

YALE PEABODY MUSEUM

P.O. BOX 208118 | NEW HAVEN CT 06520-8118 USA | PEABODY.YALE. EDU

JOURNAL OF MARINE RESEARCH

The *Journal of Marine Research*, one of the oldest journals in American marine science, published important peer-reviewed original research on a broad array of topics in physical, biological, and chemical oceanography vital to the academic oceanographic community in the long and rich tradition of the Sears Foundation for Marine Research at Yale University.

An archive of all issues from 1937 to 2021 (Volume 1–79) are available through EliScholar, a digital platform for scholarly publishing provided by Yale University Library at <https://elischolar.library.yale.edu/>.

Requests for permission to clear rights for use of this content should be directed to the authors, their estates, or other representatives. The *Journal of Marine Research* has no contact information beyond the affiliations listed in the published articles. We ask that you provide attribution to the *Journal of Marine Research*.

Yale University provides access to these materials for educational and research purposes only. Copyright or other proprietary rights to content contained in this document may be held by individuals or entities other than, or in addition to, Yale University. You are solely responsible for determining the ownership of the copyright, and for obtaining permission for your intended use. Yale University makes no warranty that your distribution, reproduction, or other use of these materials will not infringe the rights of third parties.



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.
<https://creativecommons.org/licenses/by-nc-sa/4.0/>



JOURNAL OF MARINE RESEARCH

VOLUME XI

1952

NUMBER 3

ON THE STATISTICAL DISTRIBUTION OF THE HEIGHTS OF SEA WAVES

By

M. S. LONGUET-HIGGINS

Trinity College, Cambridge¹

ABSTRACT

The statistical distribution of wave-heights is derived theoretically on the assumptions (a) that the wave spectrum contains a single narrow band of frequencies, and (b) that the wave energy is being received from a large number of different sources whose phases are random. Theoretical relations are found between the root-mean-square wave-height, the mean height of the highest one-third (or highest one-tenth) waves and the most probable height of the largest wave in a given interval of time. There is close agreement with observation.

1. INTRODUCTION

At present several different quantities are in use for describing the state of the sea: for example, the mean height of the waves, the root-mean-square height, the height of the "significant" waves (defined by Sverdrup and Munk [1947] as the mean height of the highest one-third of all the waves), the maximum height over a given interval of time, and so on. The purpose of the following is to investigate the relationship of these quantities to one another in some special cases, and especially in the case when the spectrum of the waves consists of a single narrow frequency-band.

¹ The author is indebted to the Commonwealth Fund, New York, for a Fellowship to enable him to study at the Scripps Institution of Oceanography, where this paper was prepared.

For definiteness let us consider the elevation ζ of the sea surface at a fixed point, given as a function of the time t only. Much of the following, however, will apply to any oscillatory function of a single variable: for example, to the pressure at a point on the bottom, or to the rolling motion of a ship as measured by its angular deflection from the vertical. In general we shall denote by a the amplitude of ζ , which may be defined as half the distance in level between a wave crest and the preceding trough; thus $2a$ equals the wave-height. The period, or interval between successive crests, will be denoted by τ , or $2\pi/\sigma$, where $\sigma/2\pi$ is the frequency. I denotes any interval of the t -axis, of length T , in which the variable ζ is under consideration; it is supposed that $T \gg \tau$, i. e., the interval contains a large number of complete periods. The successive values of a in the interval I may

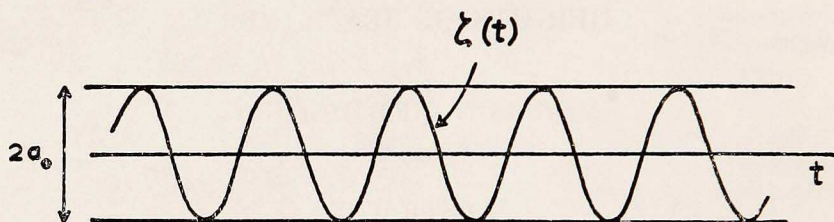


Figure 1. A simple sine-wave: definition of the wave amplitude.

be denoted by a_1, a_2, \dots, a_N . If these are arranged in descending order of magnitude, the mean value of the first pN of these, where p is a fraction between 0 and 1, will be denoted by $a^{(p)}$. Thus the amplitude of Sverdrup and Munk's "significant waves" is $a^{(1/3)}$. The mean amplitude of all of the waves is $a^{(1)}$. It is clear that $a^{(p)}$ is a decreasing, or at any rate a nonincreasing, function of p ; and if a_{\max} is the maximum value of a in the interval, we have

$$a_{\max} \geq a^{(p)} \geq a^{(1)}. \quad (1)$$

The root-mean-square amplitude \bar{a} is defined by

$$\bar{a}^2 = \frac{1}{N}(a_1^2 + a_2^2 + \dots + a_N^2). \quad (2)$$

It may easily be shown that

$$\bar{a} \geq a^{(1)}. \quad (3)$$

Since \bar{a} is of physical significance in a wide class of cases, $a^{(p)}$ will normally be expressed in terms of \bar{a} . The mode is defined as the most frequently occurring wave amplitude and will be denoted by $\mu(a)$.

Example 1. Simple sine-wave. Suppose

$$\zeta = a_0 \cos \sigma t ; \quad (4)$$

then we have a simple sine-wave of period $2\pi/\sigma$ and amplitude a_0 (see Fig. 1). All the waves are of amplitude a_0 , and therefore

$$a_{\max} = a^{(p)} = \bar{a} = \mu(a) = a_0 . \quad (5)$$

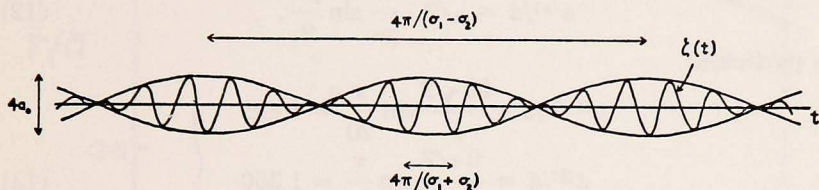


Figure 2. Combination of two sine-waves of slightly different frequency.

