SECOND INTERNATIONAL CONFERENCE ON
PORT AND OCEAN ENGINEERING UNDER ARCTIC CONDITIONS
UNIVERSITY OF ICELAND
DEPARTMENT OF ENGINEERING AND SCIENCE

STATISTICAL METHODS FOR DETERMINING EXTREME SEA STATES

K. G. Nolte
Staff Research
Engineer

Amoco Production Company

Tulsa
Oklahoma
U. S. A.

INTRODUCTION

A wealth of ocean wave data and associated statistics have been obtained in conjunction with maritime operations, see for example [1].* However, since it is good maritime practice to avoid extremely severe sea states, data and associated methods of analysis are relatively deficient for extreme sea states. Due to the expansion of petroleum development into hostile sea areas, emphasis has recently been placed on predicting the statistics of very extreme sea states, for example [2, 3, 4, 5]. Accurate extreme wave statistics are required to design offshore installations which will have the required reliability and yet avoid a severe economic penalty for overdesign.

Two recently published papers [2, 3] present statistics for extreme wave heights for the same area of the North Sea, namely 57°30'N, 3°00'E. Considerable interest has been generated in these papers because this area is in the vicinity of offshore petroleum development. In [2] the projected 100-year wave height was approximately 24 meters, while in [3] the projected 100-year wave height was 28.3 meters. This difference of over four meters is very significant when translated into design wave forces or cost of constructing an appropriate installation.

In order to understand why this difference resulted, one must consider the basic procedures for determining wave statistics. The determination of wave statistics involves three important considerations; (1) desired end use; (2) method of analysis or mathematical model; and (3) existing data base. For any statistical study, all three of these considerations should be defined and the compatibility of the three established. For extreme statistics, the desired end

* References are given at the end of the paper

use is a projection of return periods for extreme or rare events. No single mathematical model for determining extreme wave statistics has gained universal acceptance. Therefore, many different models have been employed [2, 3, 4, 5, 6]. The sufficiency of the data base has not received proper consideration in most published work on extreme waves. An exception to this statement is [4]. In light of the above discussion, the different projections in [2] and [3] are not surprising since different mathematical models and data bases were used.

The primary purpose of this paper is to investigate the consequences of using different mathematical models to project extreme wave statistics. The different models will be compared by applying them to a set of visual wave data covering approximately a 10-year period. In addition to comparing the different models by the use of data, the models will be compared on a mathematical basis.

The visual wave data used in this paper were not corrected by one of the calibration formulas suggested in [1] or [2]. Therefore, the numerical results obtained for the illustrative purposes in this paper are not to be regarded as representing actual extreme wave conditions.

MATHEMATICAL MODELS

In the following section five different mathematical models to project extreme wave statistics will be discussed. The primary purpose of considering these models is to establish return periods for waves of different heights. In order to establish correct return periods, independent data samples are required. For the purpose of extreme wave statistics, it will be assumed that individual storms, defined in some appropriate manner, are independent events. Although future data may demonstrate the inadequacy of this assumption, an assumption of this nature is required for a meaningful discussion of extreme wave statistics. In this paper, probability distributions will be symbolically expressed in summation form instead of the equivalent integral form. The summation form is used because this is the form required for the application to data.

Individual Wave Model

The first model to be considered was used in [2] and will be referred to as the individual wave model. In this model, wave records of significant wave height values are used to reconstruct the statistical distribution of individual wave heights occurring during the period of wave record. The distribution of individual heights is then used to determine the maximum wave height expected to occur over some future period of time. For example, if M waves occur during m years, then the expected maximum wave height during m years was found in [2] as the wave height having an occurrence probability of 1/M. Mathematically the distribution of individual wave heights [2] can be expressed as

$$I(H) = \int_0^\infty R(H,Hs)\,\frac{dP(Hs)}{dHs}\,dHs \qquad (1)$$

in which R(H,Hs) is the Rayleigh distribution for individual heights, H, resulting for a particular value of significant height, Hs (see Appendix A). P(Hs) is the historical distribution of the significant height.

An equivalent form of (1) for N discrete observations of Hs (i.e., Hs_i) taken at a fixed time interval is

$$I(H) = \sum_{i=1}^{N} \frac{R(H,Hs_i)}{N} \qquad (2)$$

It was suggested in [7] that Equation (1) would overestimate the occurrence of larger waves because of the positive correlation between the significant wave height and average wave period of sea states. In [7] the following form of (1) was presented which weighted the occurrence of wave heights by the inverse of the associated average period of the sea state, t,

$$I(H) = \bar{t} \iint \frac{R(H,Hs)}{t}\,p(Hs,t)\,dHs\,dt \qquad (3)$$

in which $\bar{t}$ is the long term average of the average period of individual sea states and p(Hs,t) is the joint probability density. An equivalent form of (3) for N discrete observations of Hs taken at a fixed time interval is

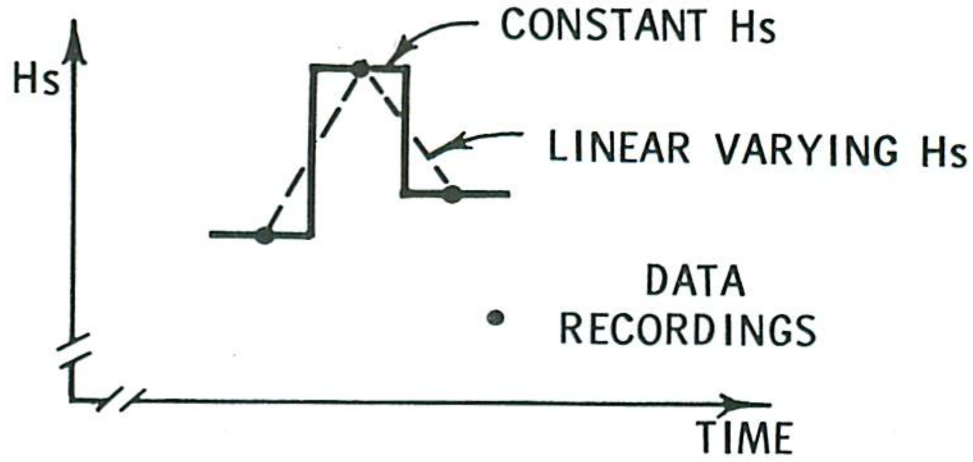
$$I(H) = \sum_{i=1}^{N} \frac{n_i\,R(H,Hs_i)}{\sum_{i=1}^{N} n_i} \qquad (4)$$

in which n_i is the number of waves which occurred during the i^{th} interval of time while the sea state was Hs_i. The value of n_i can be determined from the average wave period during the i^{th} observation.

An implicit assumption in (4) is that the sea state remains constant at the value Hs_i during the i^{th} time interval. A more realistic assumption is that the sea state varies linearly between observation times as illustrated in Figure 1. The distribution of H, corresponding to R in (4), for a varying sea state is given in (A-6) of Appendix A and will be denoted as $V(H,Hs,\alpha)$. V gives the distribution for individual wave heights for a sea state which varies linearly between an upper value of Hs and a lower value of αHs ($\alpha \leq 1$) for two observations. For such a sea state (4) becomes

$$I(H) = \sum_{i=1}^{N} n_i\,\frac{V(H,Hs_i,\alpha_i)}{\sum_{i=1}^{N} n_i} \qquad (5)$$

FIGURE 1. VARIATION OF SIGNIFICANT HEIGHT, Hs



In [2], the wave height with a return period of m years (during which M waves will occur) was taken as the value of H for which $I(H) = 1/M$. This interpretation is incorrect as briefly discussed in [7]. It is correct that, over a long term average, this value of H will occur once every m years. However, the average interval between occurrences (the return period) of this value will be greater than m years. The return period is greater than m because the occurrence of large waves tend to group together in time, that is, occur in storms. For an example of the required correction, consider that on the long term average, the height, H, is exceeded once every 10 years. Moreover, consider that during storms in which it is exceeded, the value H is exceeded twice, again on the long term average, during these storms. For this case, the return period for storms with heights exceeding H is 2 x 10 = 20 years. Thus, for determining the return period of H, individual exceedences of H are not independent. The required correction will be termed the <u>grouping correction</u>. An expression for this correction is derived in Appendix C.

<u>Most Probable Maximum Model</u>

The next model to be considered was used in [3] and will be termed the most probable maximum model. Whereas the previous model utilized the distribution for all heights between observation periods, this model utilizes only the peak value, or most probable value, Hm, of the probability density function for the maximum height between observation periods. The different statistics used by the two models are illustrated in Figure 2-a. For each observation of significant height, Hs, the corresponding value of Hm can be determined on the basis of the distribution for all heights, in terms of Hs, and the number of waves occurring between observations (Appendix A). The distribution, $Q(H)$, for N observations of Hm can be represented as

$$Q(H) = \sum_{i=1}^{N} \frac{h(Hm_i - H)}{N} \tag{6}$$

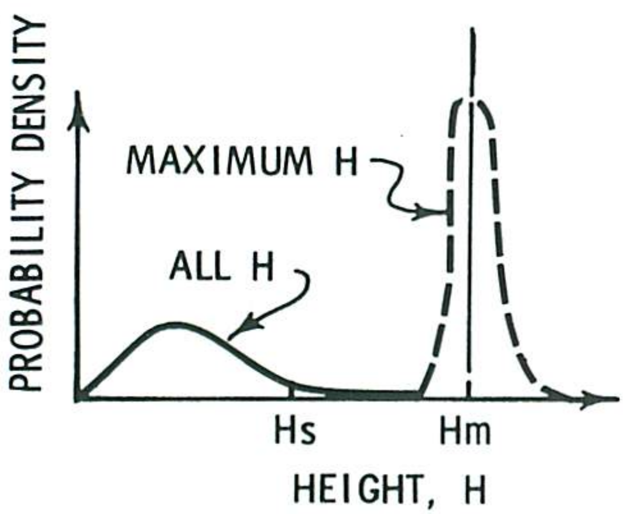
where Hm_i are the individual observations of Hm and h( ) is the Heaviside function which equals unity for a positive argument and is equal to zero for a negative argument.

In [3] the values of Hm were determined assuming Hs to be constant between observation times. As for the previous model, it is more realistic to assume that Hs varies linearly between observation times as illustrated in Figure 1. Equation (A-9) of Appendix A gives the expression for determining Hm for the case of a linearly varying Hs. If this value of Hm is denoted by $\overset{*}{Hm}$ then (6) becomes

$$Q(H) = \sum_{i=1}^{N} \frac{h(\overset{*}{Hm}_i - H)}{N} \tag{7}$$

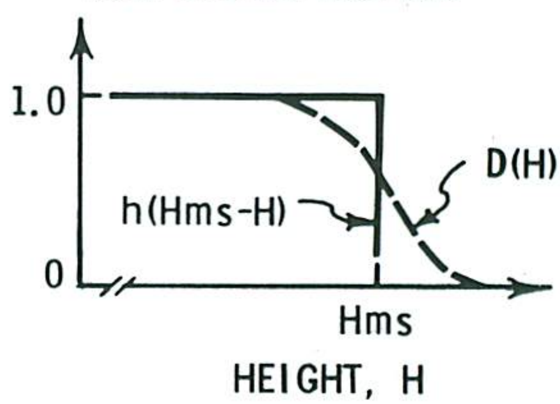
Moreover, in [3], the return period of m years was assigned to the value of H for which Q(H) = 1/M, where M is the number of observation periods in m years. Again, as in the previous model, this is an incorrect return period for H because the larger values of Hm tend to group together during storms. A grouping correction is obtained from the average number of observation periods for which Hm exceeds H during storms that contain at least one value of Hm which exceeds H.

FIGURE 2-a. PROBABILITY DENSITIES FOR WAVE HEIGHTS



Hs - SIGNIFICANT HEIGHT
Hm - MOST PROBABLE MAXIMUM HEIGHT

FIGURE 2-b. FUNCTIONS FOR STORM MODELS



Annual Maximum Model

The model which uses the statistic of the annual maximum wave height has received very limited use for extreme wave statistics even though this model is commonly used for other meteorological phenomena. In this paper the annual maximum wave height will be defined as the largest value of the most probable maximum heights for all storms occurring during a given year. This value will be denoted as Hma. The distribution of these values will be denoted as A(H).

$$A(H) = \text{prob}\ (H > Hma) \tag{8}$$

The value of height, H, with a return period of m years is the value of H yielding $A(H) = 1/m$. Because the individual values of Hma are independent, this return period does not require a grouping correction as was necessary for the previous two models. The independence of the Hma values follows from the assumption that storms are independent.

