Groups of Waves in Shallow Water

STEVE ELGAR, R. T. GUZA, AND R. J. SEYMOUR

Scripps Institution of Oceanography

Wave group statistics predicted by linear theories are compared to numerical simulations, thus determining ranges of spectral shapes for which the theories are valid. It is found that these theories are not generally valid for ocean data because of many assumptions and simplifications beyond linearity and random phase or because their range of applicability does not include the vast majority of ocean conditions. The simulations also provide quantitative information about the variability of linear wave group statistics which is useful when examining ocean field data. The simulation technique is used to show that important ocean gravity wave group statistics are not inconsistent with an underlying wave field composed of linearly superposed random waves. The majority of the field data examined were collected in 10 m depth; significant wave heights varied from about 20 to 200 cm, and the spectral shapes ranged from fairly narrow to broad $(1 < Q_p < 6)$. For the 10-m depth data, the observed mean run length, variance of run length, and probabilities of runs of a given number of waves were statistically consistent with the simulations. In contrast to the apparently linear groups observed in 10 m depth, waves in 2–3 m depth showed marked departures from the linear simulations.

1. INTRODUCTION

Groups, or sequences, of high waves are of interest to coastal engineers and naval architects. A run or group of waves is defined as a sequence of waves, the heights of which exceed a particular level [*Goda*, 1970]. There are several linear theories which predict wave group statistics, such as the mean group length, given only the energy spectrum. Comparisons of group statistics predicted by these theories with a limited set of numerical simulations have shown that the theories are restricted to a particular range of spectral shapes, some for narrow band spectra and others for broadband spectra [*Goda*, 1976]. The first part of this paper (section 2) will review these theories, some of which are shown to contain internal inconsistencies. Numerical simulations with a wide variety of test spectra show the theories generally to have a more restricted range of accuracy than was inferred by the unimodal spectral simulations of *Goda* [1976]. The simulations also provide quantitative information about the distributions, and hence the variability, of linear wave group statistics. The statistical variability of certain wave group parameters was noticed by *Goda* [1983] and is quantified in sections 3.1 and 3.3. The simulations are crucial when examining ocean field data because the majority of the ocean data considered here do not lie within the range covered by the linear theories. There is also a large (and growing) literature studying amplitude modulations (i.e., grouping) which arise as a consequence of nonlinearity (see *Yuen and Lake* [1980] for a recent review). These theories are not considered at all. The question at hand is whether observed groups for the range of conditions investigated are consistent with linear dynamics.

Previous observations of ocean group statistics have been limited to a narrow range of spectral conditions [*Chakrabarti et al.*, 1974; *Goda*, 1983], have not given the relevant spectral parameters [*Wilson and Baird*, 1970; *Rye*, 1974], or have used such short time series that the statistical stability of the observations is doubtful [*Andrew and Borgman*, 1981; *Goda*, 1983]. In spite of these limitations, these studies show rough, qualitative agreement between linear theory and observations. No instances have been found of quantitative comparisons be-

tween observed group statistics and those predicted by linear theory. The present investigation makes use of longer records, representing widely varying wave conditions, to study wave group statistics. These ocean field data are compared to theoretical results based on the measured spectrum (these are of limited utility) and to numerical simulations of linear waves having the same spectrum as the observed waves (section 3). It is seen that the fundamental assumption of linear random waves produces group characteristics not statistically inconsistent with data in water 10 m deep, although this does not imply that other statistics, such as sea surface skewness, are consistent with linear dynamics. On the other hand, as the waves shoal and become more nonlinear, the linear representation is no longer valid, and simulated linear group statistics differ substantially from those observed in the field data.

2. EXISTING THEORIES

2.1. Broad Spectra

One approach to predicting group statistics considers the wave field to be composed of a succession of discrete, independent waves, an assumption appropriate for broadband spectra. Results from the theory of runs are then employed to determine certain group statistics [*Goda*, 1970; *Nagai*, 1973]. Denote the probability that the height of a wave is greater than some critical value, H_c, as p. The probability that a sequence of waves higher than H_c contains j waves is [*Goda*, 1970]

$$P(j) = p^{(j-1)}(1 - p) \tag{1}$$

The mean length of such runs is

$$E[j] = \frac{1}{1 - p} \tag{2}$$

where E is the expected value operator and the standard deviation of run length is

$$\sigma[j] = \frac{p^{1/2}}{1 - p} \tag{3}$$

If the wave field is a linear, Gaussian process, and the underlying spectrum is narrow, then the wave heights are distributed

with a Rayleigh probability density [*Longuet-Higgins*, 1952]:

$$P(H) = \frac{2H}{H_{rms}^2} \exp\{-H^2/H_{rms}^2\} \qquad (4)$$

where H is the crest to trough wave height and H_{rms} is the root mean square wave height. Thus,

$$p = P(H > H_c) = \int_{H_c}^{\infty} \frac{2H}{H_{rms}^2} \exp\{-H^2/H_{rms}^2\}\, dH$$

$$= \exp\{-H_c^2/H_{rms}^2\} \qquad (5)$$

For the Rayleigh distribution,

$$H_{rms} = 2^{3/2}\, m_0^{1/2}$$

where m_0 is the variance of the time series [*Longuet-Higgins*, 1952], so

$$p = \exp\{-H_c^2/8m_0\} \qquad (6)$$

A common⠀⧸ selected value for the critical level is

$$H_c = 4m_0^{1/2} = H_{1/3}$$

where $H_{1/3}$ is the average of the highest 1/3 waves. Combining this with (6) and (1) yields

$$p = 0.1348 \qquad (7a)$$

$$1 - p = 0.8652 \qquad (7b)$$

$$E[j_{1/3}] = 1.16 \qquad (7c)$$

$$\sigma[j_{1/3}] = 0.42 \qquad (7d)$$

The above theory has an internal inconsistency in that a wave field cannot have both the narrow band spectrum required for the Rayleigh distribution of heights and the broad spectrum required for independent successive waves. However, the Rayleigh distribution provides a remarkably robust description of ocean wave heights [*Goda*, 1974, and many others]. Consequently, discrepancies between the mean length of runs calculated by using the above independent wave theory and those observed, either in numerical simulations or field data, may be due to dependence of successive waves, rather than to deviations from the Rayleigh distribution. The numerical simulations of *Goda* [1970] show that the best agreement of the "independent Rayleigh" model (7) does indeed occur for a broad spectrum. As the spectrum narrows, discrepancies with (7) become greater due to increasing correlations between successive waves. Waves were defined with a zero-upcrossing methods.

Kimura [1980] extended Goda's theory by allowing for correlations between successive waves, although the correlation of waves lagged by more than one successive wave was assumed to be zero. There is some limited observational support for this assumption [*Wilson and Baird*, 1970; *Rye*, 1974; *Thompson*, 1981], but the assumption must break down in the limit of a very narrow spectrum. Nevertheless, with the assumed one wave correlation, the mean length of runs of high waves can be evaluated for different values of the correlation coefficient. When the correlation is zero, the results are the same as Goda's (equation (7)). Like Goda, Kimura compares his theoretical results with wave trains numerically simulated from a specified spectral shape (target spectrum). Kimura obtains the correlation coefficient required in the theory from the numerical simulations. As the correlation increases, the mean length of run also increases, in both the theory and numerical

simulations, while Goda's theory (equation (7c)) predicts a constant mean run length. For the fairly broad spectra used, Kimura's theoretical results are quite good and extend Goda's range of accuracy to narrower spectra. However, it is not clear how well this theory will do in general because Kimura's transition matrix requires that waves lagged by more than one successive wave are independent, which does not necessarily follow from the observation that such waves are uncorrelated. Indeed, *Goda* [1983] indicates that the mean run lengths observed in ocean swell with significant correlations at lags of two and three waves are greater than Kimura's theory predicts. Furthermore, *Kimura's* [1980] scheme has the undesirable property that numerical simulations are necessary to determine the correlation coefficient which is needed in the theory. Given that numerical simulations for every spectrum are required to use Kimura's theory, it seems more direct simply to calculate the desired group statistics from the simulations.

J. A. Battjes (personal communication, 1982) points out that Kimura's results can be extended to account directly for the shape of an underlying narrow band spectrum, rather than requiring that correlation coefficients be determined from numerical simulations. Indeed, the joint probability of wave heights used by Kimura is essentially the probability density of two successive points on the height envelope of a narrow band process, derived by Rice [1944, 1945], which is expressed in terms of integral properties of the spectrum. Unfortunately, Rice's theory is only appropriate for a narrow band spectrum, and, as mentioned above, Kimura's assumption that only successive waves are correlated will itself break down if the spectrum is narrow.