Example 2. Two sine-waves. Consider the sum of two sine-waves of equal amplitude but of very slightly differing period:

$$\zeta = a_0 \cos \sigma_1 t + a_0 \cos \sigma_2 t , \quad (6)$$

say, where $|\sigma_1 - \sigma_2| \ll |\sigma_1 + \sigma_2|$. We may write

$$\zeta = 2a_0 \cos \frac{\sigma_1 + \sigma_2}{2} t \cos \frac{\sigma_1 - \sigma_2}{2} t , \quad (7)$$

showing that the resultant consists of a carrier wave $\cos \frac{\sigma_1 + \sigma_2}{2} t$, whose period is nearly the same as that of the two component waves, modulated by an envelope function $2a_0 \cos \frac{\sigma_1 - \sigma_2}{2} t$, whose period, by hypothesis, is comparatively long (see Fig. 2). The maxima and minima of ζ occur nearly on the envelope and so are nearly equal in magnitude to the magnitude of the envelope function. In the limit they are distributed at even intervals along the t -axis. Taking the interval $0 < t < \pi/(\sigma_1 - \sigma_2)$ as typical, and supposing it contains N waves, we see that the highest pN waves will be contained in the interval $0 < t < p\pi/(\sigma_1 - \sigma_2)$. The mean amplitude $a^{(p)}$ of these is given by

$$\frac{p\pi}{\sigma_1 - \sigma_2} a^{(p)} = \int_0^{p\pi/(\sigma_1 - \sigma_2)} 2a_0 \cos \frac{\sigma_1 - \sigma_2}{2} t dt , \quad (8)$$

and hence

$$a^{(p)} = 2a_0 \cdot \frac{2}{p\pi} \sin \frac{p\pi}{2} . \quad (9)$$

The root-mean-square wave-height is given by

$$\frac{\pi}{\sigma_1 - \sigma_2} \bar{a}^2 = \int_0^{\pi/(\sigma_1 - \sigma_2)} 4a_0^2 \cos^2 \frac{\sigma_1 - \sigma_2}{2} t dt, \quad (10)$$

and hence

$$\bar{a}^2 = 2a_0, \quad \bar{a} = \sqrt{2} a_0. \quad (11)$$

Thus

$$a^{(p)}/\bar{a} = \sqrt{2} \cdot \frac{2}{p\pi} \sin \frac{p\pi}{2}. \quad (12)$$

In particular

$$\left. \begin{aligned} a^{(10)}/\bar{a} &= \frac{20\sqrt{2}}{\pi} \sin \frac{\pi}{20} = 1.408 \\ a^{(4)}/\bar{a} &= \frac{6\sqrt{2}}{\pi} \sin \frac{\pi}{6} = 1.350 \\ a^{(1)}/\bar{a} &= \frac{2\sqrt{2}}{\pi} \sin \frac{\pi}{2} = 0.901 \end{aligned} \right\} \quad (13)$$

$a^{(p)}/\bar{a}$ is plotted against p in Fig. 3. We have also

$$a_{\max} = 2a_0; \quad a_{\max}/\bar{a} = \sqrt{2}. \quad (14)$$

The statistical distribution of the wave amplitudes is evidently the same as that of the envelope function, which is that of the simple cosine curve

$$r = 2a_0 \cos \theta. \quad (15)$$

See Fig. 4. The probability that a point in the interval $0 < \theta < \pi/2$ lies in any given region of width $d\theta$ is $2|d\theta|/\pi$. Hence the probability $P(r)|dr|$ that the function (15) lies between r and $r + dr$ is given by

$$P(r) |dr| = \frac{2}{\pi} |d\theta|. \quad (16)$$

Thus, when $0 < r < 2a_0$,

$$P(r) = \frac{2}{\pi} \left| \frac{d\theta}{dr} \right| = \frac{2}{\pi} \frac{1}{2a_0 \sin \theta} = \frac{2}{\pi} \frac{1}{(4a_0^2 - r^2)^{\frac{1}{2}}}. \quad (17)$$

Clearly a can never exceed $2a_0$ or $\sqrt{2}\bar{a}$. Hence the probability-distribution $P(r)$ of the wave-height a is given by

$$P(r) = \begin{cases} \frac{2}{\pi} \frac{1}{(2\bar{a}^2 - r^2)^{\frac{1}{2}}}, & (r < \sqrt{2}\bar{a}) \\ 0, & (r > \sqrt{2}\bar{a}) \end{cases} \quad (18)$$

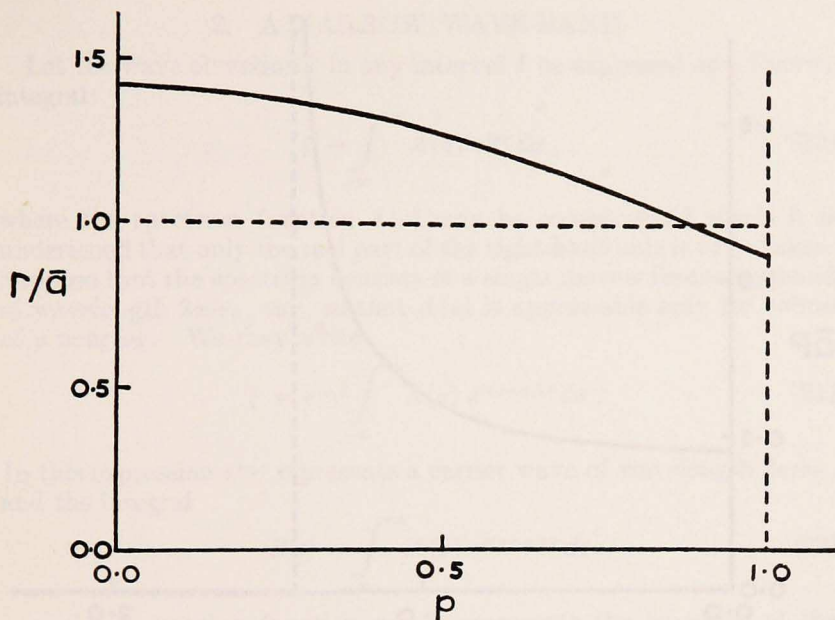


Figure 3. Graph of $a^{(p)}/\bar{a}$ as a function of p , for two sine-waves of slightly different frequency.

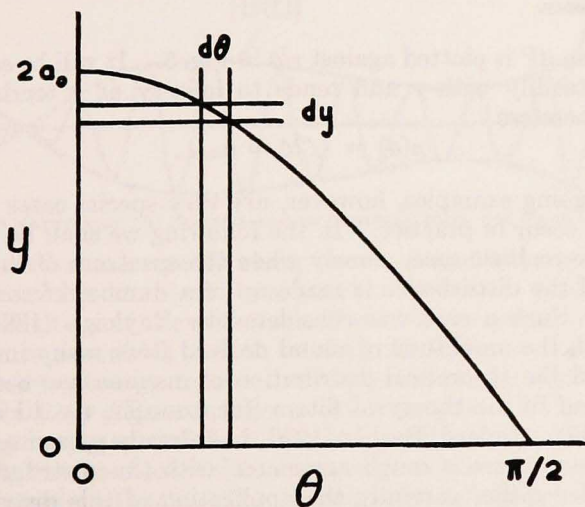


Figure 4. The curve $y = 2a_0 \cos \theta$.

