THE SOCIETY OF NAVAL ARCHITECTS AND MARINE ENGINEERS
One World Trade Center, Suite 1369, New York, N.Y. 10048

Papers to be presented at Extreme Loads Response Symposium
Arlington, VA, October 19–20, 1981

Principles of Extreme Value Statistics and their Application

Michel K. Ochi, University of Florida, Gainesville, FL

ABSTRACT

This paper presents the theoretical background of recent techniques to predict extreme loads (and responses) pertinent to ship (or marine system) operation in a seaway. The magnitude of extreme values, in general, can be estimated through different approaches, each based on a different principle. This paper outlines the three different approaches for predicting extreme values which provide information useful for design consideration of a marine system. These are,

(1) The initial probability distribution is known; for example, prediction of extreme responses of a marine system in a seaway in the short-term based on the linear superposition principle. In this case, the extreme values can be simply evaluated by analytical formulation through application of order statistics to the initial probability distribution.

(2) The initial probability distribution is unknown but observed (or computed) data are available; for example, prediction of the extreme sea state (significant wave height) or prediction of the extreme responses of a marine system in the long-term, etc. Estimation of the extreme values is carried out by approximate methods through the use of the observed (or computed) data.

(3) The initial probability distribution is unknown but observed data of the maxima only are available; for example, prediction of lifetime extreme responses of a marine system from daily observed maximum values, etc. In this case, the extreme values are obtained from the asymptotic extreme value distributions developed in order statistics.

The difference in these three prediction approaches are clearly explained, and practical examples to evaluate the extreme values following each method are presented.

INTRODUCTION

This paper is prepared to provide information necessary for understanding the probabilistic prediction of extreme values, specifically extreme responses of a marine system in a seaway.

For the design of a marine system, it is highly desirable to obtain the magnitude of the system's responses in various seas. Among others, the largest response (extreme value) which the system will experience in her lifetime is necessary to assess possible structural failure which may occur as soon as a single load exceeds the value critical for the system's structural strength. This type of failure is called the *first excursion failure* in stochastic process theory. The information on first excursion failures is necessary together with that on fatigue failures for reliability analysis of the system under random excitation.

The magnitude of extreme response of a system which is associated with the first excursion failure can be evaluated through different approaches, each based on a different principle. Hence, great care has to be given in evaluating the extreme responses depending on the prediction method one may take.

The *extreme value* is defined, in general, as the largest value expected to occur in a certain number of observations or in a certain period of time. It can be defined on a short-term basis in which the sea environment is statistically invariant (usually from 30 minutes to several hours) as well as on a long-term basis (usually for many years). In either case, however, the number of observations or a period of time have to be specified in defining the extreme value. For example, in estimating the magnitude of the extreme response of a marine system's lifetime, the number of response cycles expected in the lifetime has to be clearly specified. It is unfortunate that this important issue of the extreme value statistics has often been neglected in estimating the extreme responses of a marine system.

Prior to discussing the extreme response of a marine system, let us consider the response at every cycle of encounter with waves irrespective of its magnitude. Here, the response is a random variable, denoted by X, and it has its own probability density function, f(x), and the cumulative distribution function, F(x). These functions are often called the *initial probability density function* and the *initial cumulative distribution function*, respectively, in discussing the extreme value statistics.

The extreme response in n-wave encounters, denoted by Y_n, is also a random variable and follows its own probability law, which is different from that applicable for the response X. To avoid a possible confusion, let us write the probability density function and cumulative distribution function of the extreme response as $g(y_n)$ and $G(y_n)$, respectively. Here, the probability functions, f(x), F(x), $g(y_n)$, and $G(y_n)$ have mathematical relationships as will be shown later. Therefore, the extreme response can be easily evaluated by applying the formulation in extreme value statistics, called *order statistics*, if the initial distribution, f(x), is known.

However, this is not always the case. If the initial distribution is unknown, the extreme response can be evaluated either through approximate methods or by applying asymptotic formulations. Consequently, the estimation of the extreme values may be categorized into three areas as shown in Table 1. The details are as follows:

(1) The initial probability distribution is known

The estimation of extreme responses of a marine system in the short-term is a typical example of this case, since the initial probability distribution is usually considered to be the Rayleigh probability distribution. In this case, the probability function of the extreme values can be precisely derived by applying order statistics.

(2) The initial probability distribution is unknown

There are many practical cases of evaluating extreme responses for which the initial probability distribution is not known. In this case, the extreme responses can be evaluated either through statistical estimation of the initial distribution or by application of the asymptotic extreme value statistics. For the former approach, measured (or computed) data of the response are required, while data of the daily, monthly, or yearly measured (or observed) maxima are required for the latter approach.

(a) Measured (or computed) data are available

A typical example of this case is the prediction of the extreme responses in the lifetime of a marine system. The initial distribution of the response covering the lifetime of the system is not precisely known, but accumulation of some measured (or computed) data may be available.

Another example is the estimation of the most severe sea state (significant wave height) in the long term. Again, the probability distribution for the long-term significant wave height is not precisely known, and hence it is estimated approximately through the use of the observed data.

(b) Measured (or observed) data of the maxima are available

It is often necessary to estimate the extreme values in a specified period of time from the accumulation of daily, monthly, or yearly measured (or observed) largest values, called the maxima for brevity's sake. For example,

Table 1 Estimation of extreme values

CONDITION		ESTIMATION OF EXTREME VALUES
INITIAL PROBABILITY DISTRIBUTION IS KNOWN		EXACT METHOD
INITIAL PROBABILITY DISTRIBUTION IS UNKNOWN	MEASURED (OR COMPUTED) DATA ARE AVAILABLE	APPROXIMATE METHOD
	MEASURED (OR COMPUTED) DATA OF THE MAXIMA ARE AVAILABLE	ASYMPTOTIC FORMULATIONS

estimate the extreme ship response expected in 30 years from the data of the daily largest response measured over 3 years. In this case, the extreme responses in 30 years, for example, are evaluated by the asymptotic extreme value distributions which were originally developed by Fréchet and later were systematized by Gumbel (1958).

EXACT EVALUATION OF EXTREME RESPONSES

As stated earlier, the initial probability function of the response and the extreme probability function are mathematically related. That is,

Probability density function of extreme value:

$$g(y_n) = n[f(x)\{F(x)\}^{n-1}]_{x=y_n} \tag{1}$$

Cumulative distribution function of extreme value:

$$G(y_n) = [\{F(x)\}^n]_{x=y_n} \tag{2}$$

Various statistical properties of the extreme values can be evaluated from (1) and (2). For example, the *probable extreme value*, $\bar{Y}_n$, defined as the extreme value most likely to occur in n-observations can be obtained as the modal value of the probability density function $g(y_n)$ as shown in Figure 1.