Storm Models

In this section two models for extreme wave statistics are discussed which use the maximum wave height during storms as a statistic. One of the models uses the most probable maximum wave height during a storm as a statistic [4, 5, 6] and the other model uses the complete distribution of possible maximum heights during a storm as a statistic.

The first storm model utilizes the distribution of the most probable maximum wave heights during individual storms, Hms. This is in contrast to the most probable maximum model which employed the most probable maximum height during a fixed period of time. The distribution for this storm model is

$$S(H) = \sum_{i=1}^{N} \frac{h(Hms_i - H)}{N} \tag{9}$$

where N is the total number of storms and Hms_i is the most probable maximum height for the i^{th} storm. h( ) is the Heaviside function defined previously and illustrated in Figure 2-b.

For this model only values of Hms which exceed some level of height are included in S(H). This threshold level is taken as the definition of "storm" for the model. The return period used with (9) is governed by the annual rate of occurrence of storms exceeding the threshold level. If this rate is denoted by λ, the expected number of storms during m years is $m\lambda$. Therefore, the height, H, with a return period of m years is the value of H for which $S(H) = 1/m\lambda$. Because of the assumption that storms are independent events, no grouping correction is required for the return periods associated with this model.

Equation (9) is commonly coupled with the Poisson distribution [4,5] to obtain the following distribution

$$Sn(H) = 1-\exp[-\lambda nS(H)] \tag{10}$$

in which Sn(H) is the probability of obtaining a storm which exceeds H during n years. The Poisson distribution is used to describe the probability of the number of storms which will occur during the n years. In (10), n can be taken as unity to obtain the annual occurrence probability

$$Sa(H) = 1-\exp[-\lambda S(H)]. \tag{11}$$

The derivation of (10) or (11), resulting from the coupling of the Poisson distribution with (9), is given in (A-15) of Appendix A.

The wave height, H, with a return period of m years is found from (11) by selecting the value of H which satisfies Sa(H) = 1/m. For values of m which are greater than 10, Equations (9) and (11) give the same return periods since

$$Sa(H) = 1-\exp[-\lambda S(H)] \simeq 1-[1-\lambda S(H)] = \lambda S(H). \tag{12}$$

The second storm model is very similar to the previous model, but employs the distribution for all possible values of the maximum wave height in a storm. For example, if $D_i(H)$ represents the occurrence probability for the maximum height exceeding H during the i^{th} storm, the distribution for this model is

$$S^*(H) = \sum_{i=1}^{N} \frac{D_i(H)}{N} \tag{13}$$

in which N is the total number of storms. The similarity between the two storm models can be seen by comparing (9) and (13) with the aid of Figure 2-b.

Equivalent expressions for (10), (11), and (12) can be written employing S*(H) instead of S(H). The equivalent of (11) can be expressed as

$$\overset{*}{Sa}(H) = 1-\exp[-\lambda S^*(H)] \tag{14}$$

The return periods for (13) or (14) are found in the same manner as for (9) or (11) respectively.

It is interesting to note that the deletion of "insignificant" storms from the populations used to determine (9) or (13) does not affect the value of $\overset{*}{Sa}(H)$ of (14) or Sa(H) of (11). An "insignificant" storm is defined here as a storm which has effectively zero probability of producing a wave height greater than H. Therefore, the definition of "insignificant storm" depends on H. The fact that "insignificant storms" do not affect $\overset{*}{Sa}(H)$ can be demonstrated as follows. The annual storm frequency, λ, is equal to the number of storms, N, which occurred during the data record divided by the length of the data record in years, L, ($\lambda = N/L$). Therefore, after substitution of (13), the term $\lambda S^*(H)$ of (14) becomes

$$\lambda S^*(H) = \frac{N}{L} S^*(H) = \frac{N}{L} \sum \frac{D_i(H)}{N} = \sum_{i=1}^{N} \frac{D_i(H)}{L} \qquad (15)$$

Since (15) is independent of N and by definition $D_i(H) = 0$ for "insignificant storms", $\lambda S^*(H)$ and hence $\overset{*}{Sa}(H)$ is not dependent on "insignificant storms". However, in reality due to data records which are generally smaller than the desired return period, "insignificant storms" must be included in the distribution for $S^*(H)$ to permit enough data points for reasonable extrapolation. In fact it is most likely that the complete data base is insignificant" (by the above definition) for the 50 or 100-year wave. However, the above exercise was useful to show that the choice of the threshold level for the definition of storms does not affect the final statistics if the storms included in the sample permit a reasonable extrapolation. Even though the above demonstration concerning "insignificant" storms was for the model expressed by (14) a similar demonstration can be made for the model expressed by (11).

The two storm models discussed above differ only in the two distributions, S(H) of (9) and $S^*(H)$ of (13). However these two distributions are constructed from essentially the same function as illustrated in Figure 2-b. Therefore the two distributions, (9) and (13), and hence the two models are essentially equivalent.

Appendix B contains a demonstration that the mathematical expression for the return period of a specific height is the same for the individual wave model and the second storm model. In addition, Appendix B contains a similar demonstration that the wave height with a given return period for the most probable maximum model will be slightly less (by an amount depending on the length of the data observation interval) than the height with the same return period for the first storm model. Therefore, since the two storm models are essentially equivalent, it must follow that all four models, discussed in this paragraph, will give approximately the same wave heights for a given return period. However, it is noted that the difference between the most probable maximum model and the other models will increase as the data observation interval is decreased (Appendix B).

Since the use of different models should produce approximately the same return periods, the significant difference found in [2] and [3] should not have resulted from the use of different models.

APPLICATION OF DATA

The wave data employed in this study consisted of 10 years of visual data collected every three hours by the Norwegian Meteorological Institute from the rescue ship, Famita, stationed in the North Sea near 57°30'N, 3°00E. The data were collected during the months of October through May. This set of data was the primary data base for [2]. Appendix D contains a more detailed description of the way in which the data were employed for this study.

In the following section of the paper, the five models for extreme wave statistics are compared on the basis of this set of visual wave data.

Individual Wave Model

Figure 3 shows the distributions which result from application of the different forms of the individual wave model to the visual wave data.

The distribution, labeled A in Figure 3, is found from Equation (2) and is essentially the same form used in [2]. This distribution does not consider the correlation of significant height and average wave period for sea states. The dashed line extension in this figure denotes the extrapolation required to reach the return periods indicated. The grouping correction has not been applied to these return periods. Therefore, the uncorrected periods are denoted as 1/m years implying an average occurrence of once per m years, in contrast to an average period of m years between occurrences. Following this notation, the indicated 1/100 year wave for Curve A of Figure 3 is 27 meters. This is greater than the value of 24 meters found in [2]. However, a calibration formula was used in [2] to convert the visual wave height to the significant height. One can show that the calibration formula used in [2] would have reduced the significant height associated with a 27-meter wave by about 12 percent, and hence would have also reduced the maximum wave height, 27 meters, by about 12 percent. Therefore, if the calibration formula were applied, the 27-meter value would have been reduced by 12 percent and would have resulted in the value of 24 meters found in [2].

The curve labeled B in Figure 3 shows the distribution which results for the form of the individual wave model given by Equation (4). This form of the model includes the correlation between the significant height, Hs, and average wave period for the sea states. As seen from the figure, this distribution results in 26 meters for the 1/100 year wave. This implies that including the correlation of Hs and average wave period reduces the 1/100 year wave by 1 meter.

The curve labeled C in Figure 3 shows the distribution which results for the form of the individual wave model given by (5). This form does not assume that the value of Hs for the sea state remains constant between observation times, but varies linearly between these times as illustrated in Figure 1. In addition, this form includes the correlation of Hs and average wave period. As seen in the figure, the 1/100 year wave for this distribution is 25.2 meters or 1.8 meters less than the distribution for Equation (2) and 0.8 meters less than the

FIGURE 3. DISTRIBUTIONS FOR INDIVIDUAL WAVE MODEL

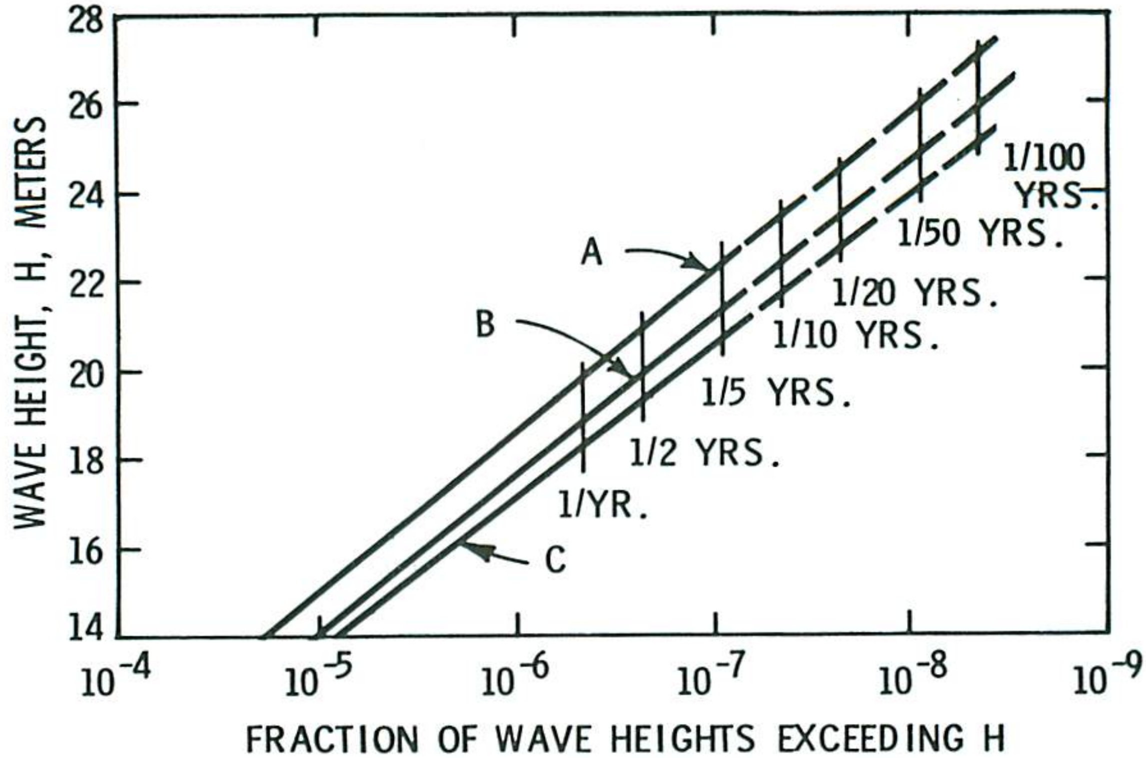


FIGURE 4. GROUPING CORRECTION

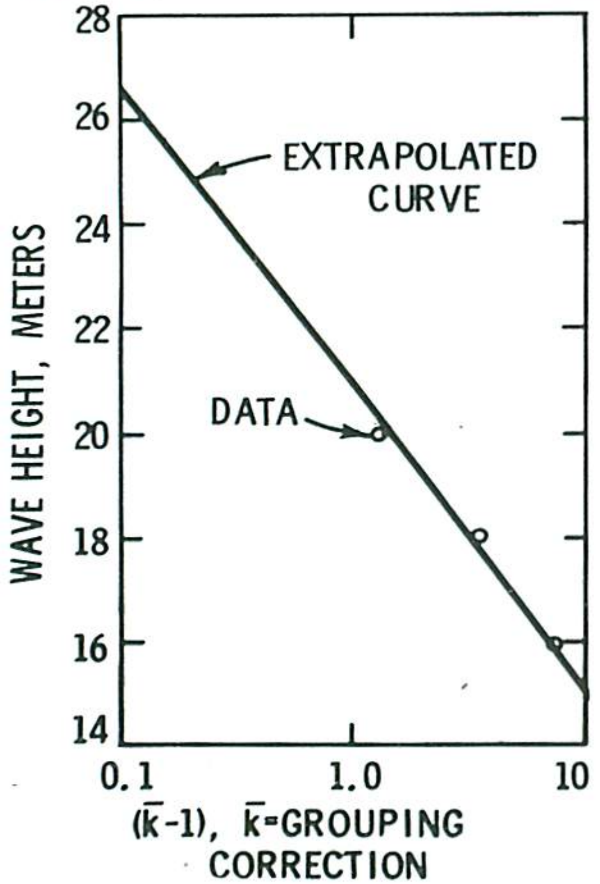
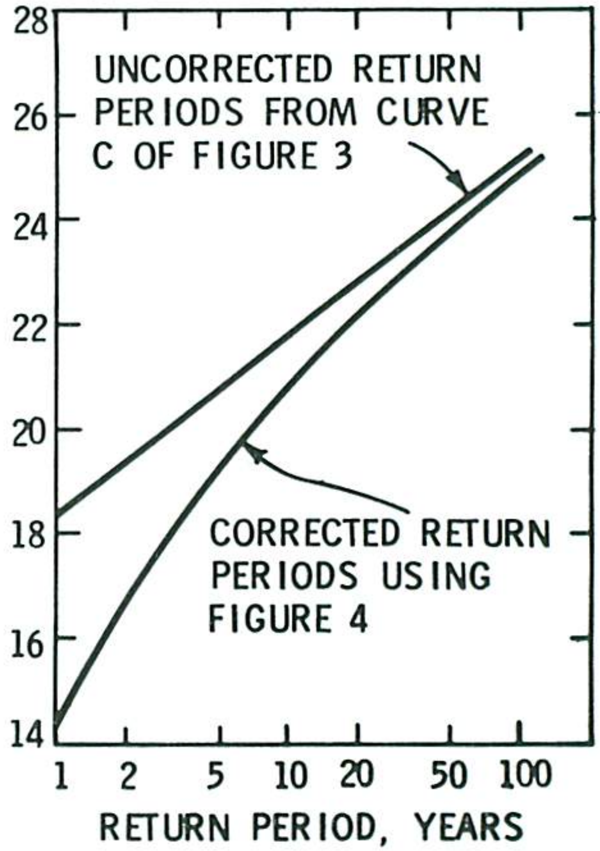


