measured from the real data. For all four plots the contour lines are set at the same levels as in figure 5.

Figure 7(*a*) shows the joint p.d.f. of data pairs $(h, T_q) = \{(H_t, 4T_1) \bigcup (H_t, 4T_2) \bigcup (H_c, 4T_3) \bigcup (H_c, 4T_4)\}$. Figure 7(*b*) shows the same as (*a*) but for whole periods instead of scaled quarter periods, that is to say, $(h, T) = \{(H_t, T) \bigcup (H_c, T)\}$. Figure 7(*c*) shows the joint density $p_{h_n}(h, T)$ and figure 7(*d*) shows $p_{\mathrm{LH}}(h, T)$, both with $\nu = 0.68$.

It is clear from these four plots that $p_{h_n}(h, T)$ shows fairly good agreement with the scaled quarter periods, whereas $p_{\mathrm{LH}}(h, T)$ suffers from a higher density of large wave periods, but neither make particularly good predictions when whole periods are considered.

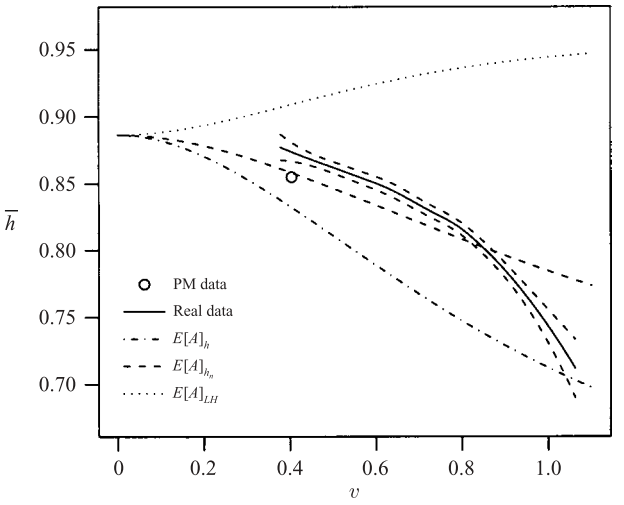

FIGURE 8. Comparisons of the predictions of $E[A](\nu)$ from equations (36), (37) and (38) with results obtained from simulated and real data.

The unrealistically high density of large wave periods predicted by $p_{\mathrm{LH}}(h, T)$ is even more pronounced than for the simulated data because of the larger value of ν in this case. The predictions for the p.d.f. of wave amplitude and whole wave period are poor for all theories but, as with the simulated data, it is suggested that $p_{h_n}(h, T)$ fits better than $p_{\mathrm{LH}}(h, T)$.

A further comparison between the various models was made by comparing the values of $E[A]$ predicted by (36), (37) and (38) with the values obtained from the real and simulated data. These comparisons are shown in figure 8. The simulated data provide one mean value of h at $\nu = 0.4041$ and, as can be seen in the figure, this point falls very close the to prediction $E[A]_{h_n}$ of (37). From the real data, values of ν and mean values of h where calculated from each of the locally stationary 20-min records. As these raw values show significant scatter, the local regression procedure *locfit* (Loader, 1999; locfit is implemented in the *R* and *S-Plus* statistical analysis packages) was used to find a non-parametric estimate of the relationship between $\overline{h}$ with ν in the real data. This local regression, along with its 95% confidence intervals, is also plotted in figure 8; it closely follows the predictions of (37) for moderately broad bandwidths.

5. Discussion

The results presented here apply to any narrowband linear random process, as do the earlier results of Longuet-Higgins (1983). The results presented here, however, account for the correlation between wave period and amplitude by adjusting the frequency at which the wave envelope is sampled. In this respect, they are an improvement on the results of Longuet-Higgins (1983) where these effects are not included. However, none of the new results provides exact distributions when the bandwidth is finite.

