Available online at www.sciencedirect.com
ScienceDirect
Ocean Engineering 34 (2007) 1631–1649
OCEAN ENGINEERING
www.elsevier.com/locate/oceaneng

Wave-height distributions and nonlinear effects

M. Aziz Tayfun[a,*], Francesco Fedele[b]

[a]Civil Engineering Department, College of Engineering and Petroleum, Kuwait University, P.O. Box 5969, Safat 13060, Kuwait
[b]School of Civil & Environmental Engineering, Georgia Institute of Technology, 210 Technology Circle, Savannah, GA 31407, USA

Received 19 May 2006; accepted 1 November 2006
Available online 12 February 2007

Abstract

Theoretical distributions proposed for describing the crest-to-trough heights of linear waves are reviewed briefly. To explore the effects of nonlinearities, these are generalized to second-order waves, utilizing quasi-deterministic results on the expected shape of large waves. The efficacy of Gram–Charlier models in describing the effects of third-order nonlinearities on the distributions of wave heights, crests and troughs are examined in detail. All models and a fifth-order Stokes–Rayleigh type model recently proposed are compared with linear and nonlinear waves simulated from the JONSWAP spectrum representative of long-crested extreme seas, and also with oceanic data gathered in the North Sea. Uncertainties arising from the variability of probability estimates derived from sample populations of limited size are considered. Ultimately, the comparisons show that nonlinearities do not have any discernable effect on the crest-to-trough heights of oceanic waves. Most of the linear models considered yield similar and reasonable predictions of the observed data trends. Gram–Charlier type distributions seem neither effective nor particularly useful in describing the statistics of large wave heights or crests under oceanic conditions. However, they do surprisingly well in predicting unusually large wave heights and crests observed in some 2D wave-flume experiments and 3D numerical simulations of long-crested narrow-band random waves.

Keywords: Cumulants; Extreme values; Gram–Charlier series; Nonlinear waves; Probability distributions; Sampling variability; Wave crests; Wave envelopes and phases; Wave heights; Wave interactions; Wave troughs

1. Introduction

Current interest in the mechanics and statistics of large waves necessitates a re-examination of various theoretical forms for describing the distributions of wave heights and crests (see e.g. Haver and Andersen, 2000; Stansell, 2004; Walker et al., 2004; Fedele and Arena, 2003, 2005; Janssen, 2005; Guedes Soares and Pascoal, 2005; Socquet-Juglard, 2005; Socquet-Juglard et al., 2005; Fedele, 2006; Onorato et al., 2004, 2005, 2006; Tayfun, 2006). Over the years, a variety of numerical, empirical and analytic wave height and crest models have been proposed. Most of these have been reviewed and compared previously (Forristall, 1984, 2000; Tayfun, 1990a, 2006). This study will first focus on an initial short list of just three analytic crest-to-trough wave-height models due to Tayfun (1981, 1990a), Naess (1985) and Boccotti (1989). Of these, Naess' model (N) has received wide popularity. This is justifiably so due to its simple functional form although previous comparisons (Tayfun, 1990a) and those to be presented here indicate that N underestimates the observed wave heights slightly. Tayfun's model (T) for large wave heights is consistently more accurate than N. But, it is totally ignored apparently because its functional form is more complex and thus less amenable to analytical and/or practical applications than N. However, Boccotti's model (B) is just about as simple as N, but has not received the attention it probably deserves either. Thus, one of the present objectives is to review these models briefly, including two obvious and simple approximations that follow from T. The variability of probability estimates derived from sample populations of limited size and its relevance in interpreting the nature of exceptionally large waves are also considered. Subsequently, all the linear

*Corresponding author. Tel.: +965 622 7814; fax: +965 481 7524
E-mail address: aziztayfun@usa.net (M.A. Tayfun).

0029-8018/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.oceaneng.2006.11.006

models are compared with simulated linear waves and oceanic data gathered during two severe storms in the northern North Sea.

Since large waves are nonlinear, the linear models considered are subsequently modified to include first the effects of second-order nonlinearities, using the extension of Boccotti's (1989, 2000) linear quasi-deterministic theory to second-order waves by Fedele and Arena (2005). The resulting second-order models, a fifth-order Stokes–Rayleigh model recently proposed by Dawson (2004) and a second-order version that follows from it are then compared with simulated nonlinear waves representative of extreme seas and also with the same North Sea data.

Nearly all past oceanic observations as well as the present measurements gathered during two exceptionally severe storms clearly indicate that wave heights are not affected by nonlinearities. However, some recent analyses by Janssen (2003, 2005), Mori and Janssen (2006), and Onorato et al. (2004, 2005, 2006) based on the nonlinear Schrödinger (NLS) equation and wave-flume experiments show that occurrences of so-called 'freak' waves and wave heights considerably larger than those typically predicted with the conventional probability laws can be explained in terms of third-order nonlinear interactions. Mori and Janssen (2006) also contend that the distribution of such wave heights is described by a modified form of the Gram–Charlier (GC) series dependent solely on the kurtosis of surface elevations. Dysthe et al. (2005), Socquet-Juglard (2005) and Socquet-Juglard et al. (2005) explore the effects of third-order nonlinearities further with a series of intriguing 3D numerical simulations based on the Dysthe equation, a higher-order form of the NLS equation modified for directional waves with large steepness and broader spectra (Dysthe, 1979; Trulsen and Dysthe, 1996; Trulsen and Stansberg, 2001; Dysthe et al., 2003). These confirm that an increased density of unusually large waves does in fact appear in nearly 2D or long-crested waves initially characterized by relatively narrow-band spectra. This tends to occur in the absence of dissipation and surface stresses, and as spectra change relatively rapidly due to modulational instabilities toward an equilibrium range proportional to ω^{-4} over high frequencies. However, the simulations representative of the more realistic short-crested waves also show clearly that similar spectral changes do not cause any discernable aberrations, and the statistical characteristics of the free surface elevations, wave heights and crests are described surprisingly well with the presently available linear and second-order probabilistic models.

The possibility that third-order nonlinearities can modify the statistical structure of surface waves dramatically under certain conditions also necessitates a re-examination of the efficacy of GC type approximations in describing the distributions of large wave heights, crests and troughs. Thus, GC models appropriate to third-order waves are considered, drawing on the formulations devised previously in Tayfun and Lo (1990) and Tayfun (1994, 2006), and extending these to wave crests and troughs. All the resulting theoretical expressions are then compared with the North Sea data, and also with some 3D simulations from Socquet-Juglard et al. (2005) and 2D wave-flume data from Onorato et al. (2004, 2005, 2006).

2. Linear waves

2.1. Definitions

Consider linear deep-water waves, and let S represent the surface spectral density as a function of angular frequency ω. Denoting the ordinary moments of S by m_j $(j = 0, 1, \ldots)$, $\sigma \equiv m_0^{1/2}$ corresponds to the root-mean-square (r.m.s.) surface elevation, and $\omega_m = m_1/m_0$, $T_m = 2\pi/\omega_m$, $\nu = \sqrt{(m_0 m_2/m_1^2) - 1}$ and $\omega_0 = \sqrt{m_2/m_0} = \omega_m\sqrt{1+\nu^2}$ define the spectral average frequency, associated wave period, spectral bandwidth and the mean zero up-crossing frequency, respectively. Further, the r.m.s. surface gradient is given by $\mu \equiv m_4^{1/2}/g$ with $g \approx 9.8\ \mathrm{m/s^2}$. In general, $\mu^2 \ll 1$ since $\mu \approx O(10^{-1})$ at most.

Next, let η and $\dot{\eta} = \partial\eta/\partial t$ represent, respectively, the surface elevation from the mean sea level and its time derivative at a fixed point as a function of time t. Scaling η with σ and $\dot{\eta}$ with $m_2^{1/2} = \sigma\omega_0$ allows their normalized autocorrelation kernels to be expressed as

$$\rho(\tau) = \langle \eta(t)\eta(t+\tau)\rangle = m_0^{-1}\int_0^\infty S(\omega)\cos(\omega\tau)\,\mathrm{d}\omega, \tag{1}$$

$$\rho''(\tau) = \langle \dot{\eta}(t)\dot{\eta}(t+\tau)\rangle = -m_2^{-1}\int_0^\infty \omega^2 S(\omega)\cos(\omega\tau)\,\mathrm{d}\omega, \tag{2}$$

where $\rho'' = \mathrm{d}^2\rho/\mathrm{d}\tau^2$. The upper (+) and lower (−) envelopes of ρ are then given by

$$\pm r(\tau) = \pm\sqrt{\rho^2 + \hat{\rho}^2}, \tag{3}$$

where

$$\hat{\rho}(\tau) = \langle \eta(t)\hat{\eta}(t+\tau)\rangle = m_0^{-1}\int_0^\infty S(\omega)\sin(\omega\tau)\,\mathrm{d}\omega, \tag{4}$$

and $\hat{\eta}$, $\hat{\rho} \equiv$ Hilbert transforms of η and ρ, respectively.

Now, define the dimensionless parameters

$$r_m = r(\tau_m), \quad a = \rho(\tau^*), \quad b = \rho''(\tau^*), \tag{5}$$

where $\tau_m = T_m/2$ for simplicity, and $\tau^* \equiv$ the time lag at which the first minimum of ρ occurs (Boccotti, 1989, 2000). These parameters can all be estimated either from a time series of surface elevations or, somewhat more accurately and just as simply from the associated frequency spectrum. For example, consider

$$S(\omega) = \frac{m_0}{\omega_p}u^{-4}\exp(-1.25u^{-4})\gamma^{g(u)}, \tag{6}$$

where $\omega_p \equiv$ spectral-peak frequency, $\gamma \equiv$ peak-enhancement coefficient, $u = \omega/\omega_p$, and $g(u) \equiv$ standard JONSWAP

exponent function of γ. The spectral shape above represents a slightly modified version of the well-known JONSWAP form so as to mimic the u^{-4} type asymptotic behavior of oceanic spectra for $u \gg 1$. The normalized kernels ρ, ρ'', envelopes $\pm r$, parameters r_m, a, b, and the characteristic times τ^* and τ_m corresponding to Eq. (6) with $m_0 \approx 10.72\ \mathrm{m}^2$, $\omega_p \approx 0.467\ \mathrm{rad/s}$ ($T_p \equiv 2\pi/\omega_p \approx 13.44$ s) and $\gamma = 3.3$ are illustrated in Fig. 1.

2.2. Statistics

The exceedance distribution functions (EDF) describing the linear crest-to-trough wave height H_1 scaled with σ have the general form

$$E = \Pr\{H_1/\sigma > h_1\} = c_0 f(h_1) \exp(-c_1 h_1^2). \tag{7}$$

The corresponding probability density functions (PDF) follow simply from $-\mathrm{d}E/\mathrm{d}h_1$. The parameters c_0, c_1 and the function f are all given in Table 1 for N, B, T, T1, T2 and the conventional Rayleigh (R) law.

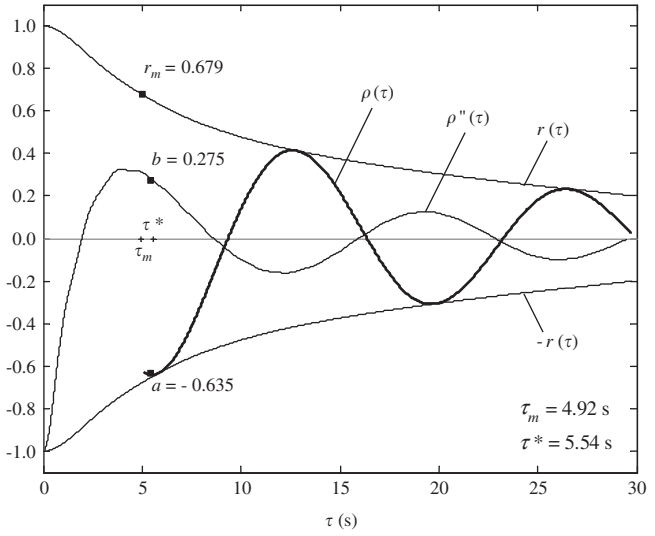

Fig. 1. Principal parameters and functions in linear wave-height models.

Table 1
Parameter definitions in different EDF models of scaled linear wave heights h_1

Model (Reference)	c_0	c_1	$f(h_1)$
N (Naess, 1985)	1	$\frac{1}{4(1-a)}$	1
B (Boccotti, 1989)	$\frac{1+b}{\sqrt{2b(1-a)}}$	$\frac{1}{4(1-a)}$	1
T (Tayfun, 1990a)	$\sqrt{\frac{1+r_m}{2r_m}}$	$\frac{1}{4(1+r_m)}$	$1+\frac{1-r_m^2}{4r_m h_1^2}$
T1 (Tayfun 1)[a]	$\sqrt{\frac{1+r_m}{2r_m}}$	$\frac{1}{4(1+r_m)}$	1
T2 (Tayfun 2)[a]	1	$\frac{1}{4(1+r_m)}$	1
R (Rayleigh)	1	1/8	1

[a]Present approximations to T.

Naess (1985) does not define the key parameter in his distribution exactly. As a result, previous studies interpreted it in a variety of slightly different forms. These include the present one in Table 1, which makes N consistent with B. Note further that only N, T2 and R are properly normalized to unity at $h_1 = 0$. Others are not since they are derived using the asymptotic behavior of either the surface itself, as in B, or its envelope, as in T and T1 both. So, the non-normalized PDFs in the table are valid over the range of relatively large waves, whereas the normalized ones are valid for $h_1 \geqslant 0$. T1 corresponds to the lower-bound EDF in Tayfun (1990a). Thus, T2 is a normalized approximation for T1, as N is for B. As $\nu \to 0$, then $r_m \to |a| \to b \to 1$, and all EDFs assume the same limit form R. In general, r_m is similar to but always slightly larger than $|a|$ so that for large h_1, T $\geqslant$ T1 > T2 > N invariably.

The mean $(h_1)_{1/n} = \langle h_1 | h_1 > (h_1)_n \rangle$, conditional on $h_1 > (h_1)_n = E^{-1}(1/n) = \sqrt{\ln(c_0 n)/c_1}$, is given by

$$(h_1)_{1/n} = (h_1)_n + \frac{1}{2} c_0 \sqrt{\frac{\pi}{c_1}}\, n\, \mathrm{erfc}(\sqrt{\ln(c_0 n)}); \quad n \geqslant 1, \tag{8}$$

where erfc $\equiv$ complementary error function. This expression is valid for all models for which $f = 1$. The conditional mean $(h_1)_{1/n}$ corresponding to T with $f \neq 1$ has a more complicated form, and is given in Tayfun (1990a).