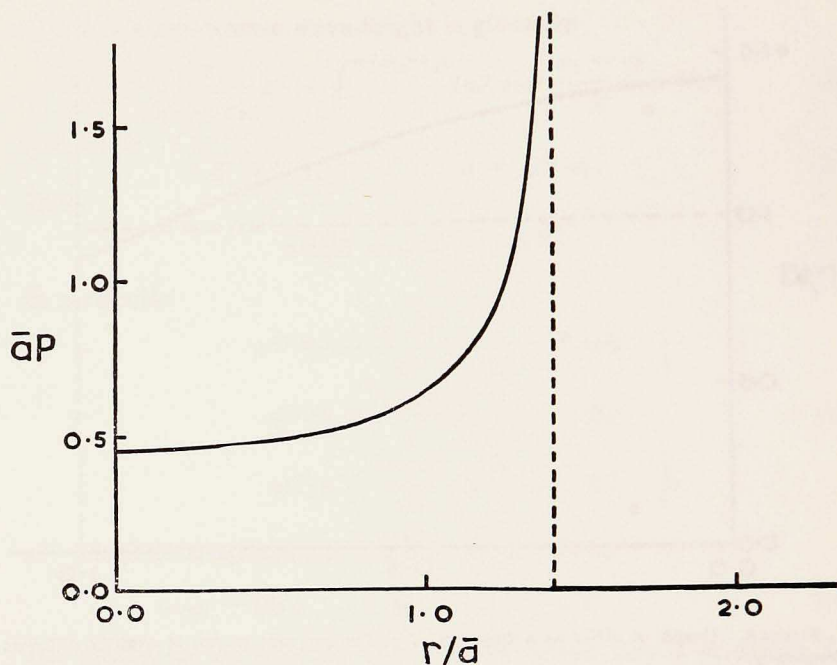


Figure 5. Frequency distribution of the wave amplitude for two sine-waves of slightly different frequency.

The function $\bar{a}P$ is plotted against r/\bar{a} in Fig. 5. It will be seen that P increases steadily with r and tends to infinity as r tends to $\sqrt{2}\bar{a}$. We have therefore

$$\mu(a) = \sqrt{2}\bar{a} = a_{\max}. \quad (19)$$

The foregoing examples, however, are very special cases which are unlikely to occur in practice. In the following we shall be concerned with a more realistic case, namely when the spectrum of the waves is narrow and the disturbance is made up of a number of random contributions. Such a case was considered by Rayleigh (1880) in connection with the amplitude of sound derived from many independent sources, and the theoretical distribution of maxima has been used in acoustics and in the theory of filters (for example, see Rice, 1944-5; Eckart, 1950). Indeed, Barber (1950) has already presented evidence that for waves there is rough agreement with this distribution. We shall consider rather carefully the application of this distribution to sea waves, find the theoretical values of $a^{(p)}/\bar{a}$ and the distribution of a_{\max}/\bar{a} , and compare the results with observation.

2. A NARROW WAVE-BAND

Let the wave elevation ζ in any interval I be expressed as a Fourier integral:

$$\zeta = \int_{-\infty}^{\infty} A(\sigma) e^{i\sigma t} d\sigma, \quad (20)$$

where the spectrum function $A(\sigma)$ may be complex and where it is understood that only the real part of the right-hand side is to be taken. Suppose that the spectrum consists of a single narrow frequency-band of wavelength $2\pi/\sigma_0$, say, so that $A(\sigma)$ is appreciable only for values of σ near σ_0 . We may write

$$\zeta = e^{i\sigma_0 t} \int_{-\infty}^{\infty} A(\sigma) e^{i(\sigma-\sigma_0)t} d\sigma. \quad (21)$$

In this expression $e^{i\sigma_0 t}$ represents a carrier wave of wavelength $2\pi/\sigma_0$, and the integral

$$B(t) = \int_{-\infty}^{\infty} A(\sigma) e^{i(\sigma-\sigma_0)t} d\sigma \quad (22)$$

is a slowly varying function which represents the envelope of the waves (see Fig. 6). As in Example 2 above, the maxima and minima

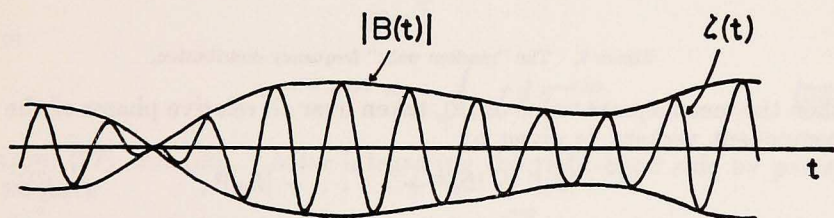


Figure 6. A disturbance $\zeta(t)$ having a narrow frequency band, and its envelope $|B(t)|$.

of ζ are spaced nearly evenly along the t -axis and are approximately equal to the value of $|B|$ at these points. It follows that the probability-distribution of the wave amplitudes is the same as the probability-distribution of $|B|$, which we shall therefore consider.

Now the wave-energy received at any point on the coast will have originated in many different places over a wide area. We may imagine that the generating area of the waves is divided into a large number of different regions, each of which will contribute to the wave-height ζ and to the envelope function B . If each region of the generating area is sufficiently large compared with a wavelength, it may be assumed that the phases of the contributions from different regions are inde-

pendent of one another. Then it is reasonable to assume that B is the sum of a very large number of small components of random phase. The probability-distribution of such a sum, which is known as the "random walk," was found by Rayleigh (1880) and has since been studied by many workers (for references, see Bartels, 1935). If the component vectors are $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_M$, i. e., if

$$B = \mathbf{b}_1 + \mathbf{b}_2 + \dots + \mathbf{b}_M, \quad (23)$$

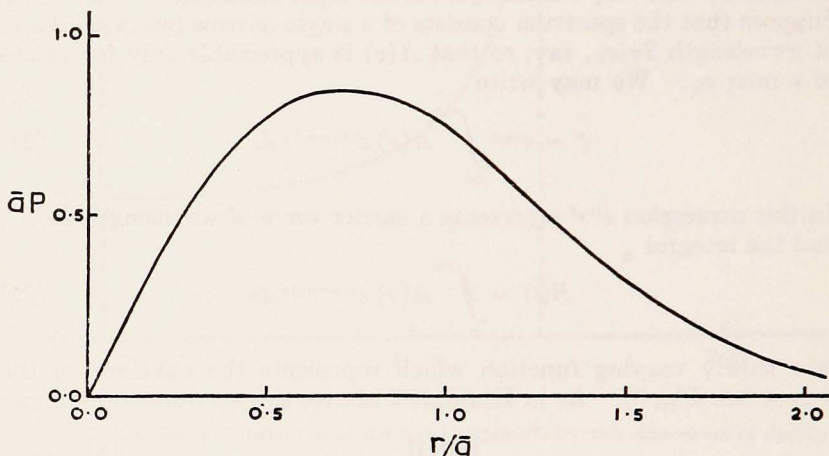


Figure 7. The "random walk" frequency distribution.