FIGURE 5. RETURN PERIODS



distribution for Equation (4). Thus, the 1/100 year wave is reduced by almost 1 meter as a result of permitting the value of Hs to vary between observation periods.

The uncorrected return periods shown on Figure 3 are designated as 1/m years implying that the indicated wave height is expected to occur once every m years on the long term average. However because of the grouping correction discussed previously, the expected or average period of time between the occurrences of storms which contain waves that exceed this indicated height is greater than m years.

The grouping correction for this model is denoted by $\overline{k}$. This correction is a function of height, H, and is the average number of times waves greater than H occur during storms which have at least one exceedence of H. An expression for $\overline{k}$ is derived in Appendix C. This expression was applied to the visual wave data and the resulting values of $\overline{k}$ are shown as data points on Figure 4. Because the data set was finite, it was necessary to extrapolate these values of $\overline{k}$ for heights above 20 meters. As is obvious from the definition of $\overline{k}$ and as indicated by the extrapolated curve on Figure 4, $\overline{k}$ has an asymptotic value of unity for large values of H. The grouping correction, given by the extrapolated curve in Figure 4, was used to obtain the corrected return periods shown in Figure 5. The corrected values were obtained by multiplying the uncorrected values, from Curve C of Figure 3, by the appropriate value of $\overline{k}$.

The wave height with a corrected return period of 100 years is found in Figure 5 to be 24.8 meters compared to the 1/100 year wave of 27 meters for the Curve A Figure 3. Thus the "100-year wave" is reduced 2.2 meters by the above discussed modifications to the model used in [2]. Because the correction for return period is more important for shorter return periods, the annual wave is reduced from 19.8 to 14.2, or 5.6 meters by applying these modifications to the model used in [2].

Most Probable Maximum Model

For all following models, the only form presented is the one in which the significant height is assumed to vary linearly between observations as illustrated in Figure 1. The following models are limited to this form because Figure 3 (for the previous model) adequately shows the difference which results between application of the two different assumptions for the variation of significant height between observations. Equation (7) represents this form for the present model. The statistic used for this model is the most probable maximum wave height, Hm, for the time period between observations. For the visual wave data used, this time period was 3 hours.

FIGURE 6. DISTRIBUTION FOR MOST PROBABLE MAXIMUM MODEL

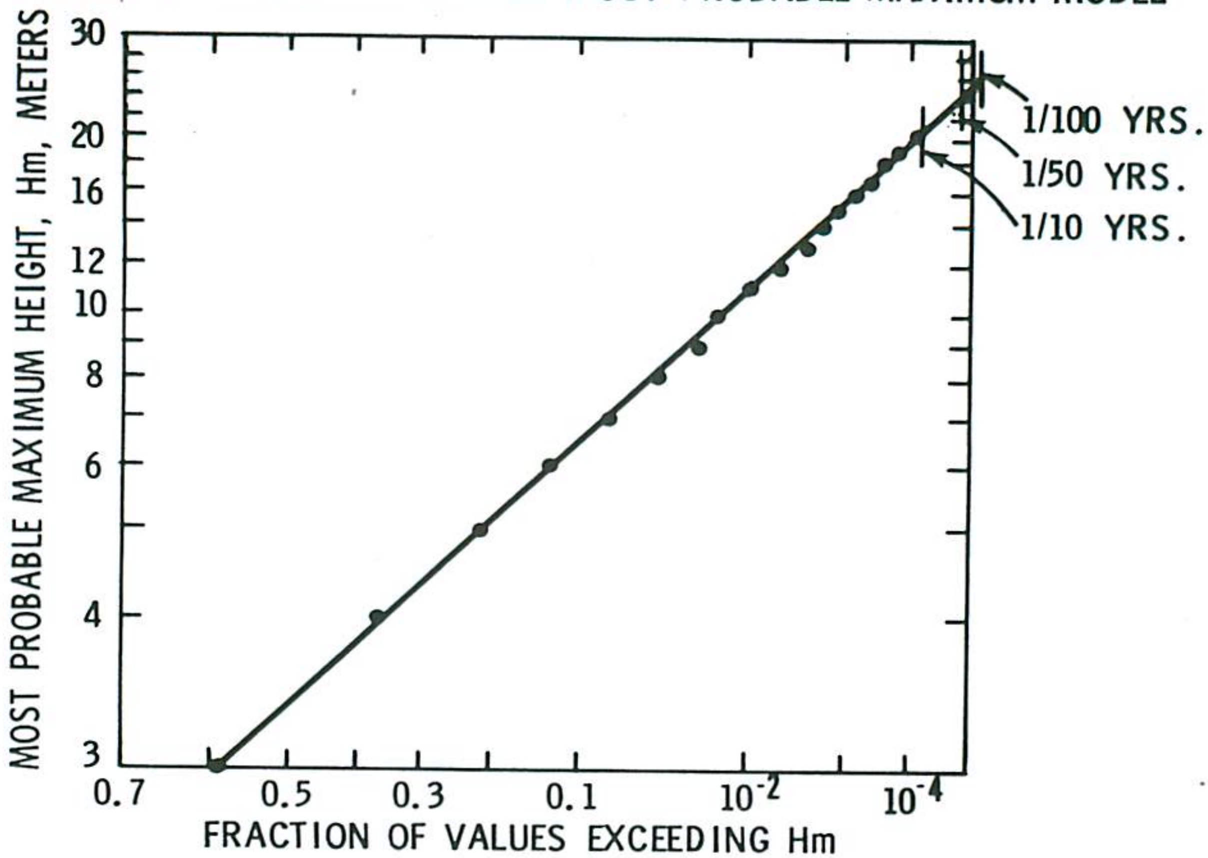


FIGURE 7. GROUPING CORRECTION

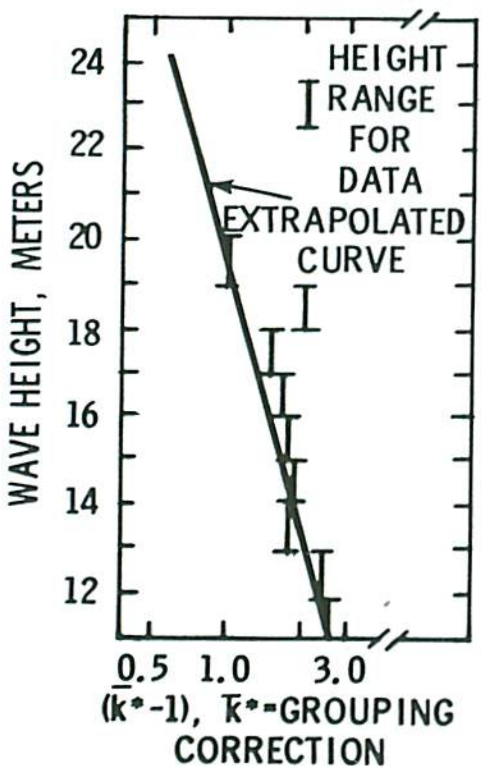


FIGURE 8. RETURN PERIODS

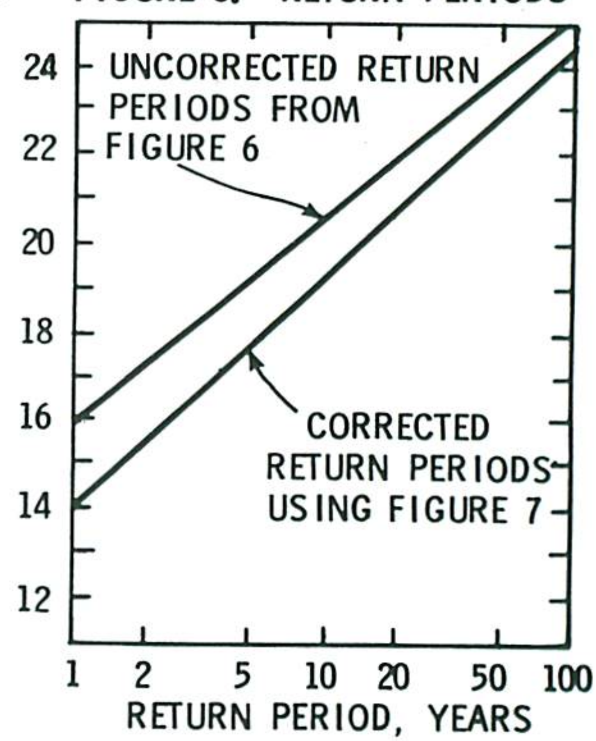


Figure 6 shows the consequence of applying (7) to the visual wave data. The data points on this figure represent the fractional portion of Hm values which exceed different height levels. Figure 6 also shows the uncorrected return periods for 100, 50, and 10 years. Figure 6 is plotted on a Weibull scale (see Appendix D) with a resulting slope of unity. Therefore, the distribution can be represented by an exponential distribution. It is noted that the portion of the distributions shown on Figure 3 for the previous model could also be represented by an exponential distribution. This follows because the distributions are represented by straight lines on a semi-log plot.