Consider now a sample population of N independent wave heights $(h_1)_j$ $(j = 1, 2, \ldots, N)$ identically described by the parent EDF in Eq. (7) with $f = 1$. Let $Z_1 = \max\{(h_1)_j;\ j = 1, 2, \ldots, N\}$. For large N, the EDF of Z_1 is given by the Gumbel distribution (see e.g. Ang and Tang, 1984)

$$\Pr\{Z_1 > z\} \equiv E_{1,N}(z) = \exp[-\exp\{-\alpha_{1,N}(z - Z_{1,\mathrm{mode}})\}], \tag{9}$$

where

$$Z_{1,\mathrm{mode}} = \sqrt{\ln(c_0 N)/c_1}, \quad \alpha_{1,N} = 2\sqrt{c_1 \ln(c_0 N)}. \tag{10}$$

The mean and standard deviation of Z_1 then follow, respectively, as

$$\langle Z_1 \rangle = Z_{1,\mathrm{mode}} + \frac{\gamma_e}{\alpha_{1,N}} = \sqrt{\frac{\ln(c_0 N)}{c_1}} + \frac{\gamma_e}{2\sqrt{c_1 \ln(c_0 N)}}, \tag{11}$$

$$\sigma_{Z_1} = \frac{\pi}{\alpha_{1,N}\sqrt{6}} = \frac{\pi}{2\sqrt{6 c_1 \ln(c_0 N)}}, \tag{12}$$

where $\gamma_e = 0.577216\ldots \equiv$ Euler's constant.

The preceding statistics are appropriate to long-crested waves and to wave heights extracted from a surface time series gathered at a fixed point. The spatial extreme value analysis of directional waves requires significant modifications by way of Piterbarg's asymptotic theorems for homogeneous Gaussian processes over multi-dimensional spaces (Piterbarg, 1996). Krogstad et al. (2004) present an excellent interpretation and applications of Piterbarg's theorems to the extreme value analyses of simulated 3D

wave fields, both linear and nonlinear. Further extensions and applications to linear and nonlinear wave heights and crests are given in Socquet-Juglard (2005), Socquet-Juglard et al. (2005) and Forristall (2005, 2006).

2.3. Stability of EDF estimates

Consider a set of n independent observations of a random variable y distributed according to the PDF and EDF given by p and E, respectively. Rank-order these into the order statistics $y_1 > y_2 > \cdots > y_n$. For the jth largest value y_j (see e.g. Borgman and Resio, 1982),

$$\Pr\{x < y_j \leqslant x + \mathrm{d}x\} = p_j(x)\,\mathrm{d}x = \frac{n!}{(j-1)!(n-j)!}E^{j-1}(1-E)^{n-j}p(x)\,\mathrm{d}x. \quad (13)$$

Now, regard E as a random variable, as it would be in estimating it from the order statistics. Since $\mathrm{d}E = -p(x)\,\mathrm{d}x$ by definition, the PDF of E follows from $p_E|\mathrm{d}E| = p_j|\mathrm{d}x|$ as

$$p_E(E) = \frac{n!}{(j-1)!(n-j)!}E^{j-1}(1-E)^{n-j}. \quad (14)$$

So, the simple moments of E are given by

$$\langle E^m\rangle = \int_0^1 x^m p_E(x)\,\mathrm{d}x = \frac{n!(m+j-1)!}{(j-1)!(m+n)!}; \quad m = 0,1,2,\ldots. \quad (15)$$

Thus, the mean and standard deviation are

$$\langle E\rangle = \frac{j}{n+1}, \quad \sigma_E = \frac{1}{n+1}\sqrt{\frac{j(n-j+1)}{n+2}} \quad (16)$$

from which the coefficient of variation of E follows as

$$\delta_E = \frac{\sigma_E}{\langle E\rangle} = \sqrt{\frac{n-j+1}{j(n+2)}} \approx \sqrt{\frac{1}{j}} \quad (17)$$

for $n \gg j$. The exact form of Eq. (17) is plotted in the upper part of Fig. 2 for $1 \leqslant j \leqslant 10^2$ and various n. For $j \leqslant 10$ and $n \geqslant 400$, δ_E is given by $\sqrt{1/j}$ quite accurately, as shown in the lower part of the same figure. It is seen that for $1 \leqslant j \leqslant 10$, $\delta_E \geqslant 30\%$ always. In particular, δ_E of the largest wave height is 100% since σ_E is as large as the estimate E itself. Clearly, the larger the sample population n, the more stable are the estimates, but the stability characteristics of the largest group of wave heights, say, the largest five or so remain persistently poor. In general, the stability of estimates with negligible bias is indicated with confidence intervals. But, the simpler alternative and the one that will be preferred here is to use the standard deviation σ_E and define the upper and lower stability bands associated with the estimate $E \approx j/(n+1)$ as $E + \sigma_E$ and $E - \sigma_E$, respectively.

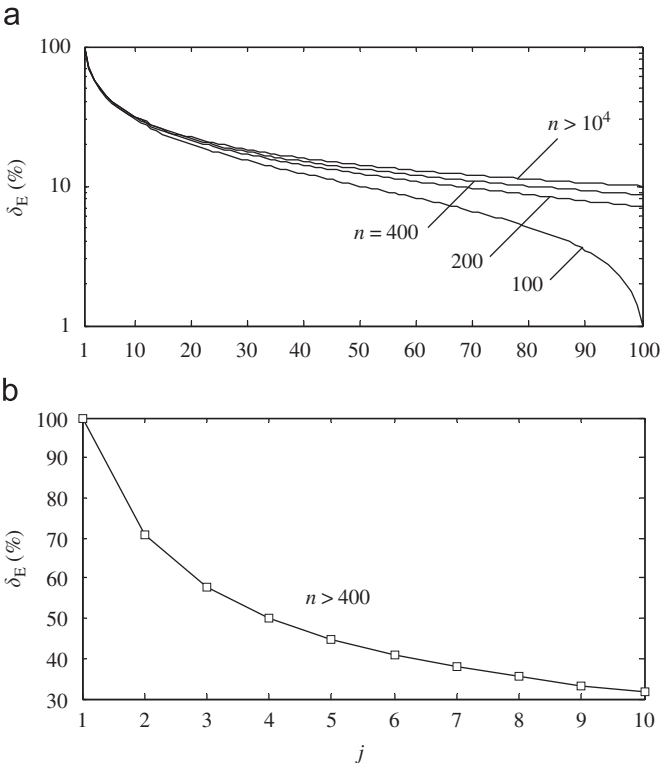

Fig. 2. Coefficient of variation δ_E of EDF estimates for jth largest wave height in n waves: (a) variation of δ_E with rank-order j and sample size n; (b) δ_E for $n > 400$ and $j \leqslant 10$.

2.4. Comparisons with linear waves from JONSWAP simulations

The simulations for 2D linear and nonlinear deep-water waves are based on

$$S(\omega) = \frac{m_0}{\omega_p}u^{-4}\exp(-1.25u^{-4})\gamma^{g(u)}w(u); \quad 0.2 \leqslant u \leqslant 10, \quad (18)$$

where $w(u) = 1$ for $u \leqslant 4$, and $w(u) = (4/u)^{\leftarrow 4}$ for $u > 4$, and all other variables and parameters are the same as in Eq. (6). The 'filter' $w(u)$ helps to simulate second-order waves whose spectral density attenuates as ω^{-4} as in Eq. (6) (Tayfun, 1990b; Tayfun and Fedele, 2006).

For the present linear comparisons, $m_0 \approx 10.15\ \mathrm{m}^2$, $\nu \cong 0.43$ and $T_p \approx 13.44$ s, and an ensemble of 40 linear and nonlinear time series sampled at 5 Hz was simulated, following an efficient approach described in Tayfun (1986) but with linear random spectral amplitudes whose r.m.s. values mimic the spectral shape in Eq. (18). These yield altogether 453,416 linear waves first and, subsequently, 471,309 nonlinear waves whose spectral density closely approximates Eq. (6) with $m_0 \approx 10.72\ \mathrm{m}^2$, $\nu \cong 0.63$ and the same T_p as before. For economy of space, all parameters relevant to the simulated linear and nonlinear wave heights, and their various theoretical predictions are inset in the figures displaying a variety of results both for the linear simulations here and for the later comparisons with nonlinear simulations and oceanic measurements. These

include the comparisons of EDFs of short and long samples of wave heights, conditional moments and the extreme value statistics.

The linear models R, T, N and the simulated EDFs are compared in Fig. 3. The predictions from B, T1 and T2 are nearly the same as T and cannot be differentiated adequately in semi-logarithmic scales. The differences between the models are seen more clearly in Fig. 4. This figure and similar ones to follow later are constructed based on an approach popularized by Forristall (1984), plotting the ratio h_1/h_R versus E, where $h_R \equiv \sqrt{-8\ln(E)}$ and represents the wave height that would be predicted by the Rayleigh law at the same E level. In this type of plots,

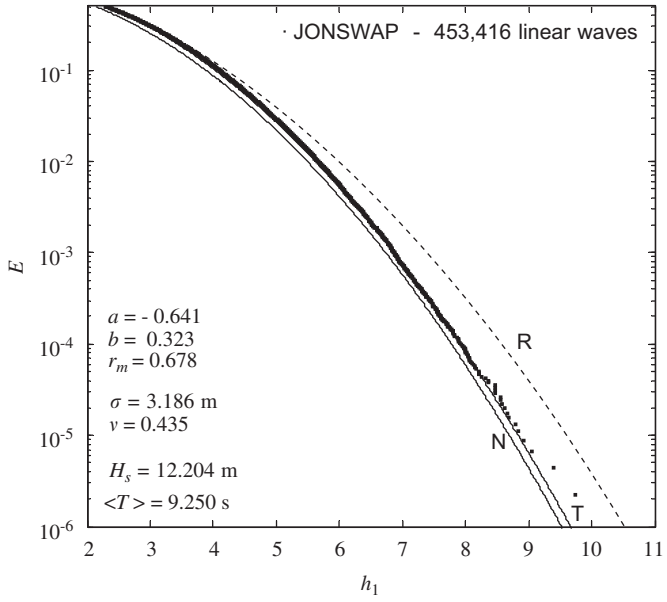

Fig. 3. Linear simulations: exceedance distribution of scaled wave heights compared to R, N, and T ($\approx$T1$\approx$B, and not shown).

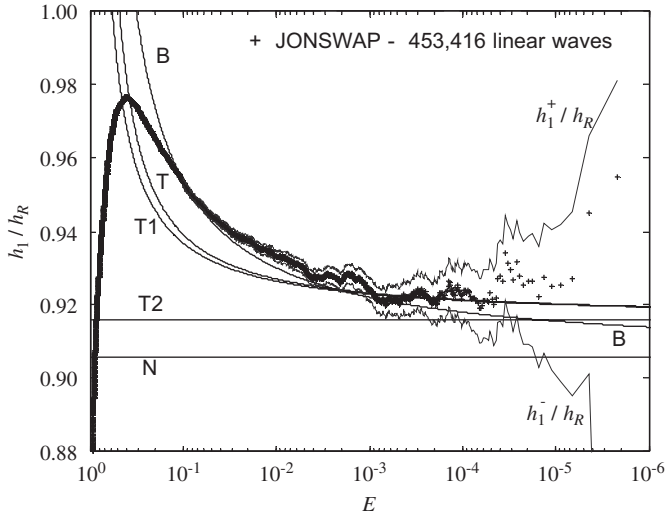

Fig. 4. Linear simulations: simulated ratio h_1/h_R compared to predictions. $h_1^{\pm}/h_R \equiv$ upper and lower stability bands.

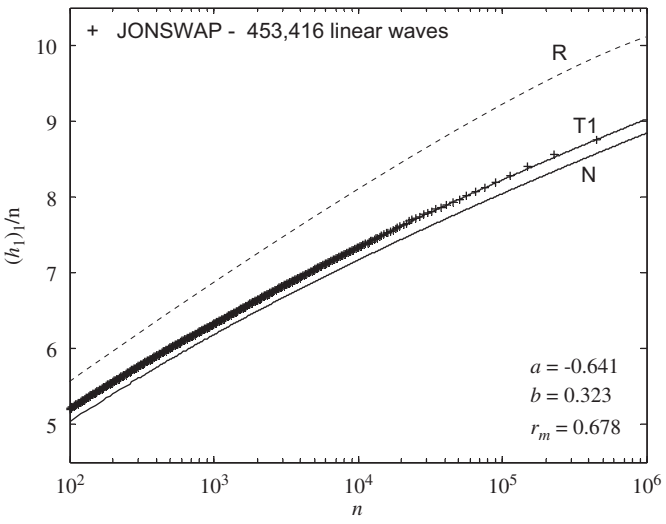

Fig. 5. Linear simulations: simulated conditional mean $(h_1)_{1/n}$ compared to predictions from linear models.

the relative stability of the observed ratios h_1/h_R is indicated in terms of the upper h_1^+/h_R and lower h_1^-/h_R bands, where $h_1^{\pm} \equiv \sqrt{-8\ln(E \pm \sigma_E)}$. Among all the models shown in Fig. 4, B describes relatively large wave heights most accurately, except toward the very extreme tail where T and T1 converge and also match the data slightly better. R over predicts the large wave heights by 7–8%, as is often noted. In contrast, N under predicts them about 2%. So does T2 also, but by less than 1%. In general, all the models except R seem to describe the simulated data reasonably well over the high-wave range with maximum errors of about 2% or less.

The theoretical conditional mean $(h_1)_{1/n}$ from Eq. (8) for $n \geqslant 1$ is compared with the simulated values for $n \geqslant 10^2$ in Fig. 5. Only the predictions from R, T1 and N are included in this figure for clarity as B, T and T2 are all nearly the same as T1. It is seen that the simulated means are accurately described by all linear models except for R and N.

The expected maximum wave height in N waves is shown in Fig. 6 for $N \geqslant 10^3$ in comparison with Eq. (11) for R, N and B. The predictions from T1 and T2, not shown for clarity, hardly differ from B. Also, for N fixed, the simulated values represent the average values of the maxima from 200 realizations, with each realization of size N randomly drawn from 453,416 waves. As expected, N under predicts the simulated data slightly. In contrast, B, and thus T1 and T2 do follow the data trend quite closely.

3. Second-order waves

3.1. Quasi-deterministic (Q-D) model

Assume that the surface elevation measured from the mean sea level at a fixed point and scaled with σ is expressed to $O(\mu)$ in the usual second-order form $\eta = \eta_1 + \eta_2$, where $\eta_1 \equiv$ first-order linear Gaussian component

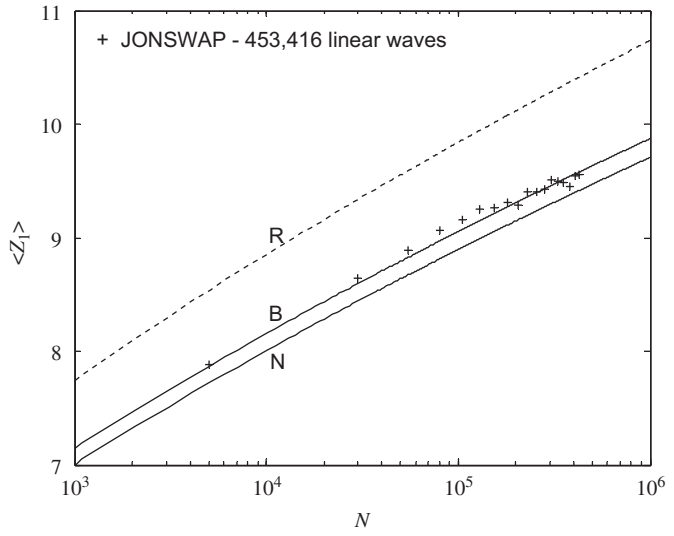

Fig. 6. Linear simulations: expected largest wave height in N waves compared to predictions from R, B ($\approx$T1$\approx$T2) and N.

of $O(1)$; and, $\eta_2 \equiv$ second-order nonlinear correction of $O(\mu)$. The first- and second-order spectral densities and the associated simple moments m_j differ slightly, but these differences are of $O(\mu^2)$ and thus neglected in the following.