then the mean square value of $|B|$, taken over all relative phases of the component vectors, is given by

$$\bar{B}^2 = |\mathbf{b}_1|^2 + |\mathbf{b}_2|^2 + \dots + |\mathbf{b}_M|^2, \quad (24)$$

and, under certain general restrictions on the size of the component vectors (see Khintchine, 1933) the probability that $|B|$ lie between r and $r + dr$ is given by

$$P(r)dr = e^{-r^2/\bar{B}^2} \cdot 2r/\bar{B}^2 dr. \quad (25)$$

Since the probability-distribution of $|B|$ equals that of the a 's, we have

$$\bar{B} = \bar{a}, \quad (26)$$

a quantity that can be estimated from observation; no detailed knowledge of the component vectors is required. Thus the probability-distribution of the a 's is given by

$$P(r) dr = e^{-r^2/\bar{a}^2} \cdot 2r/\bar{a}^2 dr = -de^{-r^2/\bar{a}^2}. \quad (27)$$

The function

$$\bar{a}P(r) = e^{-r^2/\bar{a}^2} \cdot 2r/\bar{a} \quad (28)$$

is shown in Fig. 7 (cf. Fig. 5). It will be seen that $P(r)$ is zero when $r = 0$, that it increases to a maximum, and then falls away rapidly for large values of r/\bar{a} . The total area under the curve is, of course, unity. The maximum value occurs when $r/\bar{a} = 1/\sqrt{2}$, so that the mode $\mu(a)$ is given by

$$\mu(a)/\bar{a} = \frac{1}{\sqrt{2}} = 0.717. \quad (29)$$

The chance $\varphi(r)$ that a should exceed a certain value r is given by

$$\varphi(r) = \int_r^\infty P(r)dr = e^{-r^2/\bar{a}^2}. \quad (30)$$

To find $a^{(p)}$, we note, first, that the proportion p of a 's which exceed a certain value r is equal to $\varphi(r)$, so that from (30)

$$p = e^{-r^2/\bar{a}^2}; \quad r = \left(\log \frac{1}{p} \right)^{\frac{1}{2}}. \quad (31)$$

The mean value $a^{(p)}$ of those a 's that are greater than r is given by

$$\varphi(r) a^{(p)} = \int_r^\infty r P(r) dr \quad (32)$$

or

$$e^{-r^2/\bar{a}^2} a^{(p)} = - \int_r^\infty r d e^{-r^2/\bar{a}^2} \quad (33)$$

from (27) and (30). After integrating the right-hand side by parts, we have

$$e^{-r^2/\bar{a}^2} a^{(p)} = r e^{-r^2/\bar{a}^2} + \int_r^\infty e^{-r^2/\bar{a}^2} dr, \quad (34)$$

and hence

$$\begin{aligned} a^{(p)}/\bar{a} &= r/\bar{a} + e^{r^2/\bar{a}^2} \int_r^\infty e^{-r^2/\bar{a}^2} dr/\bar{a} \\ &= \left(\log \frac{1}{p} \right)^{\frac{1}{2}} + \frac{1}{p} \int_{\left(\log \frac{1}{p} \right)^{\frac{1}{2}}}^\infty e^{-\theta^2} d\theta \end{aligned} \quad (35)$$

$$= \left(\log \frac{1}{p} \right)^{\frac{1}{2}} + \frac{1}{p} \frac{\sqrt{\pi}}{2} \left[1 - H \left\{ \left(\log \frac{1}{p} \right)^{\frac{1}{2}} \right\} \right], \quad (36)$$

where $H(\theta)$ is the probability function:

$$H(\theta) = \frac{2}{\sqrt{\pi}} \int_0^\theta e^{-\theta^2} d\theta. \quad (37)$$

Numerical values of $a^{(p)}/\bar{a}$ are given in Table I for some representative values of p . In particular, the mean $a^{(1)}$ is given by

$$a^{(1)}/\bar{a} = \frac{\sqrt{\pi}}{2} = 0.886. \quad (38)$$

TABLE I. REPRESENTATIVE VALUES OF $a^{(p)}/\bar{a}$ IN THE CASE OF A NARROW WAVE SPECTRUM

p	$a^{(p)}/\bar{a}$	p	$a^{(p)}/\bar{a}$
0.01	2.359	0.4	1.347
0.05	1.986	0.5	1.256
0.1	1.800	0.6	1.176
0.2	1.591	0.7	1.102
0.25	1.517	0.8	1.031
0.3	1.454	0.9	0.961
0.3333	1.416	1.0	0.886

The second moment of the distribution about the origin being \bar{a}^2 , by definition, we have for the second moment about the mean:

$$[\sigma(a)]^2 = \bar{a}^2 - a^{(1)2} = \bar{a}^2 \left(1 - \frac{\pi}{4}\right). \quad (39)$$

Thus the standard deviation $\sigma(a)$ is given by

$$\sigma(a)/\bar{a} = \left(1 - \frac{\pi}{4}\right)^{\frac{1}{2}} = 0.453. \quad (40)$$

$a^{(p)}/\bar{a}$ is shown as a function of p in Fig. 8, which may be compared with Fig. 3. An asymptotic formula for $a^{(p)}/\bar{a}$ when p approaches zero is found by further integration by parts in equation (35):

$$a^{(p)}/\bar{a} = \left(\log \frac{1}{p}\right)^{\frac{1}{2}} + \frac{1}{2} \left(\log \frac{1}{p}\right)^{-\frac{1}{2}} - \frac{1}{2} \cdot \frac{3}{2} \left(\log \frac{1}{p}\right)^{-\frac{3}{2}} + \frac{1}{2} \cdot \frac{3}{2} \cdot \frac{5}{2} \left(\log \frac{1}{p}\right)^{-\frac{5}{2}} \dots \dots \dots (41)$$

Thus, in Fig. 8, as p tends to 0, $a^{(p)}/\bar{a}$ tends to infinity logarithmically. However, for the validity of this result it is essential that the fraction of the sample containing the highest pN wave amplitudes a shall not be too small; otherwise the present approximation will not hold.