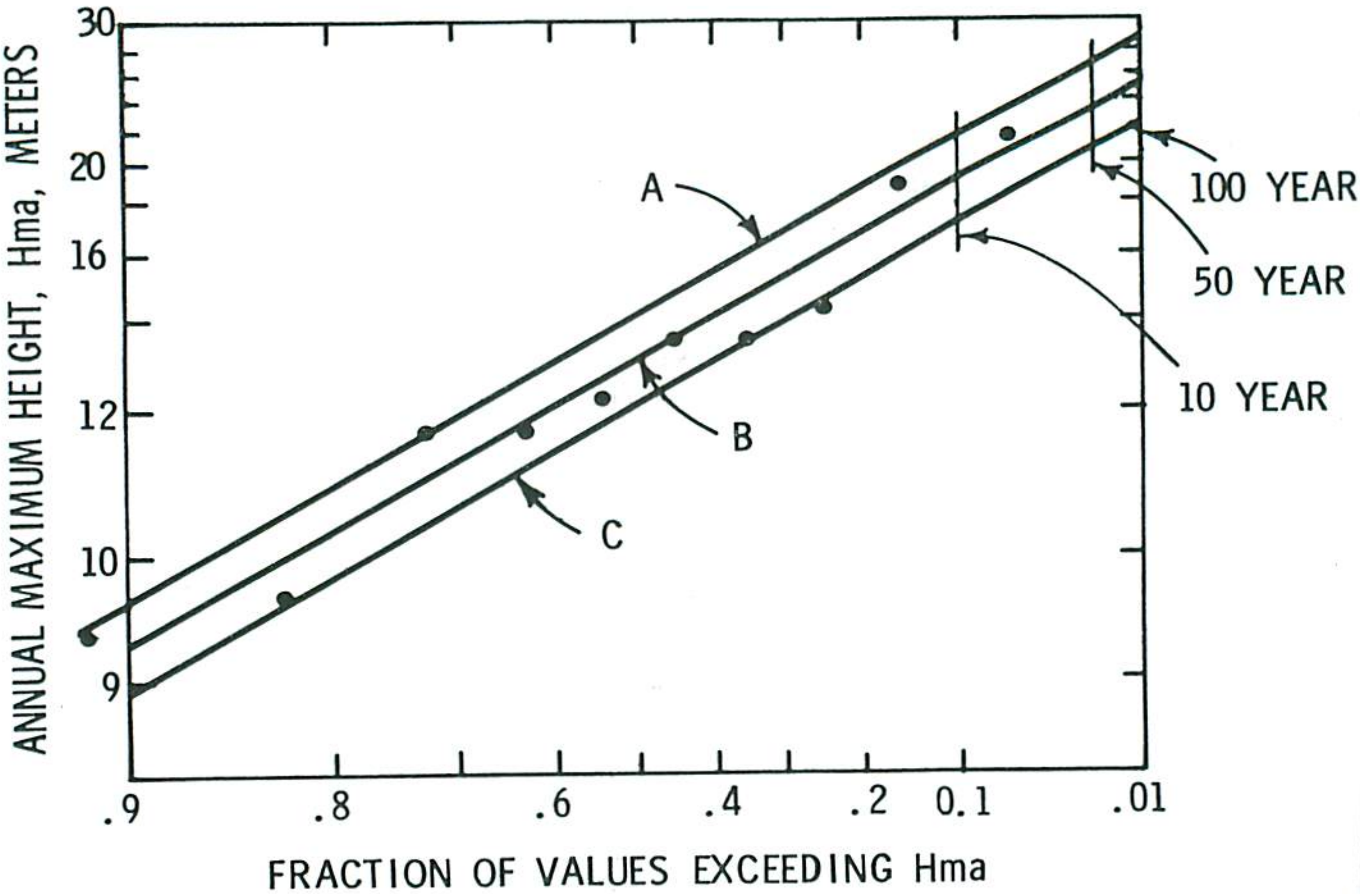
Figure 8 presents a plot of the uncorrected and corrected return periods versus height for this model. The corrected return periods were obtained by multiplying the uncorrected values by the appropriate grouping correction, $\overline{k}*$, given in Figure 7. The data points for $\overline{k}*$ in this figure were found from the visual wave data by the application of (B-14) of Appendix B. $\overline{k}*$ is a function of height, H, and is the average number of times Hm exceeds H during storms which have at least one value of Hm that exceeds H. Figure 8 shows that the wave height with a return period of 100 years is reduced from 25.2 to 24.4 meters by the grouping correction. This correction is somewhat larger than the similar correction in the previous model. The reason for the difference is that the correction for this model does not asymptotically approach unity as quickly as the correction for the previous model. The relatively slower approach to unity reflects the assumption of a linearly varying significant height, Hs, between observations. Using this variation of Hs results in two nearly equal values of Hm for the period before and after the peak observation of Hs. Thus, the largest value of Hm for any storm tends to be paired with a second, approximately equal, value of Hm.

Annual Maximum Model

The annual maximum model uses the maximum wave height during a given year as a statistic. The maximum value used is the most probable maximum value for the storm which produces the largest waves during the given year. The significant height is assumed to vary linearly between observations as illustrated in Figure 1. The procedure for calculating this value is given in Appendix A. Because the visual data set contains 10 years of data, there are 10 annual maximum values for the data set.

Figure 9 shows the distribution of these 10 values plotted on a Weibull scale. The appropriate plotting formula (Appendix D) was used for this plot. The curve denoted as B in Figure 9 will be used to represent the distribution for this model. The outer two curves, denoted as A and C, on this figure will be discussed in a later section.

FIGURE 9. DISTRIBUTION FOR ANNUAL MAXIMUM MODEL



The wave heights with return periods of 10, 50 and 100 years are designated on the figure and are 19, 23.2 and 25 meters respectively. Since these yearly maximum values are assumed to be independent, no grouping correction is required for the return periods of this model.

Storm Models

Two models based on individual storms were discussed previously. The first storm model used the most probable maximum wave height for the individual storms as a statistic. The second storm model considered all possible values for the maximum height during the individual storms instead of only the most probable value. For these models, the significant height is assumed to have varied linearly between data observations as illustrated in Figure 1.

The two distributions, representing Equations (9) and (11), for the first storm model are shown respectively as Curves A and B in Figure 10. The distributions are presented on a Weibull plot and the appropriate plotting formula (Appendix D) was used for the data points. Curve A in this figure (Equation 9) is constructed from the 10 data points for the most probable maximum height for each of the 10 most severe storms in the data set. The threshold level of height for the definition of a storm was chosen so that 10 storms exceeded this value. As demonstrated previously, this threshold level does not affect the results for the model if a sufficient number of storms is retained to permit the construction of an accurate distribution. Curve B in Figure 10 results from coupling the distribution for individual storms (Curve A) with a Poisson distribution for storm occurrence (see Appendix A). Curve B represents the annual probability of exceeding different wave heights. Because 10 years of data were used and 10 storms were retained resulting in a storm frequency of one per year, the two curves have the same values for small probabilities. It is seen from the figure that Curve B approaches the probability value 0.63 asymptotically. This asymptotic value implies that during any year, there is a 63 percent probability of a storm occurring which exceeds the threshold height selected to define a storm.

The return periods of 10, 50 and 100 years are indicated on Figure 10 and correspond to heights of 20.2, 23.5, and 24.7 meters respectively. Because it has been assumed that individual storms are independent events, these return periods require no grouping correction. In order to insure that the individual storms considered were independent, the selected definition of a storm included a threshold level to be exceeded, and also a lower level which the sea state (significant height) must fall below in order to terminate the duration of the storm. The lower value of significant height was taken as 2.6 meters which was the 10-year average value of all observations.

The second storm model was defined in terms of Equations (13) and (14). The same set of 10 storms used for the first storm model were used to construct Curve A of Figure 11 for the distribution representing Equation (13). This distribution gives the probability of exceeding a given wave height during a storm which meets the definition of a storm. Curve B on the figure results from the coupling of the Curve A with the Poisson distribution for storm occurrences. The dashed portion of these curves represents the extrapolated portion of the curves. The curves were extrapolated for heights above the most probable maximum for the most "severe" storm.

The return periods of 10, 50 and 100 years are indicated on Figure 11 for this model and correspond to heights of 20.7, 24.0 and 25.2 meters respectively. These values are 0.5 meter greater than the corresponding values for the first storm model. This difference in the similar models should be expected because the first model only considers the most probable value of the maximum height while the second model considers the complete distribution of all possible values. From examining the data, it was found that the most probable value was always

FIGURE 10. DISTRIBUTIONS FOR FIRST STORM MODEL

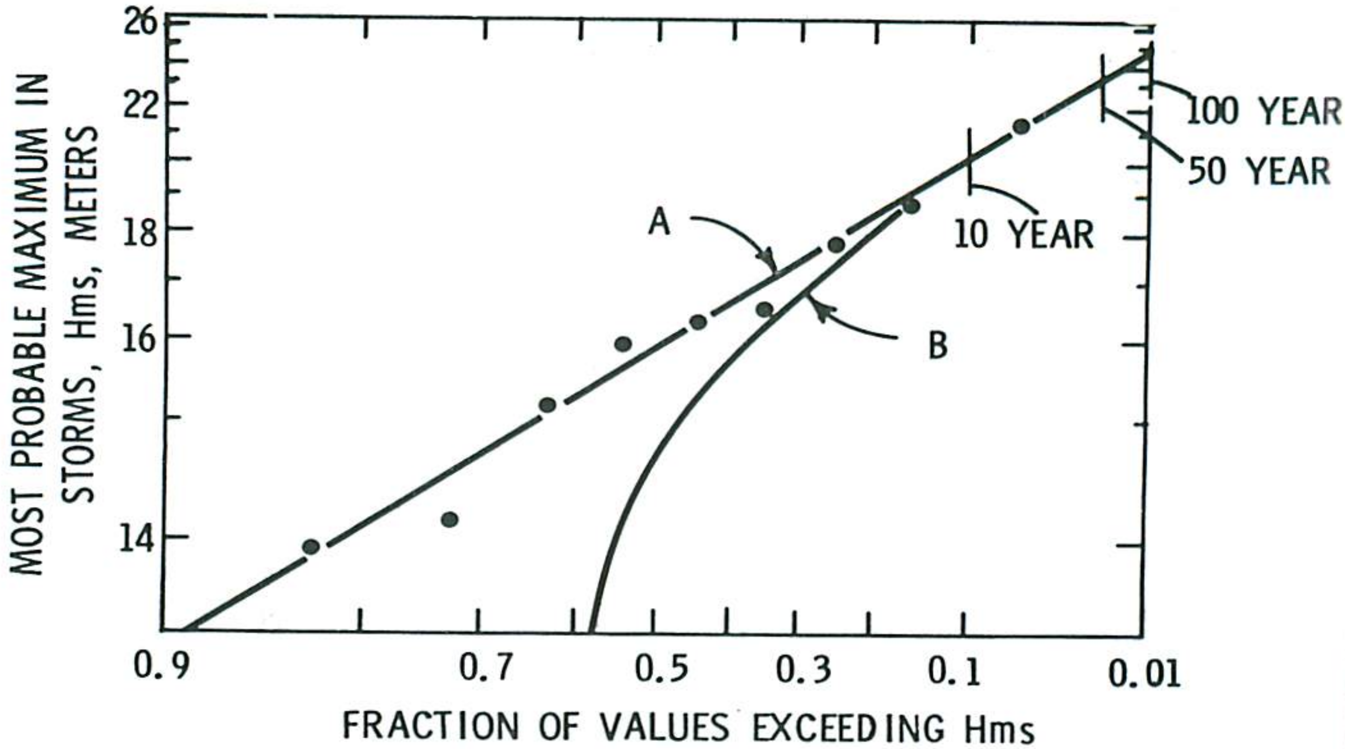
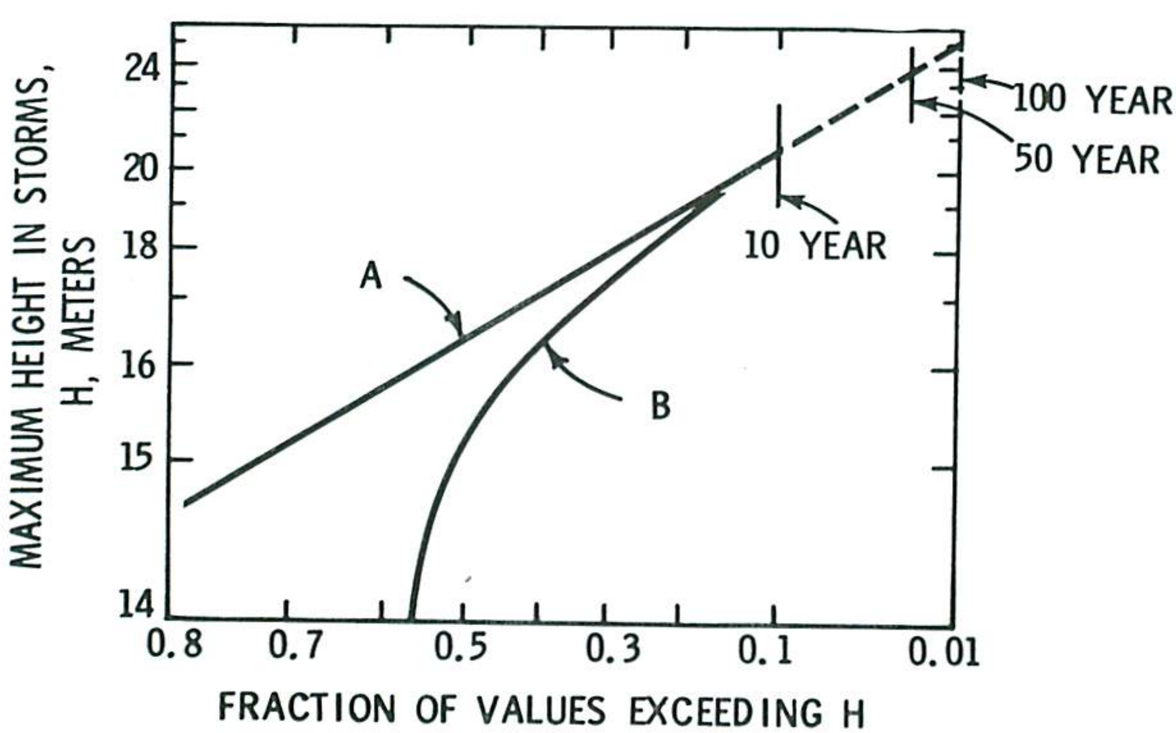


FIGURE 11. DISTRIBUTIONS FOR SECOND STORM MODEL



720

approximately a 40 percentile value. Therefore, one would expect that 60 percent of the time the most probable maximum was less than the actual maximum value. The data indicated that if the 50 percentile value were used in the first storm model instead of the most probable value, the two storm models would have given the same heights for the same return periods.

