CHAPTER 4

MAXIMUM WAVE HEIGHT PROBABILITIES FOR A RANDOM NUMBER OF RANDOM INTENSITY STORMS

L E Borgman
Professor of Geology and Statistics
University of Wyoming

ABSTRACT

A very general model is presented for the probability distribution function for wave heights in storms with time-varying intensities Some of the possible choices for functions in the model are listed and discussed Techniques for determining the "equivalent rectangular storm" corresponding to a given historically recorded storm are developed The final model formula expresses the probabilities for a random number of random length storms each with random intensities

INTRODUCTION

The probability law for the largest of N independent, identically distributed random variables is covered quite well in statistical and scientific literature Gumbel (1954) provides an excellent survey of the main elements of the theory His book (Gumbel, 1958) gives a very complete bibliography and many additional details

The application of these techniques to determine probabilities for the largest ocean wave heights in a sequence of N identically distributed and independent waves was developed by Longuet-Higgins (1952) What modifications are necessary to yield maximum wave probabilities for storms which vary in intensity with time? Furthermore, how would one obtain probabilities for the maximum wave in a random number of such time-varying storms? These questions will be considered in detail in the following

PRELIMINARY ASSUMPTIONS

The basic assumptions needed in the development are

(1) The probability distribution function

$$F_H(h) = P[H \leq h] \tag{1}$$

for wave heights is known as a function of time-varying intensity parameters Here, and in later deviations, P[ ] will denote the probability of the event indicated within the square brackets The intensity parameters in $F_H(h)$ may be the rootmean square wave height, a, if the Rayleigh distribution is used

$$F_H(h) = \begin{cases} 1 - e^{-h^2/a^2} , & \text{for } h \geq 0 , \\ 0 , & \text{for } h < 0 \end{cases} \tag{2}$$

or the r m s wave height, a, and the breaking wave height if the clipped Rayleigh distribution is used

$$F_H(h) = \begin{cases} \dfrac{1 - e^{-h^2/a^2}}{1 - 2^{-H_b^2/a^2}} , & \text{if } 0 \leq h \leq H_b \\ 0 , & \text{otherwise} \end{cases} \tag{3}$$

Another possibility is the Rice distribution outlined by Longuet-Higgins and Cartwright (1956) which depends on the r m s wave height and a parameter, ε, which is determined from the spectral density for the water level elevations

(2) It will also be assumed that each wave height is statistically independent of the heights of its neighbors This assumption is largely one of convenience The theory is much harder without it However, it has been shown theoretically that the limiting distribution for the maximum of random variables which are what is called "m-dependent" of each other is the same as the limiting distribution for independent random variables (Watson, 1954) The term, m-dependent, used here means that random variables in the sequence with more than m-1 other random variables between them are statistically independent of each other It seems reasonable to assume that a wave height is at most interdependent with the first several wave heights occurring before and after it and essentially independent with waves further back into the past or forward into the future Hence m-dependence seems reasonable for wave heights

Since the limiting distribution is the same for independent as well as m-dependent random variables, one can tentatively presume the independence assumptions for wave heights will not lead to badly incorrect conclusions It would appear that the independence assumption would lead to a conservative estimate of the maximum wave height probabilities, in any case Longuet-Higgins and others have made this same assumption and it will be made here also

(3) It will also be presumed that there is a known, or estimated, function T(t) such that for any small time interval dt, the number of waves in the interval is given by dt/T(t)

A SINGLE TIME-VARYING STORM

Consider first N identically-distributed, independent wave heights, each with probability distribution function, $F_H(h,\underline{a})$ Here $\underline{a}$ denotes the set of one or more intensity parameters which characterize the intensity of the sea conditions For this situation, let $H_1, H_2, H_3, \ldots, H_N$ be the N waves The largest wave in the sequence will be less than or equal to h if, and only if, every one of the waves are less than or equal to h Thus

$$P[\max H \leq h] = P[H_1 \leq h, H_2 \leq h, H_3 \leq h, \ldots, H_N \leq h] \quad (4)$$

But since the wave heights are assumed independent of each other,

$$P[\max H \leq h] = P[H_1 \leq h]\, P[H_2 \leq h]\, P[H_3 \leq h] \ldots P[H_N \leq h] \quad (5)$$

Finally since the N waves are taken to have the same probability distribution function

$$P[\max H \leq h] = \{F_H(h,\underline{a})\}^N$$

Suppose now that the time-varying storm can be subdivided into steps as shown in Table I below