2.2. Narrow Spectra

Rice's [1944, 1945] work on the envelope of a random function forms the foundation for an alternate approach to predicting wave group statistics. Following *Rice* [1944, 1945], let the sea surface elevation, $\eta(t)$, be represented as a sum of Fourier amplitudes and phases.

$$\eta(t) = \sum_{n=1}^{N} C_n \cos(\omega_n t - \phi_n) \qquad (8a)$$

where $N \gg 1$ and

$$C_n = (2S(f_n)\Delta f)^{1/2} \qquad (8b)$$

are the Fourier amplitudes, $S(f)$ is the energy density spectrum, $\omega_n = 2\pi f_n$, $f_n = n\Delta f$, and ϕ_n are random phase angles, uniformly distributed in $[0, 2\pi]$. Defining a midband frequency f_m, the sea surface can be expressed as

$$\eta(t) = \sum_{n=1}^{N} C_n \cos(\omega_n t - \omega_m t - \phi_n + \omega_m t)$$

$$= I_c \cos \omega_m t - I_s \sin \omega_m t \qquad (9)$$

where

$$I_c = \sum_{n=1}^{N} C_n \cos(\omega_n t - \omega_m t - \phi_n) \qquad (10a)$$

$$I_s = \sum_{n=1}^{N} C_n \sin(\omega_n t - \omega_m t - \phi_n) \qquad (10b)$$

and the envelope $R(t)$ is given by

$$R(t) = [I_c^2(t) + I_s^2(t)]^{1/2} \qquad (11)$$

For narrow band spectra the envelope is slowly varying, its probability density is Rayleigh distributed,

$$P(R) = \frac{R}{m_0} \exp\{-R^2/2m_0\} \quad (12)$$

and the wave heights are twice the envelope height. Defining $k \equiv H_c/2m_0^{1/2}$, the probability that the envelope exceeds $H_c/2$ is

$$P(R > H_c/2) = \exp\{-H_c^2/8m_0\} = \exp\{-k^2/2\} \quad (13)$$

where

$$m_n = (2\pi)^n \int_0^\infty S(f) f^n df \quad (14)$$

The frequency with which $R(t)$ crosses a level $H_c/2$ with a positive (or negative) slope is [Cramer and Leadbetter, 1967]

$$N = (\mu_2/2\pi)^{1/2} \frac{k}{m_0^{1/2}} \exp\{-k^2/2\} \quad (15)$$

where

$$\mu_{2n} = (2\pi)^{2n} \int_0^\infty (f - f_0)^{2n} S(f) \, df \quad (16)$$

and f_0 is the centroid of the energy spectrum. In a record of length L seconds, the total expected amount of time the envelope is above $H_c/2$ is $P(R > H_c/2) \, L$, and the average number of runs above $H_c/2$ is NL. Thus, the average length of time $R(t)$ is greater than $H_c/2$ is

$$T_E = \frac{L \, P(R > H_c/2)}{L \, N} = (2\pi/\mu_2)^{1/2} \, m_0^{1/2} \frac{1}{k} \quad (17)$$

The waves' frequency is one-half the frequency of zero crossings, N_0, where [Rice, 1944, 1945]

$$N_0 = (1/\pi) \, (m_2/m_0)^{1/2} \quad (18)$$

Vanmarke [1972] and Ewing [1973] argue the mean length of runs whose envelope lies above the level $H_c/2$ is therefore

$$E[l] = T_E \frac{N_0}{2} = (m_2/2\pi\mu_2)^{1/2} \, (1/k) \quad (19)$$

where l is the "number of waves." Thus, the mean length of runs of waves greater than $H_{1/3}$ is [Ewing, 1973]

$$E[l_{1/3}] = 1/2 \, (m_2 2\pi\mu_2)^{1/2} \quad (20)$$

To test data directly against (equations (19) and (20)) the envelope time series should be constructed, using the defining equations (8)–(11). The duration of envelope level crossings, normalized by the wave frequency, could then be compared to theory (equations (19) and (20)). Note that this envelope scheme defines the duration of groups differently than the commonly used method based on the number of discrete wave heights exceeding a certain level. For example, the envelope crossing method can define a particular section of wave record as a group, while the individual wave method does not detect any groups at all (Figure 1). Vanmarke [1972] points out, therefore, that (19) requires modification for the counting of discrete waves. A counting scheme based on continuous envelope theory allows (19) to predict $E[l] < 1$, in obvious contradiction to the discrete counting method, which (because there is always at least one wave in the group) must have $E[j] \geq 1$, where $E[j]$ is the mean length of runs counted dis-

cretely. Vanmarke [1972] and Goda [1976] suggest the following correction. By hypothesis, a wave group will have discrete run length j when the envelope duration (T_E) is $(j - 1)T_w < T_E < jT_w$, where T_w is the mean wave period. Therefore, the mean length of runs on a discrete basis is

$$E[j] = \sum_{j=1}^{\infty} j \, (P[jT_w] - P[(j - 1)T_w]) \quad (21)$$

where j is counted discretely and $P(jT_w)$ is the probability that an envelope duration is less than jT_w. Note that (21) correctly states that even if all envelope durations were much less than one period $(0 < T_E \ll T_w)$ then $E[j] = 1$, as opposed to (19) which would predict $E[l] \ll 1$. There is difficulty in evaluating (21), because theory predicts only the average envelope duration (17), not its distribution function. Goda therefore assumes, as a first approximation, that group lengths are Poisson distributed [Nolte and Hsu, 1972], an assumption only true for broad spectra. Using the Poisson assumption in (21) yields

$$E[j] = \frac{1}{1 - \exp(-1/E[l])} \quad (22)$$

where $E[l]$ is the expected "continuous envelope theory" run length (19). Note that (22) has the limits

$$E[j] \rightarrow 1 \quad \text{as} \quad E[l] \rightarrow 0 \quad (23a)$$

$$E[j] \rightarrow E[l] + 0.5 \quad \text{as} \quad E[l] \rightarrow \infty \quad (23b)$$

The validity of these limits can be investigated in more detail as follows. Refering to Figure 2, the duration of an envelope wave group is

$$T_E = T_A + (j - 1)T_w + T_B \quad (24)$$

where T_A is the time interval between the envelope's upcrossing of $H_c/2$ and the first wave crest, T_B is the interval between the last crest and the envelope's downcrossing of $H_c/2$, and T_w is the average wave period. Thus,

$$E[l] = E[T_E/T_w] = E[T_A/T_w] + E[j - 1] + E[T_B/T_w]$$

with $0 \leq T_A/T_w$, $T_B/T_w < 1$. For relatively broad spectra, $C_n/C_m = O(1)$ for all n in (10), and the envelope varies as rapidly as the waves. Thus in the limit of $E[l] \rightarrow 0$ and a broad spectrum, $T_A, T_B \rightarrow 0$ and $E[j] \rightarrow E[l] + 1 \rightarrow 1$ as given by (23a). However, for a narrow spectrum $(T_E \gg T_w$ and $E[l] \rightarrow \infty)$ it is clear that T_A/T_w, T_B/T_w are uniformly distributed between 0 and 1. Thus, $E[T_A/T_w] = E[T_B/T_w] = 1/2$, and $E[l] \rightarrow E[j]$, in contrast to (23b). As demonstrated in section 3, numerical simulations of narrow, analytical spectral shapes support the limit $E[j] \rightarrow E[l]$ for narrow spectra. Since (22) is based on an assumption true for broad spectra only, it is not surprising that it yields incorrect results for narrow spectra.

The only assumption leading to (19) is that the spectrum be narrow and the waves linear. Curiously, there are few, if any, comparisons of (19) (with or without discrete counting) with either ocean data or numerical simulations. This lack of comparisons apparently stems from the fact that Ewing [1973] evaluated (19) for a narrow Gaussian spectrum

$$S(f) = \sigma^{-1}(2\pi)^{-1/2} \exp\{-(f - f_0)/2\sigma^2\} \quad f_0 \gg \sigma \quad (25)$$

where f_0 is the central frequency and σ is a measure of the bandwidth ($f_0 \gg \sigma$ is required for a narrow spectrum). For
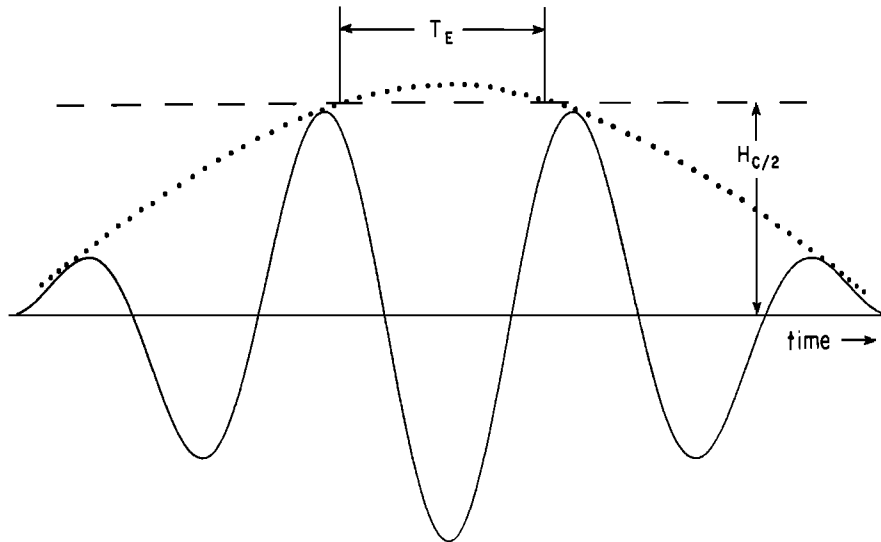
Fig. 1. Schematic of differing definitions of groups. Envelope method (dotted line) defines a group of duration T_E, while discrete wave method (solid line) does not find a group.