Comparison of Different Models

The predicted heights for 100, 50 and 10 year return periods resulting from application of the five different models are compared in the table below:

Model	Return Period		
	10 Year	50 Year	100 Year
Individual Wave	20.6	23.7	24.8
Most Probable Maximum	19.4	23.3	24.4
Annual Maximum	19.0	23.2	25.0
First Storm Model	20.2	23.5	24.7
Second Storm Model	20.7	24.0	25.2

This table shows that all five models predict the 50 and 100-year wave within 0.5 meters of the average value for the five models, and the 10-year wave within 1 meter of the average value for the five models. The appropriate grouping corrections were applied for the values given in the Table.

As discussed previously, the two storm models should give approximately the same height values; the second storm model and individual wave model should give the same height value; and the first storm model should give height values which are slightly greater than those for the most probable maximum model. These assertions agree with the results shown in the table which were obtained by applying the different models to data. Some discrepancy is expected between the assertions and the table due to the extrapolations required to obtain the return periods given in the table.

As discussed previously, the most probable maximum model is expected to diverge from the other models by an amount which increases as the data observation interval decreases. The table shows that this divergence is quite small for a data observation interval of 3 hours.

A mathematical comparison of the annual maximum model to the other models was not apparent. However, the table shows that for the set of data considered, the annual maximum model produces results which are similar to the results for the other models.

DATA BASE CONSIDERATIONS

As briefly discussed in the Introduction, the existing data base is an important consideration in extreme statistics. It is obvious that in order to project meaningful extreme statistics, a data base must be collected which is representative of severe sea states for the location of interest.

The storm models discussed above were developed for projecting extreme waves in the Gulf of Mexico. This area is subjected to tropical hurricanes which are the only type of storms that produce extreme waves in the area. Because any given location in the Gulf of Mexico will experience a hurricane only once in five to ten years on the average, the technique of historical hindcasts was evolved to produce a sufficient data base to calculate extreme wave statistics. In applyin this technique, historical data for barometric pressure and wind during hurricanes are collected, and the historical wave conditions for the hurricanes are estimated. Such hindcasts span periods up to 70 years [8,9]. These hindcasts are calibrated by instrumented wave data from hurricanes which have occurred recently [8]. Extreme wave statistics are then constructed from the hindcasts [9]. The storm models previously discussed are obvious models to appl to these hindcast data.

In a likemanner, the individual wave model and most probable maximum model evolved for the quite different wave climate in the North Sea. Whereas only one or two hurricanes occurred per decade at a Gulf of Mexico location, very severe storms occur frequently in the North Sea, accompanied by more frequent moderate storms. Because of the frequency of storm occurrence, the approach of direct wave observations on a continuous basis was undertaken to collect data bases which were used for extreme waves. In several published studies, [3] and [7] for example, extreme wave projections were made from only one winter season of observations. The use of the individual wave or most probable maximum models were natural models for observations collected on a continuous basis.

Thus different types of extreme wave models developed for different geographical areas which reflect the different wave climates and methods of data collection for these areas. However, the question arises as to the length of data base required to accurately project extreme wave statistics. Relevant input for this question can be obtained from the following discussion. It can be concluded that for the purpose of extreme statistics, no more than 10 independent samples were contained in the 10 years of visual data, even though the data contained over 13,000 separate observations taken every 3 hours. This follows because the two storm models and the annual maximum model used only 10 data samples to predict the same extreme statistics as the other two models which employed over 13,000 observations each. It is noted that all 10 storms used for the two storm models occurred in the last two years (1968 and 1969). Although all the storms occurred in these two years, the other eight years of data were required to establish meaningful return periods.

Reference 4 presented quantitative information relating to the required length of data bases for extreme wave statistics. This enlightening study used the Monte Carlo technique to randomly pick limited-size data bases from a fixed distribution of storm intensities. Uncertainty bounds were then constructed for the projected wave heights as a function of sample size. The uncertainty bounds were found to decrease at a meaningful rate until about 50 storms were included in the sample.

If at most, 10 independent extreme samples occurred during the 10 years of visual data considered in this study, at least 50 years of data may be required to reach the number of 50 samples found in [4]. This estimate can be further investigated by the use of Figure 9 which shows the distribution of annual maximum waves. In addition to the center line, this figure contains a pair of parallel lines, Curves A and C, which contain all of the data points. This set of lines can be interpreted as uncertainty bands of some unknown magnitude. The width of the band for the 100-year return period is $\pm$ 3 meters from the center line. A result from elementary statistics [10] is that the width of uncertainty bands is inversely proportional to the square root of the number of samples. Basing the $\pm$ 3 meters on ten samples, approximately 100 samples (or years of data) would be required to reduce this uncertainty to $\pm$ 1 meter, or 40 years to reduce this uncertainty to $\pm$ 1.5 meters for the 100-year wave in this area of the North Sea.

The possible need of up to 50 years of data, in contrast to one season of data used in some studies, is obviously a considerable difference. A major reason for the need of more than one season of data can be seen from Figure 12. This figure shows the 10 years of visual data collected from the Famita and 15 years of data [11] collected by Dutch lightvessels stationed at Terschellingerbank. The figure presents the data in the form of maximum annual observed wave height divided by the average for all annual maximum. The data spans 21 years and shows that the annual maximum exceeds the average only seven or 1/3 of the years. Also during the 21 years, there were two nonoverlapping seven-year periods in which the annual maximum did not exceed 1.1 times the average annual maximum. There were four years in which the maximum exceeded 1.4 times the average annual maximum. These data clearly illustrate that the extreme waves in the North Sea not only vary greatly from year to year, but also vary greatly from one five-year period to the next.

Therefore, the significantly different results found in [2] and [3] could be expected since only one winter season of data were used in [3] and at least 10 years in [2]. The consequence of using data from only one year is illustrated in Figure 13. In this figure, the most probable maximum model, used in [3], is applied to the visual data for the years 1967 and 1968. The projected 1/100

FIGURE 12. VARIATION OF ANNUAL MAXIMUM SEA STATE

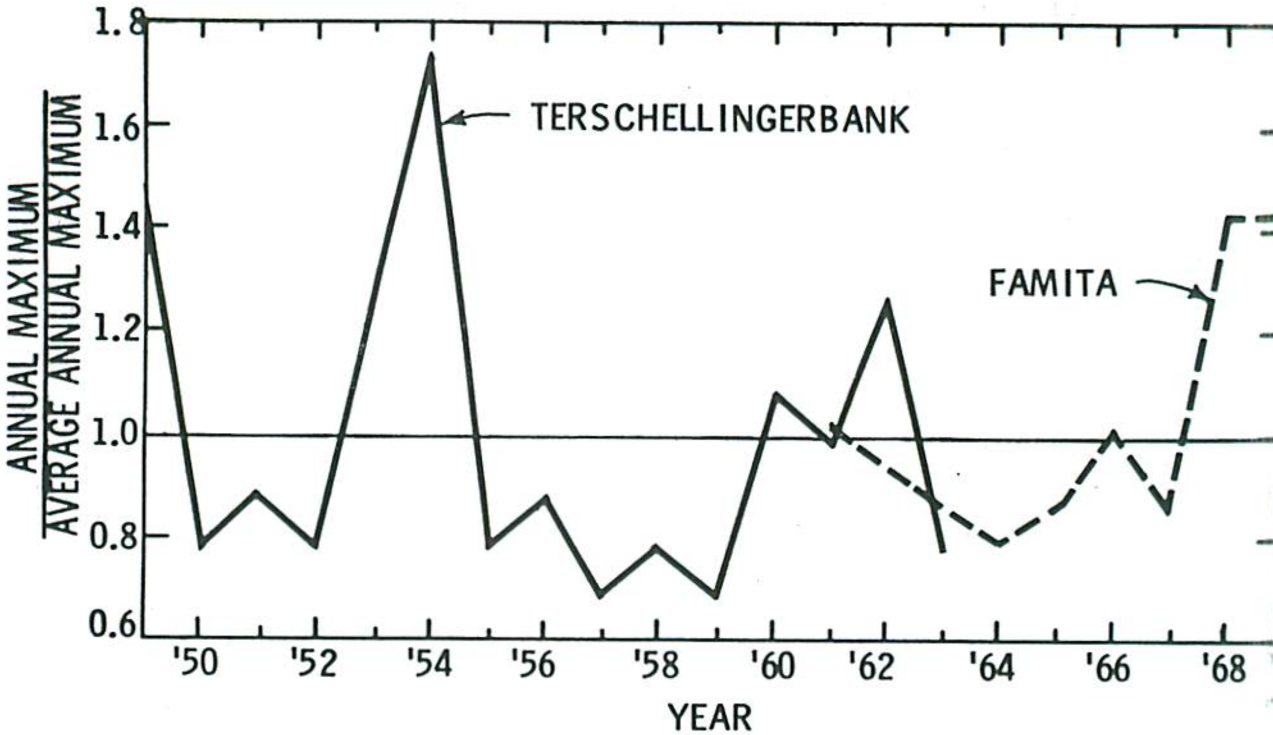
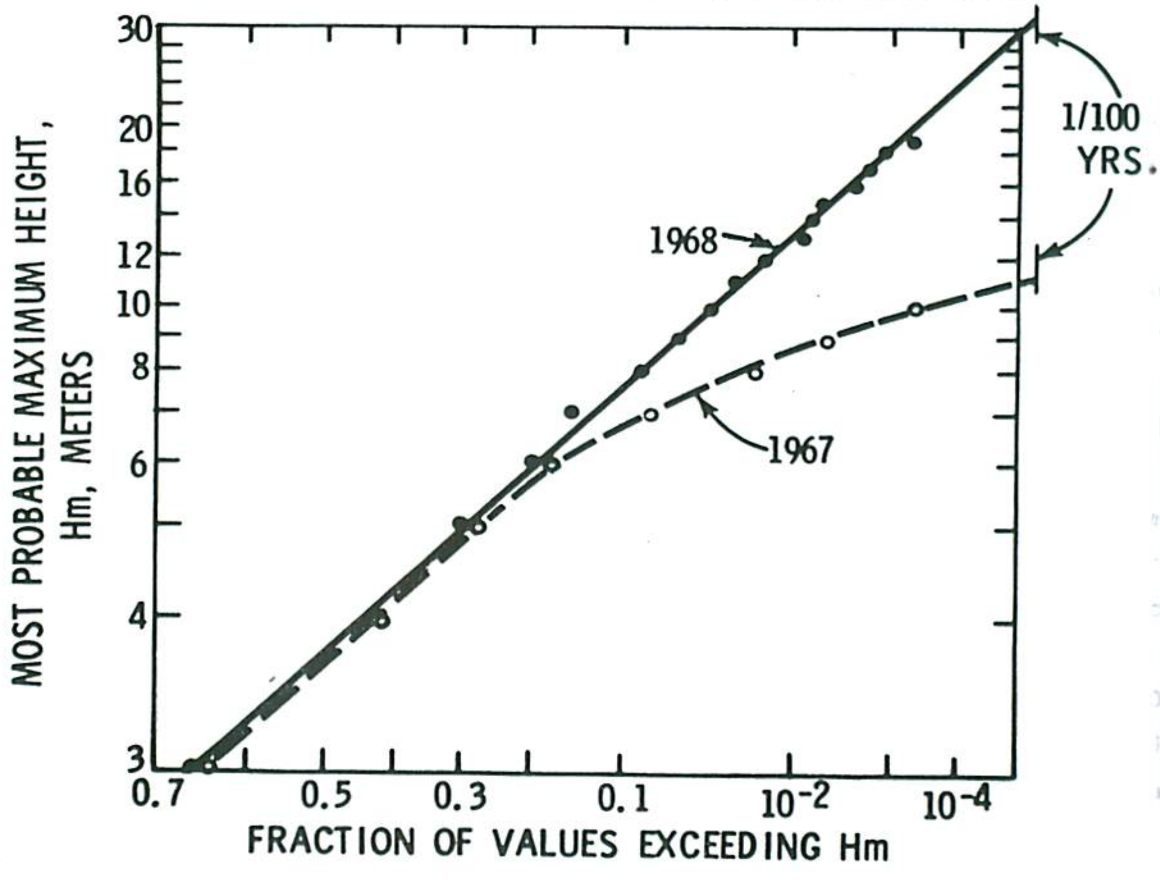


FIGURE 13. EXTRAPOLATIONS OF 1967 AND 1968 DATA



year wave for the 1968 data is 32 meters while the same projected wave for the 1967 data is only 11.5 meters. Thus the 1/100 year wave for the 1968 data is almost three times greater than that for the 1967 data. It is interesting to note that the distributions for the two years are nearly identical for the lower 80% of the observations.