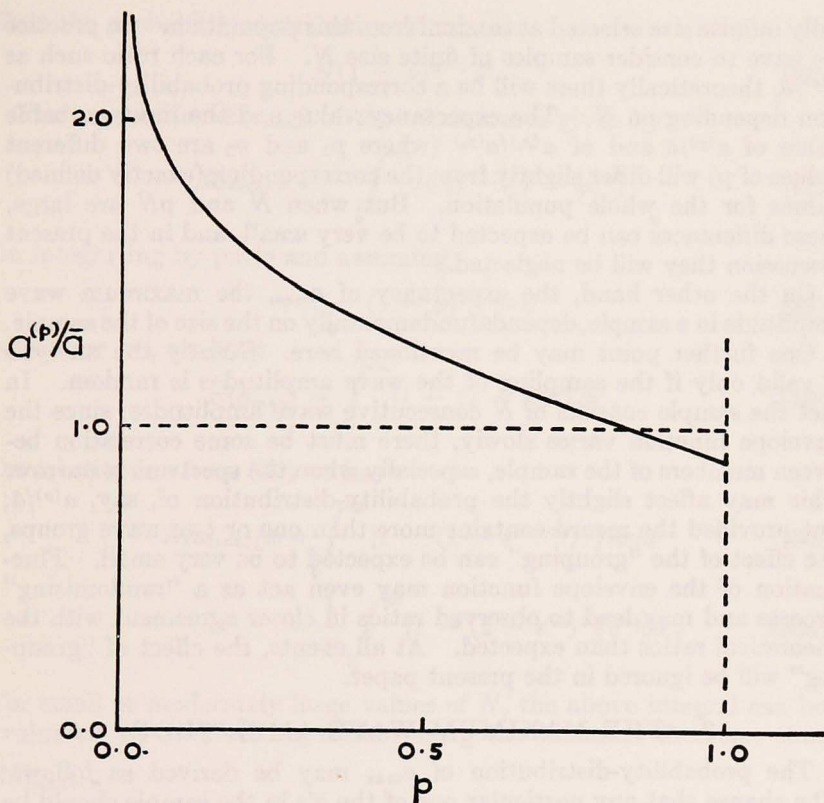


Figure 8. Graph of $a^{(p)}/\bar{a}$ as a function of p , for the narrow frequency band.

Suppose, for example, that we wish to find the expectancy $E(a_{\max})$ of the highest wave in an interval containing N waves. An approximate answer might be obtained by setting $p = 1/N$ and finding $a^{(1/N)}$. But this answer will not be exactly correct; for $a^{(1/N)}$ in fact represents the mean height of the $1/N$ th highest waves in a large sample, say a sample of size mN obtained by collecting together m samples, each containing N wave amplitudes. The m highest waves may not be distributed evenly, one in each of the samples; and if not, the mean of the highest waves, one from each group, will clearly be less than the mean of the m highest from all mN wave amplitudes together. Hence we see that the expected value of a_{\max} must be somewhat less than $a^{(1/N)}$.

Of course the ratios $a^{(p)}/\bar{a}$ found above refer to the total statistical "population" of wave amplitudes, or at least to a sample of theoretic-

cally infinite size selected at random from this population. In practice we have to consider samples of finite size N . For each ratio such as $a^{(p)}/\bar{a}$, theoretically there will be a corresponding probability-distribution depending on N . The expectancy value and the most probable value of $a^{(p)}/\bar{a}$ and of $a^{(p_1)}/a^{(p_2)}$ (where p_1 and p_2 are two different values of p) will differ slightly from the corresponding (exactly defined) values for the whole population. But when N and pN are large, these differences can be expected to be very small, and in the present discussion they will be neglected.

On the other hand, the expectancy of a_{\max} , the maximum wave amplitude in a sample, depends fundamentally on the size of the sample.

One further point may be mentioned here. Strictly the analysis is valid only if the sampling of the wave amplitudes is random. In fact the sample consists of N consecutive wave amplitudes; since the envelope function varies slowly, there must be some correlation between members of the sample, especially when the spectrum is narrow. This may affect slightly the probability-distribution of, say, $a^{(p)}/\bar{a}$; but provided the record contains more than one or two wave groups, the effect of the "grouping" can be expected to be very small. Fluctuation of the envelope function may even act as a "randomising" process and may lead to observed ratios in closer agreement with the theoretical ratios than expected. At all events, the effect of "grouping" will be ignored in the present paper.

3. THE MAXIMUM WAVE AMPLITUDE

The probability-distribution of a_{\max} may be derived as follows: The chance that any particular one of the a 's in the sample should be less than r , say, is

$$\int_0^r P(r) dr = 1 - \varphi(r), \quad (42)$$

where φ is given by equation (30). The chance that every a in the sample shall be less than r is therefore $(1 - \varphi)^N$; and the chance that at least one a shall exceed r is $1 - (1 - \varphi)^N$. The chance that the maximum a shall lie in the interval $(r, r + dr)$ is the chance that at least one a shall exceed r , minus the chance that at least one a shall exceed $r + dr$, that is,

$$- d [1 - (1 - \varphi)^N], = d (1 - \varphi)^N, \quad (43)$$

or

$$N(1 - \varphi)^{N-1} P dr, \quad (44)$$

since

$$\frac{d\varphi}{dr} = - P. \quad (45)$$

Thus the probability-distribution of a_{\max} is

$$N(1 - \varphi)^{N-1} P. \quad (46)$$

The expectation $E(a_{\max})$ of the maximum is given by

$$E(a_{\max}) = - \int_0^{\infty} r d[1 - (1 - \varphi)^N]. \quad (47)$$

On integrating by parts and assuming

$$r[1 - (1 - \varphi)^N] \rightarrow 0 \quad (48)$$

when $r \rightarrow \infty$, we have

$$E(a_{\max}) = \int_0^{\infty} [1 - (1 - \varphi)^N] dr. \quad (49)$$

On substitution from (30) we have

$$E(a_{\max})/\bar{a} = \int_0^{\infty} [1 - (1 - e^{-r^2/\bar{a}^2})^N] dr/\bar{a} \quad (50)$$

$$= \frac{1}{2} \int_0^{\infty} [1 - (1 - e^{-\theta})^N] \theta^{-\frac{1}{2}} d\theta. \quad (50a)$$

For small or moderately large values of N , the above integral can be evaluated by a direct expansion, using the binomial theorem; thus

$E(a_{\max})/\bar{a} =$

$$\frac{1}{2} \int_0^{\infty} \left[N e^{-\theta} - \frac{N(N-1)}{1.2} e^{-2\theta} + \dots (-)^{NH} e^{-N\theta} \right] \theta^{-\frac{1}{2}} d\theta, \quad (51)$$

and since

$$\int_0^{\infty} e^{-n\theta} \theta^{-\frac{1}{2}} d\theta = n^{-\frac{1}{2}} \left(-\frac{1}{2} \right)! = \left(\frac{\pi}{n} \right)^{\frac{1}{2}}, \quad (52)$$

we have

$$E(a_{\max})/\bar{a} = \frac{\sqrt{\pi}}{2} \left[\frac{N}{\sqrt{1}} - \frac{N(N-1)}{1.2} \frac{1}{\sqrt{2}} + \dots (-)^{N+1} \frac{1}{\sqrt{N}} \right]. \quad (53)$$

Table II gives the exact values of the integral for $N = 1, 2, 5, 10$ and 20 . However, we are chiefly interested in values of N of the order of 50 or more, for which the binomial coefficients in (53) become so large that computation by means of that expression becomes impracticable.