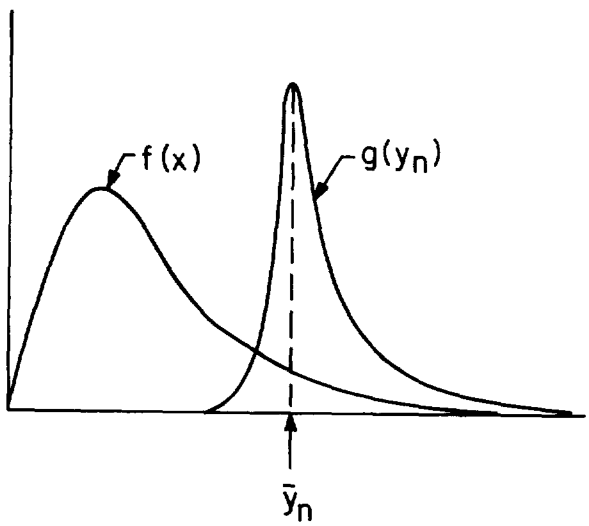

Fig. 1 Explanatory sketch of initial probability density function f(x) and extreme value probability function $g(y_n)$

That is, $\bar{Y}_n$ is given as the solution of the following equation:

$$\frac{d}{dy_n} g(y_n) = 0 \tag{3}$$

which yields,

$$f'(y_n)F(y_n) + (n-1)\{f(y_n)\}^2 = 0 \tag{4}$$

The relationship of the probable extreme value $\bar{Y}_n$ to the initial probability density function, f(x), shown in Figure 1 will be discussed in detail in the next section.

It is often assumed that waves are considered to be a narrow-band Gaussian random process and that a marine system's responses in waves are linear. Hence, the magnitude of the response follows the Rayleigh probability law which can be written in the dimensionless form by,

$$f(\xi) = \xi e^{-\frac{\xi^2}{2}} \qquad F(\xi) = 1 - e^{-\frac{\xi^2}{2}} \tag{5}$$

where, $\xi = x/\sqrt{m_o}$

x = response (amplitude)

m_o = area under the response spectrum

Thus, for the Rayleigh probability distribution, equation (4) becomes in the dimensionless form,

$$\zeta_n^2 (n e^{-\frac{\zeta_n^2}{2}} - 1) - (e^{-\frac{\zeta_n^2}{2}} - 1) = 0 \tag{6}$$

where, $\zeta_n = y_n/\sqrt{m_o}$

The second term in (6) may be neglected for a large n, and this results in the probable extreme response as,

$$\bar{\zeta}_n = \sqrt{2\,\ell n\, n} \quad \text{(Dimensionless form)}$$
$$\bar{y}_n = \sqrt{2\,\ell n\, n}\sqrt{m_o} \quad \text{(Amplitude)} \tag{7}$$

In equation (7), the extreme response is presented as a function of the number of wave encounters, n. For practical purposes, however, it may be more meaningful to express the extreme response in terms of time rather than as a function of number of wave encounters. This expression can be made by using the formulation of the average number of zero-crossing per unit time given by,

$$n = \frac{1}{2\pi}\sqrt{\frac{m_o}{m_2}} \qquad (8)$$

where, m_2 = second moment of the response spectrum

Then, the probable extreme response, $\bar{Y}_n$, is expressed as a function of time T:

$$\bar{y}_n = \sqrt{2\,\ell n\left\{\frac{(60)^2 T}{2\pi}\sqrt{\frac{m_2}{m_o}}\right\}}\,\sqrt{m_o} \qquad (9)$$

where, T = time in hours

In the foregoing derivation of the probable response, $\bar{Y}_n$, the response is assumed to be narrow-banded. If this assumption is removed, then it is extremely difficult to find a solution of (4) since the response no longer obeys the Rayleigh distribution; instead, the probability density function of the response is given as a function of the band-width parameter, ε. That is,

$$f(\xi) = \frac{2}{1+\sqrt{1-\varepsilon^2}}\left[\frac{\varepsilon}{2\pi}\, e^{-\frac{\xi^2}{2\varepsilon^2}} + \sqrt{1-\varepsilon^2}\,\xi e^{-\frac{\xi^2}{2}} \times\left\{1-\Phi\left(-\frac{\sqrt{1-\varepsilon^2}}{\varepsilon}\,\xi\right)\right\}\right] \qquad (10)$$

$$0 \leq \xi < \infty$$

where, ε = band-width parameter

$$= \sqrt{1-\frac{m_2^2}{m_o m_4}}$$

m_o = 4th moment of the spectrum

$$\Phi(u) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{u} e^{-\frac{u^2}{2}}\,du$$

The solution of equation (4) can be found approximately for the band-width parameter less than 0.9. That is, for ε < 0.9, it is possible to use the following approximation for a large number of observations, n (Ochi 1973):

$$\Phi\left(-\frac{1-\varepsilon^2}{\varepsilon}\,\zeta_n\right) \sim 0$$
$$\Phi\left(\frac{\zeta_n}{\varepsilon}\right) \sim 1 \qquad (11)$$

By neglecting terms of small order of magnitude, (4) becomes,

$$-\frac{\sqrt{1-\varepsilon^2}\,(1+\sqrt{1-\varepsilon^2})}{2}\, e^{-\frac{\xi_n^2}{2}} + n(1-\varepsilon^2)\, e^{\xi_n^2} + o(\zeta_n) = 0 \qquad (12)$$

Then, the following solution can be obtained as the dimensionless probable extreme response:

$$\bar{\zeta}_n = \sqrt{2\,\ell n\left\{\frac{2\sqrt{1-\varepsilon^2}}{1+\sqrt{1-\varepsilon^2}}\, n\right\}} \qquad (13)$$

The dimensional probable extreme response (amplitude), $\bar{Y}_n$, becomes,

$$\bar{y}_n = \sqrt{2\,\ell n\left\{\frac{2\sqrt{1-\varepsilon^2}}{1+\sqrt{1-\varepsilon^2}}\, n\right\}}\sqrt{m_o} \qquad (13\,a)$$

The dimensionless probable extreme response calculated from (13) as a function of the band-width parameter are shown in Figure 2. As can be seen in the figure, there is no significant difference in the probable extreme response for up to ε = 0.9, and the effect of ε-value of the extreme response is noticeable only for greater than 0.90, irrespective of the number of observations.

Since the range of ε-values of the marine system's response spectra spans from the order of 0.35 to 0.80, it may safely be concluded that the effect of band-width parameter can be ignored as far as the estimation of the extreme response is concerned.