Next, suppose that a wave crest occurs at time $t = t_0$. Defining $\xi_1 \equiv \eta_1(t_0)$ as the linear component of that crest, the conditional mean of $\eta_1(t_0+\tau)$, given $\eta_1(t_0) = \xi_1 \gg 1$, has the form (Lindgren, 1970, 1972; Phillips et al., 1993; Boccotti, 2000)

$$\langle \eta_1(t_0+\tau) | \eta_1(t_0) = \xi_1 \rangle = \xi_1 \rho_1(\tau), \tag{19}$$

where $\rho_1 = \langle \eta_1(t_0)\eta_1(t_0+\tau)\rangle = \rho = \langle \eta(t_0)\eta(t_0+\tau)\rangle$, correct to $O(\mu)$. The extension of this result to second-order waves by Fedele and Arena (2005) and Arena (2005) yields the conditional mean of second-order η as

$$\langle \eta(t_0+\tau) | \eta(t_0) = \xi^+ \rangle = \xi_1 \rho(\tau) + \tfrac{1}{6}\xi_1^2 \lambda(\tau), \tag{20}$$

where $\xi^+ \equiv$ second-order nonlinear crest height at t_0; and,

$$\lambda(\tau) \equiv 3\langle \eta_1^2(t_0)\eta(t_0+\tau)\rangle = \frac{3}{2\sigma^3}\int_{\boldsymbol{k}}\int_{\boldsymbol{k}'} \psi\psi'[K^+ \cos\phi^+ + K^- \cos\phi^-]\,\mathrm{d}\boldsymbol{k}\,\mathrm{d}\boldsymbol{k}', \tag{21}$$

with $\boldsymbol{k} \equiv$ wave-number vector; $\psi = \psi(\boldsymbol{k}) \equiv$ wave-number spectrum; $\phi^\pm = (\omega \pm \omega')\tau$; $\omega \equiv$ angular frequency, related to $k = |\boldsymbol{k}|$ via the usual dispersion relationship $\omega^2 = gk\tanh kd$ with $d \equiv$ mean water depth, and, $K^\pm \equiv$ second-order interaction kernels (see e.g. Forristall, 2000). It may be noticed that $\lambda(\tau) \leqslant \lambda(0) = \lambda_3 = \langle \eta^3 \rangle \equiv$ skewness coefficient of η, correct to $O(\mu)$ (Tayfun, 1994, 2006).

3.2. Wave heights and related parameters

The nonlinear crest height at t_0 follows from Eqs. (20) and (21) with $\tau = 0$ as

$$\xi^+ = \xi_1 + \tfrac{1}{6}\lambda_3 \xi_1^2. \tag{22}$$

The wave trough, say ξ^-, immediately following the crest ξ^+ at t_0 occurs at time $t = t_0 + \tau^*$, and follows similarly from Eqs. (20) and (21) with $\tau = \tau^*$ as

$$\xi^- = \xi_1 \rho(\tau^*) + \tfrac{1}{6}\lambda(\tau^*)\xi_1^2 = a\xi_1 + \tfrac{1}{6}\lambda^* \xi_1^2, \tag{23}$$

with $a = \rho(\tau^*)$ as defined before and $\lambda^* = \lambda(\tau^*)$ for brevity. Thus, the nonlinear crest-to-trough wave height scaled with σ can be written in the quadratic form

$$h \equiv \xi^+ - \xi^- = h_1 + \tfrac{1}{2}\beta h_1^2, \tag{24}$$

where $h_1 = \xi_1(1-a) \equiv$ first-order linear wave height, and

$$\beta = \frac{1}{3}\frac{\lambda_3 - \lambda^*}{(1-a)^2}. \tag{25}$$

As $\nu \to 0$ ($a \to -1$), $\beta \to 0$ and thus $h \to h_1$, leading to the Rayleigh limit R appropriate to linear wave heights.

For long-crested waves in deep water, $K^\pm = \pm|\omega^2 \pm \omega'^2|/g$ so that λ^* as given by Eq. (21) with $\tau = \tau^*$ reduces to a double integral, thus simplifying β somewhat (Fedele and Arena, 2005). In the general case, it can also be estimated from a time series of η in a more practical but somewhat approximate fashion. Evidently, the equivalent definition of $\lambda(\tau)$ in Eq. (21) requires the linear η_1 associated with η. An approximate procedure for 'linearizing' η is described in Tayfun (1983, 1984) based on the narrow-band approximation

$$\eta_1 \approx \eta - \alpha(\eta^2 - \hat{\eta}^2), \tag{26}$$

where $\alpha \ll 1$ and represents a parameter chosen so as to satisfy

$$\langle \eta_1^3 \rangle \approx \langle \eta^3 \rangle - 3\alpha(\langle \eta^4 \rangle - \langle \eta^2 \hat{\eta}^2 \rangle) = 0, \tag{27}$$

correct to $O(\alpha)$. Since the moments on the right-hand side can be estimated from the observed series of η and its Hilbert transform, it follows that

$$\alpha \approx \frac{1}{3}\frac{\langle \eta^3 \rangle}{\langle \eta^4 \rangle - \langle \eta^2 \hat{\eta}^2 \rangle}. \tag{28}$$

Numerical computations with simulated nonlinear series and oceanic measurements indicate that the preceding approach, though rather simplistic, works reasonably well and thus circumvents the four-fold integration implied by Eq. (21) in the more general case of directional seas.

In theory, h_1 has the form implied by B and N only. However, all the models in Table 1 can be used for the linear wave heights just as well, including all Ts if the parameters λ^* and $-a$ are replaced with $\lambda_m = \lambda(\tau_m)$ and r_m in Eq. (25). These alternatives will hereafter be differentiated as β_B for β based on λ^* and a, and as β_T for β based on λ_m and r_m, respectively. Computations based on the JONSWAP type spectra with variable bandwidth show that β_B and β_T are nearly the same, and that $\beta_B \cong \beta_T \approx 0.003-0.022$. The effects of nonlinearities on the surface statistics such as simple moments, cumulants, and on the distributions of surface elevations, wave heights and crests are most pronounced in long-crested waves such as those simulated numerically or mechanically in wave

flumes (Tayfun, 1994, 2006; Forristall, 2000). Thus, the larger β values typify long-crested waves with broad spectra, and the smaller values imply relatively narrow spectra.

From Eq. (24),

$$\frac{h}{h_1} = 1 + \frac{1}{2}\beta h_1. \tag{29}$$

If $\beta \approx 0.003-0.004$ as a typical narrow-band case, then $h/h_1 \approx 1.015-1.02$ for $h_1 \approx 10$. Thus, the effect of second-order nonlinearities is likely to be rather small in narrow-band seas, being at most 2% or less even for exceptionally rare waves with heights as large as 10σ. The situation is only slightly different for waves with broader spectra, as will be seen later. Nonetheless, it is clear at this point that the preceding model coupled with R as a model for h_1 cannot possibly explain unusually large waves in terms of second- or higher-order nonlinearities.

3.3. Statistics

The EDF of h would be given by

$$E(h) = c_0 f(h_1)\exp(-c_1 h_1^2); \quad h_1 = (-1+\sqrt{1+2\beta h})/\beta. \tag{30}$$

The corresponding PDF then follows from $-(\mathrm{d}E/\mathrm{d}h_1)(\mathrm{d}h_1/\mathrm{d}h)$. The conditional mean $h_{1/n} = \langle h|h>h_n\rangle$ is given in this case by

$$h_{1/n} = (h_1)_{1/n} + \frac{\beta}{2c_1}[1+\ln(c_0 n)] \tag{31a}$$

where

$$h_n = E^{-1}(1/n) = (h_1)_n + \frac{\beta}{2}(h_1)_n^2. \tag{31b}$$

This result is valid for all models with $f=1$ in Table 1. Further, the distribution of $Z = \max\{h_j;\ j=1,2,\ldots,N\}$ is of the same Gumbel form as in Eq. (9), with Z_1 replaced by Z. Thus,

$$Z_{\mathrm{mode}} = Z_{1,\mathrm{mode}}(1+\tfrac{1}{2}\beta Z_{1,\mathrm{mode}}),$$
$$\alpha_N = \alpha_{1,N}(1-\beta Z_{1,\mathrm{mode}}) + O(\beta^2), \tag{32}$$

and lead to the mean and standard deviation of Z given by

$$\langle Z\rangle = \langle Z_1\rangle + \frac{1}{2c_1}\beta[\gamma_e + \ln(c_0 N)] + O(\beta^2), \tag{33}$$

$$\sigma_Z = \sigma_{Z_1}[1+\beta\sqrt{\ln(c_0 N)/c_1}] + O(\beta^2). \tag{34}$$

As expected, these also converge to the linear results in Eqs. (11) and (12) in the limit as $\beta \to 0$.

3.4. Stokes–Rayleigh models

In a recent study, Dawson (2004) derived a nonlinear wave-height distribution, using the fifth-order deterministic Stokes theory as a model. Although the Stokes analogy in this case is entirely heuristic in a probabilistic context, it is worthwhile to review Dawson's distribution briefly for completeness and to explore a second-order model that readily follows from it. In the present notation, the EDF of scaled wave heights in Dawson's model (D5) is given by

$$E(h) = \exp(-2\alpha_{\mathrm{D}}^2 h^2)\left[1 - \frac{3}{16}\mu_{\mathrm{D}}^2 h^2 - \frac{107}{3072}\mu_{\mathrm{D}}^4 h^4\right], \tag{35}$$

where $\alpha_{\mathrm{D}} \equiv \sigma/H_s$; $\mu_{\mathrm{D}} = \sigma\bar{\omega}^2/g \equiv$ steepness parameter; $\bar{\omega} = 2\pi/\langle T\rangle$; and, H_s and $\langle T\rangle$ represent, respectively, the significant wave height and mean zero up-crossing period derived from actual measurements. As the latter two statistics require a wave-by-wave analysis before Eq. (35) can be used, D5 is not a predictive model in the same sense as the other models considered here. Further, setting $\mu_{\mathrm{D}} = 0$ leads to the EDF of second-order wave heights in the form (D2)

$$E(h) = \exp(-2\alpha_{\mathrm{D}}^2 h^2). \tag{36}$$

The conditional moments and the extreme value statistics associated with D5 cannot be obtained in a closed form. Dawson (2004) obtains D5 by series inversion from the Stokes fifth-order wave model and retains only terms of $O(\mu_{\mathrm{D}}^4)$ in the exponent of the EDF, as in Eq. (35). Actually, the distribution of wave heights and all the related conditional and extreme value statistics can be derived from Dawson's (2004) fifth-order model without resorting to series inversion and other approximations. However, preliminary comparisons here showed that the resulting statistics do not represent the simulated and observed data trends as well as D5 does. Finally, it is noted that in contrast with D5, D2 has the same simple form as Eq. (1). So, all the conditional and extreme value statistics associated with it readily follow from Eqs. (8), (11) and (12) by setting $c_0 = 1$ and $c_1 = \alpha_{\mathrm{D}}\sqrt{2}$.

3.5. Comparisons with nonlinear waves from JONSWAP simulations

The comparisons of the EDF estimates for the crest-to-trough heights of 471,309 simulated waves are shown in Fig. 7. The quasi-deterministic extensions of the linear models are identified with the suffix *Q-D*. For clarity of presentation, this figure shows R, D2, D5 and only N *Q-D* as the best apparent fit to large wave heights among all the second-order *Q-D* models considered. Evidently, both D2 and D5 tend to under predict the simulated data. Further, D2 is not only simpler than D5, but it also seems to describe the simulated data somewhat better than D5 does.

The effects of sample size n on the EDF estimates and their stability are illustrated in Fig. 8 for the simulation series #13, which contains the largest scaled wave height 9.94 in Fig. 7. The upper part of the figure shows the hourly section of the series with 411 waves including the largest wave, and the lower part is for all 11,835 waves in the series. Both cases also include R as a reference.

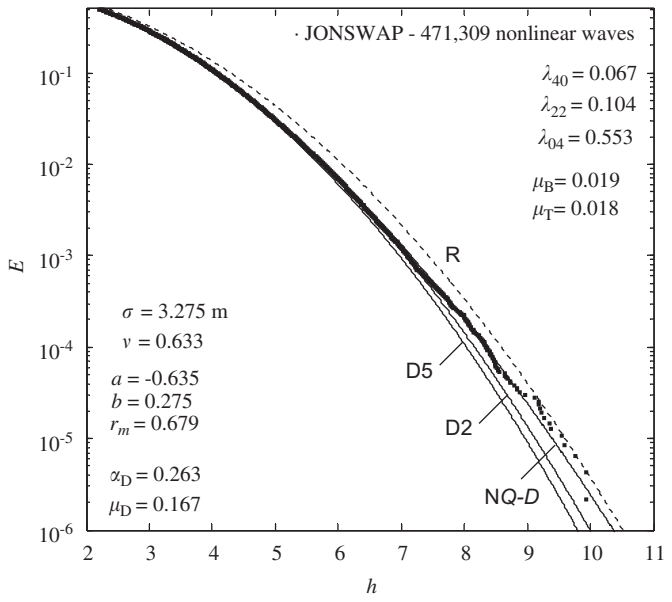

Fig. 7. Nonlinear simulations: exceedance distribution of scaled wave heights compared to R, N *Q-D*, D2 and D5.

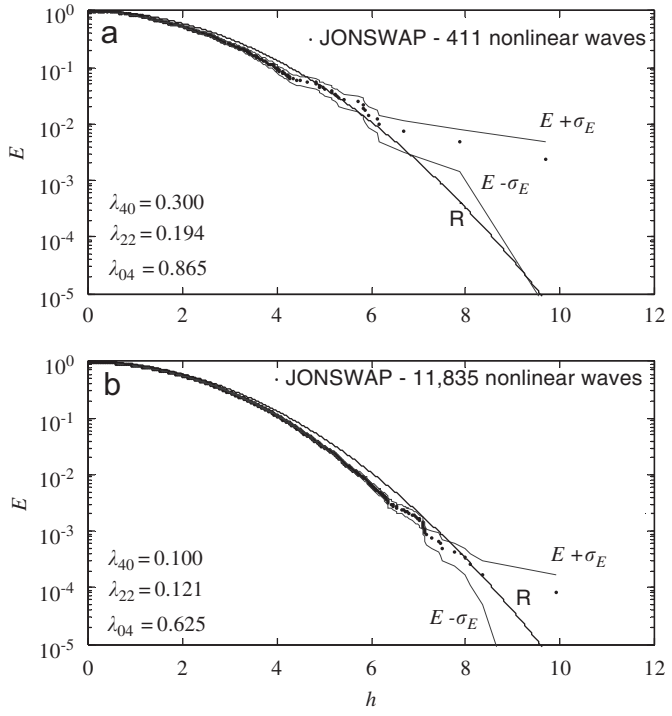

Fig. 8. Effects of sample size n on EDF estimates and their stability: (a) $n = 411$ waves from 1-h nonlinear series #13 containing max h = 9.94 in Fig. 7; (b) $n = 11{,}835$ waves from series #13 as a whole ($\approx$29.2 h). $E \pm \sigma_E \equiv$ upper and lower stability bands.