The conclusion is drawn from the above discussions that although the occurrence of moderately large waves is similar from year to year in the North Sea, the occurrence of extreme waves is quite variable from year to year, and that a very long data base may be required to accurately determine the extreme wave statistics.

In light of the above discussion, the conclusion is reached that the hindcast technique is the only possible source of the long term data base required to accurately determine extreme wave statistics for the North Sea. This conclusion is amplified by the possibility that the extreme statistics will vary greatly from one area to another in the North Sea. This variation is expected because of the variability of the bathymetry and its effect on wave propagation, in addition to, the effect of land sheltering on wave generation. Although the existing data base for accurately measured extreme waves is meager and limited to isolated locations, there exists a wealth of historical data for barometric pressure and winds during past North Sea storms. A wave hindcast of historical storms could utilize the existing wave and meteorological data [8] to expand the data base for extreme waves to cover a period of at least 50 years and the complete expanse of the North Sea. Hindcasts do not replace wave measurement programs, but instead complement the measurement programs. Accurately measured wave data are required for final tuning and calibration [8] of hindcast programs. During this phase of the hindcast program, existing measured data become a part of the resulting data base. Because of the need for measured data in a hindcast program, wave measurements should continue even after a hindcast program is initiated.

A major reservation to the use of hindcasts is accuracy. In order to discuss the accuracy of hindcasts, an objective standard must be selected. One possible standard is the commonly used procedure, for example see [3] or [12], to estimate the significant height for instrumented wave measurements. The estimated root mean square, rms, error associated with this procedure is 13 percent [13]. For an extreme sea state with a significant height on the order of 15 meters, this procedure would produce an rms error of 2 meters in the estimated value of significant height. Therefore, an objective standard by which to judge the accuracy of hindcasts is an rms error of 2 meters. Inspection of the comparisons made in [8] and [14] of measured and hindcast waves, indicates that hindcasts with this accuracy standard appear feasible for the North Sea. If sufficient measured data is available to permit the uncertainties in the hindcasts to be quantified, this uncertainty can rationally be incorporated into the statistical procedures for determining the extreme wave statistics.

SUMMARY

The investigation reported in this paper was initiated to determine why significantly different results were reported in [2] and [3] for the extreme wave statistics at the same location in the North Sea. Possible reasons for the difference were that different statistical models and different data bases were used. Therefore, this paper investigated the models used in [2] and [3] along with three other models. These models were compared on a mathematical basis and on the basis of application to an actual set of wave data. In addition, the importance of the time span for data bases was examined.

In the course of the investigation of statistical models, the following conclusions were reached.

1. The two models used in [2] and [3] require a grouping correction in order to obtain correct return periods. In addition, the model used in [2] requires the inclusion of the effect of the correlation between the significant height and the average wave period for sea states.

2. After the appropriate corrections were made, these two models were found to predict similar heights for the same return period on the basis of a comparison of mathematical form and when applied to a common data base.

3. These two models were also found to give predictions similar to those of three other models considered. This conclusion was based on comparisons of mathematical form and projections from a common data base.

4. The projected extreme wave heights were reduced approximately one meter by incorporating into the models the realistic assumption that the significant height varies linearly between observations instead of remaining constant.

Although the model used in [3] produced results similar to the other models for the data base considered, this model is expected to diverge from the other models for data collected at a shorter interval than the three-hour interval data considered in this investigation. Data collected at less than a three-hour interval is desirable to accurately record peak wave conditions in storms.

The following conclusions were reached concerning the data base for extreme wave statistics.

1. The extreme wave conditions in the North Sea vary greatly from year to year, and severe storms may not occur for periods of up to seven years.

2. Projections of the 100-year wave height from one year of data can vary between consecutive years by a factor of three.

3. At most 10 independent events, relating to extreme wave statistics, were contained in the 10 years of visual data. In order to obtain 50 independent events, a data base of approximately 50 years may be required for extreme wave statistics in the North Sea.

4. The technique of hindcasting waves in historical storms will be required to achieve a 50-year data base.

5. Uncertainties associated with hindcasts may be no more than those inherent in some accepted techniques for analyzing instrumented wave data. If uncertainties in hindcasts are quantified by comparison with measured data, these uncertainties can be rationally included in the statistical procedures for determining extreme wave statistics.

From the above summary, the conclusion was reached that the significant difference between the projections of extreme wave heights in [2] and [3] did not result from the use of different models, but resulted because of the use of an inadequate data base, one winter season, in [3]. However, there also exist limitations to the projections obtained in [2]. These limitations result from not considering the correlation of significant height with average wave periods for sea states, not including a group correction for return periods, and the inherent uncertainties in visual wave data.

The general conclusion of this investigation is that the choice of a statistical model is not as important, for extreme wave statistics, as the need for an adequate data base. This conclusion assumes that the appropriate physical reality is incorporated in the selected model, such as, consideration of the correlation between significant height and average wave period, and the grouping into storms of large waves or observations containing large waves. After a sufficient data base is collected, on the order of 50 years, further consideration must be given to the validity of the assumption that individual storms are independent events. In addition, the data should be examined for the importance of cyclic weather patterns [19] on the return period of extreme wave heights.

ACKNOWLEDGEMENTS

The author wishes to express appreciation to Det Norske Veritas for furnishing the visual wave data used in this study. In addition, the author appreciates the valuable discussions concerning extreme waves with Dr. F. H. Hsu.

REFERENCES

1. Hogben, N. and Lumb, F. E.: <u>Ocean Wave Statistics</u>; National Physical Laboratory; Her Majesty's Stationery Office; London (1967).

2. Pedersen, B.: "Prediction of Long-Term Wave Conditions with Special Emphasis on the North Sea"; Proceedings, The First International Conference on Port and Ocean Engineering Under Arctic Conditions (Vol. II); Trondheim (1971).

3. Draper, L. and Driver, J. S.: "Winter Waves in the Northern North Sea at 57°30'N 3°00'E, Recorded by M. V. Famita"; Proceedings, The First International Conference on Port and Ocean Engineering Under Arctic Conditions (Vol. II); Trondheim (1971).

4 Petrauskas, C. and Aagaard, P. M.: "Extrapolation of Historical Storm Data for Estimating Design Wave Heights"; 1970 Offshore Technology Conference, Houston; SPE Journal 11 (1970).

5. Jahns, H. O. and Wheeler, J. D.: "Long-Term Wave Probabilities Based on Hindcasting of Severe Storms"; 1972 Offshore Technology Conference, Houston; Journal of Petroleum Technology, Vol. XXV (1973).

6. Durning, P. J.: "Prediction of Maximum Wave Height from Historical Data"; Preprints, Offshore Technology Conference, Houston (1971).

7. Battjes, J. A.: "Long-Term Wave Height Distribution at Seven Stations Around the British Isles": N.I.O. Internal Report No. A.44, National Institute of Oceanography, Wormley (1970).

8. Hsu, F. H.: "Hindcast Storm Waves for Compilation of Wave Statistics"; Second Annual Meeting of Society of Petroleum Engineers of AIME, London (1973).

9. Besse, C. P. and LeBlanc, N. F.: "An Application of Oceanographic Data in Offshore Structural Design"; Transactions of Society of Petroleum Engineers of AIME, Vol. 240 (1967).

10. Lindgren, B. W. and McElrath, G. W.: <u>Introduction to Probability and Statistics</u>; The Macmillan Company, New York (1961).

11. Dorrestein, R.: <u>Wind and Wave Data of Netherlands Lightvessels Since 1949</u>; Royal Netherlands Meteorological Institute (1967).

12. Dattatri, J.: "Waves Off Mangalore Harbor - West Coast of India"; Proceedings of the ASCE, Journal of the Waterways, Harbors and Coastal Engineering Division, Vol. I (1973).

13. Draper, L.: "The Analysis and Presentation of Wave Data - A Plea for Uniformity"; Proceedings, 10th Conference on Coastal Engineering, Vol. I, Tokyo (1966).

14. Wilson, B. W., Chakrabarti, S. K; and Feldhausen, P. H.: "Hindcasts of Deep-Water Wave Energy Spectra for Hurricane Audry of 1971"; Preprints, Offshore Technology Conference, Houston (1973).

15. Rice, S. O.: "Mathematical Analysis of Random Noise"; Selected Papers on Noise and Stochastic Processes; Nelson Wax, Editor; Dover Publications, Inc., New York (1954).

16. Longuet-Higgins, M. S.: "On the Statistical Distribution of the Heights of Sea Waves"; Journal of Marine Research, Vol XI (1952).

17. Borgman, L. E.: "Risk Criteria"; Proceedings of the ASCE, Journal of the Waterways and Harbors Division, Vol. III (1963).

18. Borgman, L. E.: "Probability for Highest Wave in Hurricane"; Proceedings of the ASCE, Journal of the Waterways, Harbors and Coastal Engineering Division, Vol. II (1973).

19. Russel, L. and Schueller, G.: "Probabilistic Models for Texas Gulf Coast Hurricane Occurrences"; Preprints, Offshore Technology Conference, Houston (1971).

APPENDIX A: Distributions for Wave Heights

This appendix will consider the distributions for the individual wave heights and the maximum height for sea states specified in terms of the significant height, Hs. The distributions will relate to sea states for which Hs is considered constant and for which Hs is assumed to vary linearly between discrete values.

The Rayleigh distribution has been used almost exclusively for the distribution of wave heights. The basis for the use of the Rayleigh distribution comes from the basic work on stochastic processes by Rice [15] and the extension to sea waves by Longuet-Higgins [16]. The Rayleigh distribution can be expressed in terms of the significant height, Hs, by

$$R(H,Hs) = \exp(-2H^2/Hs^2) \qquad \text{(A-1)}$$

for the probability that any wave height will exceed H.

The distribution for the maximum value of H contained in a train of n waves was introduced in [16], and can be expressed as

$$M(H) = 1-[1-R(H,Hs)]^n \qquad \text{(A-2)}$$

The most probable maximum value, Hm, is the value of H which produces the largest value of the density function associated with (A-2), see Figure 2-a. An asymptotic expression for Hm was derived in [16] and can be expressed as

$$Hm/Hs = \sqrt{(\ln n)/2} \qquad \text{(A-3)}$$

where $\ln$ denotes the logarithm of base e.