TABLE I

FINITE STEP APPROXIMATION TO THE TIME-VARYING STORM

Time Interval	Number of Waves	Intensity	Interval Width	Period of Waves
t_0 to t_1	N_1	$\underline{a}_1$	Δt_1	T_1
t_1 to t_2	N_2	$\underline{a}_2$	Δt_2	T_2
t_2 to t_3	N_3	$\underline{a}_3$	Δt_3	T_3
t_{m-1} to t_m	N_m	$\underline{a}_m$	Δt_m	T_m

The probabilities for the maximum wave in the entire storm will be the product of the probabilities for each of the steps

$$P[\max H \leq h] = \prod_{j=1}^{m} P[\max H \leq h \text{ in the } j^{th} \text{ step}]$$

$$= \prod_{j=1}^{m} \{F_H(h,\underline{a}_j)\}^{N_j}$$

It is being assumed that the waves within a step change intensity sufficiently slowly so that, to a fair approxkmation, they may be taken as being identically distributed

It follows that the natural logarithm of P[max H ≤ h] can be written

$$\log P[\max H \leq h] = \sum_{j=1}^{m} N_j \log F_H(h,\underline{a}_j) \tag{8}$$

If

$$N_j = \Delta t_j / T_j \tag{9}$$

is substituted into eq (8), one gets

$$\log P[\max H \leq h] = \sum_{j=1}^{m} (1/T_j) \log F_H(h,\underline{a}_j)\Delta t_j \tag{10}$$

Now let $m \to \infty$ and max $\Delta t_j \to 0$ By the usual definition of an integral,

$$\log P[\max H \leq h] = \int_{t_0}^{t_m} [1/T(t)] \log F_H(h,\underline{a}(t))dt \tag{11}$$

It is presumed that the integrand is continuous and uniformly bounded so that the stepwise expression in eq (10) becomes eq (11) in the limit

A SERIES APPROXIMATION

The distribution function $F_H(h,\underline{a}(t))$ will be expanded in a power series about some convenient value h_0 Several possibilities for h_0 are the breaking wave height, H_b, (if the maximum wave is probably going to be close to breaking), and the "expected" probable maximum, V, (which will be defined later) Thus, let

$$\log F_H(h,\underline{a}(t)) = b_0(t) + b_1(t)(h-h_0) + b_2(t)(h-h_0)^2 + \tag{12}$$

It is presumed that the distribution function is differentiable to the required order so that

$$b_0(t) = \log F_H(h_0,\underline{a}(t))$$

$$(1')b_1(t) = \frac{d}{dh} \log F_H(h,\underline{a}(t)), \qquad h=h_0 \tag{13}$$

$$(2')b_2(t) = \frac{d^2}{dh^2} \log F_H(h,\underline{a}(t)), \qquad h=h_0$$

etc

If eq (12) is substituted into eq (11), one gets

$$\log P[\max H \leq h] = B_0 + B_1(h-h_0) + B_2(h-h_0)^2 + \tag{14}$$

with

$$B_k = \int_{t_0}^{t_m} [b_k(t)/T(t)]dt, \qquad k=0,1,2, \tag{15}$$

The evaluation of B_0, B_1, B_2 to whatever number of terms is desired gives a convenient representation of the probability of maximum height as

$$P[\max H \leq h] = \exp\{B_0 + B_1(H-h_0) + B_2(h-h_0)^2 + \quad \} \tag{16}$$

Presumably the first few terms would be sufficient for most situations since the higher order derivatives for most distribution functions become negligible as h grows large

THE COMBINATION OF SEVERAL STORMS

Another advantage of the representaion in eq (14) is that it facilitates the determination of probabilities for the maximum for the combined wave heights in several storms This is under the supposition that the same h_0 has been used for all the storms

Let $P_r[\max H \leq h]$ denote the probability that the maximum wave in the r^{th} storm is less than or equal to h Suppose there are R storms to be considered Then the probabilities for the maximum wave in the combined set of wave heights would be

$$\log P[\max H \leq h] = \sum_{r=1}^{R} \log P_r[\max H \leq h] \tag{17}$$

The function of h expressed by $P_r[\max H \leq h]$ can be evaluated from eq (11) for each storm Alternatively, let B_{kr} be the B_k value from eq (15) for the r^{th} storm Then if the same h_0 was used for each storm, one may write

$$\log P[\max H \leq h] = \sum_{r=1}^{R} B_{0r} + \sum_{r=1}^{R} B_{1r}(h-h_0) + \sum_{r=1}^{R} B_{2r}(h-h_0)^2 + \tag{18}$$