this particular spectral form

$$\frac{m_2}{\mu_2} = \frac{(\sigma^2 + f_0^2)}{\sigma^2} \qquad (26a)$$

$$Q_p \equiv \frac{2\int_0^\infty fS^2(f)\, df}{m_0^2} = \frac{f_0}{\pi^{1/2}\sigma} \qquad (26b)$$

Thus

$$Q_p = (m_2/\pi\mu_2)^{1/2} \qquad (26c)$$

and

$$\varepsilon \equiv (1 - m_2^2/m_0 m_4)^{1/2} \approx (6/\pi)^{1/2} Q_p^{-1} \qquad (26d)$$

where Q_p is a spectral peakedness parameter [*Goda*, 1970] and ε is a spectral width parameter [*Cartwright and Longuet-Higgins*, 1956]. Substitution of (26c) into (19) yields

$$E[l] = \frac{Q_p}{2^{1/2}k} \qquad (27)$$

Although (27) is asymptotically correct for a narrow Gaussian spectrum as $Q_p \to \infty$, it is not necessarily accurate for other spectral shapes. For example, Figure 3 shows m_2/μ_2 as a function of Q_p for the six spectral shapes described in Table 1. The solid line is (26c) for a narrow Gaussian spectrum, which is used to derive (27). Each spectrum was normalized by $S(f_p)$, where f_p is the frequency of the spectral peak, and was numerically integrated to obtain the required moments. Note that m_2/μ_2 is indeed asymptotically related to Q_p for the Gaussian spectrum and some of the other spectral shapes. However, the asymptotic Gaussian limit is not valid in general, as is most clearly demonstrated by the double-peaked spectra, which contribute most of the points below the Gaussian asymptote for $Q_p > 10$. All the spectral shapes diverge from the asymptote when m_2/μ_2 is small. In particular, the ocean spectra considered here do not indicate any relationship between m_2/μ_2 and Q_p. Thus, the practice [*Ewing*, 1973; *Chakrabarti et al.*, 1974] of comparing group statistics (from both simulations and ocean data) with the Gaussian asymptotic form (27) introduces errors not inherent in (19). Furthermore, it is shown in section 3.2 that it is difficult to precisely define the range of spectral shapes for which (19) is valid.

Ewing [1973] suggested that the narrow band theory could be modified to treat broad band spectra. Ewing modified the frequency of envelope zero crossings (N, (15)) by the factor q_1/q_0, where q_0, q_1 are the cumulative probabilities of a given local maxima exceeding a given value, for narrow (q_0) and broad (q_1) spectra (q_0, q_1 are given by *Cartwright and Longuet-Higgins* [1956]). With no other changes, Ewing derived the following expressions for $E[l]$, valid for $H_c > 2\, m_0^{1/2}$.

$$E[l] \approx (1 - \varepsilon^2)^{-1/2}\, (m_2/2\pi\mu_2)^{1/2}\, 2m_0^{1/2}/H_c \qquad 0 \le \varepsilon < 1 \qquad (28a)$$

$$E[l] \approx (m_2/\mu_2)^{1/2} \qquad \varepsilon \to 1 \qquad (28b)$$

Here, ε is the spectral width parameter defined by (26c). No theoretical basis for this derivation has been found. As mentioned above, for broad band spectra the envelope (11) fluctuates as rapidly as the waves, and thus a smoothly varying modulation of an underlying carrier wave is simply not a relevant concept. Ewing's broadband results (28) can only be viewed as heuristic. Notice that $m_2/\mu_2 \ge 1$ for all spectral shapes, hence (28) predicts $E[l] \ge 1$ for $\varepsilon \to 1$ (a broad band spectrum). However, as mentioned in the discussion of discrete counting schemes, $E[l] \to 0$ is a possible limit for broad spectra. This inconsistency in the limit $E[l] \to 0$, and the numerical simulations (section 3.2), indicate that (28) is, in general, unsatisfactory.

Nolte and Hsu [1972] combined the assumption of a Poisson distribution for group durations (implying a broadband process [*Goda*, 1976]) with a narrow band theory relating the mean group duration ($\bar\tau$) to spectral moments. When tested against a single 90-min data run, using only the measured spectral shape to determine $\bar\tau$, the model-data comparison was apparently poor. Much better agreement was obtained when the observed distribution of wave heights was used to calculate $\bar\tau$. The method of Nolte and Hsu has therefore not produced accurate group statistics, given only the wave spectrum. Analogous to *Kimura*'s [1980] narrow band scheme discussed above, detailed examination of the time series itself is necessary to determine parameters needed for predictions of group statistics.
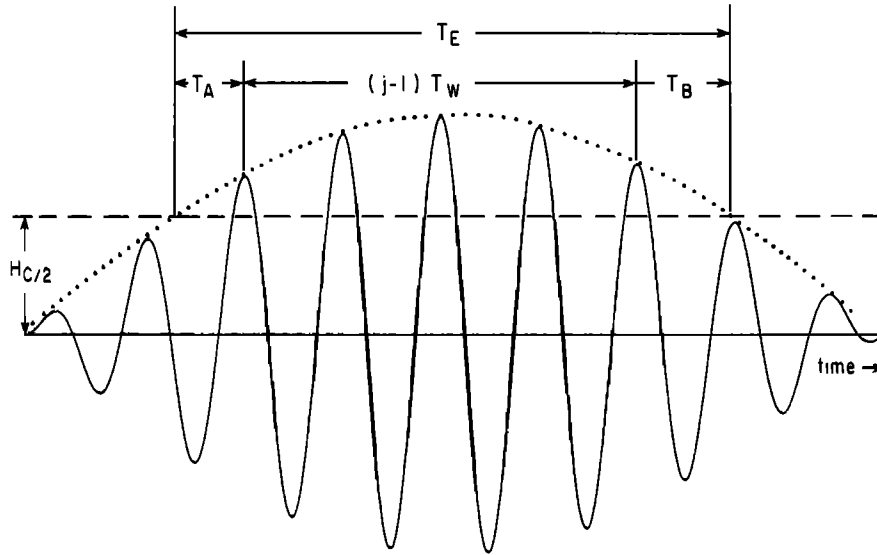
Fig. 2. Definition sketch for discrete counting scheme correction. See text for explanation of symbols.

The conclusions of this review of existing linear wave group theories are the following:

1. There are no internally consistent models which yield group statistics given only the shape of a broad energy spectrum. However, Goda's [1970] model, which predicts a constant mean group length (7c), independent of spectral shape, has been shown [Goda, 1976] to be in good agreement with data from "broad" spectra.

2. Vanmarke's [1972] and Ewing's [1973] result for narrow spectra (19) is on solid theoretical grounds. No correction for discrete counting should be used for narrow spectra.

3. There are several theories [Nolte and Hsu, 1972; Kimura, 1980] which may yield improved predictions of group statistics, but these require detailed examination of the wave time series itself, rather than proceeding directly from spectral information.

3. RESULTS

The preceding section has reviewed some relevant theories. To proceed with detailed comparisons to ocean data, it is necessary to be somewhat more quantitative about the parameter range over which the analytic results are accurate. Following Goda [1970], this is accomplished by numerically generating time series which have arbitrary spectral shape and the desired linearity and random phases. Ranges of validity of the analytic results are then established, where possible, by comparison to the simulations. The original plan was to next compare ocean field data to the analytic results for the cases in which the analytic accuracy is good. However, it developed that the analytic forms required either very broad or very narrow spectra for accurate results, conditions not met by these (or most other) ocean data. Consequently, the data were compared directly to numerical simulations of the observed spectra. Although considerably more time consuming than using analytic theories, the use of simulations has the added benefit of allowing comparisons in a statistical framework.

Observed group statistics, such as the mean length of runs, can reflect nonlinearities in the wave field. Since the theories reviewed above are based on the fundamental assumption of a linear, Gaussian process, discrepancies between theory and data might be attributed to deviations from linearity of the

waves. However, when faced with discrepancies between ocean data and approximate theories, it is difficult to determine whether the cause of the differences is due to inaccuracies in the approximate theories within a linear framework, or due to nonlinearities in the wave field. For instance, observed runs longer than Goda's theory predicts may be due to neglected correlations of waves (i.e., narrow spectra), rather than nonlinear effects.
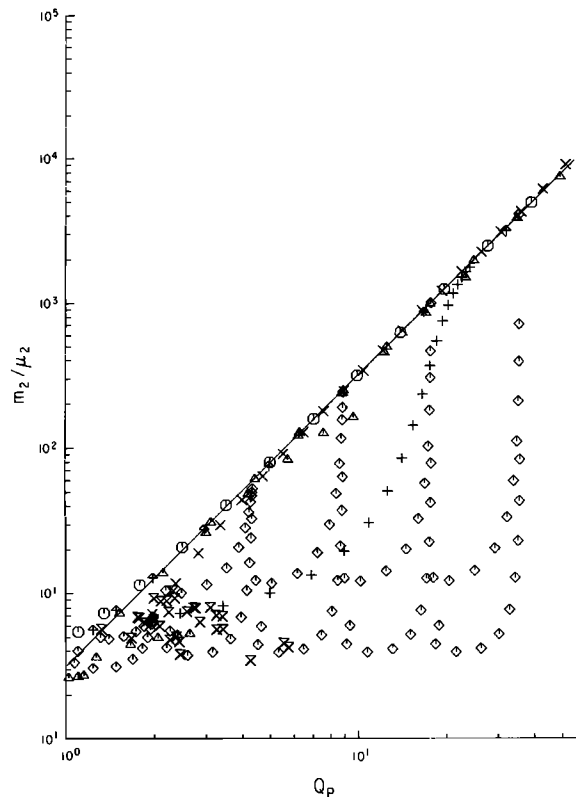

Fig. 3. m_2/μ_2 versus Q_p. Solid line is the Gaussian asymptote (equation 26d). The six spectral shapes described in Table 1 are displayed as follows: circle, Gaussian; triangle, Pierson-Moskowitz; plus, JONSWAP; cross, Ochi-Hubble; diamond, double peaked; arrow, tophat; cross with overbar, ocean field data.