An asymptotic expression for large values of N may be found as follows: write

$$\theta_0 = \log N; \quad e^{-\theta_0} = \frac{1}{N}; \quad \theta = \theta_0 + \theta', \quad (54)$$

say, then

$$(1 - e^{-\theta})^N = \left(1 - \frac{e^{-\theta'}}{N}\right)^N \doteq e^{-e^{-\theta'}}, \quad (55)$$

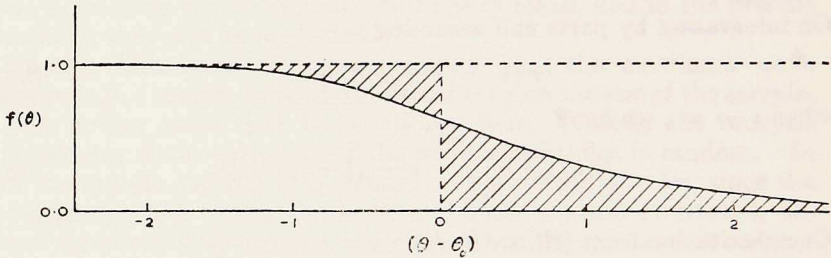


Figure 9. Graph of the function $f(\theta) = (1 - e^{-\theta})^N$ when N is large.

with errors of order $1/N$. Thus we see that the first function in the integrand of (43), i. e.,

$$f(\theta) = 1 - (1 - e^{-\theta})^N, \quad (56)$$

has a rather sudden drop from 1 to 0 in the neighborhood of θ_0 (see Fig. 9). We have, therefore,

$$\begin{aligned} I &= \frac{1}{2} \int_0^{\theta_0} \theta^{-\frac{1}{2}} d\theta + R \\ &= \theta_0^{\frac{1}{2}} + R, \end{aligned} \quad (57)$$

where R is a remainder of order $\theta_0^{-\frac{1}{2}}$ at most; in fact

$$\begin{aligned} R &= \frac{1}{2} \theta_0^{-\frac{1}{2}} \left[- \int_{-\infty}^0 e^{-e^{-\theta'}} d\theta' + \int_0^{\infty} (1 - e^{-e^{-\theta'}}) d\theta' \right] + R' \\ &= \frac{1}{2} \theta_0^{-\frac{1}{2}} \left[- \int_1^{\infty} e^{-\alpha} \frac{d\alpha}{\alpha} + \int_0^1 (1 - e^{-\alpha}) \frac{d\alpha}{\alpha} \right] + R' \\ &= \frac{1}{2} \gamma \theta_0^{-\frac{1}{2}} + R', \end{aligned} \quad (58)$$

where R' is of order $\theta_0^{-\frac{3}{2}}$ at most and γ is Euler's constant ($= 0.5772,2$; see Whittaker and Watson, 1950: 236). Thus

$$E(a_{\max}) / \bar{a} = (\log N)^{\frac{1}{2}} + \frac{1}{2} \gamma (\log N)^{-\frac{1}{2}} + O(\log N)^{-\frac{3}{2}}. \quad (59)$$

TABLE II. VALUES OF $E(a_{\max})/\bar{a}$ AND $\mu(a_{\max})/\bar{a}$ FOR DIFFERENT VALUES OF N , FOR A NARROW SPECTRUM

N	$(\log N)^{\frac{1}{2}}$	$E(a_{\max})/\bar{a}$		$\mu(a_{\max})/\bar{a}$
		exact expression	asymptotic expression	
1	0.000	0.886	—	0.707
2	0.833	1.146	—	1.030
5	1.269	1.462	—	1.366
10	1.517	1.676	1.708	1.583
20	1.731	1.870	1.898	1.778
50	1.978	—	2.124	2.010
100	2.146	—	2.280	2.172
200	2.302	—	2.426	2.323
500	2.493	—	2.609	2.509
1,000	2.628	—	2.738	2.642
2,000	2.757	—	2.862	2.769
5,000	2.918	—	3.017	2.929
10,000	3.035	—	3.130	3.044
20,000	3.147	—	3.239	3.155
50,000	3.289	—	3.377	3.296
100,000	3.393	—	3.478	3.400

The above equation may be compared with equation (41) for $a^{(p)}/\bar{a}$ when $p = 1/N$. We see that $E(a_{\max})$ differs from $a^{(1/N)}$ in the second term of the asymptotic expansion. Since $\frac{1}{2}\gamma = 0.28861$, we have $E(a_{\max}) < a^{(1/N)}$ as expected. For large N , however, $E(a_{\max})$ still increases like $(\log N)^{\frac{1}{2}}$ and therefore tends to infinity with the length of the interval, though very slowly. Values of the asymptotic expression for $E(a_{\max})/\bar{a}$ are given in Table II for values of N ranging from 10 to 100,000. It will be seen that in this range $E(a_{\max})/\bar{a}$ increases only from about 1.7 to about 3.5. The asymptotic expression may be compared with the exact expression for $N = 10$ and 20. The differences in the two cases are 0.032 and 0.028 respectively, or about 2%. For $N \geq 50$ the error in the asymptotic expression is almost certainly less than 0.03; for large values of N the error may be expected to diminish proportionally to $(\log N)^{-\frac{1}{2}}$.

The "most probable" value of a_{\max} , which we shall denote by $\mu(a_{\max})$, is given simply by the maximum value of the probability-distribution (46), i. e., the maximum of

$$N(1 - e^{-r^2/\bar{a}^2})^{N-1} e^{-r^2/\bar{a}^2} \cdot 2r/\bar{a}. \quad (60)$$

Writing

$$\theta = r^2/\bar{a}^2, \quad (61)$$

we see that we must have

$$\frac{d}{d\theta} \left[\theta^{\frac{1}{2}} e^{-\theta} (1 - e^{-\theta})^{N-1} \right] = 0, \quad (62)$$

and hence

$$\frac{1}{2\theta} - 1 + \frac{(N-1)e^{-\theta}}{1-e^{-\theta}} = 0 \quad (63)$$

or

$$Ne^{-\theta} = 1 - \frac{1}{2\theta} (1 - e^{-\theta}) \quad (64)$$

or

$$\theta = \log N - \log \left[1 - \frac{1}{2\theta} (1 - e^{-\theta}) \right] \quad (65)$$

$$= \log N + 0 \left(\frac{1}{\theta} \right) \quad (66)$$

when N is large. Thus

$$\theta = \log N + 0 (\log N)^{-1}, \quad (67)$$

and the most probable value of a_{\max} is given by

$$\mu(a_{\max})/\bar{a} = \theta^{\frac{1}{2}} = (\log N)^{\frac{1}{2}} + 0(\log N)^{-\frac{3}{2}}. \quad (68)$$

Thus there is no term in $(\log N)^{-\frac{1}{2}}$. Starting with the approximate value $(\log N)^{\frac{1}{2}}$, one may find closer approximations to $\mu(a_{\max})/\bar{a}$ by applying Newton's method of successive approximation (see, for example, Whittaker and Robinson, 1932: 84) in equation (65). Values of $\mu(a_{\max})/\bar{a}$ so found are given in the last column of Table II.