That is, the B_{kr} can be added storm by storm If B_k is redefined as $\sum_{r=1}^{R} B_{kr}$ then eq (16) gives the distribution function for the maximum height in the combined set of waves

The above development is appropriate for hindcasting the probabilities for maximum heights in storms whose fundamental time-varying intensities were measured or are known from other considerations What about probabilities for future periods of time, say the next hundred years? One could take the historical record as given by Wilson (1957) and determine B_0, B_1, B_2 for each storm Then the probability density jointly for (B_0, B_1, B_2) could be estimated from the data and used to make the extension to the future

This procedure appears to have grave disadvantages in that $(B_0, B_1,$

B_2) are not "intuitive" quantities whose meanings are easy to interpret One runs the risk of making mistakes because the unreasonableness of values arising apparently from the data are not recognized A more trustworthy procedure would appear to be to shift over to intuitively interpretable values

To fill this need, the concept of an "equivalent rectangular storm" will be introduced A rectangular storm is defined to be one in which the intensity, wave period, and distribution function for the height of a single wave remain constant during the duration of the storm The "equivalent rectangular storm" corresponding to a given historical storm will be that rectangular storm which leads to the same values of B_0, B_1, and B_2 as the historical storm The constants for the "equivalent rectangular storm" will have intuitive meaning in characterizing the severity of the storm and in making predictions for the future

PROBABILITIES FOR A RECTANGULAR STORM

A special development will be made for the maximum wave height in a rectangular storm as related to the intensity parameters $\underline{a}$ and the number of waves, N Let $w(h,\underline{a})$ be defined for N independent, identically distributed random wave heights, each with distribution function $F_H(h)$, as

$$w(h,\underline{a}) = N[1 - F_H(h,\underline{a})] \tag{19}$$

Then the distribution function for the maximum value may be written approximately (Cramer, 1946, p 28 6, eq 28 6 2, Borgman, 1961, pp 3296 - 3297, see eq (6)) for large values of N as

$$P[\max H \leq h] = \{F(h,\underline{a})\}^N = \left(1 - \frac{w(h,\underline{a})}{N}\right)^N \approx e^{-w(h,\underline{a})} \tag{20}$$

Hence

$$\log P[\max h \leq h] \approx N[1 - F_H(h,\underline{a})] \tag{21}$$

Gumbel (1954, p 13, eq 2 11) defines the "expected" largest value, V, of a variate to be the value V which satisfies the equation

$$w(V,\underline{a}) = N[1 - F_H(V,\underline{a})] = 1 \tag{22}$$

This has a physical interpretation in that $1 - F_H(V,\underline{a})$ is the probability, P[H>V] Multiplying this probability by N gives the expected number of times wave heights will exceed V in the N occurrences Hence V is that value such that on the average there will be exactly one exceedance in the N wave heights

From eq (22)

$$N = [1 - F_H(V,\underline{a})]^{-1} \tag{23}$$

This can be inserted into eq (21) to give the approximation

$$\log P[\max H \leq h] \approx \frac{1 - F_H(h,\underline{a})}{1 - F_H(V,\underline{a})} \tag{24}$$

Now suppose that, paralleling eq (12), one expands $F_H(h)$ in a power series about h_0

$$F_H(h) = C_0 + C_1(h - h_0) + C_2(h - h_0) + \tag{25}$$

Then keeping only the terms to second order

$$\begin{aligned} \log P[\max H \leq h] &\approx \frac{1 - C_0 - C_1(h - h_0) - C_2(h - h_0)^2}{1 - C_0 - C_1(V - h_0) - C_2(V - h_0)^2} \\ &= B_0^* + B_1^*(h - h_0) + B_2^*(h - h_0)^2 \end{aligned} \tag{26}$$

with

$$\begin{aligned} B_0 &= (1 - C_0)/[1 - C_0 - C_1(V - h_0) - C_2(V - h_0)^2] \\ B_1^* &= -C_1/[1 - C_0 - C_1(V - h_0) - C_2(V - h_0)^2] \\ B_2^* &= -C_2/[1 - C_0 - C_1(V - h_0) - C_2(V - h_0)^2] \end{aligned} \tag{27}$$

The value of $\underline{a}$, N, and V for the equivalent rectangular storm will be determined by equating B_0^*, B_1^*, and B_2^* to the B_0, B_1, and B_2 respectively given by eq (15) for the historical storm Thus, to second order, the equivalent rectangular storm will be producing the same probabilities for maximum wave heights as did the historical storm