TABLE 1. Spectral Forms Used in Simulations

Spectral Form (Normalized by $S(f_p)$)	Name	Reference
$S(f) = \exp\{-(f - f_p)^2/(2\sigma^2)\}$; σ is a width parameter	Gaussian	*Ewing* [1973]
$S(f) = (f/f_p)^{-m} \exp\{-m/n((f/f_p)^{-n} - 1)\}$ $n, m > 0$	Pierson-Moskowitz	*Goda* [1979]
$S(f) = (f/f_p)^{-5} \exp\{-5/4((f/f_p)^{-4} - 1)\}\gamma^{\exp\{-(f/f_p - 1)^2/(2\sigma^2)\} - 1}$		
$\gamma \geq 1, \sigma = 0.07$ if $f < f_p$, $\sigma = 0.09$ if $f > f_p$	Jonswap	*Goda* [1979]
$S(f) = (f/f_p)^{-(4\lambda + 1)} \exp\{-((4\lambda + 1)/4)((f/f_p)^{-4} - 1)\}$;		
λ is a width parameter	Ochi-Hubble	*Andrew and Borgman* [1981]
$S(f) = E[f, f_1] + cE[f, f_2]$;		
$E[f, f_i]$ is an Ochi-Hubble spectrum with		
$f_1 = f_p, f_2 > f_1, 0 < c < 1$	Double Peak	*Andrew and Borgman* [1981]
$S(f) = \begin{vmatrix} 1 & f_{\text{low}} < f < f_{\text{high}} \\ 0 & \text{otherwise} \end{vmatrix}$		
$f_{\text{high}} = Kf_{\text{low}}, f_{\text{low}} = 2f_p/(K + 1)$, $K > 1$	Tophat	

Below, first the simulation method is described (section 3.1), then comparisons between simulations and analytic results are discussed (section 3.2), and finally group statistics from simulations are compared with the observations (section 3.3).

3.1. Simulations

The fundamental assumption of linear waves is that the sea surface $\eta(t)$ can be represented as a linear combination of waves with random phases, as expressed in (8). An alternative expression is

$$\eta(t) = \sum_{n=1}^{N} a_n \cos \omega_n t + b_n \sin \omega_n t \qquad (29)$$

where a_n, b_n are independent Gaussian distributed random variables with zero mean and variance $S(f_n)\Delta f$.

Simulations using (8), which will be referred to as a random phase scheme, have spectra that always exactly match the target spectrum ($S(f)$), while simulations with (29), a random Fourier coefficient scheme, have spectra with a statistical variation about $S(f)$. Both methods were implemented and, as expected, yield nearly identical results, as detailed below. For comparing simulations with analytic results, the target spectra were a variety of standard spectral shapes, which are described in Table 1. For simulation-field data comparisons, $S(f)$ was the measured spectrum. Integral properties of the field spectra were calculated from smoothed versions of the measured spectra, with 32 degrees of freedom.

For the random phase scheme (8), Fourier coefficients (8b) were coupled with random phases produced by a numerical random number generator. The random number generator internally maintains two independent 32-bit generators. The first of these is a congruential generator, while the second is a Tausworthe, or shift-register, generator. The output from the two generators is combined, producing a random number, uniformly distributed between 0 and 1 [*Gross*, 1979]. An inverse Fourier transform of the unsmoothed spectrum results in a simulated time series with the identical spectral shape as the observed time series, but with random phases.

To obtain random Fourier coefficients (29), Gaussian distributed, zero mean, unit variance random variables were generated, and then multiplied by $(S(f_n)\Delta f)^{1/2}$ producing new Fourier amplitudes with the desired properties [*Andrew and Borgman*, 1981]. Again, an inverse Fourier transform yields a simulated time series.

Rice [1944, 1945] invoking the central limit theorem, points

out that both representations (8) and (29) will yield the same statistics in the limit as $N \to \infty$. Nevertheless, both forms were used in the simulations because there is some question about bias in spectral moments and width parameters using the two different methods [*Tuah and Hudspeth*, 1982].

Each set of random phases or coefficients, via the simulations described above, produce a time series whose properties are statistics which fluctuate about mean values. Clearly, one realization of such a simulation is not sufficient to determine representative values of the parameters of interest for a particular target spectrum. Previous studies [*Tuah and Hudspeth*, 1982, *Andrew and Borgman*, 1981] have reported results apparently based on a single realization of similar simulation schemes. Of course, these results suffer from a lack of statistical stability. In this study, for any particular target spectrum, 100 simulated time series were produced, each with its own set of random phases or Fourier coefficients.

Individual wave heights were determined by using the zero-upcrossing definition and were considered to belong to a group of high waves if the crest to trough distance exceeded a critical level. For the purposes of this study, the critical exceedence level was set equal to $4m_0^{1/2}$. For a Rayleigh distribution, this critical level is theoretically equal to the average of the highest one third of the waves, and from the Rayleigh distribution (6), about 13.5% of the waves belong to a group. Some investigators [e.g., *Rye*, 1973; *Thompson*, 1981] have pointed out that ocean wave records contain many high frequency "bumps," some of which may result in a small wave apparently interrupting a run of high waves. For apparently aesthetic reasons, these authors remove the small waves or "bumps," using various smoothing schemes in the time domain. This ad hoc smoothing of the time series vastly complicates comparisons with other studies. Furthermore, these changes in the time domain are not accompanied by respective changes in the frequency domain, unlike standard filtering operations. Consequently, established relationships between spectral and time domain parameters are no longer valid.

Each time series was 8192 s long. The mean period of the ocean data was about 10 s, a value also used in the analytical spectral forms. There were about 800 waves per simulated time series, and about 80,000 waves per target spectrum. The mean length of runs of waves greater than the significant height in the ocean data varies from 1 to almost 2.5. Thus, the number of groups in each ocean time series is between 30 and 100, and there are between 3,000 and 10,000 simulated groups per target spectrum.

Results for the same target spectrum were compared, using both random phase and random Fourier coefficient schemes. Of course, statistics such as spectral moments, $E[j]$, $\sigma[j]$, and Q_p produced by each scheme cannot be compared on a single realization basis. However, when averaged over 100 realizations, most statistics from each scheme are nearly identical. For example, the ensemble average of the first five spectral moments about the origin and about the centroid ($m_n, \mu_m, n = 1, 2, \cdots 5$) differ from the corresponding target spectral values by less than 1%. Similarly, the ensemble averaged values of $E[j]$ and $\sigma[j]$ obtained from each scheme are in close agreement. On the other hand, as is demonstrated theoretically in the appendix, Q_p was found to be biased when obtained from unsmoothed spectra produced with the random coefficient scheme.

Probability density functions (pfd's) of the number of waves per group were calculated from the ensembles of simulated time series. The pdf's for the time series resulting from the spectra in Figure 4 are displayed in Figure 5. The spectra shown in Figure 4 represent a (visually) narrow band ocean spectra (February 2, $Q_p = 5.6$, $m_2/\mu_2 = 4.7$) and a quite broadband ocean spectra (February 15, $Q_p = 1.3$, $m_2/\mu_2 = 5.7$).

Valid comparisons of theory and ocean data to the simulations require that the simulations are statistically stable. Since the wave groups produced by the 100 simulated realizations are only a sample of the population of groups associated with each target spectrum, it is therefore necessary to obtain confidence limits for the simulated statistics.