A comparison of results in Figs. 7 and 8 clearly illustrates that analyses based on limited data can easily lead to false conclusions on the true nature of waves observed in a short record. Clearly, the largest wave height 9.94 does not seem out of place in Fig. 7 as it does in Fig. 8, particularly in the upper part based on a rather short sample of just 411 waves. Thus, an outlier in a small sample population that does not seemingly follow an established law as in Fig. 8 can in reality be a relatively rare realization of a typical population, as in Fig. 7. Forristall (2005) also emphasizes this very point amply under a more realistic setting, using similar comparisons between small and large populations of waves gathered from an offshore platform in the Gulf of Mexico during Hurricane Ivan.

The simulated ratio h/h_R is compared to the predictions from all nonlinear models in Fig. 9. The linear simulations of Fig. 4 are also replicated in this figure for contrast with the nonlinear results. It is observed that second-order effects tend to make the wave heights 2–4% larger. The trend of the nonlinear data is best predicted by D2 over the mid- to high-wave range and, apparently, by N *Q-D* toward the extreme tail. Somewhat surprisingly, D5 seems to predict the linear wave heights reasonably well, but it under predicts the nonlinear heights.

The conditional mean $h_{1/n}$ from Eq. (31) for $n \geqslant 1$ is compared with the simulated values for $n \geqslant 10^2$ in Fig. 10. Only the predictions from R ($\approx$ B *Q-D*), N *Q-D* and D2 are included in this figure. The simulated data seem to follow D2 for $n \leqslant 10^3$ approximately, and N *Q-D* for $n > 10^3$.

The expected maximum wave height in N waves is compared in Fig. 11 for $N \geqslant 10^3$ with the predictions from Eq. (33) for R, B *Q-D* ($\approx$T1 *Q-D* $\approx$ T2 *Q-D*) and N *Q-D*. As in the linear case, for N fixed, the simulated values represent the average values of the maxima from 200 realizations, with each realization of size N randomly chosen from 471,309 waves. The simulations are described reasonably well by N *Q-D*. Discrepancies do appear for $N > 10^5$ and toward the extreme tail where larger samples are needed for estimating the expected maxima more reliably.

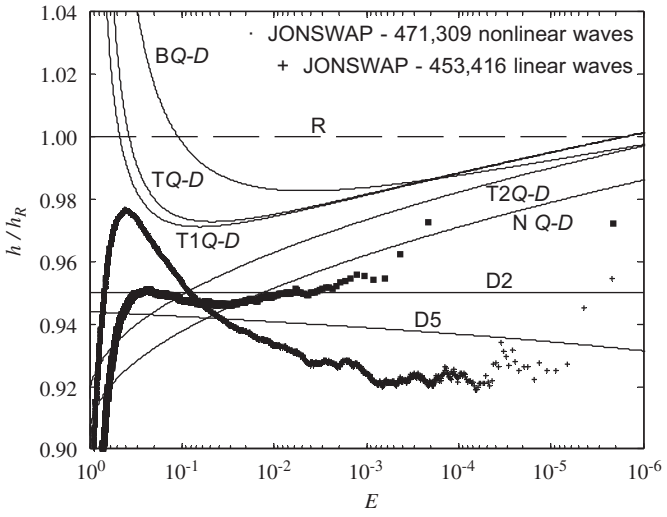

Fig. 9. Nonlinear (and linear) simulations: simulated ratio h/h_R compared to predictions from nonlinear models.

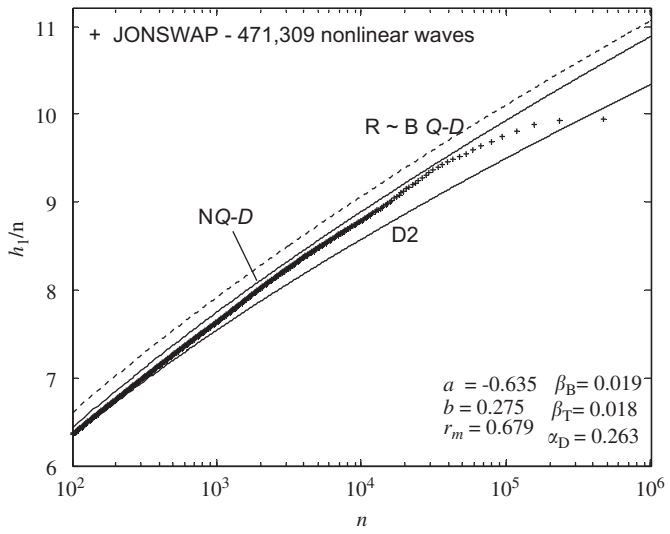

Fig. 10. Nonlinear simulations: simulated conditional mean $h_{1/n}$ compared to predictions from R ($\approx$B *Q-D*), N *Q-D* and D2.

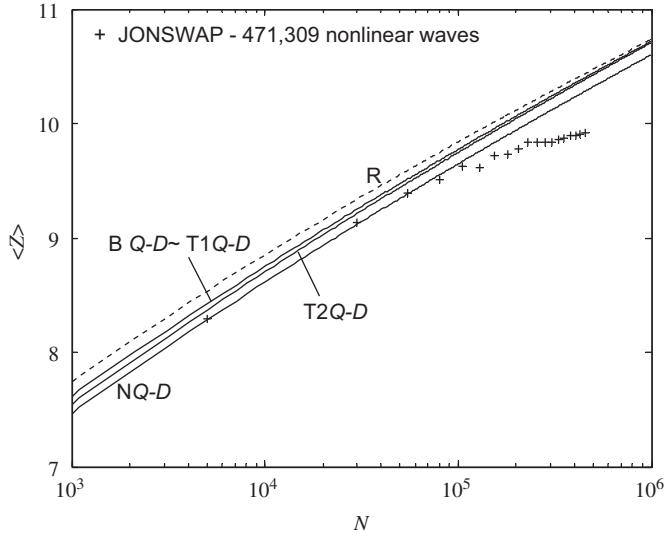

Fig. 11. Nonlinear simulations: expected largest wave height in N waves compared to predictions from R, B *Q-D* ($\approx$T1 *Q-D* $\approx$T2 *Q-D*) and N *Q-D*.

4. Third-order nonlinearities and GC models

4.1. Wave envelopes

GC models represent approximations that arise from the Taylor series expansion of the characteristic function of random variables whose probability structure is not known exactly. Their applications and relevance to weakly nonlinear sea-surface elevations and a variety of related properties were first described by Longuet-Higgins (1963, 1964). However, his applications do not include nonlinear wave envelopes and phases. The GC approximations for the joint and marginal distributions of these were formulated much later in Tayfun and Lo (1990). The third- and fourth-order surface cumulants included in the Tayfun–Lo model are $O(\mu)$ and $O(\mu^2)$, respectively. Thus, the formulation is theoretically valid for waves characterized with second- and third-order nonlinearities, although Tayfun and Lo elaborate it explicitly for second-order long-crested waves. Its application to second-order deep-water waves in the most general case is described in Tayfun (1994), correct to $O(\mu)$.

To review the underlying ideas in the Tayfun–Lo formulation briefly, let $\eta = \xi\cos\theta$ and $\hat{\eta} = \xi\sin\theta$, where $\xi = (\eta^2 + \hat{\eta}^2)^{1/2} \equiv$ wave envelope scaled with σ; and, $\theta = \tan^{-1}(\hat{\eta}/\eta) \equiv$ wave phase with a non-uniform distribution over an interval of 2π. In the original formulation, the wave envelope is scaled with its r.m.s. value $\sigma\sqrt{2}$, instead of just σ as in the present case. And, a key point is that if η is statistically homogeneous, then η and $\hat{\eta}$ are orthogonal, viz.

$$\langle\eta\hat{\eta}\rangle = \pi^{-1}\int_{-\infty}^{\infty}\langle\eta(t)\eta(t-\tau)\rangle\tau^{-1}\,\mathrm{d}\tau = 0 \tag{37}$$

since the integrand is odd. This allows the joint PDF of η and $\hat{\eta}$ to be expanded as

$$p_{\eta\hat{\eta}} = \frac{1}{2\pi}\mathrm{e}^{-(\eta^2+\hat{\eta}^2)/2}\left[1 + \sum_{j=0}^{3}\frac{\lambda_{(3-j)j}}{(3-j)!j!}H_{3-j}(\eta)H_j(\hat{\eta}) + \sum_{j=0}^{4}\frac{\lambda_{(4-j)j}}{(4-j)!j!}H_{4-j}(\eta)H_j(\hat{\eta})\right], \tag{38}$$

where

$$\lambda_{mn} = \langle\eta^m\hat{\eta}^n\rangle/\sigma^{m+n}; \quad m+n=3, \tag{39}$$

$$\lambda_{mn} = \langle\eta^m\hat{\eta}^n\rangle/\sigma^{m+n} + (-1)^{m/2}(m-1)(n-1); \quad m+n=4, \tag{40}$$

represent the third- and fourth-order normalized joint cumulants of η and $\hat{\eta}$, respectively; and, $H_j(x) \equiv j$th-order Hermite polynomial. For x arbitrary, $H_0 = 1$, $H_1 = x$, $H_2 = x^2 - 1$, $H_3 = x^3 - 3x$ and $H_4 = x^4 - 6x^2 + 3$.

The joint PDF of ξ and θ follows from a change of variables in the form $p_{\xi\theta} = \xi p_{\eta\hat{\eta}}$. In particular, integrating Eq. (38) with respect to θ yields the marginal PDF of ξ by itself as

$$p_{\xi}(x) = x\mathrm{e}^{-x^2/2}\left[1 + \frac{\Lambda}{64}(x^4 - 8x^2 + 8)\right], \tag{41}$$

where

$$\Lambda = \lambda_{40} + 2\lambda_{22} + \lambda_{04}. \tag{42}$$

Notice that the third-order joint cumulants λ_{mn} for which $m+n=3$, and the fourth-order λ_{31} and λ_{31} do not appear in the preceding results since all terms that contain these cumulants drop out in integrating Eq. (38) with respect to θ.

The EDF corresponding to Eq. (41) is

$$E_{\xi}(x) = \int_x^{\infty} p_{\xi}\,\mathrm{d}\xi = \mathrm{e}^{-x^2/2}\left[1 + \frac{\Lambda}{64}x^2(x^2-4)\right]. \tag{43}$$

To $O(\mu^0)$ and $O(\mu)$, ξ is Rayleigh-distributed since Λ is $O(\mu^2)$ and ignored. This is consistent with the general second-order random wave model in Tayfun (1994). For the same model,

$$\lambda_{03} = \lambda_{21} = 0, \quad \lambda_{30} = 3\lambda_{12}, \tag{44}$$

as a general result correct to $O(\mu)$.

At the next order of approximation, the fourth-order cumulants come into play. The nature of the latter statistics is not known in the most general case. Various approximations appropriate to so-called quasi-Gaussian and long-crested narrow-band waves are possible, but such approximations tend to lack generality. For second-order long-crested waves, Tayfun and Lo (1990) express all three joint cumulants in Eq. (42) to $O(\mu^2\nu)$. In the present notation, they are given by

$$\lambda_{40} = 12(1-2\gamma_0)\mu_m^2, \quad \lambda_{22} = 4(1-\gamma_0)\mu_m^2, \quad \lambda_{04} = 12\mu_m^2, \tag{45}$$

where $\mu_m \equiv \sigma k_m = \sigma\omega_m^2/g = O(\mu)$ and $\gamma_0 = O(\nu)$ is a non-negative dimensionless parameter, defined explicitly elsewhere (Tayfun, 1986; Tayfun and Lo, 1990). It reflects the effect of the spectral bandwidth on the fourth-order cumulants. For long-crested waves with JONSWAP type spectra, $0 < \gamma_0 < \frac{1}{2}$. Hence,

$$\lambda_{40} < 3\lambda_{22} < \lambda_{04}. \tag{46}$$

As $\nu \to 0$, $\gamma_0 \to 0$ also, leading to $\lambda_{40} = 3\lambda_{22}$ and $\lambda_{04} = \lambda_{40}$, correct to $O(\mu^2)$. These in turn simplify Eq. (42) to

$$\Lambda \equiv \Lambda_{\text{app}} = 8\lambda_{40}/3. \tag{47}$$

Finally, if Eqs. (41) and (45) are combined, the resulting expression will be identical to the wave envelope PDF given in Eq. (48) in Tayfun and Lo (1990), except for apparent differences due to the additional factor $\sqrt{2}$ used in scaling the envelope heights with $\sigma\sqrt{2}$ in their case.

4.2. Wave heights

Let wave heights be measured as 2ξ, as suggested some years ago (Tayfun, 1983). In essence, 2ξ represents an upper bound to crest-to-trough wave heights. Its PDF and EDF follow from Eqs. (41) and (43) simply as

$$p_{2\xi}(x) = \frac{x}{4}\mathrm{e}^{-x^2/8}\left[1 + \frac{\Lambda}{1024}(x^4 - 32x^2 + 128)\right], \tag{48}$$

$$E_{2\xi}(x) = \mathrm{e}^{-x^2/8}\left[1 + \frac{\Lambda}{1024}x^2(x^2 - 16)\right]. \tag{49}$$

Generally, 2ξ differs appreciably from the crest-to-trough definition h because it ignores the variation of the wave envelope over the time interval between a wave crest and the following trough in a typical zero up-crossing cycle. That variation is $O(\nu)$ to the leading order, and can be rather significant for relatively broad-band oceanic waves. However, if $\nu \to 0$ then $h \cong 2\xi$. Assuming further that $\lambda_{40} \cong 3\lambda_{22}$ and $\lambda_{04} \cong \lambda_{40}$ as in the case of long-crested second-order waves, Eq. (49) becomes

$$E_h(x) = \mathrm{e}^{-x^2/8}\left[1 + \frac{\lambda_{40}}{384}x^2(x^2 - 16)\right]. \tag{50}$$

Mori and Yasuda (2002) also draw on the Tayfun–Lo formulation, albeit without an appropriate acknowledgement or a correct citation. Their interpretations of the nature of some joint cumulants are also somewhat flawed and thus disagree with the results in Tayfun and Lo (1990). However, these turn out to be inconsequential as they all drop out of the eventual results, and the expression more recently given in Mori and Janssen (2006) is identical to Eq. (50) above, referred to as a modified Edgeworth–Rayleigh (MER) distribution. Clearly, MER either assumes narrow-band second-order waves or requires that $\lambda_{04} = \lambda_{40}$ and $\lambda_{40} = 3\lambda_{22}$. The GC model of Eqs. (48) and (49) does neither. Thus, it is more general than and thus preferable to MER.