The equations to be solved are

$$\begin{aligned} D &= 1 - C_0 - C_1(V - h_0) - C_2(V - h_0)^2 \\ B_0 &= (1 - C_0)/D \\ B_1 &= -C_1/D \\ B_2 &= -C_2/D \end{aligned} \tag{28}$$

Here, h_0 is regarded as a previously selected (and thus known) value to expand about Now the ratios

$$\begin{aligned} R_1 &= (B_1/B_0) = -C_1/(1 - C_0) \\ R_2 &= (B_2/B_0) = -C_2/(1 - C_0) \end{aligned} \tag{29}$$

can be computed from the values of B_0, B_1, and B_2 If R_1 and R_2 are substituted into the expression for B_0, one gets

$$B_0 R_2 (V - h_0)^2 + B_0 R_1 (V - h_0) + (B_0 - 1) = 0 \tag{30}$$

Hence V can be determined from eq (30) as a quadratic solution

Now $F_H(h,\underline{a})$ is typically a monotone decreasing function of storm intensity for fixed h That is, a higher storm intensity normally means that there is a larger probability of exceeding the fixed h value or a smaller probability of being less than or equal to that h value But eq (23) states that

$$F_H(V,\underline{a}) = 1 - \frac{1}{N} \tag{31}$$

Hence, the storm intensity can be determined from the value of N which is usually known, approximately at least, from other considerations If the intensity of $\underline{a}$ is a vector, reasonable interrelations between the components of $\underline{a}$ must be imposed

In summary the computational procedure for determining V and $\underline{a}$ for the rectangular storm is as follows

(1) Calculate R_1 and R_2 from eq (29)

(2) Determine V from eq (30)

(3) Compute $\underline{a}$ from eq (31) and the value of N

PROBABILITY GENERATING FUNCTIONS

In developing the probabilities for the maximum height in a random number of random length and random intensity storms, it will be natural to introduce various probability generating functions A probability generating function for a random variable N is defined to be the infinite series

$$G_N(s) = \sum_{n=0}^{\infty} P[N=n]\ s^n \tag{32}$$

These functions have closed form for many probability laws (Borgman, 1961, p 3305, eq (21) - (27)) Two examples of particular usefulness are the probability generating functions for the Poisson and the negative binomial probability laws (Williamson and Bretherton, 1963, pp 9 - 10)

$$\text{Poisson} \quad P[N=n] = e^{-\lambda}\lambda^n/n! \tag{33}$$

$$G_N(s) = \exp[-\lambda(1-s)] \tag{34}$$

$$\text{Negative binomial} \quad P[N=n] = \binom{r+n-1}{n} p^r q^n \tag{35}$$

$$G_N(s) = p^r(1 - qs)^{-r}, \quad p + q = 1 \tag{36}$$

The mean and variance of the Poisson is λ The corresponding mean and variance of the negative binomial are respectively

$$\text{mean} = rq/p \tag{37}$$

$$\text{variance} = rq/p^2 \tag{38}$$

$$\text{where } p + q = 1 \tag{39}$$

The negative binomial parameters, p, q, and r, can be estimated from the mean $\bar{N}$ and variance $(N) = s^2$ by the method of moments as

$$\hat{p} = \bar{N}/s^2 \tag{40}$$

$$\hat{q} = 1 - p \tag{41}$$

$$\hat{r} = \bar{N}\hat{p}/\hat{q} \tag{42}$$

PROBABILITIES FOR A RANDOM LENGTH STORM

Suppose a rectangular storm has a random length N and fixed intensity, $\underline{a}$ What is the probability law for the maximum wave height in the storm? Let $G_N(s)$ be the probability generating functions for N

By eq (21), the approximate probability law for H given a particular value of N = n is

$$P[\max H \leq h \mid N=n] \approx \{\exp[1 - F_H(h,\underline{a})]\}^n \tag{43}$$

Then for a random number of waves

$$P[\max H \leq h] \approx \sum_{n=0}^{\infty} P[\max H \leq h \mid N=n]\, P[N=n]$$

$$= \sum_{n=0}^{\infty} P[N=n]\, \{\exp[1 - F_H(h,\underline{a})]\}^n \tag{44}$$

$$= G_N(\exp[1 - F_H(h,\underline{a})]) \tag{45}$$

A comparison of eq (44) with eq (32) will justify substituting the exponential for the argument s of the probability generating function

In practice one could use the guessed values of $\bar{N}$ and s^2 together with the negative binomial probability law to determine the function $G_N(s)$ Alternatively another probability generating function could be used

PROBABILITIES FOR RANDOM LENGTH AND RANDOM INTENSITY STORMS

If $\underline{a}$ is also random, then eq (45) must be regarded as a probability given that intensity = $\underline{a}$ Let