Kendall and Stewart [1967, chap. 30, section 56] present expressions for the confidence limits which relate the probability of the true probability distribution function falling within a specified bandwidth (BW) around the sample probability distribution to the size of the total sample. Following Kendall and Stewart [1967, chap. 30, section 55] it is readily shown that the bandwidth for the probability density function of the number of waves per group is given by
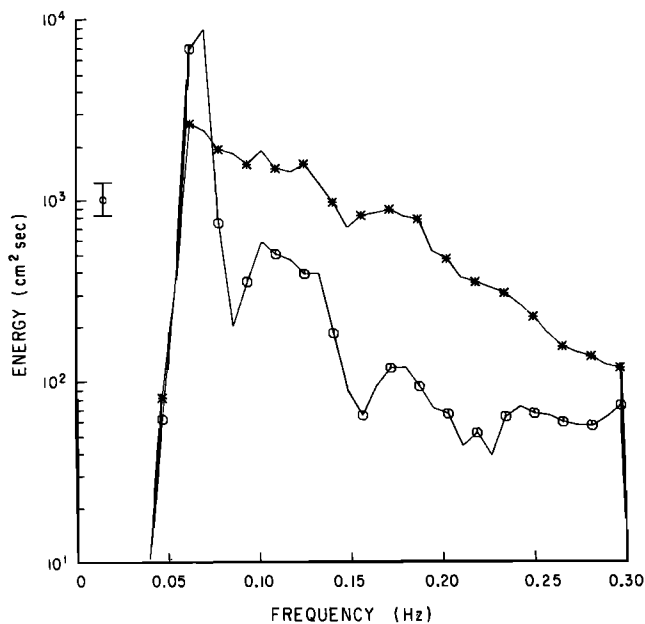
$$BW = (2.72/n^{1/2}) \qquad (30)$$

Fig. 4. Power spectral density of sea surface elevation in water 10 m deep. Circle, narrow band, February 2; asterisk, broadband, February 15. The spectra have 128 degrees of freedom, and the 90% confidence interval is indicated by the bars.
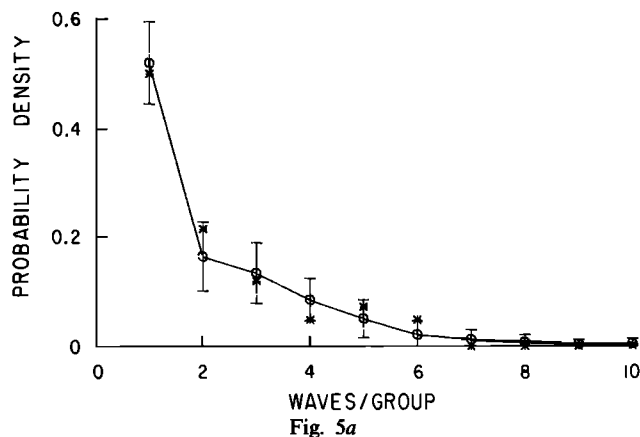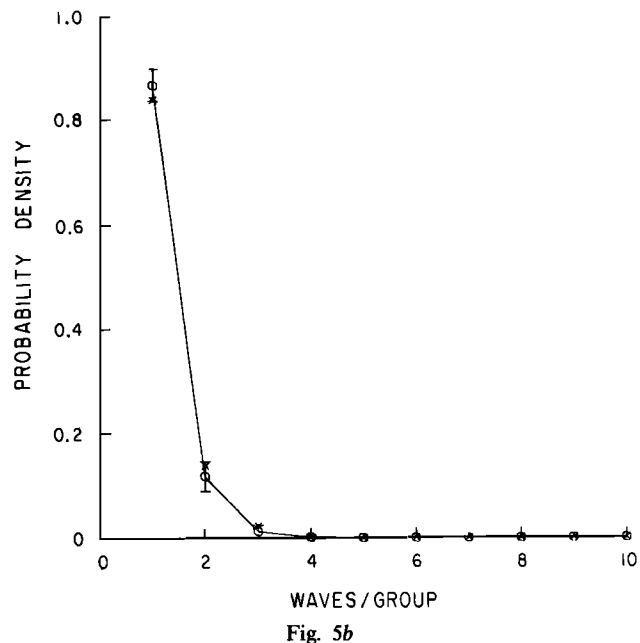
Fig. 5. Probability density function of the number of waves per group corresponding to the spectra in Figure 4; circle, simulations; asterisk, ocean field data. (a) narrow band, February 2 (42 groups were observed in the field data); (b) broadband, February 15 (74 groups were observed in the field data). Bars indicate ± 1 standard deviation of simulated values.

with probability $= 0.95$, where n is the total number of groups.

For example, on February 2, there were 4000 groups in the simulated time series. The corresponding 95% probability level bandwidth for the $P(j)$ is 0.04. Using the $P(j)$ from the simulations it is seen that the true $P(1) = 0.5 \pm 0.04$, or within 8% of the sample value. Similarly, the true $P(2)$ is within 26% of the sample value, and so on. Since the data for February 2 contained the fewest number of groups (i.e., largest value of $E[j]$) of the ocean data, a bandwidth $= 0.04$ is a maximum value for the field data. Moreover, as discussed in Kendall and Stewart [1967] the value for BW given by (30) is always an overestimate.

To lend further confidence to the statistical stability of the estimates, 1000 realizations were run for both February 2 and February 15, and the resulting $P(j)$ compared to those from 100 realization ensembles. Taking the ensemble of 1000 realizations as "true," the 100 realization ensembles were well within the expected 95% confidence limits. Indeed, the differ-
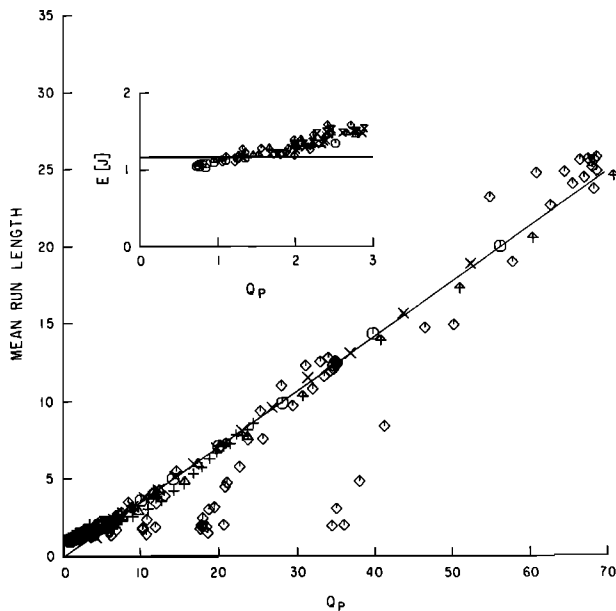
Fig. 6. Mean length of runs ($E[j]$) greater than H-1/3 versus Q_p. The solid line is for a narrow Gaussian spectral shape (27). Inset shows behavior of $E[j]$ for small values of Q_p. Solid line in inset is (7c). Symbols are described in Figure 3.

ence between the "true" and "sample" $P(j)$ was less than 10% for $j \leq 4$.

Clearly, as $P(j)$ becomes small, the relative bandwidth beomes large. However, in those cases where the distribution is not well estimated ($j \geq 4$), the values of $P(j)$ are very small, as can be seen in Figure 5. Consequently, those $P(j)$ for large j have negligible effect on integral properties of the pdf's such as $E[j]$. For example, for the February 2 simulations, 1000 realizations yielded $E[j] = 2.19$, while 100 realizations yielded $E[j] = 2.17$, while for February 15 both values of $E[j]$ were equal to 1.15. The conclusion is that simulations with 100 realizations are extensive enough to estimate the true value of $E[j]$ within a few percent.

For each ensemble of 100 simulations, the distribution of each $P(j), j = 1, 2, 3$ was compared with a Gaussian distribution. According to a χ^2 test, the $P(j)$ distributions did not deviate from a Gaussian distribution any more than could be expected 10% of the time for a truly Gaussian process. The distribution of $P(j)$, $j \geq 4$ is also Gaussian if the number of realizations which contain runs of length ≥ 4 is sufficiently large (about two thirds of the 100 simulations). For example, February 2, a narrow band day, has many realizations (almost 100) with runs of four waves, and $P(4)$ is Gaussian distributed according to a χ^2 test. On the other hand, February 15, a broadband day, has very few realizations with a run as large as four waves, and $P(4)$ is not Gaussian. The Gaussian character of $P(j)$, for small j, is useful in ocean data comparisons considered in section 3.3.

3.2. Simulation-Analytic Theory Comparisons

As discussed above, previous simulation work [*Goda*, 1970, 1976] has shown (7) to be accurate for broad spectra (although broad was only loosely defined), while (19) [*Vanmarke*, 1972; *Ewing*, 1973] must be valid for very narrow spectra (also loosely defined) since it is derived exactly. To compare ocean data to linear analytic theory (to test if ocean wave group statistics are consistent with linear theory, for example), the

data must lie within the range of applicability of the theory. In an attempt to quantify these ranges of applicability (i.e., to determine the parameter ranges for which (7) and (19) are valid), test spectra were generated for the spectral shapes in Table 1. Double peaked spectra were generated as the linear superposition of two Ochi-Hubble spectra, each with variable width, energy, and peak frequency.

Figure 6 shows the average run length $E[j]$ as a function of Q_p, with the Gaussian asymptote ((27) with $k = 2$) shown. Figure 7 plots $E[j]$ against m_2/μ_2. Since (20) and (27) are valid for narrow spectra, and $E[j] = E[l]$ in this limit, neither equation is altered for discrete counting. As Figure 7 demonstrates, for spectra without high frequency structure (e.g., Gaussian, Pierson-Moskowitz, Ochi-Hubble, tophat spectral forms), (20) models the simulations quite well for large values of m_2/μ_2. Applying the correction for discrete counting (22) to (20) offsets the theoretical curve for small m_2/μ_2 by imposing a minimum value of 1 on the corrected equation. A large value of m_2/μ_2 is associated with high $E[l]$ (Figure 7); however, the converse is not true (i.e., large values of $E[l]$ also occur for small m_2/μ_2). This is especially true for spectra with high frequency structure, such as the double-peaked and JONSWAP spectra. Indeed, the addition of a small, high frequency bump in an otherwise narrow spectrum will cause μ_2 to increase, and in turn m_2/μ_2 to decrease, often dramatically, even though $E[l]$ is hardly effected. Thus, although a large value of m_2/μ_2 is necessary for a spectrum to be narrow, it does not appear to be sufficient for (20) to be valid. Unfortunately, no simple shape parameter has been found which identifies a spectrum for which (20) is valid (higher order moments are clearly involved).