The effect on the wave amplitude of the small loops in the phase plane that do not encircle the origin is not accounted for in the present theory. The wave amplitudes associated with these loops may be about half the loop radius, but we used the distance from the origin to the wave envelope in the $(\eta, \hat{\eta})$ phase plane to represent

this wave amplitude. Thus, as illustrated in figure 3, in the region of these loops the wave amplitude is not the same as the envelope amplitude. Because of this, even if we sample the wave envelope at the actual zero down-crossing frequency as opposed to the local frequency, the resulting distribution would not be the distribution of wave amplitudes. Clearly, for a general broadband approach one cannot use the wave envelope to represent the wave amplitude.

Also, it is clear that Longuet-Higgins' assumption of a local wave frequency that is always positive and slowly varying is not realizable in practice, even for narrow bandwidths. For any finite bandwidth there will be some regions where $\dot{\phi}$ is negative, and within less than a wave cycle it will return to typical positive values. There will be local maxima corresponding to such regions and so approaches to deriving the Rayleigh distribution for amplitudes based on the assumption of one maxima per zero up crossing (for example, Newland 1984), will also be in error for any finite bandwidth. Equations (18) and (27) are useful in this context as they give a bandwidth related measure of the error involved in the assumption of non-negative $\dot{\phi}$. For much real data, however, the region in which the local wave frequency is negative is small and so the number of small loops, and the related error, will be comparatively small. For example, from (27) the proportion of waves that are associated with $\dot{\phi} < 0$ is about 3.6% for $\nu = 0.4$, 7.1% for $\nu = 0.6$ and 11% for $\nu = 0.8$. One would expect, therefore, the models presented here to be best suited to the moderate bandwidths that are frequently observed in Metocean data.

When compared with measurements of wave amplitude from simulated data and waveheight from real data we find that, overall, the new distributions presented here provide better fits than either the Rayleigh distribution or the bandwidth dependent distribution of Longuet-Higgins (1983). Clearly, sampling from the wave envelope at the local wave frequency leads to an improved distribution for wave amplitudes. Likewise, the new joint distributions for wave amplitude and period provide a somewhat better fit to both the simulated and real wave data than does the joint distribution of Longuet-Higgins (1983).

In the limit of the bandwidth tending to zero our marginal distributions, $p_h(A)$ and $p_{h_n}(A)$, and that of Longuet-Higgins (1983), $p_{\mathrm{LH}}(A)$, all converge to the Rayleigh distribution. But it seems from our analysis that as ν moves away from zero the density $p_{\mathrm{LH}}(A)$ changes shape in a way that is the opposite to the shape change measured from real or simulated data. This is seen in the upper pairs of plots in both figures 4 and 6. Compared to the Rayleigh density the density of the measured data of finite bandwidth shows an increase for small wave amplitudes and a decrease for wave amplitudes around the mode of the distribution. This is predicted by the both $p_h(A)$ and $p_{h_n}(A)$; but contrasts, however, with the prediction of $p_{\mathrm{LH}}(A)$ which shows a decrease for small amplitudes and an increase for amplitudes around the mode. These differences between the distributions lead to differences in predictions of mean waveheight, $E[A]$, as a function of bandwidth. The different predictions for $E[A]$ are plotted in figure 8. They clearly show that $p_{\mathrm{LH}}(A)$ gives an increases in $E[A]$ with increasing ν, whereas $p_h(A)$ and $p_{h_n}(A)$, and the measurements from the simulated and real data, all show $E[A]$ decreasing with increasing ν. Our samples of real wind-driven storm waves from the northern North Sea are found to have quite a large range of bandwidths: from $\nu \approx 0.4$ to 1.06. At $\nu \approx 0.8$ the Rayleigh distribution over predicts the mean waveheight of this sample of real data by about 7% and the Longuet-Higgins distribution over predicts by about 12%. Clearly, $p_{h_n}(A)$ provides the best fit, whereas $p_h(A)$ significantly under predicts the mean height and so is not recommended for practical use.

As seen from the p.d.f. and p–p plots in figure 4, for the simulated data the measured density of wave amplitude is fitted best by different marginal density functions in different amplitude regions: for the region below the mode the best fit is given by $p_h(A)$; for the region above the mode, by $p_{h_n}(A)$; and for the region of largest wave amplitudes, by the Rayleigh density. Over the whole probability range, however, the best fit is provided by $p_{h_n}(A)$. As can be seen in figure 6, the best overall fit for the real data is also provided by $p_{h_n}(A)$.