Next, let us express the probable extreme response given in (13a) which is applicable for any non-narrow-band spectrum in terms of time instead of

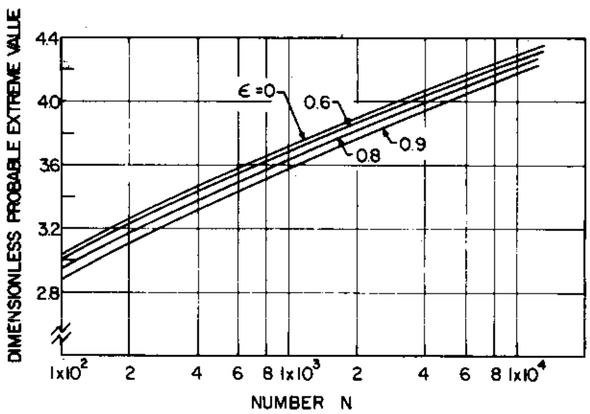

Fig. 2 Probable extreme value $\bar{\zeta}_n$ (amplitude) as a function of ε

number of observations, n. The expected number of positive maxima per unit time is given by,

$$n = \frac{1}{4\pi}\left(\frac{1+\sqrt{1-\varepsilon^2}}{\sqrt{1-\varepsilon^2}}\right)\sqrt{\frac{m_2}{m_o}} \qquad (14)$$

Then, the probable extreme response becomes,

$$\bar{y}_n = \sqrt{2\,\ell n\left\{\frac{(60)^2\,T}{2\pi}\sqrt{\frac{m_2}{m_o}}\right\}}\sqrt{m_o} \qquad (15)$$

It is noted that the above equation is exactly the same as that for narrow-band spectra given in (9). This leads to an important conclusion that the magnitudes of the extreme responses in a specified period of time are the same irrespective of the band-width parameter of a response spectrum. However, it should be remembered that the number of peaks for a non-narrow-band spectrum is larger than that for a narrow-band spectrum during the same period of time.

The probable extreme response, $\bar{Y}_n$, may be interpreted as being most likely to occur, since it is the value for which the probability density function of the extreme response, $g(y_n)$, has a peak (see Figure 1). It should be noted, however, that the chance of occurrence of a response higher than the probable extreme response, $\bar{Y}_n$, is rather high; hence it is not appropriate to consider $\bar{Y}_n$ for the design of marine systems in a seaway.

In order to amplify the above statement, let us evaluate the probability that the extreme response (in the dimensionless form) will exceed the probable extreme value. Equations (2) and (13), together with the approximation given in (11), yield,

$$\lim_{n\to\infty} \Pr\{\text{Extreme response} > \bar{\zeta}_n\} = 1 - G(\bar{\zeta}_n)$$

$$= \lim_{n\to\infty}\left[1 - \{F(\bar{\zeta}_n)\}^n\right]$$

$$= \lim_{n\to\infty}\left[1 - \left(1 - \frac{2\sqrt{1-\varepsilon^2}}{1+\sqrt{1-\varepsilon^2}}\, e^{-\frac{\bar{\zeta}_n^2}{2}}\right)^n\right]$$

$$= \lim_{n\to\infty}\left\{1 - \left(1 - \frac{1}{n}\right)^n\right\}$$

$$= 1 - e^{-1} = 0.632 \qquad (16)$$

Equation (16) implies that there is a 63.2 percent chance that the largest response will exceed $\bar{Y}_n$. This probability is extremely high; hence, it is highly desirable from a design consideration to predict the extreme response for which the probability of being exceeded is very small. In other words, choose a very small number, α, which may be called the *risk parameter*, and obtain the extreme response ζ_n (in dimensionless form) for which the following relationship holds (Ochi 1973);

$$\int_o^{\hat{\zeta}_n} g(\zeta_n)\,d\zeta_n = \{F(\hat{\zeta}_n)\}^n = 1 - \alpha \qquad (17)$$

where, $$F(\hat{\zeta}_n) = \int_o^{\hat{\zeta}_n} f(\zeta_n)\,d\zeta_n$$

Considering that α is small and n is large, we have,

$$F(\zeta_n) = (1-\alpha)^{\frac{1}{n}} \sim 1 - \frac{\alpha}{n} + o(\alpha^2) \qquad (18)$$

Then, with the aid of the assumption given in (11), we have,

$$\hat{\zeta}_n = \sqrt{2\,\ell n\left(\frac{\sqrt{1-\varepsilon^2}}{1+\sqrt{1-\varepsilon^2}}\,\frac{2n}{\varepsilon}\right)} \quad \text{for } \varepsilon \leq 0.9 \qquad (19)$$

For a narrow-band assumption $\varepsilon = 0$, (19) is reduced to,

$$\hat{\zeta}_n = \sqrt{2 \ln \frac{n}{\alpha}} \qquad (20)$$

The extreme response for design consideration with $\alpha = 0.01$ calculated from (19) and (20) is shown in Figure 3 as a function of the band-width parameter. As can be seen in the figure, there is no appreciable difference in the extreme wave amplitude for ε up to 0.9. The difference in the magnitude of extreme response is noticeable for ε greater than 0.90; however, the ε-value is usually less than 0.8 in practice, as stated earlier in connection with Figure 2. Hence, the narrow-band assumption is acceptable for the extreme response for design consideration of marine systems. Thus, analogous to equation (15), the design extreme response can be written as a function of time by,

$$\hat{y}_n = \sqrt{2 \ln\left\{\frac{(60)^2 T}{2\pi\alpha}\sqrt{\frac{m_2}{m_o}}\right\}}\sqrt{m_o} \qquad (21)$$

for amplitude

The risk parameter, α, involved in (21) is at the designer's discretion. In choosing the α-value, it is noted that the number of encounters for a marine system with a particular sea severity in the lifetime has to be considered even though the design extreme value given in (21) is applied to the short-term prediction. For example, suppose we evaluate the design extreme value in a particular sea severity with 99 percent confidence, i.e., $\alpha = 0.01$. If the system will encounter seas of this severity 20 times in her lifetime, then it is necessary to divide the risk parameter $\alpha = 0.01$ by 20 (i.e. $\alpha = 5 \times 10^{-4}$) for evaluation in order to maintain the 99 percent assurance in the prediction.

The effect of the α-value on the magnitude of design extreme values of the midship bending moment of the MARINER in seas of significant wave height 4.6 m (15 ft) and 10.7 m (35 ft) are shown in Figure 4. The family of six-parameter wave spectra consisting of eleven members in each sea severity (Ochi 1976) is used in this computation, and the largest values in each sea severity are plotted in the figure. The solid circles are plotted in the figure are the extreme values evaluated with $\alpha = 0.01$ but taking into consideration of the number of occurrences of each sea state in the North Atlantic in 30 years. The figure indicates that the design extreme values do not increase substantially with increasing α-value.