The expression (A-3) can be shown to be equivalent to the value of H which is expected to occur one time during n waves; that is

$$R(Hm,Hs) = 1/n \qquad \text{(A-4)}$$

The most probable maximum height, Hm, during a storm or a fixed length of time, is commonly [3, 4, 5, 6] the parameter used in extreme wave statistic analyses. In order to determine Hm from either (A-3) or (A-4) a value of Hs must be selected which is considered constant for a period of time which determines n. However, actual sea states are not constant with time, but change from one observation time to the next. For practical considerations, the period of time between observations is selected to be relatively long, for example, 3 hours. Therefore, sea states can change significantly between observation times and the selection of an equivalent constant value of Hs and corresponding n is not straightforward and can introduce a bias in the resulting values of Hm. One way of avoiding this possible bias and the necessity of selecting equivalent values of Hs and n, is to assume that the value of Hs varies linearly between observation times. This is equivalent to connecting the individual observations of Hs with straight line segments as illustrated in Figure 1. For two consecutive observations, denoting the larger value of significant height as Hs and the smaller value as αHs ($\alpha<1$), the distribution for the individual heights, H, between the two observations is

$$V(H,Hs,\alpha) = \int_{\alpha Hs}^{Hs} R(H,h)\left(\frac{dh}{\int_{\alpha Hs}^{Hs} dh}\right) = \int_{\alpha Hs}^{Hs} \frac{[\exp(-2H^2/h^2)]dh}{Hs(1-\alpha)} \qquad \text{(A-5)}$$

$$= 1 + \sum_{m=1}^{\infty} \frac{(-2)^m}{m!} \left(\frac{H}{Hs}\right)^{2m} \frac{[\alpha^{-(2m-1)}-1]}{(2m-1)(1-\alpha)} \qquad \text{(A-6)}$$

In the above expressions, h denotes values of Hs, and the series representation of the exponential function was used for exp( ) in the integration. The above infinite series converges at an efficient rate on a digital computer for cases which satisfy $(H/Hs)/\alpha < 3$. For cases not meeting this condition, the integration is computed over two ranges, αHs to α'Hs and α'Hs to Hs, where α' is selected as (H/Hs)/3. Since the contribution to the integral by the first range is effectively zero, the distribution in (A-5) can be evaluated by the following expression for cases in which $\alpha<(H/Hs)/3$

$$V(H,Hs,\alpha) = \frac{1-\alpha'}{1-\alpha} V(H,Hs,\alpha') \qquad \text{(A-7)}$$

$$\alpha' = (H/Hs)/3$$

The distribution of H for M+1 consecutive observations, which will contain M time periods between observations, can be expressed in terms of Hs_i, α_i, and m_i (number of waves) for the i^{th} time period,

$$V^*(H) = \frac{\sum_{i=1}^{M} m_i V(H,Hs_i,\alpha_i)}{\sum_{i=1}^{M} m_i} \qquad \text{(A-8)}$$

The distribution V* can be applied to a storm which consists of M observations of Hs.

The distributions V and V* derived above for time varying sea states are equivalent to R, for constant sea states. Therefore, expressions for the distribution of the maximum wave height or the most probable maximum wave height can be found for V and V* in the same manner as the equivalent expressions were found for R, namely (A-2) and (A-4).

The most probable maximum height, Hm, occurring for n waves satisfies the implicit relationship

$$V(Hm, Hs, \alpha) = 1/n \tag{A-9}$$

Because of the nature of (A-9), an iteration process must be used to determine Hm. Figure 14 shows the dependence of Hm on n and α. In addition, the figure shows that the curves for different values of α are nearly parallel to the curve for $\alpha = 1$ which is given by (A-3). Because of this behavior, an empirical correction can be applied to (A-3) which gives a very good first approximation to Hm and permits the expression in (A-9) to be satisfied with sufficient accuracy in only two iterations. Figure 14 shows curves for α values greater than 0.8. Significant height levels, less than 0.8 times the maximum value of Hs, effectively do not influence the statistics of the maximum height value. Therefore, for any case with α less than 0.8, the value of Hm can be found from the $\alpha = 0.8$ curve if the number of waves is selected as the number which occurred when Hs exceeded the 0.8 level.

The distribution for the most probable height, Hms, during a storm composed of M linearly varying segments for Hs and containing a total of n waves is found from (A-8) by the implicit expression

$$V^*(Hms) = 1/n \tag{A-10}$$

For the purpose of this study, the values of Hms for each storm was found by inspection of a plot of V* versus H for the storm.

The distribution for all possible values of the maximum height for a storm composed of M linearly varying segments for Hs and containing a total of n waves, is found from (A-8) as

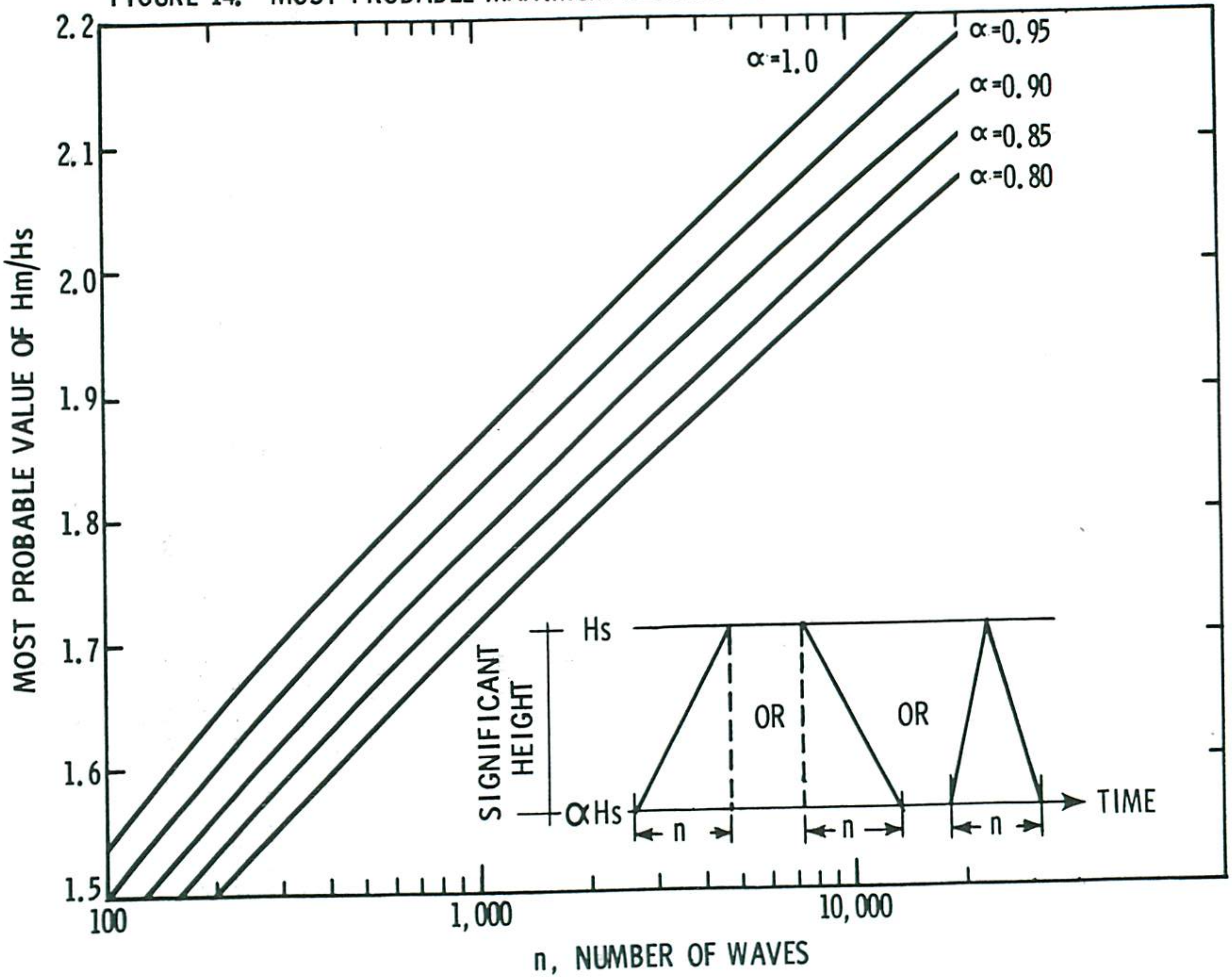
$$D(H) = 1-[1-V^*(H)]^n \tag{A-11}$$

The number of waves which occur between two observation periods is calculated using the length of time between observations and the mean of the two observed values for the average wave period.

The storm models, represented by Equations (10) and (14), employ the Poisson distribution for probability of storm occurrences during n years. The application of the Poisson distribution for this purpose was discussed in [17]. A brief derivation is given below for the application of the Poisson distribution to extreme wave statistics.

If λ is the average annual frequency of storm occurrences, the probability of k storms occurring during n years is

FIGURE 14. MOST PROBABLE MAXIMUM HEIGHTS FOR LINEARLY VARYING SEA STATES



$$P(k) = \exp(-n\lambda)\,\frac{(n\lambda)^k}{k!} \tag{A-12}$$

In addition, if Q(H) is the probability of not exceeding H during a storm, the probability of not exceeding H during n years is

$$\begin{aligned} Un(H) &= P(0) + P(1)Q(H) + P(2)Q^2(H) + \ldots \\ &= \sum_{k=o}^{\infty} P(k)Q^k(H) = \exp(-n\lambda) \sum_{k=o}^{\infty} \frac{[n\lambda Q(H)]^k}{k!} \\ &= \exp(-n\lambda)\ \exp[n\lambda Q(H)] = \exp\{-n\lambda[1-Q(H)]\} \end{aligned} \tag{A-13}$$

The probability of exceeding H during n years is then

$$Sn(H) = 1-Un(H) = 1-\exp\{-n\lambda[1-Q(H)]\} \tag{A-14}$$

In (A-13), the term $P(k)Q^k(H)$ is the combined probability of the occurrence of k storms and the nonoccurrence of H during the k storms. The term, $[1-Q(H)]$ in (A-14), is the probability of exceeding H during a storm, and therefore is equivalent to S(H) of Equation (10) or S*(H) of Equation (14). Thus (A-14) can be written as

$$Sn(H) = 1-\exp[-n\lambda S(H)] \tag{A-15}$$

Since the completion of the investigation reported here, a paper has been published [18] in which an "extremal Rayleigh distribution" is derived for the maximum waves in storms with a time varying significant height. The implications of [18] as related to the present study have not been investigated.

APPENDIX B: Comparison of Mathematical Form of Models

In this appendix, it will be demonstrated that the individual wave model and the second storm model give the same return periods, and that the most probable maximum model and the first storm model give approximately the same return periods. For convenience it will be assumed that the data bases are large enough that the resulting distributions can be represented in a discrete form and that extrapolation is not required. Although this is rarely the case in reality, the conclusion achieved in this appendix should also be applicable for limited data bases.

The distribution used by the individual wave model can be expressed in terms of Equations (4) or (5) as

$$I(H) = \frac{\sum_{i=1}^{K} k_i X_i}{\sum_{i=1}^{N} k_i} = \sum_{i=1}^{K} \frac{k_i X_i}{\mu L} \qquad \text{(B-1)}$$

in which K is the number of observations taken during L years of data collection and μ is the number of data observations per year. In addition, k_i is the number of individual waves associated with the i^{th} observation and X_i is the distribution for the occurrence of H associated with the i^{th} observation (X = R for (4) or V for (5)). The range of interest of I(H) will be restricted to those values of H which exceed the threshold level for the definition of a storm. Therefore, for any value of H in this range, the only distributions for H, the X_i, which effectively contribute to I(H) correspond to X_i's which result for periods of storms. In addition the remaining entries of $k_i X_i$ in the summation, which contribute to I(H), can be divided into groups corresponding to the individual storms. If the partial sum of the $k_i X_i$ for a given storm is formed and divided by the corresponding partial sum of the k_i, the resulting quantity (denoted by P) is the distribution for all waves during the storm (see (A-8) of Appendix A). That is

$$P = \frac{\sum k_i X_i}{\sum k_i} = \frac{\sum k_i X_i}{n} \quad \text{or} \quad nP = \sum k_i X_i \qquad \text{(B-2)}$$

for which the range of summation is only over those entries during a given storm and n is the total number of waves which occurred during the storm. Denoting the partial sum, nP, for the j^{th} storm by $n_j P_j$, letting N equal the total number of storms during the L years of record, and Ho be the storm threshold level, (B-1) can be rewritten as

$$I(H) = \sum_{j=1}^{N} \frac{n_j P_j(H)}{L\mu}, \text{ for } H>Ho \qquad \text{(B-3)}$$

in which the dependence of P on H is emphasized.