As seen in figures 5 and 7, the new joint probability models seem to predict the amplitudes and quarter periods for real data tolerably well, perhaps indicating that $\pi/2|\dot{\phi}|$ is a good surrogate for the quarter period. However, this is not well borne out by the marginal density plots in figures 4 and 6. One is left to conclude that $\dot{\phi}$ varies too rapidly over the time interval of a real wave to be considered as a good surrogate for time period, and other approaches need to be considered. At least the marginal distributions presented here, unlike that of Longuet-Higgins, can be integrated directly to give the correct mean periods.

Whilst the joint probability and marginal distributions presented here are developed from random linear theory, they may be expected to give reasonable predictions of waveheight up to second order. This is because, in the theory of second-order surface gravity waves, the increased amplitude of the crest is exactly offset by the reduced amplitude of the trough. The general predictive success of the upper tail of the Rayleigh distribution may be similarly attributed. The real wave data presented here are from severe storms and are significantly nonlinear, yet, as stated previously, it is clear from the p–p plots in figure 6 that the new marginal p.d.f. for waveheight, $p_{h_n}(h)$, gives quite a good overall fit to these data.

Interestingly, the density $p_{\mathrm{LH}}(h)$ predicts a higher occurrence of large waves than Rayleigh as the bandwidth increases from zero, whereas the new densities $p_h(h)$ and $p_{h_n}(h)$ predict a lower occurrence. The differences in the bandwidth-dependent predictions for the extreme value distributions between the various models only becomes significant for very high values of v. For $v = 1$ the difference in the positions of the modes of the extreme value densities is just 6%. Resulting variations in estimates of extreme loading would likewise be affected: for Morison inertia, which is linear in H, the loadings differ by about 6%; for Morison drag, which is quadratic in H, the difference will be around 12%. Furthermore, as there are more small waves and less waves of moderate height in the proposed distributions, estimates of wave-induced fatigue damage would be affected to a greater degree.

Unlike the joint p.d.f. of Longuet-Higgins, the new models do not predict extremely long wave periods at around the average amplitude. They do, however, predict periods longer than those observed in either the real or simulated data. This is of practical significance when considering low frequency motions associated with mooring systems of ships and floating offshore platforms.

6. Conclusions

Joint and marginal p.d.f.s have been derived for envelope amplitudes and local zero-crossing wave periods of a finite bandwidth Gaussian process sampled at the local frequency. In attempting to obtaining a p.d.f. for wave amplitude from envelope amplitude the logic of sampling the wave envelope at the local wave frequency is undeniably an improvement over the constant sampling rate implicit in the Rayleigh and Longuet-Higgins (1983) distributions. Thus, for moderately broadband spectra

the new joint and marginal p.d.f.s are shown to provide improved fits. The density $p_{h_n}(h)$ was found to give the best overall fit for both the p.d.f. of wave amplitudes measured from simulated data and that of wave heights measured from real data. Similarly, the density $p_{h_n}(h, T)$ was found to give the best overall fit for the joint p.d.f. of wave amplitudes and scaled quarter periods measured from simulated data, and also for the joint p.d.f of wave heights and scaled quarter periods measured from real data. The joint density $p_{h_n}(h, T)$ was found to be a definite improvement over the joint density, $p_{\mathrm{LH}}(h, T)$, of Longuet-Higgins (1983); the most important difference being the better fit in the region of the high wave-period tail. This was found to be the case for both simulated and measured wave data.

However, the approach of obtaining the p.d.f. of wave amplitude from that of the wave envelope has limitations which have been discussed. These limitations are particularly evident when comparing the measured p.d.f. of scaled zero-crossing quarter periods with that of local zero-crossing wave periods predicted by the models. The p.d.f. of local zero-crossing wave period does not appear to provide any obvious improvement over that of Longuet-Higgins (1983). The limitations of the envelope approach are also evident in the comparisons which show that none of the models gave good predictions of the measured joint p.d.f.s of waveheight and total period.