It is of interest to note the effect of time duration on the magnitude of extreme values. Figure 5 shows an example of the effect of ship operation

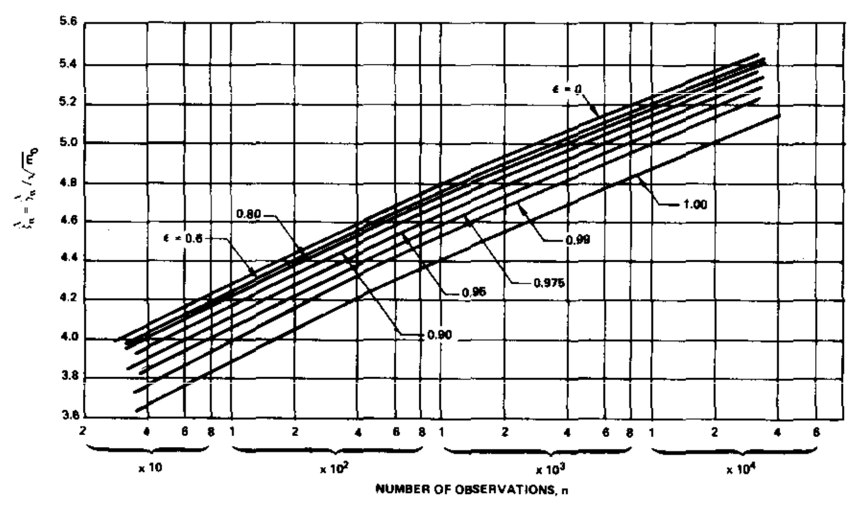

Fig. 3 Extreme wave $\hat{\zeta}_n$ (amplitude) for design consideration as a function of number of observations for various band-width parameter ε ($\alpha = 0.01$) (from Ochi 1973)

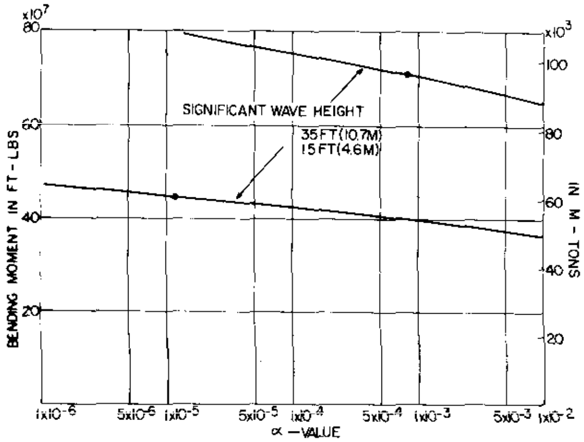

Fig. 4 Design value of the midship bending moment of the MARINER as a function of coefficient α (from Ochi 1977)

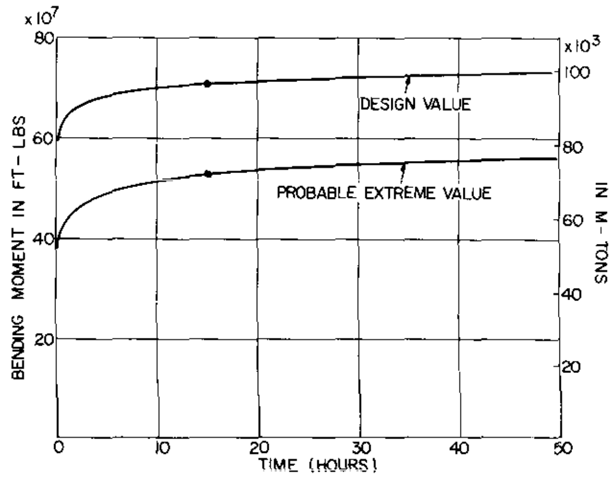

Fig. 5 Probable extreme value and design value of the midship bending moment of the MARINER in sea of significant wave height 35 ft (10.7 M) as a function of ship operation time (from Ochi 1977)

time on extreme values. The figure pertains to a sea of significant wave height of 35 ft (10.7 m), and the upper bound values using the six-parameter wave spectrum are plotted. As can be seen in the figure, the magnitude of extreme values, both probable as well as the design extreme values, increase significantly during the first several hours and thereafter increase very slowly with time. This is the general trend of the extreme values in all sea severities.

An example of the design extreme values of the wave-induced bending moment for various sea severities are shown in Figure 6. The computations are made on the MARINER, and the curves shown in the figure are the upper bound curves obtained by using the two-parameter and six-parameter families of wave spectra. As can be seen in the figure, the results of the computation using the two families of wave spectra agree well up to seas of significant height of 30 ft (9.2 m). For seas of significant wave height over 30 ft (9.2 m), however, the design extreme values evaluated using the two-parameter spectrum are higher than those using the six-parameter spectrum. It is of interest to note here that Russo and Sullivan (1953) gives the values of hogging bending moment of 85.3×10^7 ft-lbs (116×10^3 m-tons) calculated following the classic standard procedure of assuming the ship to be statically supported on a trochoidal wave equal in length to the length of the ship and having a height equal to one-twentieth

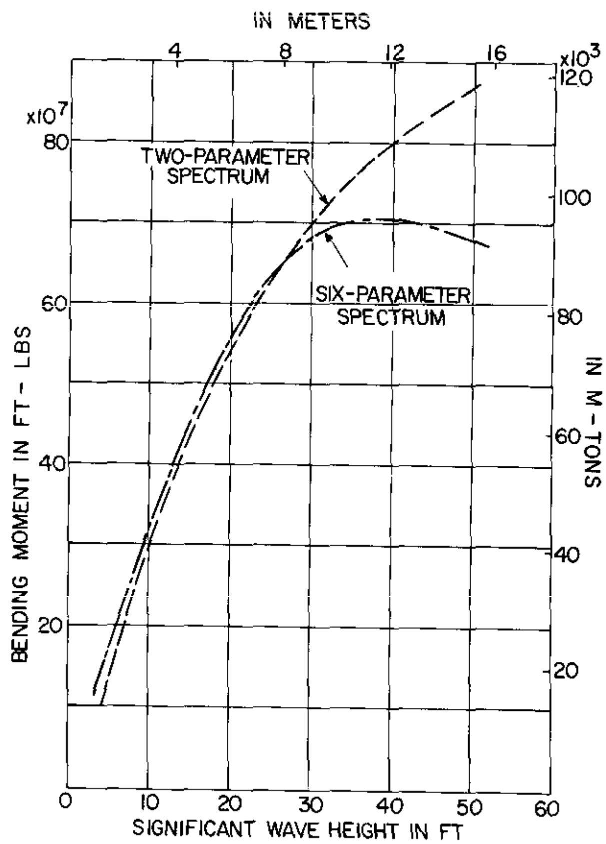

Fig. 6 Design value of the midship bending moment of the MARINER (from Ochi 1977)

of the length.