Figure 3 indicates that for the JONSWAP, Pierson-Moskowitz, and many double-peaked spectra, as well as the ocean data considered here, $m_2/\mu_2 < 10$ and $Q_p < 5$. For such low values of m_2/μ_2 and Q_p (20) and (27) are not valid (Figures 6 and 7). Similarly, although values of m_2/μ_2 are not commonly reported, Q_p from other ocean data ($Q_p < 2.5$, [*Chakrabarti et al.*, 1974]; $Q_p < 4$ [*Goda*, 1976]; $Q_p < 4.5$ [*Su et al.*, 1982]; $Q_p < 6$ [*Goda*, 1983]) indicate that (20) and (27) are not applicable for most ocean spectra. That is, both analytic forms (20)
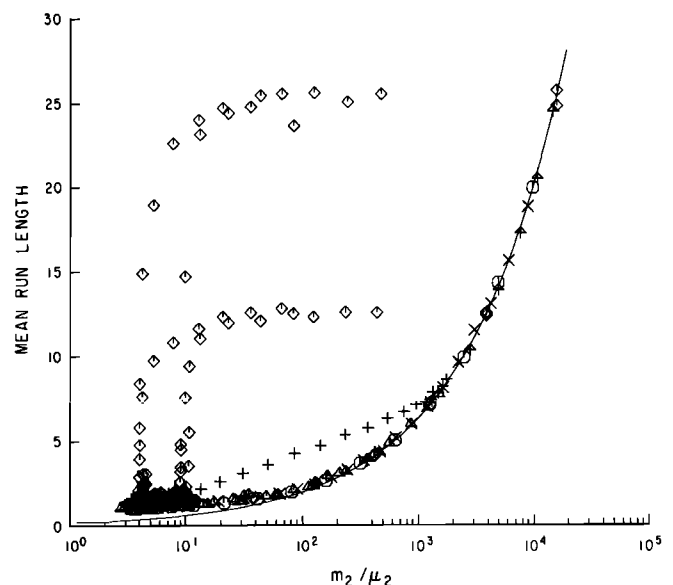

Fig. 7. Mean length of runs greater than H-1/3 versus m_2/μ_2. Solid line is narrow band theory (20). Symbols are described in Figure 3.

and (27) are unsuitable for testing linearity of typical ocean groups. On the other hand, for broad spectra the simulations indicate agreement ($\pm 10\%$) with Goda's prediction of mean run length (7) for $Q_p < 2$. However, it is important to note that $E[j]$ does not asymptotically approach (7) as Q_p approaches 0. Instead, $E[j]$ continues to decrease below the value predicted by (7) as Q_p decreases below 1.25. For these very low values of Q_p, the assumption of Rayleigh distributed wave heights probably starts to break down. This behavior of $E[j]$ for low values of Q_p is shown in the inset of Figure 6, where the solid line is (7). Similarly, the variance of run lengths was calculated from the ensemble of simulations for each target spectrum. For $Q_p < 2$, the simulations are in agreement ($\pm 10\%$) with Goda's prediction (7d). Thus, on the basis of the present simulations, spectra with $Q_p < 2$ are identified as very broad spectra.

The simulations indicate considerable variation of $E[j]$ for constant m_2/μ_2 or Q_p in the range of $m_2/\mu_2 < 100$ and $Q_p > 2$. Much of the field data falls within this range. Thus, it is inappropriate to compare field observations with numerical simulations based on an analytic spectral form with coincidentally the same Q_p. There is substantial scatter in Figure 6, even though each point represents the mean from 100 realizations. That is, identical values of Q_p do not imply identical group statistics, even in the limit of many realizations.

The scatter in plots of mean run length versus Q_p increases considerably for ocean data, which are essentially statistics produced by only one realization. For example, as discussed in section 3.3, $E[l]$ is approximately Gaussian distributed, and each observation of $E[l]$ is subject to variation consistent with the appropriate Gaussian distribution. Therefore, although the mean length of runs may be correlated with Q_p (Figure 6), the current practice of correlating wave group statistics with Q_p is essentially nonproductive. There is little to be learned from outlying points; that is, points not close to the linear regression between wave group statistics and Q_p are not necessarily due to nonlinearities in the field data. Consequently, in order to test the field data against a linear hypothesis, simulations of each observed spectrum must be performed, and the field observations examined for statistical consistancy with the simulations.

3.3. Simulation-Field Data Comparisons

The field data were obtained at Santa Barbara, California, during the Nearshore Sediment Transport Study experiment conducted in January and February 1980 [Gable, 1981]. The time series used here were obtained from bottom-mounted pressure sensors, located along a line perpendicular to the beach, extending from approximately 1 m depth to 10 m depth.

To obtain reasonable confidence limits on the statistics of lengths of runs greater than the significant wave height, $H_{1/3}$, it is clear that data runs containing many groups are necessary. Record lengths of 8192 s (2.27 hours) were used here, typically containing ≈ 800 waves. The pressure gages were continuously sampled at 2 Hz, for up to 5 hours daily during a 4-week period. Those data selected for processing were checked for stationarity by using a "run" test [Bendat and Piersol, 1971] after breaking the record into 32 sections of 256 s. If any one of the parameters checked (m_0, mean length of runs, Q_p, frequency of spectral peak) showed a trend inconsistent with random fluctuations, indicating possible nonstationarity, those data were rejected. Thirty time series were
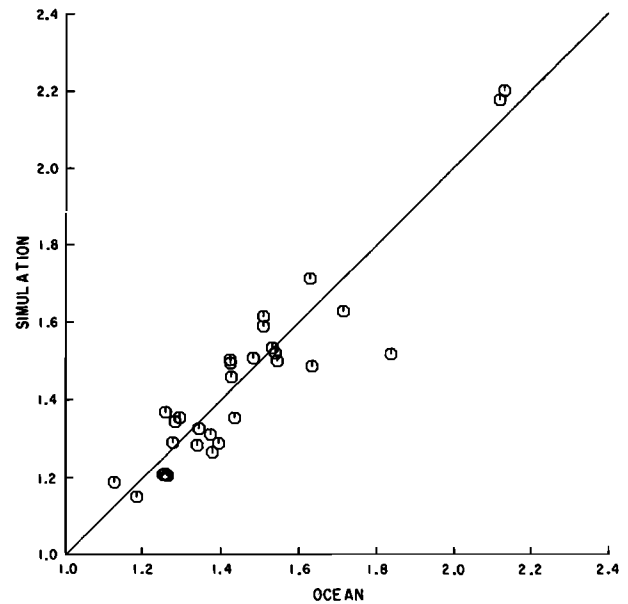

Fig. 8. Mean length of runs greater than H-1/3 from the numerical simulations versus mean length of runs greater than H-1/3 from the ocean field data. The 45° solid line indicates agreement between simulations and ocean field data.

finally chosen for analysis. The spectra of bottom pressures were converted to sea surface elevation by using linear theory, tidal trends were removed, and the data were band-pass filtered. Data from the deepest pressure sensors, in approximately 10 m depth, were band-pass filtered between 0.04 and 0.3 Hz, while the time series obtained from the other pressure gauges, located in 4 m or less, were filtered by passing those frequencies between 0.04 and 0.5 Hz.

The data cover a wide range of conditions, including very narrow (by ocean standards) and quite broadband spectral shapes (Figure 4), and significant wave heights between 20 and 200 cm. With typical peak periods from 8 to 20 s, the wave steepness (product of significant amplitude and wave number of the spectral peak) is in the range 0.006–0.1. Some days were characterized by swell from distant storms, others by local seas, and a few had multi-peaked spectra, representing different combinations of sea and swell.

Since the question at hand is whether or not ocean observations are consistent with the linear simulations, several different comparisons of ocean and simulated ocean data were made. One such comparison is of the mean length of runs. Figure 8 shows that the mean run length from the ocean data and the simulations are visually well correlated. A χ^2 test for the 100 realizations of each target spectrum showed $E[j]$ to be Gaussian distributed, thus about 68% of the simulated $E[j]$ fell within 1 standard deviation of the mean. Consequently, each ocean value of $E[j]$ can be examined to see if it comes from the same (Gaussian) population as the simulations, using a standard normal deviate test, where the simulations provide the variance of $E[j]$ for each spectrum. Individual ocean $E[j]$ deviated from the mean $E[j]$ of the simulations as would be expected for a Gaussian distribution, with 77% of the ocean $E[j]$ falling within 1 standard deviation of the simulation mean.