$f_{\underline{I}}(\underline{a})$ = probability density for $\underline{a}$

Then

$$P[\max H \leq h] = \int_{-\infty}^{\infty} P[\max H \leq h \mid \underline{I}=\underline{a}]\ f_{\underline{I}}(\underline{a})\ d\underline{a}$$

$$\approx \int_{-\infty}^{\infty} G_N(\exp[1-F_H(h,\underline{a}])\ f_{\underline{I}}(a)\ d\underline{a} \qquad (46)$$

PROBABILITIES FOR A RANDOM NUMBER OF RANDOM LENGTH AND RANDOM INTENSITY STORMS

The final complication is to introduce a probability law for the number of storms, K, which may occur In the time interval for which predictions are made Let $G_K(s)$ be the corresponding probability generating function By the identical same argument leading to eq (45),

$$P[\max H \leq h] = \sum_{k=0}^{\infty} P[\max H \leq h \mid K=k]\ P[K=k]$$

$$= \sum_{k=0}^{\infty} \left\{ \int_{-\infty}^{\infty} G_N(\exp[1-F_H(h,\underline{a})])\ f_{\underline{I}}(\underline{a})\ d\underline{a} \right\}^k P[K=k] \qquad (47)$$

or

$$P[\max H \leq h] = G_K\left(\int_{-\infty}^{\infty} G_N(\exp[1-F_H(h,\underline{a})])\ f_{\underline{I}}(\underline{a})\ d\underline{a} \right) \qquad (48)$$

The number of waves in a given storm may depend on $\underline{a}$ Hence the formula can be made a little more general by introducing the conditional probability generating function for N given $\underline{a}$ This final version of the formula would be

$$P[\max H \_ h] = G_K\left(\int_{-\infty}^{\infty} G_{N|\underline{a}}\ (\exp[1-F_H(h,\underline{a})])\ f_{\underline{I}}(\underline{a})\ d\underline{a} \right) \qquad (49)$$

SOME FINAL COMMENTS

(1) The application of the above formula will obviously require a digital computer and detailed analysis of the historical data for the particular location of interest

(2) The negative binomial appears to be the best choice for the two probability generating functions although, at least for Gulf of Mexico hurricanes, there is some basis for using the simpler Poisson probability generating function for $G_K(s)$

(3) The possible choices for $F_H(h,\underline{a})$ were discussed at the beginning of the paper Without more detailed information, the Rayleigh distribution appears to be as good a choice as any (Goodnight and Russell, 1963)

(4) The choice of $f_{\underline{I}}(\underline{a})$ would have to depend strongly on the analysis of historical data or on meteorological considerations Hence it is hard to make a guess as to a reasonable choice However, a form of the gamma density would seem to be a good first guess

(5) In this whole discussion, the randomness of wave period has been ignored A more adequate model would certainly include this source of variation

(6) An alternative approach to the maximum wave height might be made through the statistical theory of maxima and minima of a random function Unfortunately, when such an approach is attempted, theoretical difficultires arise very quickly Information on wave crest elevation probabilities can be obtained, however, by the random function type of analysis

ACKNOWLEDGMENT

The research reported was supported in part by the Chevron Oil Field Research Company under a research gift to the University of California, Berkeley, and in part, by the Coastal Engineering Research Center, U S Army Corps of Engineers under Contract DACW-72-69-C-0001 The author gratefully acknowledges their financial assistance in the study

REFERENCES

Borgman, L E (1961) The frequency distribution of near extremes, Journal of Geophysical Research, 66, pp 3295-3307

Cramer, Herald (1946) Mathematical Methods of Statistics, Princeton University Press, Princeton, New Jersey

Goodnight, R C and Russell, T L (1963) Investigation of the statistics of wave heights, Jour Waterways and Harbors Div , ASCE, WW2, paper 3254, pp 29-54

Gumbel, E J (1954) Statistical theory of extreme values and some practical applications, Applied Math Series 33, National Bureau of Standards, U S Govt Printing Office, Washington, D C

Gumbel, E J (1958) Statistics of Extremes, Columbia Univ Press,N Y

Longuet-Higgins, M S (1952) On the statistical distribution of the heights of sea waves, Jour Marine Res , 11, pp 245-266

Longuet-Higgins, M S , and Cartwright, D E (1956) The statistical distribution of the maxima of a random function, Proc Royal Soc , A, 237, pp 212-232

Watson, G S (1954) Extreme values in samples from m-dependent stationary stochastic processes, Ann Math Statistics, 25, pp 798-800