The chance that a_{\max} shall not exceed $(\log N)^{\frac{1}{2}}$ by more than a given amount may be found similarly. We have seen that the probability-distribution of a_{\max} is given by

$$\frac{d}{dr} (1 - \theta)^N = \frac{d}{dr} (1 - e^{-r^2/\bar{a}^2})^N. \quad (69)$$

If we define r_0 by the equations

$$r_0^2/\bar{a}^2 = \theta_0 = \log N; \quad e^{-r_0^2/\bar{a}^2} = \frac{1}{N}, \quad (70)$$

we have

$$\frac{d}{dr} \left(1 - \frac{e^{-(r^2-r_0^2)/\bar{a}^2}}{N} \right)^N = \frac{d}{dr} e^{-e^{-(r^2-r_0^2)/\bar{a}^2}} \quad (71)$$

approximately. The probability that a will be less than r is therefore

$$e^{-e^{-(r^2-r_0^2)/\bar{a}^2}}, \quad (72)$$

and the probability that a will be greater than r is

$$1 - e^{-e^{-(r^2-r_0^2)/\bar{a}^2}} \quad (73)$$

4. DISCUSSION

The chief assumptions used in the derivation of the theoretical distribution $P(r)$ in sections 2 and 3 are: (a) that the frequency spectrum consist of a single narrow frequency band; and (b) that the waves be considered as the sum of a large number of contributions, all of about the same frequency, and of random phase. Let us consider under what conditions these assumptions should be satisfied.

In the first place, the analysis will not apply to regular trains of waves produced by a simple organized mechanism: for example, the waves generated by a paddle in a model wave tank, or the transverse waves produced by a ship; or the wave-height distribution resulting from the interference of two wave trains of the same amplitude but of slightly differing wavelength, as in Example 2 (p. 247). In the open sea such examples constitute very special and most unlikely cases.

The present analysis is meant to apply to wind-generated waves. Since the dimensions of a storm area are large compared with the wavelengths we are considering, it is fairly safe to suppose that the phases of contributions from different parts of the storm area are random. However, the range of frequencies may not be narrow; if there are two distinct sources of wave-energy, for example a distant storm and local winds, there may well be two distinct frequency-bands in the spectrum. The most satisfactory conditions would be represented by a single storm at a great distance (compared with the dimensions of the storm); for, in the course of propagation, different frequencies in the spectrum, being propagated with different velocities, will become spread out in space, and over a short interval of time only a narrow range will be present. It must be assumed that the time of recording is not too long, so that in this time the frequency and amplitude of the waves do not change significantly. On the other hand the time must be long enough to ensure that the sample of wave-heights is sufficiently representative; this requires that at least several "groups" of waves be included in the record.

The method of recording may affect the apparent frequency spectrum of the waves. For example, if the waves are recorded by measurement of pressure on the bottom, as is now usual, the high frequencies, which are attenuated rapidly with depth, will be damped out relative to the lower frequencies, and the corresponding frequency spectrum will therefore be narrower. Hence the present analysis may apply more closely to a record of pressure on the bottom than to the actual surface elevation. In fact, the free surface, if viewed very closely, will usually show a large number of short steep waves, which may constitute maxima and minima of the wave elevation but which

are not normally of interest to us; for example, they would not affect the rolling or heaving motion of a ship. Strictly speaking, we should consider only that modification of the spectrum which is relevant to the purpose in hand. A ship itself acts as a resonant filter, which will amplify those components in the spectrum which are close to its natural period. The present analysis, therefore, might very well be applied to the angular deflection of a ship's mast from the vertical; from an analysis of the rolling motion over a few minutes, one could easily compute the maximum roll that is likely to be encountered during, say, the next hour, assuming that the sea conditions remained constant; for, in the notation in Sections 2 and 3, one could estimate \bar{a} with fair accuracy from observation over the shorter interval and hence find $E(a_{\max})$ or $\mu(a_{\max})$ over the proposed longer interval.

However, an important restriction should be mentioned here. One of the conditions implied in assumption (b) above is that the contributions from different parts of the generating area should be superposable; that is, the mechanical system we are dealing with should be linear. This assumption can be shown to be valid for low waves in deep water; but clearly it will not hold for waves approaching the maximum height. For this reason the analysis could not be applied to surf or to waves in the open sea which are nearly at their maximum height. Nor could it be applied to the rolling motion of a ship when this is large enough for nonlinear terms to become important or when the rolling is so great that the ship may capsize.

With these restrictions in mind let us compare some previous observations with the theoretical results of Sections 2 and 3.

5. COMPARISON WITH OBSERVATION

Munk (1944),² in an analysis of wave records taken at the Scripps Institution, California, compared the mean height $H^{(3/10)}$ of the highest 30% of the waves with the mean height $H^{(1)}$ of all of the waves. He found

$$\frac{H^{(3/10)}}{H^{(1)}} = \frac{0.49}{0.32} = 1.53.$$

The theoretical value from Table I is given by

$$\frac{a^{(3/10)}}{a^{(1)}} = \frac{a^{(3/10)}/\bar{a}}{a^{(1)}/\bar{a}} = \frac{1.454}{0.886} = 1.64.$$

Seiwell (1948) found that in two different localities in the Atlantic (off Cuttyhunk Island and off Bermuda) the ratio of $a^{(4)}$ to $a^{(1)}$ was

² See also Snodgrass (1951) where these and other unpublished observations are summarized.

1.57. Putz (1950) studied 25 wave records from five different localities; the mean values of the ratios $a^{(4)}/a^{(1)}$ can be found from table 1 of his paper. They are respectively 1.59 (Oceanside, 4 records); 1.66 (Point Sur, 15 records); 1.66 (Hecata Head, 2 records); 1.55 (Guam, 3 records) and 1.54 (Point Arguello, 1 record). The mean of Putz's observations, weighted according to the number of records considered, is 1.63. The theoretical value, from Table I of this paper, is given by

$$\frac{a^{(4)}}{a^{(1)}} = \frac{1.416}{0.886} = 1.60,$$

which is in fairly close agreement.