To test if the collection of ocean $E[j]$ were statistically consistent with the simulated $E[j]$, Student's t statistic for paired data was calculated. This test essentially examines
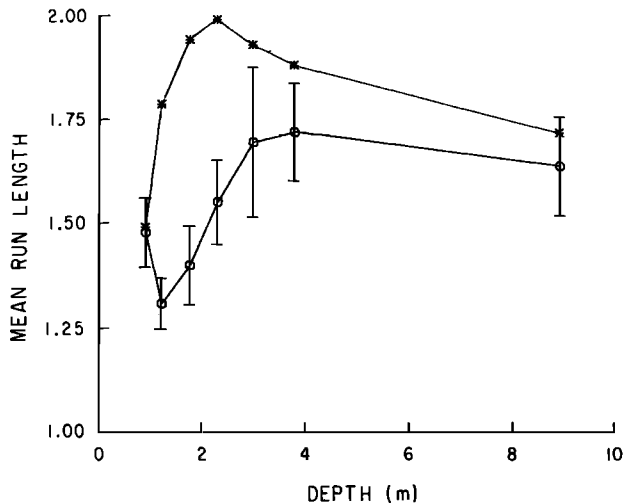
Fig. 9. Mean length of runs greater than H-1/3 versus depth of water. Asterisk, ocean field data; circle, simulations, bars indicate ± 1 standard deviation of simulated values. Data were obtained February 3.

whether or not the two treatments (ocean observation and numerical simulation) of the same data (the target spectrum) produce the same result $(E[j])$. The t statistic obtained was $t = 0.67$, which will be exceeded about 50% of the time due to random fluctuations. Hence, the hypothesis that the ocean mean group lengths are statistically consistent with linear wave theory cannot be rejected.

Similarly, the variances of run length obtained from the ocean data were compared with the corresponding variances produced by the simulations. The ratios of the square of run length coefficients of variation (standard deviation normalized by the mean, ocean and simulated) for each target spectrum were compared to tabulated values of Fisher's F distribution with the appropriate number of degrees of freedom. The ratios exceeded the tabulated values no more often than would be expected due to random fluctuations if the two populations (ocean and simulated) were the same. For example, one of the 30 ratios exceeded a value expected to occur 2% of the time due to random fluctuations, and two ratios exceeded a value expected 10% of the time. Hence, the hypothesis that the variances of run length come from the same statistical population cannot be rejected.

A comparison of the pdf's of number of waves per group was made, as shown in Figure 5. The statistical reliability of the simulation based pdf's was discussed in Section 3.2. For the $P(j)$ that are Gaussian distributed, 68% of the simulated $P(j)$ are expected to fall within 1 standard deviation of the mean. Similarly, if the ocean data are consistent with the linear simulations, 68% of the ocean $P(j)$ should also fall within 1 standard deviation of the mean. About 70% of the ocean data do so for those cases where the simulated $P(j)$ was considered to be Gaussian distributed.

A χ^2 test was used to test if the entire collection of 30 pdf's from ocean data belonged to the same statistical population as the corresponding collection of simulated pdf's. The χ^2 value obtained (with 79 degrees of freedom) has about a 12% chance of occurring by random chance if the two collections come from the same population. In light of the results discussed above, this is not considered strong enough evidence to reject the hypothesis of linearity. There are certainly no gross discrepancies between the observed and simulated pdf's (Figure

5, for example). All the above statistical tests were applied to both the random phase simulations and the random Fourier coefficient simulations, with negligible differences. The values presented are from the random phase simulations.

4. EFFECTS OF SHOALING

The results presented so far indicate that the assumption of a linear, Gaussian process, as expressed by (8) or (29), produces statistics consistent with observations of ocean wave groups in 10 m depth. However, as waves shoal they are expected to become more nonlinear. Consequently, a linear representation, such as (8) or (29), should not necessarily produce wave group statistics consistent with observations of shoaled waves. That this is the case is shown in Figure 9, based upon data collected on February 3. Values of $E[j]$, both observed and simulated (from the measured spectra at the appropriate depth), are shown as a function of depth, from 10 m depth through the breaking point (about 2 m) and into 1 m of water. The simulated mean run length varies during shoaling because of substantial changes in the observed spectrum. Simulations and observations are similar in 10 m depth, but as the waves shoal, the observed $E[j]$ becomes much greater than linear theory predicts. The ocean $E[j]$ remains higher than the corresponding values from the simulations until the waves break, when field data are once again consistent with the simulations (Figure 9). This trend occurs in most of the data sets, under varying wave conditions, and is the subject of ongoing research.

The pdf for the February 3 data observed in about 2 m depth is presented in Figure 10 (compare to 10 m data in Figure 5). The field data have a small value of $P(1)$ relative to the linear simulations. As j increases, $P(j)$ of the field data becomes greater than the corresponding $P(j)$ of the simulated time series. For the particular pdf shown in Figure 10, a χ^2 test to determine if the field data come from the same population as the simulations produces $\chi^2 = 32$ with 3 degrees of freedom. (Individual class intervals were used for $j \leq 3$, and the remaining values were pooled.) A value this large can occur due to random fluctuations less than 0.1% of the time. Furthermore, a normal random deviate test of the mean length of runs also yields a value much larger than would be
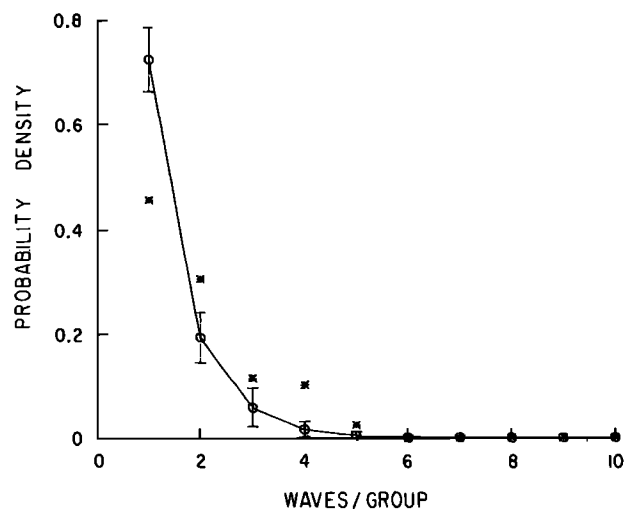

Fig. 10. Probability density function of number of waves per group corresponding to data observed in 2 m depth that are displayed in Figure 9. Symbols are described in Figure 5. (87 groups were observed in the field data).

expected 0.1% of the time. Thus, the hypothesis that the 2 m depth ocean observation comes from the same linear, random wave population as the simulations must be rejected.

5. CONCLUSIONS

There are two theoretical models which predict the mean length of runs of ocean waves greater than some critical height, given only the power spectral density. The differences between the models stem from initial assumptions about the wave field and result in different regions (with respect to spectral shape) of applicability. Goda's prediction (7) which assumes the individual waves are independent, Rayleigh distributed random variables, is most accurate for broadband spectra. However, it has an internal inconsistency, so there is no a priori range of validity. On the other hand, Vanmarke's and Ewing's result (equations (19) and (20)), based on Rice's envelope theory, appears to require a narrow band spectrum with no high frequency structure.

Extensive numerical simulation of time series from various spectral forms indicates agreement ($\pm 10\%$) with Goda's prediction of a constant mean run length for spectra with $Q_p < 2$ (Figure 6). Unfortunately, no simple spectral shape parameter (such as m_2/μ_2 or Q_p) was found which indicates quantitatively the region of validity of (20)). Furthermore, Q_p and m_2/μ_2 are not, in general, unique functions of each other. Therefore, it is not valid to transform one of these spectral parameters into the other.

For spectral shapes similar to those found in the ocean, none of the theories adequately predicts mean run length. Consequently, theories given by (7) and (19) could not be used to test if observed values of mean run length ($E[j]$) are consistent with linear dynamics (i.e., (8) or (29)). Owing to both the inherent statistical variability of wave group parameters and the inadequacy of Q_p as a definitive group parameter, it is not instructive to correlate wave group statistics with spectral parameters such as Q_p. Instead, ocean $E[j]$ was compared with those produced from linear simulations of time series with the same power spectral density. The results from 30 ocean time series, each of 8192 s length, show no compelling statistical difference between observations of $E[j]$ and linear simulations in 10 m depth. Similar results were found for the variances of run length and the probability density functions of the number of waves per group. On the other hand, disagreement with linear dynamics was found for shoaled waves in 2 m depth.