An envelope-based definition of wave heights that is statistically consistent with the crest-to-trough definition is given by

$$h_m = \xi(t) + \xi(t + T_m/2), \tag{51}$$

where T_m stands for the spectral average period, as in Section 2.1. It is easily shown that $h_m = 2\xi + O(\nu) = h + O(\nu^2)$ and that it represents the definition that leads to T (Tayfun, 1981, 1990a). Thus, for linear or second-order nonlinear waves, empirical estimates of wave-height EDFs derived from wave-envelope series via Eq. (51) should be consistent with models such as T, T1, T2 and their nonlinear *Q-D* forms, respectively. In analyzing his 3D nonlinear simulations, Socquet-Juglard (2005) employs a definition of wave heights similar to h_m correctly, instead of 2ξ. He also prefers to use the spectral peak period T_p. Since $T_p > T_m$ in general, the resulting population of observed wave heights can display a relative increase in the mid range and a corresponding deficiency toward the high-wave extreme, as seen in some EDF plots of Socquet-Juglard (2005). The opposite tends to occur with T_m. The correct choice really depends on the specific application considered. For instance, if the distribution of wave heights over high waves is of interest, the optimal choice would be $T^* = \langle T|h \gg 1\rangle \equiv$ mean zero up-crossing wave period, conditional on $h \gg 1$. In general, $T_m < T^* < T_p$, and T^* is approximated reasonably well by (Tayfun, 1993)

$$T^* \equiv \langle T|h \gg 1\rangle \cong [1 + \nu^2(1+\nu^2)^{-3/2}]T_m. \tag{52}$$

The quantitative accuracy of GC type approximations can be unpredictable. In practice, inaccuracies can arise not only from the unstable nature of fourth-order cumulants and thus the difficulty of estimating them from observational data reliably, but they can also arise from the possibility that a particular GC approximation assumed may be inappropriate to the physical situation considered. As an example, consider Fig. 12 showing the variability of fourth-order cumulants derived as half-hourly running averages at 1 min intervals from the linear and nonlinear

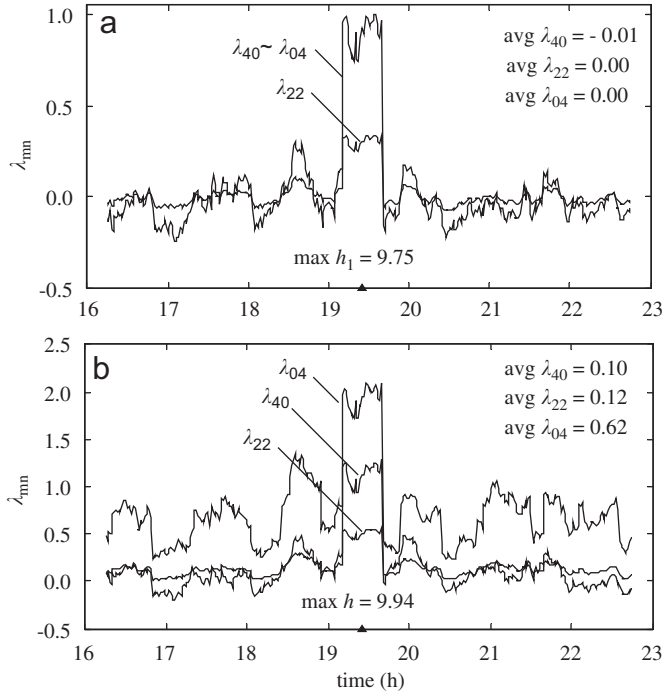

Fig. 12. Variations of λ_{40}, λ_{22} and λ_{04} in (a) linear and (b) nonlinear simulations. Both cases are from series #13 and represent $\frac{1}{2}$-h running averages at 1-min intervals. Insets: overall averages from series #13 as a whole ($\approx$29.2 h).

simulation series #13. This figure is based on two 6.5-h coincident segments containing the largest waves. The total duration of both series is about 29.2 h actually. In theory, $\lambda_{40} = \lambda_{22} = \lambda_{04} = 0$ for linear waves since η and $\hat{\eta}$ are jointly Gaussian. The overall averages $\lambda_{40} \cong -0.01$ and $\lambda_{22} = \lambda_{04} \cong 0.00$ derived from the complete linear series #13 satisfy this requirement and are thus consistent with the Gaussian norms for the most part. But, the local averages vary widely and tend to get amplified by high waves, as clearly seen around the largest waves with $\max h = 9.75$ in the linear case and $\max h = 9.94$ in the corresponding nonlinear case. Note further that the overall averages $\lambda_{40} \cong 0.10$, $\lambda_{22} \cong 0.12$ and $\lambda_{04} \cong 0.62$ derived from the nonlinear series #13 as a whole are not in conformity with or even suggestive of the narrow-band limits $\lambda_{04} = \lambda_{40}$ and $\lambda_{40} = 3\lambda_{22}$ quite simply because the nonlinear simulations for which $\nu \cong 0.63$ do not represent narrow-band waves. But, they confirm the inequalities in Eq. (46). Under such conditions, crest-to-trough wave heights h and the associated 2ξ can differ significantly. This is clearly seen in Fig. 13, showing the EDF estimates derived from the nonlinear series #1 in a comparison with the predictions from the GC model of Eq. (49) with $\Lambda \approx 0.015$ ($\lambda_{40} \approx 0.11$, $\lambda_{22} \approx 0.12$, $\lambda_{04} \approx 0.61$), Eq. (50) with $\lambda_{40} \approx 0.11$, and N *Q-D*. Evidently, the simulated crest-to-trough heights follow the N *Q-D* model quite closely. In contrast, the wave heights derived from the corresponding 2ξ series are much larger. And, they tend to follow the GC model of Eq. (49) better, but not with consistent quantitative accuracy over the entire range of simulations.

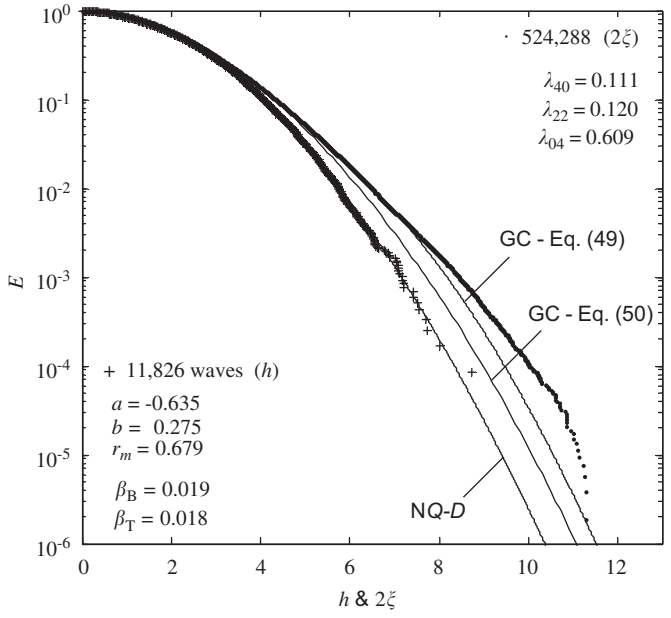

Fig. 13. Nonlinear simulation series #1: exceedance distribution of 11,826 crest-to-trough wave heights (h) and 524,288 envelope heights (2ξ) compared to GC, R and N *Q-D*.

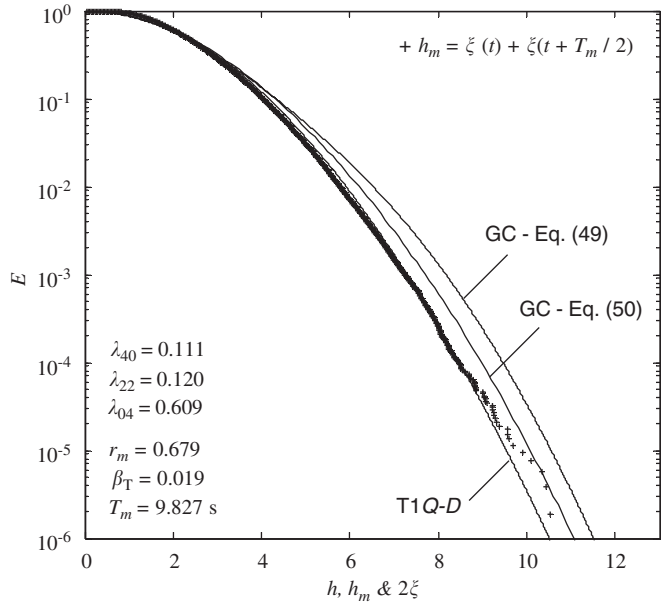

Fig. 14. Nonlinear simulation series #1: EDF of 528,263 scaled wave heights h_m derived from wave-envelope series compared to GC and T1 *Q-D* ($\approx$T *Q-D*).

In comparison, the GC prediction from Eq. (50) does rather poorly. For the same series, Fig. 14 shows the EDF estimates derived from the associated wave envelope via Eq. (51) with $T_m \cong 9.83$ s in a comparison with T1 *Q-D* ($\approx$T *Q-D*) and the same GC predictions as in Fig. 13. In this case, the simulations compare with T1 *Q-D* more favorably, as expected.

Fourth-order joint cumulants and thus Λ are typically estimated from a surface time series and its Hilbert

transform. Alternately, Λ can also be derived from the corresponding wave envelopes because if Eq. (43) is valid, then the mean of ξ assumes the form

$$\langle\xi\rangle = \sqrt{\frac{\pi}{2}}\left(1 - \frac{\Lambda}{64}\right). \tag{53}$$

Given $\langle\xi\rangle$, the alternate estimate for Λ follows from this simply as

$$\Lambda = 64\left(1 - \sqrt{\frac{2}{\pi}}\langle\xi\rangle\right). \tag{54}$$

A similar consideration in terms of higher moments $\langle\xi^j\rangle$ for $j>1$ is not particularly useful since $\langle\xi^2\rangle = 2$ and does not depend on Λ, and the higher moments are less stable than $\langle\xi\rangle$.

4.3. Wave crests and troughs

The upper ($+\xi$) and lower ($-\xi$) wave envelopes represent bounds to the surface elevations, in particular, to wave crests and troughs. Clearly, $+\xi$ and $-\xi$ are symmetrical with respect to the mean water level, and no third-order joint or marginal cumulant representing the effect of second-order nonlinearities to $O(\mu)$ appears in Eqs. (41) and (43). Thus, the resulting probability structure does not correctly describe the vertical asymmetry of the nonlinear surface in terms of higher crests and shallower troughs. In theory, the upper and lower envelope distributions appropriate to wave crest and trough segments can be derived from the joint distribution of ξ and θ, conditional on $\eta>0$ and $\eta\leqslant 0$, respectively. To $O(\mu)$, this approach was first elaborated in Tayfun (1994). It was later corrected and applied to oceanic waves in Al-Humoud et al. (2002). Further applications to shallow-water waves are more recent and given in Cherneva et al. (2005). There is no difficulty in extending the same approach to $O(\mu^2)$ with the joint PDF $p_{\xi\theta}$ in the present case, but this will not be pursued here as the quantitative accuracy of the resulting distributions tend to be somewhat poor (Al-Humoud et al., 2002; Cherneva et al., 2005). This is largely because such distributions tend to describe all local maxima and minima over wave crests and troughs, respectively, instead of just their global values. An alternate but somewhat elaborate third-order theory for describing large wave heights and crests is described in Fedele and Tayfun (2006) based on the concept of stochastic wave groups.

Consider instead a slightly different but a far simpler approach for introducing the surface skewness into Eq. (43) directly. To $O(\mu)$, the representation for the second-order upper wave envelope is given by (Tayfun, 1980, 2006)

$$\xi^+ = \xi_1 + \tfrac{1}{2}\mu^*\xi_1^2, \tag{55}$$

where $\xi_1 \equiv$ the first-order wave envelope, Rayleigh-distributed with the EDF $\exp(-\xi_1^2/2)$, and, $\mu^* = O(\mu)$ and represents a parameter for wave steepness. In theory, $\mu^* \equiv \lambda_{30}/3$, thus rendering Eq. (55) identical to Eq. (22). In general though, μ^* can be defined in a variety of slightly different forms (Tayfun, 2006). For instance, if the principal interest is in the distribution of wave crests over large waves, then $\mu^* = \sigma\omega_m^2/g$ as a somewhat conservative upper bound. More accurately,

$$\mu^* \equiv \mu_{F2} = 16\frac{\alpha_2^3}{\beta_2}\Gamma\left(\frac{3}{\beta_2}\right) - \frac{1}{4}\sqrt{\frac{\pi}{2}}, \tag{56}$$

where α_2 and β_2 represent the dimensionless parameters in Forristall's Weibull fit to the crests of long-crested random waves in deep water (Forristall, 2000). To $O(\mu)$, the EDF of ξ^+ is then given by

$$E_{\xi^+} = \exp\left[-\frac{1}{2\mu^{*2}}(-1 + \sqrt{1 + 2\mu^* x})^2\right]. \tag{57}$$

The modified GC EDF for the upper wave envelope and appropriate to nonlinear wave crests follows to $O(\mu^2)$ from simply replacing the leading term $\exp(-x^2/2)$ in Eq. (43) with Eq. (57) as

$$E_{\xi^+} = \exp\left[-\frac{1}{2}\left(\frac{-1 + \sqrt{1 + 2\mu^* x)}}{\mu^*}\right)^2\right]\left[1 + \frac{\Lambda}{64}x^2(x^2 - 4)\right]. \tag{58}$$

No further modifications are needed in this expression since Λ is $O(\mu^2)$ already.

The modified GC EDF of wave troughs follows the same approach by considering first the second-order narrow-band representation for nonlinear wave troughs, viz.

$$\xi^- = \xi_1 - \tfrac{1}{2}\mu^*\xi_1^2. \tag{59}$$

To avoid negative amplitudes, which arise when $\xi_1 > 2/\mu^*$, this expression is inverted to rewrite it as

$$\xi_1 = \xi^-(1 + \tfrac{1}{2}\mu^*\xi^-). \tag{60}$$

The corresponding EDF is then given by

$$E_{\xi^-} = \exp[-\tfrac{1}{2}x^2(1 + \tfrac{1}{2}\mu^* x)^2]. \tag{61}$$

Thus, the modified GC EDF for the lower wave envelope coincident with wave troughs similarly follows to $O(\mu^2)$ from replacing the leading term $\exp(-x^2/2)$ in Eq. (43) with the preceding EDF as

$$E_{\xi^-} = \exp\left[-\frac{1}{2}x^2\left(1 + \frac{1}{2}\mu^* x\right)^2\right]\left[1 + \frac{\Lambda}{64}x^2(x^2 - 4)\right]. \tag{62}$$

Henceforth, Eqs. (57) and (61) describing crest and trough amplitudes to $O(\mu)$ will be referred to as the NB models, and Eqs. (58) and (62) modified with the NB models for the same variables correct to $O(\mu^2)$ as the NB–GC models.