The uncorrected return period of m* years is found from (B-3) by equating I(H) to 1/M, for which M is the total number of waves during m years. Since M = μm, the height, H, with an uncorrected return period m* is found from

$$I(H) = 1/M = 1/\mu m^* = \sum_{j=1}^{N} \frac{n_j P_j(H)}{L\mu} \qquad \text{(B-4)}$$

or $$1/m^* = \sum_{j=1}^{N} \frac{n_j P_j(H)}{L}$$

The grouping correction which m* must be multiplied by to obtain the correct return period is found from (C-9) of Appendix C, and equals

$$\bar{k}(H) = \frac{\sum_{j=1}^{N} n_j P_j(H)}{\sum_{j=1}^{N} \{1-[1-P_j(H)]^{n_j}\}} \tag{B-5}$$

After m* of (B-4) is multiplied by $\bar{k}(H)$ to obtain the corrected return period m, the following results

$$1/m = 1/\bar{k}(H)m^* = \frac{1}{L}\sum_{j=1}^{N} \{1-[1-P_j(H)]^{n_j}\} \tag{B-6}$$

Since P_j is the distribution of the individual wave heights for the j^{th} storm, and n_j is the number of waves, the expression in { } on the right-hand side of (B-6) is the distribution for the maximum wave height (see Appendix A) during the j^{th} storm. This is the same distribution that was denoted as D for the second storm model. Therefore denoting this distribution by D results in

$$1/m = \sum_{j=1}^{N} \frac{\{1-[1-P_j(H)]^{n_j}\}}{L} = \sum_{j=1}^{N} \frac{D_j(H)}{L} \tag{B-7}$$

Equation (15) for the second storm model is

$$\lambda S^*(H) = \sum_{i=1}^{N} \frac{D_i(H)}{L} \tag{B-8}$$

and the value of H with a return period of m for this model is

$$S^*(H) = 1/\lambda m$$

or $$\lambda S^*(H) = 1/m = \sum_{i=1}^{N} \frac{D_i(H)}{L} \tag{B-9}$$

Therefore the expression for the corrected return period for the individual wave model (B-7) is the same as the expression for the return period for the second storm model (B-9).

The most probable maximum model utilizes the distribution for the most probable maximum wave height, Hm, which occurred during the time between data observations. This distribution for the individual Hm_i values is given in Equation (6) or (7) as

$$Q(H) = \sum_{i=1}^{K} \frac{h(Hm_i - H)}{K} = \sum_{i=1}^{K} \frac{h(Hm_i - H)}{\mu L} \tag{B-10}$$

In this expression, h( ) is the Heaviside function which equals unity for a positive argument and zero for a negative argument. Also in (B-10), K is the total number of data observations contained in the data record of length L years, and μ is the number of data observations per year.

The range of interest of Q(H) will be restricted to those values of H which exceed the threshold level, Ho, for the definition of a storm. Therefore by the definition of h( ), the observations, Hm_i, which are less than Ho do not contribute to this restricted range of Q(H). If there are I values of Hm_i greater than Ho, (B-10) can be expressed as

$$Q(H) = \sum_{i=1}^{I} \frac{h(Hm_i - H)}{\mu L} \quad \text{for} \quad H>Ho \tag{B-11}$$

Since the I values of Hm_i in (B-11) must occur during storms and assuming there are N storms, the I values can be divided into N groups representing the storms. If the partial sums for these N groups of $h(Hm_i-H)$ are formed, and if the value of the partial sum for the j^{th} group is evaluated and denoted as $s_j(H)$, (B-11) can be written as

$$Q(H) = \sum_{j=1}^{N} \frac{s_j(H)}{\mu L} \quad \text{for} \quad H>Ho \tag{B-12}$$

In (B-12), the sum of the $s_j(H)$ is equal to the total number of times Hm exceeded the value H during the data record.

The uncorrected return period, m*, for the height H is found from (B-12) as

$$1/\mu m^* = Q(H) = \sum_{j=1}^{N} \frac{s_j(H)}{\mu L} \tag{B-13}$$

or $$1/m^* = \sum_{j=1}^{N} \frac{s_j(H)}{L}$$

The grouping correction parameter which must multiply m* to give the correct return period, m, is

$$\overline{k}^*(H) = \sum_{j=1}^{N} \frac{s_j(H)}{n(H)} \tag{B-14}$$

where n(H) is the number of storms in the data record which contained Hm_i values that are greater than H. Applying this correction parameter to (B-13) results in

$$1/m = 1/\bar{k}^*(H)m^* = \frac{n(H)}{L} \tag{B-15}$$

The distribution used by the first storm model is given in Equation (9) as

$$S(H) = \sum_{i=1}^{N} \frac{h(Hsm_i - H)}{N} \tag{B-16}$$

The return period, m, for this distribution is found from

$$1/m\lambda = \frac{1}{m(N/L)} = S(H) = \sum_{i=1}^{N} \frac{h(Hsm_i - H)}{N} \tag{B-17}$$

or $$1/m = (1/L) \sum_{i=1}^{N} h(Hsm_i - H)$$

in which λ is the average annual frequency of storms and is equal to the number of storms in the record divided by the record length. In (B-17) the sum of $h(Hsm_i-H)$ is the number of storms with a most probable maximum height greater than H. From (B-17) and (B-15), it is seen that the return periods for the firs storm model would be the same as that for the most probable maximum model if

$$n(H) \stackrel{?}{=} \sum_{i=1}^{N} h(Hsm_i - H) \tag{B-18}$$

However, it must be expected that the right-hand side of (B-18) is larger than the left-hand side because the number of storms with a most probable maximum height exceeding some value of H should be larger than the number of storms which contain a period of time (equal to the observation interval) during which the most probable maximum height exceeds H. This follows because the most prob value depends on the number of waves (see Figure 14) and the number of waves during a storm is larger than the number of waves during a fixed time interval falling within the time span of the storm. Therefore, if the time interval between observations is decreased in order to obtain a more detailed set of wav data, the difference in the two models would increase. However, the most probabl value is not a strong function of the number of waves and changes by less than 5 percent if the number of waves is doubled. In addition, the portion of a stor which contributes to the statistics of the maximum wave height, is limited in time to the peak region of the storm (Appendix A). Therefore it should be expe that the two quantities in (B-18) differ only slightly and that the wave height with a given return period will only be slightly larger (on the order of 5 perce for the storm model (B-16) than for the most probable maximum model (B-10).

APPENDIX C: Grouping Correction

In this appendix the grouping correction is derived for the individual wave model. For this model, the grouping correction which must be applied to the return period associated with a given height, H, can be described as the average number of waves which exceed H during storms which contain at least one wave that exceeds H.

The most appropriate distribution for the number of waves, k, which exceed H during a storm is the binomial distribution [5]. However for the special conditions which are associated with extreme waves, the Poisson distribution is an adequate approximation [10] for the binomial distribution. These special conditions are the small probability of occurrence for the larger waves during a storm and the large number of waves of all heights which occur during a storm.

The Poisson probability-density function [10] for the number of waves, k, which exceed H during a storm, with an expected number of m waves exceeding H, is

$$p(k,m) = \exp(-m)\,\frac{m^k}{k!} \tag{C-1}$$

The expected number of waves, m, is found by multiplying the probability of a wave greater than H, P(H), by the total number of waves, n. That is, $m = nP(H)$.

The probability that at least one wave exceeds H can be expressed as

$$p(k>0,m) = 1-p(0,m) = 1-\exp(-m) \tag{C-2}$$

in which $p(0,m)$ is the probability that no waves exceed H. For a finite number of storms, N, for which the expected number of exceedences for the i^{th} storm is m_i, the density function for k from this population of storms is

$$p^*(k) = \frac{1}{N}\sum_{i=1}^{N} p(k,m_i) = \frac{1}{N}\sum_{i=1}^{N} \exp(-m_i)\,\frac{m_i^k}{k!} \tag{C-3}$$

Likewise for this population of N storms, the probability that at least one wave greater than H occurs during a storm is

$$p^*(k>0) = \frac{1}{N}\sum_{i=1}^{N} p(k>0,m_i) = \frac{1}{N}\sum_{i=1}^{N} [1-\exp(-m_i)] \tag{C-4}$$

The conditional probability density function [10] for the number of waves which exceed H, given that at least one wave exceeds H, for this population of storms is found from (C-3) and (C-4) as

$$\frac{p^*(k)}{p^*(k>0)} \tag{C-5}$$

The average value of k, denoted by $\overline{k}$, for (C-5) is

$$\overline{k} = \sum_{k=1}^{\infty} \frac{kp^*(k)}{p^*(k>0)} = \frac{1}{p^*(k>0)} \sum_{k=1}^{\infty} kp^*(k) \tag{C-6}$$

in which $p^*(k>0)$ was taken out of the summation because of its independence of k. By using the result that

$$\sum_{k=1}^{\infty} k \exp(-m_i) \frac{m_i^k}{k!} = m_i \exp(-m_i) \sum_{k=0}^{\infty} \frac{m_i^k}{k!} = m_i \exp(-m_i) \exp(m_i) = m_i$$

and inserting the expressions for $p^*(k)$ and $p^*(k>0)$ into (C-6) and interchanging the order of summation of i and k, results in

$$\overline{k}(H) = \frac{\sum_{i=1}^{N} m_i}{\sum_{i=1}^{N} [1-\exp(-m_i)]} = \frac{\sum_{i=1}^{N} n_i P_i(H)}{\sum_{i=1}^{N} [1-\exp(-n_i P_i(H))]} \tag{C-7}$$

in which the dependence of $\overline{k}$ on H is emphasized.

The expression (C-7) gives the average occurrence of the exceedence of H in storms with at least one exceedence of H, for a population of N storms which individually contain n_i waves and a probability of exceeding H of $p_i(H)$.

The denominator of (C-7) can be written in a different form by using the approximation

$$\exp(-nP) \simeq [1-P]^n \tag{C-8}$$

The validity of this approximation is based on the same assumptions which permitted the Poisson distribution to be used for the binominal distribution at the beginning of this appendix. The left-hand side of (C-8) is seen from (C-2) to be the Poisson probability of zero occurrences of heights greater than H, while the right-hand side of (C-8) is the binominal probability for zero exceedences. By the manipulation of logarithms, the exponential function, and the series representation of the exponential function, it can be shown that the ratio of the left-hand side to the right-hand side of (C-8) is $1 + 0(nP^2)$. The notation $0(x)$ implies the order of magnitude of x. For extreme waves $P = 0(1/n)$, see for example (A-10). Therefore the ratio of the terms in (C-8) is $1 + 0(n/n^2) = 1 + 0(1/n)$. n for storms is $0(10^3)$, which implies that the ratio is $1 + 0(10^{-3})$, and that the approximation in (C-8) is very good for extreme storm waves.

Using (C-8) in (C-7) results in

$$\overline{k}(H) = \frac{\sum_{i=1}^{N} n_i P_i}{\sum_{i=1}^{N} \{1-[1-P_i(H)]^{n_i}\}} \qquad \text{(C-9)}$$

APPENDIX D: Description of Data

The set of data employed by this study consisted of visual wave observations collected every 3 hours during the months of October through April for the time period of 1959 through 1969. These data were collected by the Norwegian Meteorological Institute from the rescue ship, Famita, stationed in the Norwegian sector of the North Sea near 57°30'N,3°00'E. In this paper the visual wave height estimate was used for the significant height without applying a correction formula.

Since the set of data consisted of 3-hour observation intervals, potentially 240 observations could be taken during a 30-day month, or 1440 during the six-month observation period. In this paper, it was assumed that any missing observations occurred randomly. This assumption permits the missing observations to be ignored without any statistical consequence. Observations were not made during the three-month period of October through December for the years 1959 and 1960. For the purpose of this paper, the remaining three months of observation for each of these two years were combined to form six months of observations, or a period equal to the observation period for the other years.

In addition, it will be assumed that the six-month period of October through March is sufficient to represent the extreme waves for a complete year. This assumption is reasonable because significantly more severe wave conditions occur during these six months than the remaining six months of the year. As a result of this assumption, and by combining the 1959 and 1960 data, the data set will be considered as ten complete years of wave data for the purpose of extreme wave statistics.

The visual wave data were recorded to the nearest one-half meter, with most of the larger values recorded as whole meter values. The data therefore tended to be lumped at values of whole meters. In order to permit a smooth distribution of the data, each data value was randomly distributed around the reported value by a uniform distribution with a width of ± 0.5 meters.

The distributions appropriately describing the data for the different models were plotted on a Weibull distribution scale. This scale permits many distributions to be represented by a straight line. A resulting straight line is a necessity for extrapolating the distributions to determine extreme statistics. The expression for the Weibull distribution [4] is

$$P(X>Xo) = \exp\left[-\left(\frac{Xo-A}{B}\right)^k\right] \qquad (D\text{-}1)$$

which gives the probability that the variable X is greater than the value Xo. In (D-1), A, B and k are constants. The straight line for (D-1) results on a log-log plot of (X-A) versus $-\ell n(P)$. For the plot, the value of A is selected to yield a straight line and k is the slope of the resulting line. The value of B is equal to the magnitude of (X-A) which yields P = 1/e.

The appropriate plotting formula for (D-1) was derived in [4] and depends on the value of k. This formula was used to construct Figures 9 and 10.

749