APPROXIMATE ESTIMATION OF EXTREME RESPONSES

In the previous section, exact information of the initial probability distribution was necessary for estimating the extreme response by applying order statistics. In practice, however, the information for the initial distribution is often not precisely known. For example, the probability distribution of the long-term response of a marine system is not known. It may be the Weibull distribution or the log-normal distribution. However, there is no scientific basis for selecting any specific probability distribution function to characterize the long-term responses as contrasted to the short-term responses. In other words, we have to estimate the extreme responses from the accumulation of the observed (or computed) data over a sufficiently long period of time without precise information of the initial probability distribution.

One way to overcome this difficulty is to evaluate the extreme response by an approximate method that is applicable for any probability distribution if certain conditions are met. The principle of this approximation method is as follows:

Let the initial cumulative distribution function, F(x), be in the form of

$$F(x) = 1 - e^{-q(x)} \tag{22}$$

where, q(x) is a positive real-valued function that satisfies the conditions required for F(x) being a cumulative distribution function. Then, equation (4) becomes,

$$\frac{q''(y_n)}{\{q'(y_n)\}^2}\left\{1 - e^{-q(y_n)}\right\} + n\,e^{-q(y_n)} - 1 = 0 \tag{23}$$

Since the first term is small in comparison with other terms for large n, (23) yields,

$$e^{-q(y_n)} = \frac{1}{n} \tag{24}$$

Thus, the probable extreme response, $\bar{Y}_n$, for large n is given by,

$$\bar{y}_n = q^{-1}(\ell n\, n) \tag{25}$$

where, $q^{-1}(\ )$ is the inverse function of q(x).

For evaluating the extreme response in practice, however, it is not necessary to know the function q(x) in (22). As can be seen from (22) and (24), the probable extreme response for large n is given as the x-value in (22) for which the probability of exceeding x is equal to 1/n. That is,

$$1 - F(\bar{y}_n) = \frac{1}{n} \tag{26}$$

Equation (26) implies that the probable extreme response expected to occur in n-observations can be evaluated from the initial cumulative distribution and the number of wave encounters involved. Although the initial cumulative distribution function is not precisely known, the function can be constructed from the observed (or computed) data. For this, the data are often fitted by some known distribution such as a Weibull distribution or log-normal distribution, etc., and the extreme value in a desired period of time is estimated based on this presumed distribution. Sometimes, the extreme value is determined by simply extending the cumulative distribution function without fitting the data to a particular probability distribution function.

Although many examples are available for estimating the extreme responses of marine systems by applying this approximation method, let us consider the estimation of the extreme sea severity (significant wave height) using the data presented in Table 2 (Bouws 1978). The table shows the frequency of occurrence of significant wave height measured in the North Sea. A total of 5,412 measurements were made in 3 years.

Figure 7 shows the cumulative distribution function of the significant wave height plotted on the log-normal probability paper, while Figure 8 shows the same data plotted on the Weibull probability paper. It appears from comparison of Figures 7 and 8, that the data are better represented by the Weibull distribution, and hence the estimation of the extreme significant wave height in 50 years or 100 years is made most commonly by extending the

Table 2 Significant wave height data obtained from measurements in the North Sea (53.5°N, 4°E) (from Bouws 1978)

SIGNIFICANT WAVE HEIGHT (M)	NUMBER OF OBSERVATIONS
0 – 0.5	1,280
0.5 – 1.0	1,549
1.0 – 1.5	1,088
1.5 – 2.0	628
2.0 – 2.5	402
2.5 – 3.0	192
3.0 – 3.5	115
3.5 – 4.0	63
4.0 – 4.5	38
4.5 – 5.0	18
5.0 – 5.5	21
5.5 – 6.0	7
6.0 – 6.5	8
6.5 – 7.0	2
7.0 – 7.5	1
TOTAL	5,412 in 3 Years

straight line given in Figure 8.

However, extreme care has to be given in interpreting the data plotted on the probability paper. If we take a close look at the data plotted in Figures 7 and 8, the data are not satisfactorily represented over the entire range of significant wave

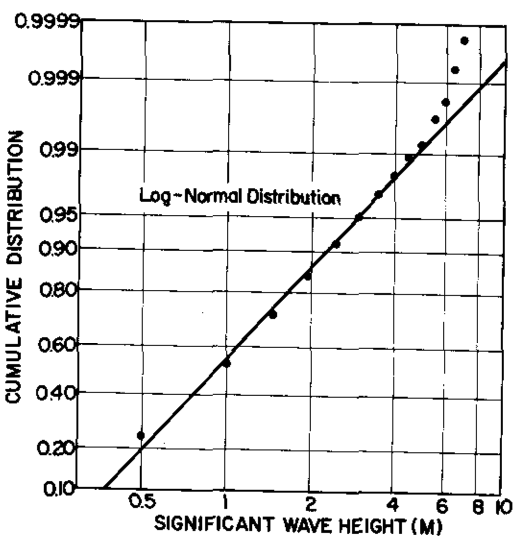

Fig. 7 Cumulative distribution of significant wave height plotted on log-normal probability paper (data from Bouws 1978)

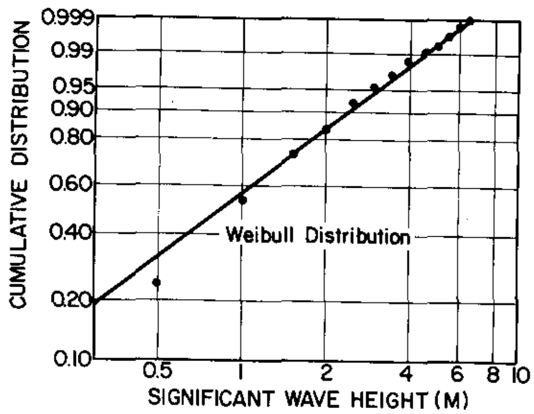

Fig. 8 Cumulative distribution function of significant wave height plotted on Weibull probability paper (data from Bouws 1978)

height by either the log-normal probability distribution or the Weibull probability distribution. That is evident in Figure 9 in which the comparison between the histogram and the two probability distribution functions are shown.