5. Comparisons with North Sea data

The data utilized for the present comparisons are in the form of fixed-point surface measurements gathered from the TERN platform in 167 m depth in the northern North Sea during two severe storms, one in January 1992 and the

other in January 1993. The duration of the 1992 measurements is 8 h, comprising three segments separated by 2-h intervals. The 1993 data represent continuous measurements, 9 h in total duration. Both data have been sampled at 5.12 Hz. Forristall (2000) elaborates the quality and statistical nature of both of these measurements, and refers to them as Tern 92 and Tern 93a. Here, they will be designated as TERN 92 and TERN 93, respectively.

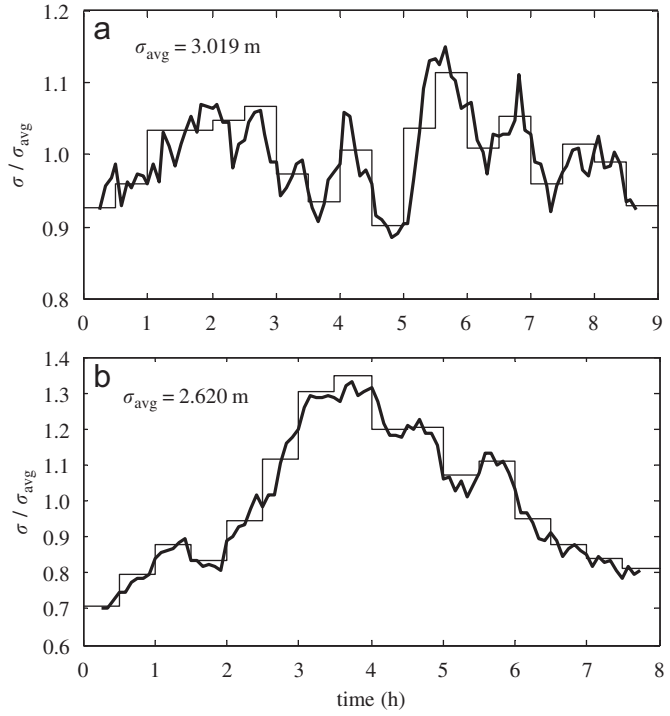

Fig. 15. Temporal variation of σ/σ_{avg} in (a) TERN 93 and (b) TERN 92 derived from $\frac{1}{2}$-h running averages at 1-min intervals (thick), and corresponding $\frac{1}{2}$-h segmental values (thin) used for scaling crest-to-trough wave heights, crests and troughs. Insets: $\sigma_{avg} \equiv$ overall averages.

The temporal variations of σ/σ_{avg}, where $\sigma_{avg} \equiv$ the overall r.m.s. value derived from each data set as a whole, are displayed in the upper part of Fig. 15 for TERN 93 and in the lower part for TERN 92. Evidently, neither record represents a stationary sea state, particularly TERN 92. To compensate for non-stationarity at least partially, all wave heights, crests and troughs extracted via wave-by-wave analyses from both data are scaled with the half-hourly segmental r.m.s. values shown in Fig. 15. This process yields a total of 6153 zero up-crossing waves for the two data sets combined, hereafter referred to as TERN 93 + 93.

The nature of the cumulants of interest to the present analysis, how these relate to one another and the comparisons between their estimates derived by using half-hourly running averages at 5 min intervals from the surface time series and/or from the corresponding envelope series are all briefly summarized in Figs. 16 and 17 for the continuous TERN 93 measurements. The corresponding results for TERN 92 are similar and thus not shown. It is noticed in Fig. 16 that $\lambda_{30}/\lambda_{12} = 3.000$. The same result follows from TERN 92 also. Because this particular relationship is known to be valid for second-order waves and correct to $O(\mu)$ (Tayfun, 1994), it gives the first indication that third-order nonlinearities of $O(\mu^2)$ may not have any significance in TERN 92 and TERN 93, although both cases represent rather severe directional seas with significant wave heights larger than 14 m during the peak storm conditions. On average, $\lambda_{40} = -0.065$, $\lambda_{22} = -0.029$, and $\lambda_{04} = -0.107$ for TERN 93 as a whole. As for TERN 92, $\lambda_{40} = 0.096$, $\lambda_{22} = 0.044$, and $\lambda_{04} = 0.167$ as overall averages. Finally, for the TERN 92 + 93 composite, $\lambda_{40} = 0.0234$, $\lambda_{22} = 0.0089$, and $\lambda_{04} = 0.0290$. All these also suggest that third- or higher-order nonlinearities are not particularly significant for the TERN data.

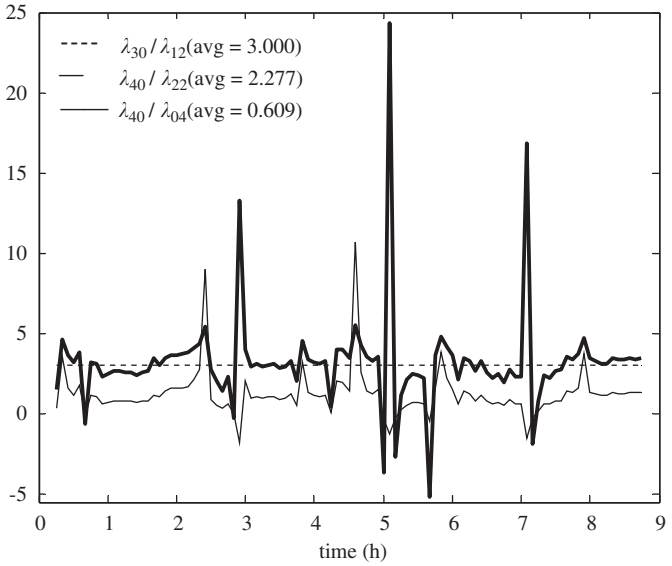

Fig. 16. TERN 93: ratios of third- and fourth-order cumulants derived from $\frac{1}{2}$-h running averages at 5-min intervals. Inset: overall averages.

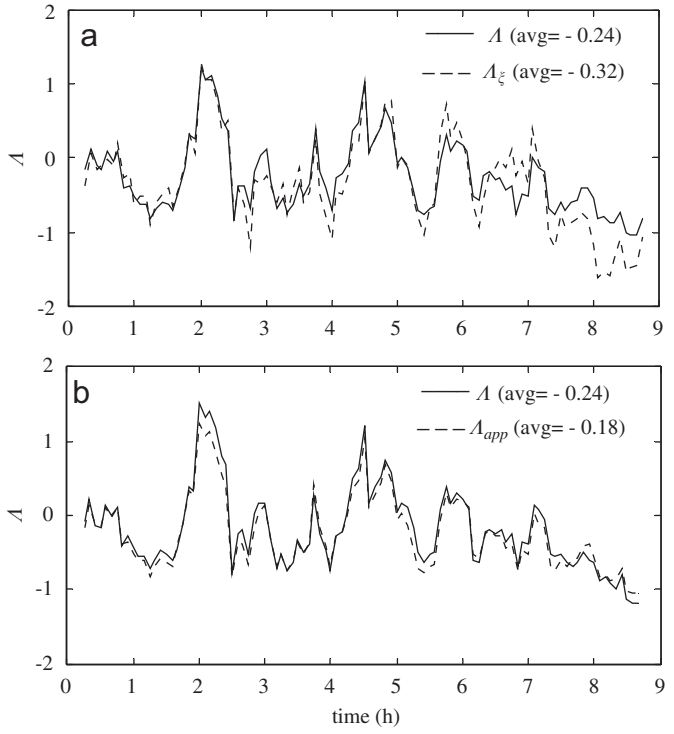

Fig. 17. TERN 93: (a) $\frac{1}{2}$-h running-average estimates of $\Lambda = \lambda_{40} + 2\lambda_{22} + \lambda_{04}$ derived from time series of η and $\hat{\eta}$ in comparison with Λ_ξ from Eq. (54). (b) Same Λ estimates compared to Λ_{app} from Eq. (45). Insets: overall averages.

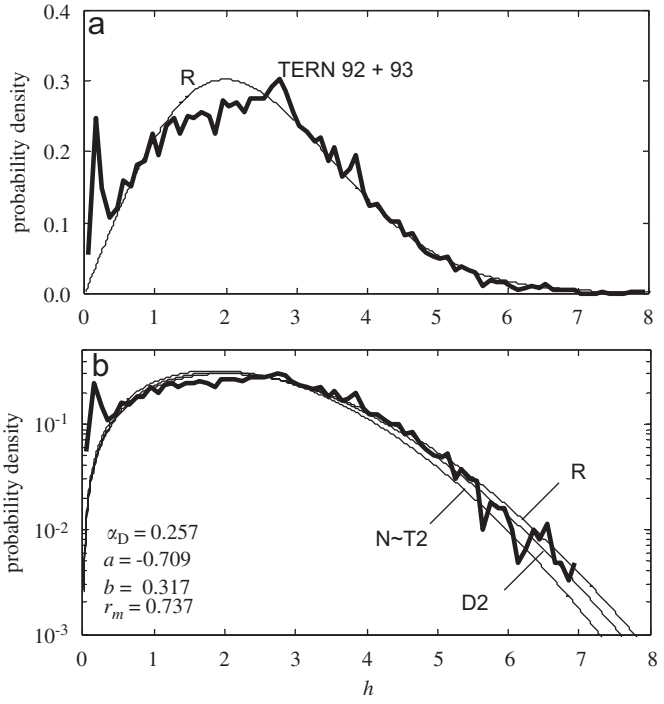

Fig. 18. TERN 92+93: probability density of scaled wave heights compared to theoretical models: (a) data versus R in linear scales; (b) same data versus R, D2, N and T2 in semi-logarithmic scales.

For TERN 93, the comparisons of $\Lambda = \lambda_{40} + 2\lambda_{22} + \lambda_{04}$ derived from the time series of η and $\hat{\eta}$, and from the corresponding mean $\langle\xi\rangle$ of the wave-envelope series via Eq. (54) are shown in the upper part of Fig. 17. The lower part of the same figure shows a similar comparison between the same Λ estimates and Λ_{app} from Eq. (47). Both sets of comparisons support the validity of the approximations Λ_ξ and Λ_{app} for Λ reasonably well for most of TERN 93. However, the comparisons between the corresponding overall averages appear to contradict this conclusion.

The comparison between the PDF estimates for 6153 crest-to-trough wave heights in TERN 92+93 and R is shown in the upper part of Fig. 18, using linear scales to emphasize the bimodal nature of the observed PDF. The semi-logarithmic comparison of the same PDF estimates in the lower part of Fig. 18 with R, D2, N and T2 is intended to show the tail-end behavior over high waves more clearly. In this case, all the theoretical forms considered appear to describe the trend of the observed PDF reasonably well for $h > 3$.

The comparison between the EDF estimates of scaled crest-to-trough wave heights in TERN 92+93 and the predictions from the GC model of Eq. (49), R, B and T is in Fig. 19, excluding other linear and nonlinear models for clarity. The comparison that includes all the present linear models as well as D2, D5 and GC is shown in Fig. 20 more clearly in terms of the observed ratio h/h_{R}. Both cases

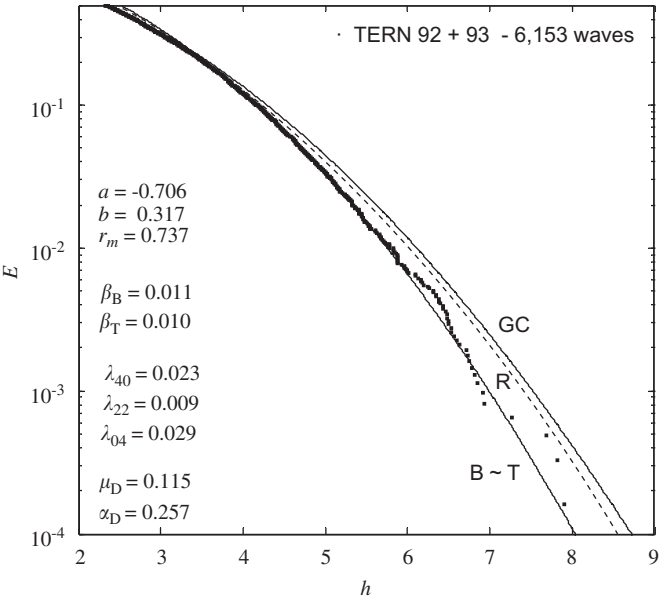

Fig. 19. TERN 92+93: exceedance distribution of scaled crest-to-trough wave heights compared to GC [Eq. (49)], R and B $\approx$ T.

indicate that neither R nor GC represents the data well at all. The observed EDF is best described by B, if allowance is made for the high variability of the largest ten or so values of the observed ratio h/h_{R}. Since the data trend is so well predicted by B, no other second-order nonlinear model of *Q-D* type is included in these comparisons.

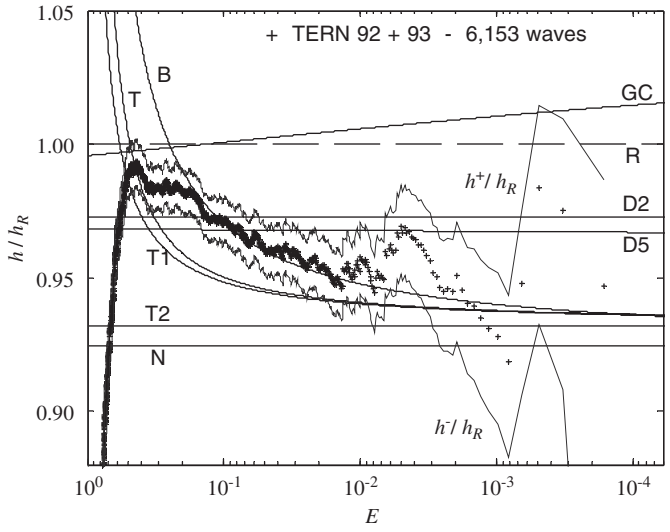

Fig. 20. TERN 92+93: observed ratio h/h_R compared to predictions from linear models and from nonlinear GC [Eq. (49)], D2 and D5. $h^{\pm}/h_R \equiv$ stability bands.