Wiegel (1949) has studied wave records from three different localities off the Pacific Coast of the United States. He found, in the three cases, that

$$\frac{a^{(1/10)}}{a^{(1)}} = 1.27, 1.30, 1.30 \text{ (mean value 1.29)}.$$

The individual estimates showed little scatter. Wiegel remarks, "Even more surprising was the fact that almost every point was within plus or minus ten per cent of this value (1.29)." The theoretical value, from Table I, is given by

$$\frac{a^{(4)}}{a^{(1)}} = \frac{1.800}{1.416} = 1.27,$$

which again is in quite close agreement.

Wiegel also compared the maximum wave from three 20-minute records each day with the mean height of the highest one-third waves. His observed values were equivalent to

$$\frac{E(a_{\max})}{a^{(4)}} = 1.85, 1.91, 1.85 \text{ (mean value 1.87)}.$$

Assuming a mean wave period of 12 seconds, we have

$$N = \frac{60 \text{ minutes}}{12 \text{ seconds}} = 300,$$

for which we find from the asymptotic formula (59):

$$\frac{E(a_{\max})}{\bar{a}} = 2.504.$$

Thus, theoretically,

$$\frac{E(a_{\max})}{a^{(4)}} = \frac{2.504}{1.416} = 1.77.$$

A possible explanation of the observed value of $E(a_{\max})/a^{(4)}$ being greater than the theoretical value is as follows. Suppose that during the day the state of the sea, as represented by the r. m. s. wave amplitude \bar{a} , underwent a change. If, during one of the three 20-minute records, the r. m. s. wave amplitude is much larger than, say twice, that in the other two, then the maximum wave amplitude will almost certainly be found in that record, and the expected value of the maximum will not be much less than twice the expected value if the wave characteristics had not changed; for $E(a_{\max})/\bar{a}$ is not much different for a 20-minute interval than for a 60-minute interval. On the other hand, $a^{(4)}$ will be multiplied by approximately 4/3. Therefore $E(a_{\max})/\bar{a}$ will be increased by about $2 \div 4/3$ or by 3/2. The same tendency is true in general.

Darbyshire (1953) has shown that in a 20-minute wave record the maximum wave-height is about twice the "equivalent wave-height," which is defined by him as the height of the uniform train of waves which would have the same total energy as the actual waves. On our present assumption that the spectrum of the waves contains only a single narrow band, the equivalent wave-height equals the root-mean-square wave height \bar{a} ; for, each wave in the record is approximately a sine-wave of the same length, and the energy per wavelength is proportional to a^2 . It may be more appropriate, in this case, to consider the most probable value $\mu(a_{\max})$ of the highest wave rather than the expectancy $E(a_{\max})$. For a mean wave period of 12 seconds we should have

$$N = \frac{20 \text{ minutes}}{12 \text{ seconds}} = 100,$$

and so from Table II

$$\mu(a_{\max})/\bar{a} = 2.17.$$

However, we see from Table II that $E(a_{\max})$ is only slightly greater. For longer wave periods, N , and therefore $\mu(a_{\max})/\bar{a}$, would be slightly less; the rather slow change in $\mu(a_{\max})$ with N would account for the success of the empirical rule irrespective of the period of the waves.

In examples quoted above, the discrepancy between theory and observation is in all cases less than 8%, and in some cases it is smaller still. In view of the somewhat strict assumptions made in deriving the theoretical probability-distribution, this agreement is surprisingly close; and it may indicate that the probability-distribution does not depend very critically upon the narrowness of the wave spectrum. For most practical purposes the theoretical values of $a^{(p)}/\bar{a}$ and $E(a_{\max})/\bar{a}$ can be used with confidence; thus, if one of the quantities \bar{a} , $a^{(1)}$, $a^{(4)}$, $a^{(1/10)}$ or $E(a_{\max})$ is known, the others may be estimated immediately.

The present discussion suggests that waves much higher than the average are likely to be extremely rare, for a given state of the sea. Table II shows that even in a time interval containing 100,000 waves, which for 10-second waves would be about $11\frac{1}{2}$ days, the most probable value of the highest wave is less than $3\frac{1}{2}$ times the root-mean-square value; of course it is unlikely that the waves would remain statistically constant throughout this interval. Equation (73) also shows that there is an extremely small chance that the most probable value of a_{max} will be greatly exceeded. The general conclusion then appears that changes in the strength of the wind or other generating forces are more important in producing variability in the wave amplitude than is the statistical variation of the waves at any one time.

REFERENCES

- BARBER, N. F.
1950. Ocean waves and swell. Lecture published by Institution of Electrical Engineers, London. 22 pp.
- BARTELS, J.
1935. Random fluctuations, persistence and quasi-persistence in geophysical and cosmical periodicities. *Terr. Magn. atmos. Elect.*, 40; 1-60.
- DARBYSHIRE, J.
1953. The generation of waves by wind. *Philos. Trans., A.* (in press).
- ECKART, C. H. (ed.)
1950. Principles of underwater sound. National Defense Research Council, Summary Technical Reports, Division 6, vol. 7.
- KHINTCHINE, A.
1933. Asymptotische Gesetze der Wahrscheinlichkeitsrechnung. *Ergebn. Math.*, 2 (4): 1-77; Berlin, Julius Springer.
- MUNK, W. H.
1944. Proposed uniform procedure for observing waves and interpreting instrument records. Scripps Institution of Oceanography, Wave Report 26 (unpublished).
- PUTZ, R. R.
1950. Wave height variability; prediction of distribution function. Institute of Engineering Research, Berkeley, Technical Report HE-116-318 (unpublished).
- RAYLEIGH, LORD
1880. On the resultant of a large number of vibrations of the same pitch and of arbitrary phase. *Phil. Mag.*, 10: 73-78.
- RICE, S. O.
1944-45. The mathematical analysis of random noise. *Bell Syst. tech. J.*, 23: 282-332; 24: 46-156.
- SEIWELL, H. R.
1948. Results of research on surface waves of the western North Atlantic. *Pap. phys. Oceanogr. Met.*, 10 (4).

SNODGRASS, F. E.

1951. Wave recorders. Proc., 1st Conference on Coastal Engineering, pp. 69-81, published by Council on Wave Research, University of California, 1951 (ed., J. W. Johnson).

SVERDRUP, H. U. AND W. H. MUNK

1947. Wind, sea and swell: theory of relations for forecasting. Report N. 601, U. S. Navy Hydrographic Office.

WHITTAKER, E. T. AND G. ROBINSON

1932. Calculus of Observations. 2nd ed., Blackie, London. 395 pp.

WHITTAKER, E. T. AND G. N. WATSON

1950. Modern Analysis. 4th ed., Cambridge University Press. 608 pp.

WIEGEL, R. L.

1949. An analysis of data from wave recorders on the Pacific Coast of the United States. Trans. Amer. geophys. Un., 30: 700-704.