As can be seen in Figure 9, the extreme significant wave height may be estimated from neither the log-normal or Weibull distribution but from the extension of data points by applying

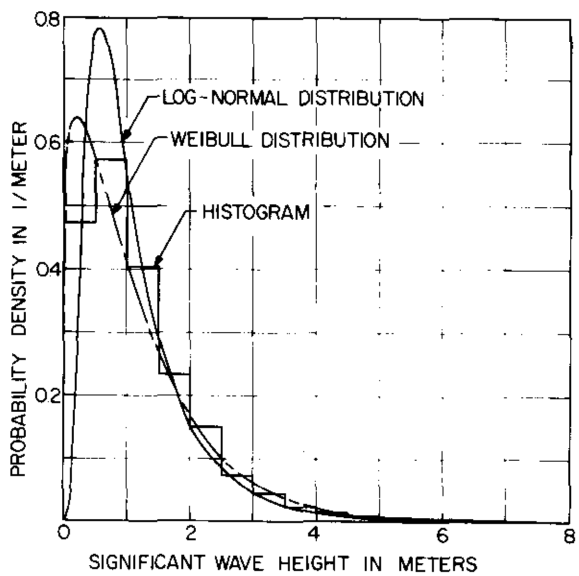

Fig. 9 Comparison between histogram of significant wave height and log-normal and Weibull probability density functions (data from Bouws 1978)

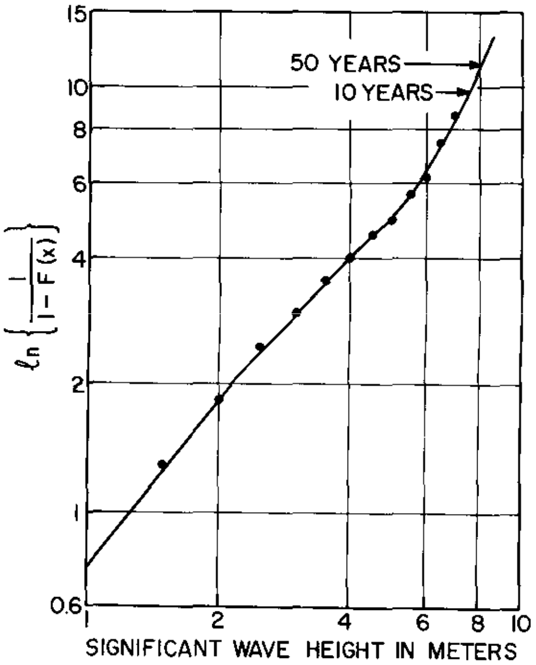

Fig. 10 Estimation of extreme significant wave height from the extension of data points (data from Bouws 1978)

equation (26). The extension of data point is shown in Figure 10. The vertical scale of the figure is the inverse of the lefthand side of (26), which is often called the *return period*, in logarithmic form. Since the number of measurements of significant wave height is 5,412 in three years, the number of significant wave heights expected in ten years will be 18,040. By taking the logarithms of this number, it is obtained from the figure that the extreme significant wave height expected in 10 years will be 7.5 meters. Similarly, the extreme significant wave height expected in 50 years can be estimated as 8.0 meters.

The approximate method presented in the foregoing can be applicable for any probability distribution; however, the method has a drawback. That is, the extreme wave height expected to occur in the future is determined by extending the line, taking into account the higher wave heights that are extremely unreliable data. For a more precise estimation, it is highly desirable to extend the cumulative distribution function representing the data over the entire range of the values. Emphasis should not be given to the representation of data points higher than the cumulative distribution of 0.99.

To achieve a precise representation of the data, a function q(x) in (22) may be expressed as a combination of an exponential and a power of the wave height (Ochi and Whalen 1980):

$$q(x) = a\,x^{m}\,e^{-px^{k}} \qquad (27)$$

The parameters in q(x) are determined numerically by a nonlinear least squared fitting procedure. The form used in this minimization procedure is given by,

$$G = \ell n\{-\ell n\,(1-F)\} = \ell n\,a + m\,\ell n\,x - px^{k} \qquad (28)$$

The parameters are optimized such that the sum of the squared values of the difference between G in (28) and the corresponding data values becomes minimal. Once the parameter values are determined, the extreme wave height can be evaluated from (26). This method is applied to significant wave height data shown in Table 2. For this example, values of four parameters involved in (27) are obtained as,

$a = 0.980$

$m = 1.101$

$p = 0.181$

$k = -1.328$

The cumulative distribution function obtained by using these values are plotted on log-normal probability paper as shown in Figure 11, together with data points. A similar presentation plotted on Weibull paper is shown in Figure 12. As can be seen in these figures, this cumulative distribution function represents very well the data over the entire range of the values, and therefore it can be used for estimating the extreme significant wave height. The result is shown in Figure 13. The comparison between the extreme significant wave height thus estimated and those estimated based on the log-normal and the Weibull probability distributions shows that the log-normal probability distribution substantially overestimates the extreme significant wave height, while the Weibull probability distribution underestimates it on the order of 10 percent for this example.

For design consideration of a marine system, a certain margin above the probable extreme value is required, and this can be obtained following the same concept as was discussed in the development of (20). That is, by choosing a small α, the extreme value for design

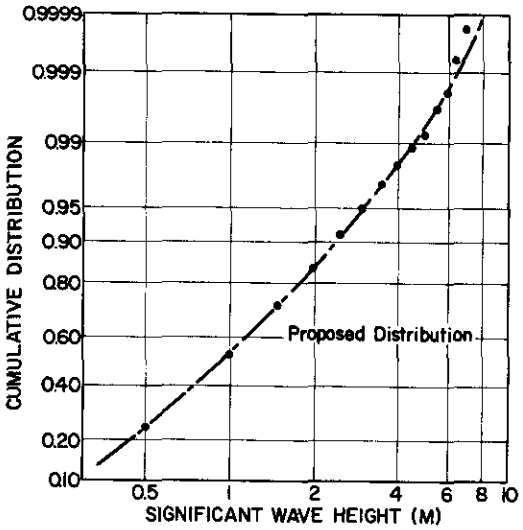

Fig. 11 Cumulative distribution function based on Equation (27) plotted on log-normal probability paper (data from Bouws 1978)

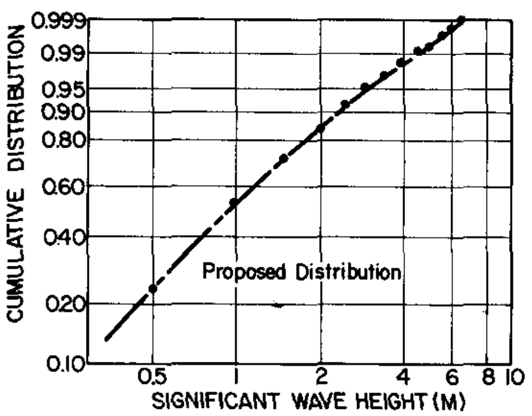

Fig. 12 Cumulative distribution function based on Equation (27) plotted on Weibull probability paper (data from Bouws 1978)

consideration, $\hat{Y}_n$, can be obtained from

$$1 - F(\hat{y}_n) = \frac{\alpha}{n} \tag{29}$$

As an application of the approximate methods discussed in this section, let us evaluate the design extreme values of the transverse wave-induced force acting on the cross-beam bridge structure on a semi-submersible offshore platform shown in Figure 14. For evaluating the lifetime response of the ocean platform, the following various parameters have to be considered:

(i) Frequency of occurrence of various sea severities, P_i
(ii) Frequency of occurrence of