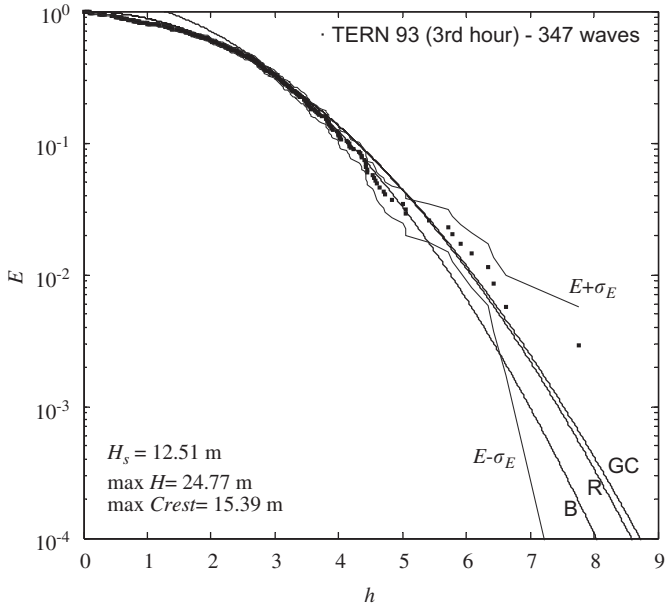

Fig. 21. Effects of sample size n on EDF estimates and their stability: $n = 347$ waves from third hourly segment of TERN 93 containing max $H = 24.77$ m compared to R and B. $E \pm \sigma_E \equiv$ stability bands.

The effects of sample size n on the EDF estimates and their stability are shown in Fig. 21 for $n = 347$ waves extracted from the third hourly segment of TERN 93 containing the largest of all waves in TERN 92+93. The rest of the figure contains the same theoretical predictions as those in Fig. 18 for comparison. Clearly, Fig. 21 shows that some large wave heights, specifically the largest eight are underestimated rather noticeably by B. The same wave heights also overshoot even the more conservative predictions from both R and GC. The specifics of the largest wave nearly conform to some heuristic definitions of a 'freak' wave. It has a relative frequency of occurrence of 1 in 347 in Fig. 21. In contrast, Fig. 19 for the same wave yields a frequency of 1 in 6153 waves, a far more reliable and realistic estimate but widely different from the previous one. All this really illustrates, as the Hurricane Ivan example in Forristall (2005) does but only somewhat more dramatically, that the statistically unstable nature of estimates based on short records can lead to false conclusions about the nature and the relative frequency of occurrence of unusually large waves. This is a point of concern for the comparisons in Fig. 3 of Mori and Yasuda (2002) and also in Fig. 1 of Mori and Janssen (2006). Both cases rely on wave-flume data comprising less than 450 waves to demonstrate that the observed wave-height EDFs deviate noticeably from R due to third-order nonlinearities. But, the comparisons based on much larger samples from oceanic measurements also given in Fig. 4 of Mori and Yasuda (2002) clearly contradict this, with the observed data showing by and large the same deviations from R as in the present TERN results.

The linear and nonlinear conditional means from Eqs. (8) and (31) are shown in Fig. 22 in a comparison with the estimates from TERN 92+93 for $n \geqslant 10$. Only the predictions from R, D2, B and B *Q-D* are shown in this figure since B$\approx$ all Ts and B *Q-D* $\approx$all T *Q-D*'s and are thus not shown for clarity. N *Q-D* overestimates the data trend somewhat, falling slightly below B *Q-D*, and N underestimates it and lies just below the B model and the observed data. Evidently, the data trend compares very favorably with the linear predictions $(h_1)_{1/n}$ that follow from Eq. (8) for B ($\approx$T$\approx$T1$\approx$T2). Some discrepancy appears toward the extreme tail as data become sparse.

For TERN 92+93, the expected maxima of N wave heights for $N \geqslant 10^2$ are compared in Fig. 23 with the linear predictions from Eqs. (11) for R, B ($\approx$T1), T2, N and D2 as a nonlinear example. The predictions from second-order nonlinear *Q-D* models tend to overestimate the TERN

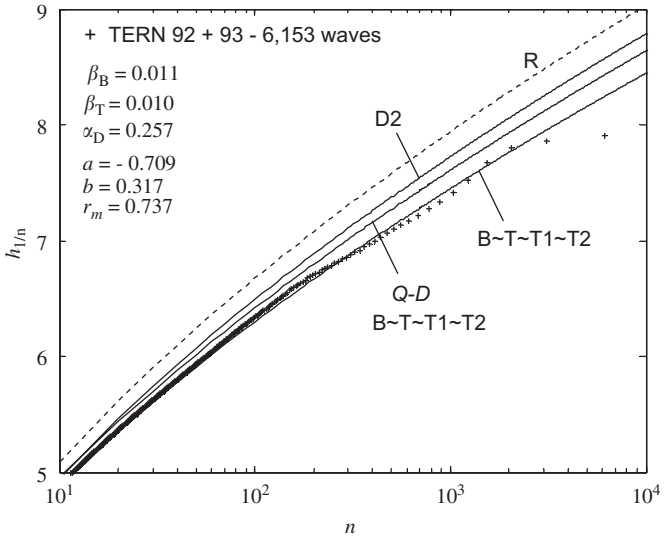

Fig. 22. TERN 92+93: observed conditional means compared to predictions from linear and nonlinear models.

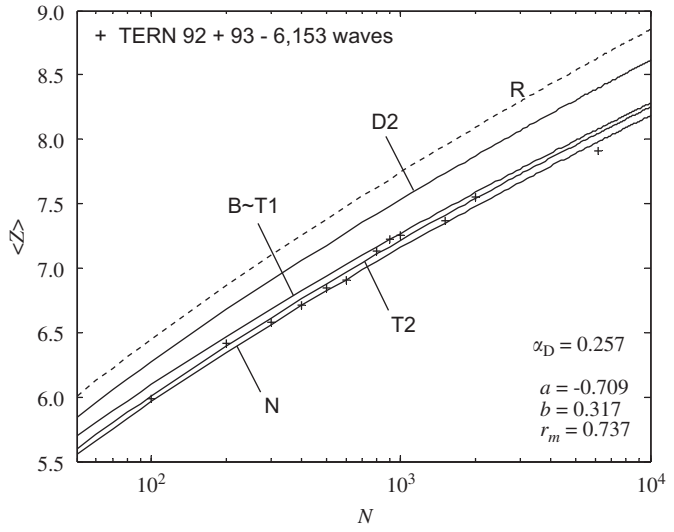

Fig. 23. TERN 92+93: expected largest wave height in N waves compared to predictions from linear R, B, T1, T2, N and from second-order D2.

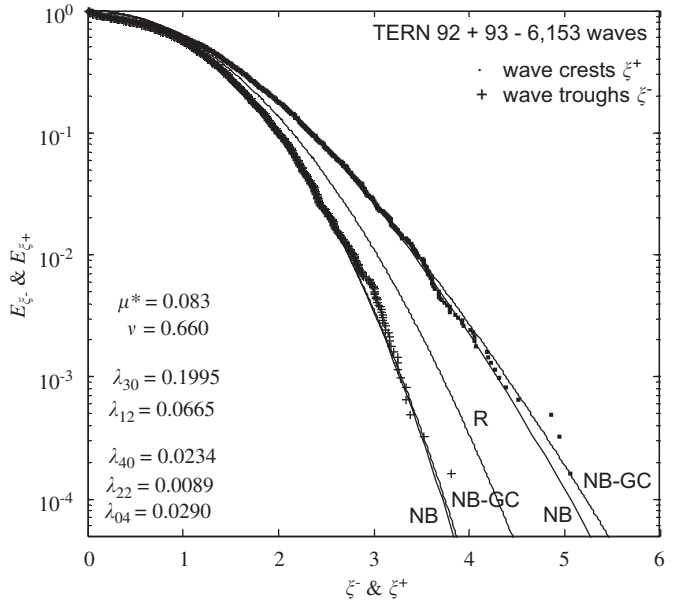

Fig. 24. TERN 92+93: exceedance distribution of scaled wave crest and trough amplitudes compared to NB [Eqs. (57) and (61)], NB-GC [Eqs. (58) and (62)], and R.

92+93 values rather noticeably, as D2 does, and are thus excluded from the comparison. The data trend seems to be described reasonably well by all linear theoretical predictions from B, T1, T2 and N.

The comparisons of the EDF estimates for the scaled wave crest and trough amplitudes in TERN 92+93 with the theoretical predictions from the NB models of Eqs. (57) and (61), and from the NB–GC models of Eqs. (58) and (62) are shown in Fig. 24. Clearly, NB and NB–GC predictions are nearly the same for wave troughs. The NB–GC model over predicts the crest heights over large waves slightly, but this is largely inconsequential in this case because of the large variability of the estimates toward the extreme tail. So, these results also confirm that third- or higher-order nonlinearities do not affect the statistics of wave crest and troughs in TERN 92+93. The observed deviations from R appear to be entirely due to second-order bound harmonics, and are thus described quite accurately with the second-order NB model.

6. Comparison with 3D simulations (Socquet-Juglard et al., 2005)

Dysthe et al. (2005), Socquet-Juglard (2005) and Socquet-Juglard et al. (2005) performed 3D numerical simulations based on the Dysthe equation (Dysthe, 1979; Trulsen and Dysthe, 1996), a modified form of the NLS equation appropriate to directional deep-water waves characterized by large steepness and broader spectra. The simulation results clearly show that second-order nonlinearities are dominant for realistic directional distributions, and that the surface-elevation and crest-height statistics are described surprisingly well with the second-order narrow-band model (Tayfun, 1980). They also show that third-order four-wave interactions largely affect long-crested narrow-band waves, causing their statistics to deviate from the conventional linear and second-order models of R and NB type. Case C in Fig. 9 of Socquet-Juglard (2005) demonstrates such a case for the statistics of wave crests. It is reproduced in Fig. 25 here and compared with the predictions from R, NB from Eq. (57) and NB–GC from Eq. (58). In this case, $\mu^* \approx 0.07$, $\lambda_{40} \approx 0.4$, and it is assumed that $\lambda_{22} \cong \lambda_{40}/3$ and $\lambda_{04} \cong \lambda_{40}$ since no estimates for λ_{22} and λ_{04} are given. It is evident that, whereas the simulated data trend deviates from R and NB rather

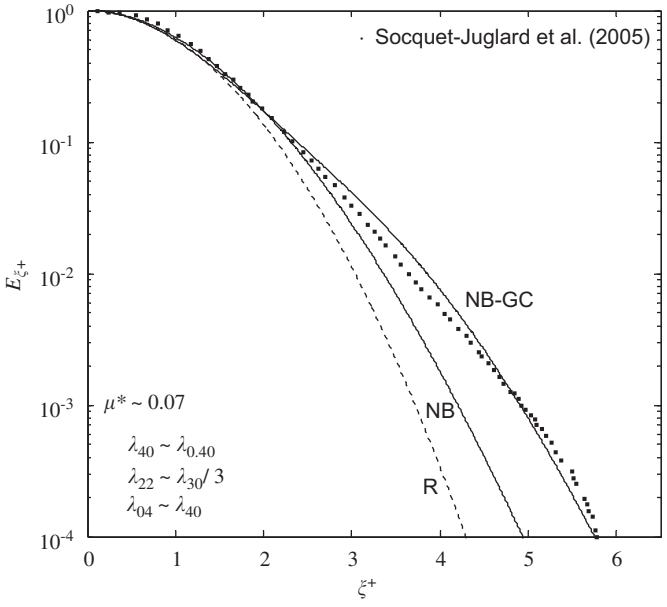

Fig. 25. Exceedance distribution of scaled wave crests from Fig. 9-case C in Socquet-Juglard et al. (2005) compared to R, NB [Eq. (57)] and NB-GC [Eq. (58)].

noticeably, it is described reasonably well by the present NB–GC model.

7. Comparisons with wave-flume data (Onorato et al., 2004, 2005)

The analyses and comparisons in Onorato et al. (2004, 2005, 2006), Janssen (2005), and Mori and Janssen (2006) with 2D waves mechanically generated in wave flumes indicate that wave heights and crests significantly larger than those typically observed under oceanic conditions do occur, provided that the waves initially generated at the wave maker are sufficiently narrow-band. The observed distributions of such wave heights and crests deviate substantially from both the conventional Rayleigh law and the second-order NB model. This is explained in terms of the Benjamin–Feir type modulational instabilities associated with the NLS equation. It appears that whenever the Benjamin–Feir index is sufficiently large, the third-order four-wave quasi-resonant interactions between free modes become at least as significant as the second-order interactions due to bound modes, thus amplifying wave heights and crests further and well beyond the typical trends observed under oceanic conditions. All this should be reflected eventually in the third- and fourth-order marginal and joint cumulants, but the emphasis is placed without exception on λ_{40} only. As none of the above-referenced studies provide any data or estimates on λ_{22} and λ_{04}, it is not possible to ascertain if the approximations $\lambda_{40} \cong 3\lambda_{22}$ and $\lambda_{04} \cong \lambda_{40}$ really hold. But, this will be so assumed in what follows as in the previous comparison with the 3D simulations from the Dysthe equation.

As was mentioned previously, the wave-flume data reported in Mori and Yasuda (2002) and Mori and Janssen (2006) are rather sparse for reliable comparisons. In contrast, those reported in Onorato et al. (2004, 2005, 2006) represent relatively large samples. As a particularly striking case, the EDF estimates of scaled wave crests in Fig. 2 of Onorato et al. (2005) are reproduced and shown in Fig. 26 here together with the predictions from R, NB wave-crest model of Eq. (57) and the corresponding NB–GC model from Eq. (58). It is seen that the observed wave crests deviate dramatically from R and NB, but they do appear to be described by NB–GC reasonably well.

A somewhat similar result on the EDF estimates for wave heights from the same wave-flume experiments is given in Fig. 6 of Onorato et al. (2004). Using the present notation, it is replicated in Fig. 27 here and compared to R and GC from Eq. (49). The estimates $\lambda_{40} \cong 0.80$ and $\lambda_{40} \cong 1.00$ inset in Figs. 26 and 27, respectively, are from Fig. 2 in Onorato et al. (2004). Further, note that the predictions from Eqs. (49) and (50) are the same in this case because of the assumption that $\lambda_{40} \cong 3\lambda_{22}$ and $\lambda_{04} \cong \lambda_{40}$. All this aside, it is seen that, whereas the observed data overshoots R dramatically again as in the previous case of Fig. 26, it is described by the GC model quite favorably. This result

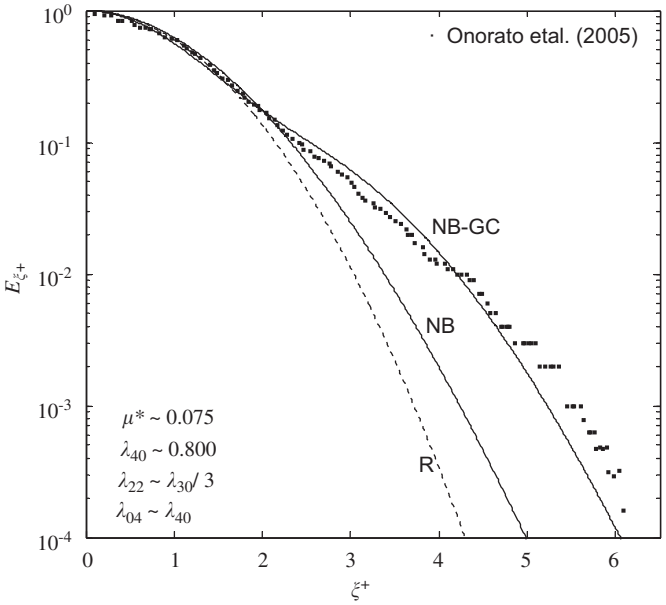

Fig. 26. Exceedance distribution of scaled wave crests from Fig. 2 in Onorato et al. (2005) compared to R, NB [Eq. (57)] and NB-GC [Eq. (58)].