The authors are grateful to Dr Stan Zachary for helpful discussions, to the UK Engineering & Physical Sciences Research Council (EPSRC grant no. GR/M15439) for funding of Paul Stansell and to TotalFinaElf who provide support for Julian Wolfram. The views expressed, however, are those of the authors' and do not necessarily coincide with those of either of the above-mentioned organizations.

Appendix A. The equivalence of equation (4) and Longuet-Higgins' (1975) equation (A 16)

Equation (4) is identical to Longuet-Higgins' (1975) equation (A 16), rewritten here for convenience as

$$p_{A^*\dot{\phi}^*_{\mathrm{LH}}}(A^*, \dot{\phi}^*_{\mathrm{LH}}) = \frac{A^{*2}}{\sqrt{2\pi\mu^*_2\mu^*_0}} \exp\left(\frac{-A^{*2}}{2\mu^*_0}\right) \exp\left(\frac{-A^{*2}\dot{\phi}^{*2}_{\mathrm{LH}}}{2\mu^*_2}\right), \tag{A 1}$$

where

$$\dot{\phi}^*_{\mathrm{LH}} = \dot{\phi}^* - \overline{\omega^*},$$
$$\overline{\omega^*} = \frac{m^*_1}{m^*_0},$$

and where the central spectral moments are defined by

$$\mu^*_n = \int (\omega^* - \overline{\omega^*})^n S^*(\omega^*)\,\mathrm{d}\omega^*, \qquad n = 0, 1, 2, \ldots,$$

giving

$$\mu^*_0 = m^*_0,$$
$$\mu^*_2 = m^*_2 - \frac{m^{*2}_1}{m^*_0}.$$

See also the discussion around (11.91) in Ochi (1990, p. 297).

REFERENCES

Arhan, M. K., Cavanié, A. & Ezraty, R. 1976 Etude théorique et expérimentale de la relation hauteur-période des vagues de tempete. ifp 24191. *Technical Report*. Centre National pour l'Exploitation des Oceans.

Brodtkorb, P. A., Johannesson, P., Lindgren, G., Rychlik, I., Rydén, J. & Sjö, E. 2000 Wafo – a matlab toolbox for analysis of random waves and loads. In *Proc. 10th Intl Offshore and Polar Engineering Conference*, vol. 3, pp. 343–350. The International Society of Offshore and Polar Engineers.

Cartwright, D. E. & Longuet-Higgins, M. S. 1956 The statistical distribution of the maxima of a random function. *Proc. R. Soc. Lond.* A **237**, 212–232.

Cavanié, A., Arhan, M. & Ezraty, R. 1976 A statistical relationship between individual heights and periods of storm waves. In *Behaviour of Offshore Structures* (*BOSS*) *'76*, pp. 354–363.

Loader, C. 1999 *Local Regression and Likelihood*. Springer.

Longuet-Higgins, M. S. 1975 On the joint distribution of the periods and amplitudes of sea waves. *J. Geophys. Res.* **80**(18), 2688–2694.

Longuet-Higgins, M. S. 1983 On the joint distribution of wave periods and amplitudes in a random wave field. *Proc. R. Soc. Lond.* A **389**, 241–258.

Newland, D. E. 1984 *An Introduction to Random Vibrations and Spectral Analysis* (2nd edn). Longman Group Ltd.

Ochi, M. K. 1990 *Applied Probability and Stochastic Processes in Engineering and Physical Sciences*. Wiley.

Ochi, M. K. 1998 *Ocean Waves: The Stochastic Approach*. Cambridge University Press.

Podgórski, K., Rychlik, I. & Machado, U. E. B. 2000 Exact distributions for apparent wave in irregular seas. *Ocean Engng* **27**, 979–1016.

Rice, S. O. 1944 Mathematical analysis of random noise. *Bell System Tech. J.* **23**, 282–332.

Rice, S. O. 1945 Mathematical analysis of random noise. *Bell System Tech. J.* **24**, 46–108.

Silverman, B. W. (ed.) 1986 *Density Estimation for Statistics and Data Analysis*. Chapman and Hall.

WAFO 2000 Version 2.0.3. URL http://www.maths.lth.se/matstat/wafo/.