APPENDIX

It was noted above that the value of Q_p obtained from the random Fourier coefficient scheme is biased, while spectral moments m_n are not. This can be demonstrated as follows. Recall the Fourier representation of the sea surface

$$\eta(t) = \sum_{n=1}^{N} a_n \cos \omega_n t + b_n \sin \omega_n t \qquad (A1)$$

where a_n, b_n are independent Gaussian distributed random variables with zero mean and variance $S(f_n)\Delta f$. Here, $S(f_n)$ is the true spectral density at frequency $f_n = n\Delta f$. From a finite length time series only the estimates $\tilde{a}_n, \tilde{b}_n$ of the true values of a_n, b_n can be obtained. The random variable

$$\tilde{Z}_n^2 \equiv (\tilde{a}_n^2 + \tilde{b}_n^2)/S(f_n)\Delta f \qquad (A2)$$

is chi square distributed, with 2 degrees of freedom. The probability density function of a chi square distributed random

variable, with v degrees of freedom is

$$f(x) = \begin{cases} \dfrac{1}{2^{v/2}\Gamma(v/2)} x^{(v/2-1)} \exp\{-x/2\}\, dx & x > 0 \\ 0 & x \le 0 \end{cases} \qquad (A3)$$

where Γ is the gamma function. Thus it is easily verified that

$$\begin{aligned} E[x] &= v \\ E[x^2] &= v^2 + 2v \end{aligned} \qquad (A4)$$

Since the estimated value of $S(f_n)$ is $(\tilde{a}_n^2 + \tilde{b}_n^2)/2\Delta f$ the expected value of the rth spectral moment is

$$\begin{aligned} E[m_r] &= E\left[(2\pi)^r \sum_{n=1}^{N} (n\Delta f)^r \frac{(\tilde{a}_n^2 + \tilde{b}_n^2)}{(2\Delta f)}\right] \\ &= (2\pi)^r \sum_{n=1}^{N} (n\Delta f)^r E[\tilde{Z}_n^2] \frac{S(f_n)\Delta f}{(2\Delta f)} \\ &= \frac{v}{2}(2\pi)^r \sum_{n=1}^{N} (n\Delta f)^r S(f_n) \\ &= \frac{v}{2} m_r = m_r \end{aligned} \qquad (A5)$$

because $v = 2$. If several realizations of the process, j, say, are available, or if neighboring frequency components are merged together, v will increase, but the above result (A5) remains the same. In this case the average

$$\tilde{T}_n^2 = \frac{1}{j} \sum_{k=1}^{j} (\tilde{a}_{n_k}^2 + \tilde{b}_{n_k}^2)/(S(f_n)\Delta f) \qquad (A6)$$

may be used in place of $\tilde{Z}_n^2$, yielding

$$\begin{aligned} E[m_r] &= (2\pi)^r \sum_{n=1}^{N} (n\Delta f)^r \frac{1}{j} E\left[\sum_{k=1}^{j} \frac{(\tilde{a}_{n_k}^2 + \tilde{b}_{n_k}^2)}{(2\Delta f)}\right] \\ &= (2\pi)^r \sum_{n=1}^{N} (n\Delta f)^r S(f_n) \frac{v}{2j} \end{aligned} \qquad (A7)$$

using (A4) it equals m_r because $v = 2j$. Next consider the expected value of Q_p,

$$\begin{aligned} E[Q_p] &= E\left[2 \sum_{n=1}^{N} (n\Delta f) \frac{[\tilde{a}_n^2 + \tilde{b}_n^2]^2}{(4(\Delta f)^2 m_0^2)}\right] \\ &= \frac{2}{m_0^2} \sum_{n=1}^{N} (n\Delta f) E[(\tilde{Z}_n^2)^2] \frac{S^2(f_n)}{4} \\ &= \left[\frac{2}{m_0^2} \sum_{n=1}^{N} (n\Delta f) S^2(f_n)\right]\left[\frac{(v^2 + 2v)}{4}\right] = 2Q_p \qquad (A8) \end{aligned}$$

because $v = 2$. Similarly, ensemble averaging over j realizations yields

$$\begin{aligned} E[Q_p] &= E\left[\frac{2}{m_0^2} \sum_{n=1}^{N} (n\Delta f) \frac{1}{j^2} \sum_{k=1}^{j} \frac{(\tilde{a}_{n_k}^2 + \tilde{b}_{n_k}^2)^2}{(4\Delta f^2)}\right] \\ &= \left[\frac{(2v + v^2)}{(4j^2)}\right] Q_p = (2/v + 1)\, Q_p \qquad (A9) \end{aligned}$$

using (A4) because $v = 2j$.

Thus, the value of Q_p estimated from a finite time series is biased. If the random Fourier coefficient scheme is employed to simulate a smoothed target spectrum, and no additional smoothing is performed on the individual realizations produced by the simulations, the mean value of Q_p obtained by averaging the individual Q_p from each realization will be twice

the target spectrum's Q_p. On the other hand, ensemble averaging the simulated spectra, and then calculating Q_p from the ensemble will produce a better estimate of the true value of Q_p, in accordance with (A9). These theoretical conclusions were verified by the simulations. Similarly, the bias of Q_p in spectra from ocean swell was noticed by *Goda* [1983]. Figure 14 of that paper displays Q_p values calculated by two different averaging schemes. The ratio of the average Q_p from an ensemble of five or six individual spectra to the value of Q_p for the spectrum obtained by averaging the individual spectra is quite close to $2/v + 1$ (A9). *Goda* [1983, Figure 14] also notes that Q_p decreases as the number of degrees of freedom in each spectrum increases, no matter which averaging procedure is used. Of course, as v is increased, individual peaks in the spectrum are smoothed, and hence Q_p decreases. Indeed, in the limit of the maximum possible number of degrees of freedom the spectrum is reduced to a single value, and $Q_p = 1$. Thus, spectra that are "undersmoothed" yield a Q_p which is biased high, while "oversmoothed" spectra give a Q_p which is biased low.

REFERENCES

Andrew, M. E., and L. E. Borgman, Procedures for studying wave grouping in wave records from California Coastal Data Collection Program, report, U.S. Army Corps of Eng., San Francisco, Calif., November 1981.

Bendat, J. S., and A. G. Piersol, *Random Data: Analysis and Measurement Procedures*, John Wiley, New York, 1971.

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

Chakrabarti, S. K., R. H. Snider, and P. H. Feldhausen, Mean length of runs of ocean waves, *J. Geophys. Res.*, *79*, 5665–5667, 1974.

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

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

Gable, C. G. (Ed.), Report on data from the Nearshore Sediment Transport Study experiment at Leadbetter Beach, Santa Barbara, California, January–February, 1980, *IMR Ref. 80-5*, Univ. of Calif., Inst. of Mar. Res., Location, 1981.

Goda, Y., Numerical experiments on wave statistics with spectral simulation, report, vol. 8, pp. 3–57, Port and Harbor Res. Inst., Nagase, Yokosuka, Japan, 1970.

Goda, Y., Estimation of wave statistics from spectral information, paper presented at Proceedings International Symposium on Ocean Wave Measurement and Analysis, New Orleans, La., September 1974.

Goda, Y., On wave groups, paper presented at Proceedings First International Conference on the Behaviour of Off-Shore Structures (BOSS '76), Trondheim, Norway, Aug. 2–5, 1976.

Goda, Y., A review on statistical interpretation of wave data, report, vol. 18, pp. 5–32, Port and Harbour Res. Inst., Nagase, Yokosuka, Jpn., 1979.

Goda, Y., Analysis of wave grouping and spectra of long-travelled swell, report, vol. 22, pp. 3–41, Port and Harbour Res. Inst., Nagase, Yokosuka, Japan, 1983.

Gross, A. M., *Programs for Digital Signal Processing*, IEEE Press, New York, 1979.

Kendall, M. G., and A. Stuart, *The Advanced Theory of Statistics*, vol. 2, *Inference and Relationship*, Hafner Publ., New York, 1967.

Kimura, A., Statistical properties of random wave groups, paper presented at Proceedings Seventeenth Coastal Engineering Conference, Am. Soc. of Civil Eng., Sydney, Australia, March 23–28, 1980.

Longuet-Higgins, M. S., On the statistical distributions of heights of waves, *J. Mar. Res.*, *11*, 245–266, 1952.

Longuet-Higgins, M. S., The statistical analysis of a random, moving surface, *Phil. Trans. R. Soc. London Ser. A*, *249*, 321–387, 1957.

Nagai, K., Runs of the maxima of the irregular sea, *Coastal Eng. Jpn.*, *16*, 13–18, 1973.

Nolte, K. G., and F. H. Hsu, Statistics of ocean wave groups, paper presented at Proceedings Offshore Technology Conference, Dallas, Texas, 1972.

Rice, S. O., The mathematical analysis of random noise, *Bell Sys. Tech. J.*, *23*, 282–332, 1944.

Rice, S. O., The mathematical analysis of random noise, *Bell Sys. Tech. J.*, *24*, 46–156, 1945.

Rye, H., Wave group formation among storm waves, paper presented at Proceedings 14th Coastal Engineering Conference, Am. Soc. of Civil Eng., Copenhagen, June 24–28, 1974.

Su, M. Y., M. T. Bergin, and S. L. Bales, Characteristics of wave groups in storm seas, paper presented at Proceedings Ocean Structural Dynamics Symposium, Oregon State Univ., Corvallis, Oregon, 1982.

Thompson, E. F., Nonrandom behavior in field wave spectra and its effect on grouping of high waves, Dissertation, The George Washington Univ. School of Eng. and Appl. Sci., Washington, D. C., 1981.

Tuah, H., and R. T. Hudspeth, Comparisons of numerical random sea simulations, *J. Waterway Port Coastal Ocean Div. Am. Soc. Civil Eng.*, *108*, 569–584, 1982.

Vanmarcke, E. H., Properties of spectral moments with applications to random vibration, *J. Eng. Mech. Div., Am. Soc. Civil Eng.*, *98*, 425–446, 1972.

Wilson, J. R., and W. F. Baird, A discussion of some measured wave data, paper presented at Proceedings Twelfth Coastal Engineering Conference, Am. Soc. of Civil Eng., Washington, D. C., September 1970.

Yuen, H. C., and B. M. Lake, Instabilities of waves on deepwater, *Ann. Rev. Fluid Mech.*, *12*, 303–334, 1980.

S. Elgar, R. T. Guza, and R. J. Seymour, Mail code A-022, Scripps Institution of Oceanography, La Jolla, CA 92093.