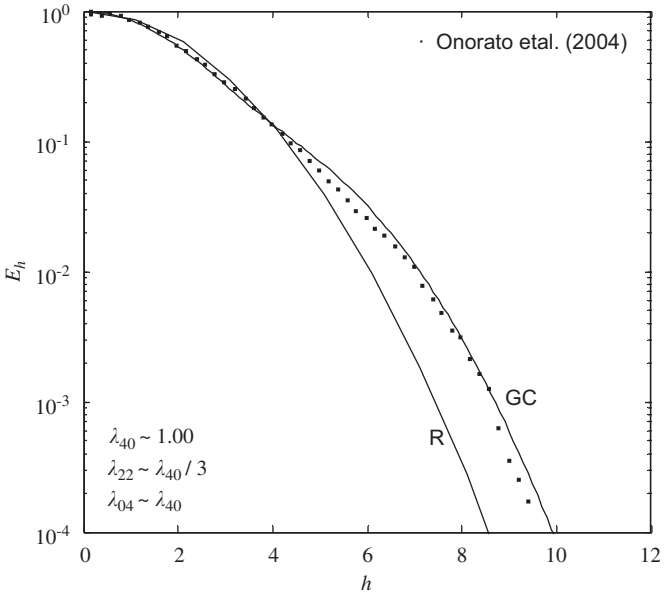

Fig. 27. Exceedance distribution of the scaled wave heights from Fig. 6 in Onorato et al. (2004) compared to GC [Eq. (49) or Eq. (50)] and R.

supports the Mori–Janssen contention that the distribution of wave heights is described well by MER, the modified form of the GC series dependent only on λ_{40}, provided that waves are long-crested and narrow-band.

8. Conclusions

The results from the present 2D linear simulations suggest that the models of Boccotti (1989) and Tayfun (1990a) yield similar predictions, and describe the data trends over relatively large waves quite accurately. The

widely popular Naess (1985) model under predicts the linear wave heights somewhat, as in past comparisons (Tayfun, 1990a). Both approximations T1 and T2 of the Tayfun (1990a) model are just as simple as the Naess model, and they also describe the linear wave heights slightly and consistently better.

The 2D nonlinear simulations for narrow-band long-crested seas lead to a re-distribution of wave heights toward slightly higher waves. For this extreme and plausibly rather unrealistic case, the nonlinear models considered do not describe the simulated data trends quite satisfactorily. Of all the nonlinear extensions considered, the quasi-deterministic form of the Naess model and Dawson's second-order Stokes–Rayleigh model seem to best describe the simulated wave heights. However, the analysis of the TERN data does not indicate any similar nonlinear effects whatsoever, and the observed wave heights are described extremely well by Boccotti's linear model. This result and those in Forristall (2005) decisively demonstrate that second- and higher-order nonlinearities do not affect the heights of large oceanic waves. And, an unusually large or freak wave in a short record that does not seem to follow an established probability law can usually be explained as a relatively rare occurrence in the same record when observed sufficiently longer.

Mori and Janssen (2006) and Onorato et al. (2004, 2005, 2006) show that the heights of 2D narrow-band waves generated in wave flumes can be amplified rather noticeably due to third-order nonlinear interactions. In particular, Onorato et al. (2006) illustrate that the conventional linear and second-order probability laws do not explain the observed statistics of wave heights and crests in the presence of such interactions. Evidently, the third-order GC models considered here seem to describe such cases reasonably well. Undoubtedly, further comparisons would be needed to ascertain if similarly favorable comparisons can be replicated under different conditions. More significant, however, is whether if the third-order nonlinearities and thus the GC type models have any real relevance to oceanic waves. The analyses of the TERN data indicate that this is not the case at all, and that the observed statistics are described extremely well by the linear models for wave heights and the second-order narrow-band models for wave crests and troughs. The nature of results from third-order directional simulations of Dysthe et al. (2005), Socquet-Juglard (2005) and Socquet-Juglard et al. (2005) lends strong support to this conclusion. Nevertheless, further comparisons with even larger populations of oceanic measurements of the sea surface as a function of both space and time may be needed to resolve this issue unequivocally and once for all.

Acknowledgments

The authors are indebted to G.Z. Forristall, Ocean Engineering, Inc., Camden, ME, USA, for the TERN data. They also thank M. Onorato and H. Socquet-Juglard for complimentary and/or in-press copies of their publications, K. Dysthe for valuable comments and, in particular, H. Krogstad for providing the authors with some key background material and insight on a variety of concepts.

References

Al-Humoud, J., Tayfun, M.A., Askar, H., 2002. Distribution of nonlinear wave crests. Ocean Engineering 29, 1929–1943.

Ang, H.-S.A., Tang, W.H., 1984. Probability Concepts in Engineering Planning and Design. II: Decision, Risk, and Reliability. Wiley, New York.

Arena, F., 2005. On non-linear very large sea wave groups. Ocean Engineering 32 (11–12), 1311–1331.

Boccotti, P., 1989. On mechanics of irregular gravity waves. Atti Accademia Nazionale dei Lincei, Memorie VIII, vol. 19, pp. 111–170.

Boccotti, P., 2000. Wave Mechanics for Ocean Engineering. Elsevier Science, Oxford.

Borgman, L.E., Resio, D.T., 1982. Extremal statistics in wave climatology. In: Proceedings on Topics in Ocean Physics, Italian Physical Society. Bologna, Italy, pp. 439–471.

Cherneva, Z., Petrova, P., Andreeva, N., Guedes Soares, C., 2005. Probability distributions of peaks, troughs and heights of wind waves measured in the Black Sea coastal zone. Coastal Engineering 52, 599–615.

Dawson, T.H., 2004. Stokes correction for nonlinearity of wave crests in heavy seas. Journal of Waterway, Port, Coastal, and Ocean Engineering 130 (1), 39–44.

Dysthe, K.B., 1979. Note on a modification to the nonlinear Schrödinger equation for application to deep water waves. Proceedings of the Royal Society of London, Series A 369, 105–114.

Dysthe, K., Trulsen, K., Krogstad, H., Socquet-Juglard, H., 2003. Evolution of a narrow-band spectrum of random surface gravity waves. Journal of Fluid Mechanics 478, 1–10.

Dysthe, K., Socquet-Juglard, H., Trulsen, K., Krogstad, H.E., Liu, J., 2005. Freak waves and large-scale simulations of surface gravity waves. In: Proceedings of the 14th 'Aha Huliko'a Hawaiian Winter Workshop. University of Hawaii, pp. 91–99.

Fedele, F., 2006. Extreme events in nonlinear random seas. Journal of Offshore Mechanics and Arctic Engineering 128, 11–16.

Fedele, F., Arena, F., 2003. On the statistics of high nonlinear random waves. In: Proceedings of the 13th International Offshore and Polar Engineering Conference. ISOPE, Honolulu, Hawaii, USA, pp. 17–22.

Fedele, F., Arena, F., 2005. Weakly nonlinear statistics of high random waves. Physics of Fluids 17 (1) Paper No. 026601.

Fedele, F., Tayfun, M.A., 2006. Extreme events and stochastic wave groups. In: 25th International Conference on Offshore Mechanics and Arctic Engineering, ASME, Hamburg, Germany, Paper No. OMAE 2006-92527.

Forristall, G.Z., 1984. The distribution of measured and simulated wave heights as a function of spectral shape. Journal of Geophysical Research 89 (C6), 547–552.

Forristall, G.Z., 2000. Wave crest distributions: observations and second-order theory. Journal of Physical Oceanography 38 (8), 1931–1943.

Forristall, G.Z., 2005. Understanding rogue waves: are new physics really necessary? In: Proceedings of the 14th 'Aha Huliko'a Hawaiian Winter Workshop. University of Hawaii, pp. 29–35.

Forristall, G.Z., 2006. Maximum crest heights over an area and the air gap problem. In: 25th International Conference on Offshore Mechanics and Arctic Engineering, ASME, Hamburg, Germany, Paper No. OMAE 2006-92022.

Guedes Soares, C., Pascoal, R., 2005. On the profile of large ocean waves. Journal of Offshore Mechanics and Arctic Engineering 127, 306–314.

Haver, S., Andersen, O.A., 2000. Freak waves: rare realizations of a typical population or typical realizations of a rare population? In: Proceedings of the 10th International Offshore and Polar Engineering Conference. ISOPE, Seattle, USA, pp. 123–130.

Janssen, P.A.E.M., 2003. Nonlinear four-wave interactions and freak waves. Journal of Physical Oceanography 33 (4), 863–884.

Janssen, P.A.E.M., 2005. Nonlinear four-wave interactions and freak waves. In: Proceedings of the 14th 'Aha Huliko'a Hawaiian Winter Workshop on Rogue Waves. University of Hawaii, pp. 85–90.

Krogstad, H.E., Liu, J., Socquet-Juglard, H., Dysthe K.B., Trulsen, K., 2004. Spatial extreme value analysis of nonlinear simulations of random surface waves. In: Proceedings of the 23rd International Conference on Offshore Mechanics and Arctic Engineering, ASME, Vancouver, British Columbia, Canada, Paper No. OMAE 2004-51336.

Lindgren, G., 1970. Some properties of a normal process near a local maximum. Annals of Mathematical Statistics 4 (6), 1870–1883.

Lindgren, G., 1972. Local maxima of Gaussian fields. Arkiv för Matematik 10, 195–218.

Longuet-Higgins, M.S., 1963. The effect of nonlinearities on statistical distributions in the theory of sea waves. Journal of Fluid Mechanics 17, 459–480.

Longuet-Higgins, M.S., 1964. Modified distributions for slightly nonlinear variables. Radio Science Journal of Research 68D (9), 1049–1062.

Mori, N., Janssen, P.A.E.M., 2006. On kurtosis and occurrence probability of freak waves. Journal of Physical Oceanography 36, 1471–1483.

Mori, N., Yasuda, T., 2002. A weakly non-Gaussian model of wave height distribution for random wave train. Ocean Engineering 29, 1219–1231.

Naess, A., 1985. On the distribution of crest to trough wave heights. Ocean Engineering 12 (3), 221–234.

Onorato, M., Osborne, A.R., Serio, M., Cavaleri, L., Brandini, C., Stansberg, C.T., 2004. Observations of strongly non-Gaussian statistics for random sea surface gravity waves. Physical Review E 70, 1–4.

Onorato, M., Osborne, A.R., Serio, M., 2005. On deviations from Gaussian statistics for surface gravity waves. In: Proceedings of the 14th 'Aha Huliko'a Hawaiian Winter Workshop on Rogue Waves. University of Hawaii, pp. 79–83.

Onorato, M., Osborne, A.R., Serio, M., Cavaleri, L., Brandini, C., Stansberg, C.T., 2006. Extreme waves, modulational instability and second-order theory: wave flume experiments on irregular waves. European Journal of Mechanics B/Fluids 25 (5), 586–601.

Phillips, O.M., Gu, D., Donelan, M., 1993. Expected structure of extreme waves in a Gaussian sea. Part I: theory and SWADE buoy measurements. Journal of Physical Oceanography 23, 992–1000.

Piterbarg, V.I., 1996. Asymptotic methods in the theory of Gaussian processes and fields. AMS Translations of Mathematical Monographs, vol. 148. Providence, Rhode Island.

Socquet-Juglard, H., 2005. Spectral evolution and probability distributions of surface ocean gravity waves and extreme events. DSc. Thesis, Department of Mathematics, University of Bergen, Norway.

Socquet-Juglard, H., Dysthe, K., Trulsen, K., Krogstad, H.E., Liu, J., 2005. Probability distributions of surface gravity waves during spectral changes. Journal of Fluid Mechanics 542, 195–216.

Stansell, P., 2004. Distributions of freak wave heights measured in the North Sea. Applied Ocean Research 26, 35–48.

Tayfun, M.A., 1980. Narrow-band nonlinear sea waves. Journal of Geophysical Research 85 (C3), 1548–1552.

Tayfun, M.A., 1981. Distribution of crest-to-trough wave heights. Journal of Waterway, Port, Coastal, and Ocean Engineering 107 (WW3), 149–158.

Tayfun, M.A., 1983. Frequency analysis of wave heights based on wave envelope. Journal of Geophysical Research 88 (C12), 7573–7587.

Tayfun, M.A., 1984. Nonlinear effects on the distribution of amplitudes of sea waves. Ocean Engineering 11 (3), 245–264.

Tayfun, MA., 1986. On narrow-band representation of ocean waves. I. Theory and 2. Simulations. Journal of Geophysical Research 91 (C6), 7743–7759.

Tayfun, M.A., 1990a. Distribution of large wave heights. Journal of Waterway, Port, Coastal, and Ocean Engineering 116 (6), 686–707.

Tayfun, M.A., 1990b. High-wave-number/frequency attenuation of wind-wave spectra. Journal of Waterway, Port, Coastal, and Ocean Engineering 116 (3), 381–398.

Tayfun, M.A., 1993. Joint distribution of large wave heights and associated periods. Journal of Waterway, Port, Coastal, and Ocean Engineering 119 (3), 261–273.

Tayfun, M.A., 1994. Distributions of wave envelope and phase in weakly nonlinear random waves. Journal of Engineering Mechanics 120 (5), 1009–1025.

Tayfun, M.A., 2006. Statistics of nonlinear wave crests and groups. Ocean Engineering 33 (11–12), 1589–1622.

Tayfun, M.A., Fedele, F., 2006. Wave-height distributions and nonlinear effects. In: Proceedings of the 25th International Conference on Offshore Mechanics and Arctic Engineering, ASME, Hamburg, Germany, Paper No. OMAE 2006-92019.

Tayfun, M.A., Lo, J.-M., 1990. Non-linear effects on wave envelope and phase. Journal of Waterway, Port, Coastal, and Ocean Engineering 116 (1), 79–100.

Trulsen, K., Dysthe, K.B., 1996. A modified nonlinear Schrödinger equation for broader bandwidth gravity waves on deep water. Wave Motion 24, 281–289.

Trulsen, K., Stansberg, C.T., 2001. Spatial evolution of water surface waves: numerical simulation and experiment of bichromatic waves. In: Proceedings of the 11th International Offshore and Polar Engineering Conference. ISOPE, Stavanger, Norway, pp. 71–77.

Walker, D.A.G., Taylor, P.H., Taylor, R.E., 2004. The shape of large surface waves on the open sea and the Draupner New Year wave. Applied Ocean Research 26, 73–83.