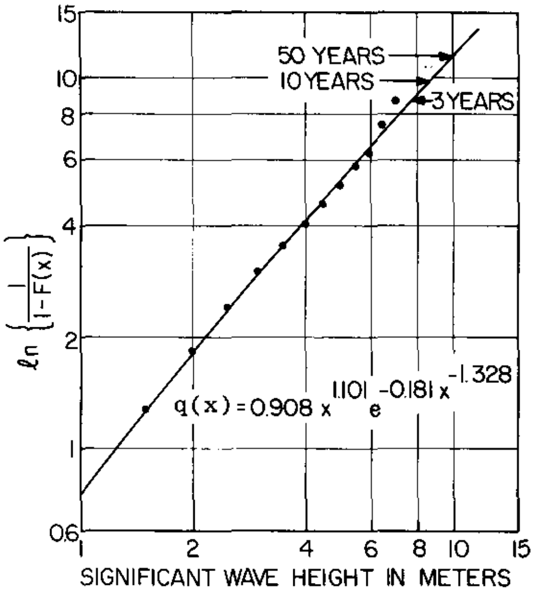

Fig. 13 Prediction of extreme wave height by using the probability function based on Equation (27) (data from Bouws 1978)

various heading to waves, P_j
(iii) Frequency of encounter with various wave spectral shapes, P_k
(iv) Expected number of cycles of response for each given sea, heading, and wave spectral shape.

Then, the probability density function for the long-term response can be written as follows:

$$f(x) = \frac{\sum_i \sum_j \sum_k n_* p_i p_j p_k f_*(x)}{\sum_i \sum_j \sum_k n_* p_i p_j p_k} \tag{30}$$

where, $f_*(x)$ is the probability density function for the short-term response, n_* is the average number of responses per unit time of the short-term response.

The total number of responses expected in the lifetime is given by,

$$n = \sum_i \sum_j \sum_k n_* p_i p_j p_k \, T \times (60)^2 \tag{31}$$

where, T is the total exposure time to sea.

Computation of the extreme response given in (30) is carried out by using the family of six-parameter wave spectra with the risk parameter $\alpha = 0.01$. Since the magnitude of the transverse force acting on the cross-beam structure is maximum in beam seas, it may be of interest to see the difference, if any,

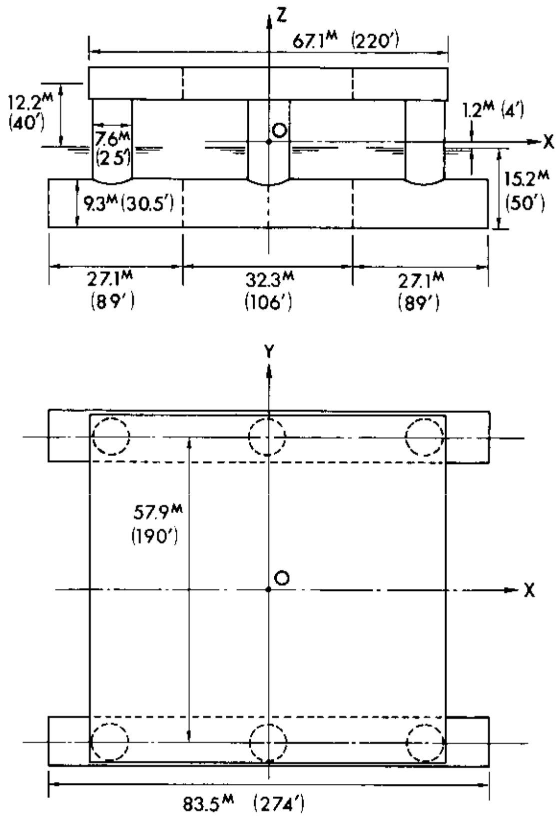

Fig. 14 Semisubmersible-type ocean platform used for computations

between the predicted long-term design extreme values including all headings, sea severities, and spectral shapes and those predicted in various sea severities and spectral shapes in beam seas only. The significant difference in these two long-term computations is the number of responses expected in the lifetime; 6.146 x 10^7 for all headings versus 2,078 x 10^7 for beam seas only in 20 years operation.

Figure 15 shows the design extreme values evaluated for these two long-terms (Ochi and Wang 1979). As can be seen in the figure, the design extreme value evaluated from the lifetime probability distribution including all headings and that evaluated from the beam seas only are both 4,800 tons, if the difference in the number of encounters for each case is taken into consideration.

Next, the extreme responses evaluated through the long-term prediction approach will be compared with those evaluated through the short-term prediction approach. It may appear that estimation through the long-term prediction approach is superior to that through the short-term prediction approach, since it deals with the accumulation of all responses. However, in reality, the long-term estimation includes a considerable percentage of small responses in relatively mild seas, which do not contribute to the extreme design value. The results of computations have shown that the design extreme value obtained through the short-term approach is 4,700 tons (Ochi and Wang 1976) in seas of significant wave height of 16 m (52.5 ft.), the value is very close to that obtained through the long-term approach. As this example shows the predicted extreme values through the long-term and short-term approaches are nearly equal if the difference in the number of encounter with waves for each case is taken into consideration. It is noted that the estimation procedure of the extreme values through the short-term prediction approach is extremely simple in comparison with that through the long-term approach. Hence, the short-term prediction approach appears to be adequate as far as estimation of extreme values is concerned.

ESTIMATION OF EXTREME RESPONSES BY ASYMPTOTIC FORMULAE

As was stated in the Introduction, it is possible to predict the largest responses expected to occur in the lifetime of a marine system from the measured (or observed) maxima. The measured (or observed) maxima defined here is the largest value that is measured (or observed) during a certain period of time; every 6 hours, every 12 hours, a day, etc. The estimation is based on the asymptotic distribution of the extreme values developed by Gumbel (1958). The significant feature of the asymptotic distribution is that the probability function for the maxima

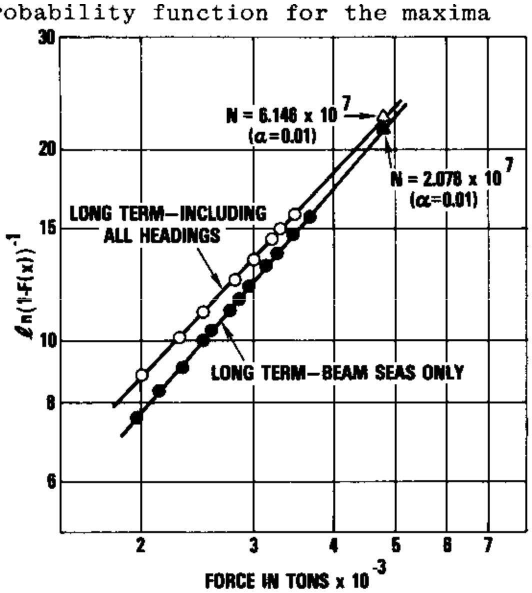

Fig. 15 Comparison of design extreme values of the transverse force computed including all headings to waves and that computed in beam seas only (from Ochi and Wang, 1979)

reduces to the same type irrespective of the initial probability distribution. The underlying principle of the asymptotic distribution is as follows:

Let us assume that the initial cumulative distribution function is in the form given in (22). By using the relationship given in (24), we can write the initial cumulative distribution function as,

$$F(x) = 1 - \frac{1}{n} e^{q(\bar{y}_n) - q(x)} \tag{32}$$

Then, from the definition given in (1), the cumulative distribution function of the extreme value for large n becomes,

$$G(y_n) = \lim_{n\to\infty} \left\{1 - \frac{1}{n} e^{q(\bar{y}_n) - q(x)}\right\}^n$$
$$= e^{-\exp\{q(\bar{y}_n) - q(y_n)\}} \tag{33}$$

Since the probability distribution function of extreme values, Y_n, is much more concentrated around its modal value, $\bar{Y}_n$, in comparison with the initial probability density function, the term $q(\bar{y}_n) - q(y_n)$ may be expanded by the Taylor series. Then, by neglecting higher order terms, (33) becomes

$$G(y_n) = e^{-\exp\{-q'(\bar{y}_n)(y_n - \bar{y}_n)\}} \tag{34}$$

Here, neither $q'(y_n)$ nor $\bar{y}_n$ are known in reality; hence, let us express (34) in the following form,

$$G(z) = e^{-\exp\{-z\}} \tag{35}$$

where,

$$Z = q'(\bar{y}_n)(y_n - \bar{y}_n) \tag{36}$$

The mean and variance of the random variable Z can be obtained from (35) as,

$$E[z] = \gamma \text{ (Euler's constant, 0.577)}$$
$$Var[z] = \pi^2/6 \tag{37}$$

Then, from (36) and (37), we can derive the following relationship

$$q'(\bar{y}_n) = \frac{\pi/\sqrt{6}}{\sqrt{Var[y_n]}}$$
$$\bar{y}_n = E[y_n] - \frac{\gamma}{q'(\bar{y}_n)} \tag{38}$$

Since the mean and variance, $E[y_n]$ and $Var[y_n]$, respectively, can be obtained from the data, the cumulative distribution function of the extreme value given in (34) can be evaluated with the aid of (38).

In summary, the cumulative distribution function of the extreme value can be written in the following form:

$$G(y_n) = e^{-e^{-\alpha(y_n - u)}} \qquad -\infty < y_n < \infty \tag{39}$$

where,

$$\alpha = \frac{\pi/\sqrt{6}}{\sqrt{Var[y_n]}}$$
$$u = E[y_n] - \frac{\sqrt{6}}{\pi}\gamma\sqrt{Var[y_n]}$$

Equation (39) is the so-called Gumbel's *Asymptotic extreme value distribution.*

The method to estimate the extreme responses expected in the lifetime of a marine system from equation (39) is the same procedure as discussed in connection with equations (26) and (29). Since the concept of the asymptotic distribution of the extreme value deals with the maxima, Y_n, observed in certain period of time, the cumulative distribution function, $G(y_n)$, given in (39) can be considered as the initial probability distribution. Hence, the lifetime probable extreme value is determined as the $\bar{y}_n$-value which satisfies

$$1 - G(\bar{y}_n) = \frac{1}{n} \tag{40}$$

where, n is the number of the maxima expected to occur in the lifetime, and it is estimated from the observed data (see the example shown in Figure 18). Similarly, the design value with the risk parameter, α , can be evaluated as the y_n-value which satisfies

$$1 - G(\hat{y}_n) = \frac{\alpha}{n} \qquad (41)$$

The sample space of the distribution given in equation (39) is unlimited; however, there is another distribution called *Asymptotic extreme value distribution with upper limit* which may also be applicable for the analysis of observed maxima. The distribution is given by (Gumbel 1958),

$$G(y_n) = e^{-\left(\frac{w - y_n}{w - v}\right)^k} \qquad -\infty < y_n < w \qquad (42)$$

where, w = upper limit

k = positive constant linked to the limit value, w

v = y_n-value for which $G(y_n) = e^{-1} = 0.368$.

The derivation of this asymptotic distribution will not be given here; however, it may suffice to state that it is derived by applying the Taylor series expansion to the initial probability distribution function in the neighborhood of w.

The values of the two parameters, w and k, can be determined from the following two moments of Y_n:

$$E[y_n] = w - (w - v)\,\Gamma\left(1 + \frac{1}{k}\right)$$

$$E[y_n^2] = w^2 - 2w(w - v)\,\Gamma\left(1 + \frac{1}{k}\right) + (w - v)^2\,\Gamma\left(1 + \frac{2}{k}\right) \qquad (43)$$

The values of w and k can also be determined graphically by choosing the cumulative distribution function, $G(y_n)$, for two different y_n-values. It should be noted that the upper limit value, w, thus determined from the data is merely the limit value of y_n for $n \to \infty$, which is unrealistically large value, and hence it cannot be used for design consideration.

The probable extreme value as well as the extreme value for design should be evaluated by the same procedure as presented in connection with the asymptotic extreme value distribution with unlimited sample space.

The method to estimate the lifetime design value through the asymptotic extreme value formulae is applied to the results of extensive full scale trials carried out on eight SL-7 vessels on SEA-LAND service (Fair and Booth 1979). During these trials, the strain gage recorder was installed at midship of each ship, and the maximum peak to maximum trough stresses which occurred during every 4-hour sampling period were recorded.

As an example of analysis which can be made from these data, the extreme responses expected in the lifetime of the SL-7 will be estimated based on the following two data:

(a) One-year data obtained on SEA-LAND MCLEAN in the Atlantic. A total of 1,802 observations.

(b) Five-data-years obtained in the Atlantic. A total of 12,319 observations.

It is noted that the results of analysis have shown that the responses observed in the Atlantic service routes are larger than those in the Pacific service routes. Therefore, the design extreme value may be estimated from analysis of data in the Atlantic.

Figures 16 and 17 show the comparisons of the cumulative distribution functions of these data and the asymptotic extreme value formulations given in (39) and (42). Figure 16 shows the comparisons using the one-year data, while Figure 17 shows those for the five-data years.

The lifetime extreme responses can be estimated from the relationship given in equation (40) together with the information on the number of observations. The results are shown in Figures 18 and 19. As can be seen in these figures, the lifetime extreme response of the SL-7 estimated from the one-year data is very close to that estimated from the five-data-years (in which many sister ships' data are involved), if the difference in the number of observations in each case is taken into consideration.

For instance, from equation (39), the design extreme value (peak-to-trough stress) in 30 years with the risk parameter $\alpha = 0.01$ is 64.0 kpsi (45.4 kg/mm^2) from the one-year data as compared with 60.0 kpsi (42.6 kg/mm^2) from the five-data-years.

It can also be seen in these figures that the extreme design estimate based on the asymptotic formulation with upper bound is approximately 17 percent

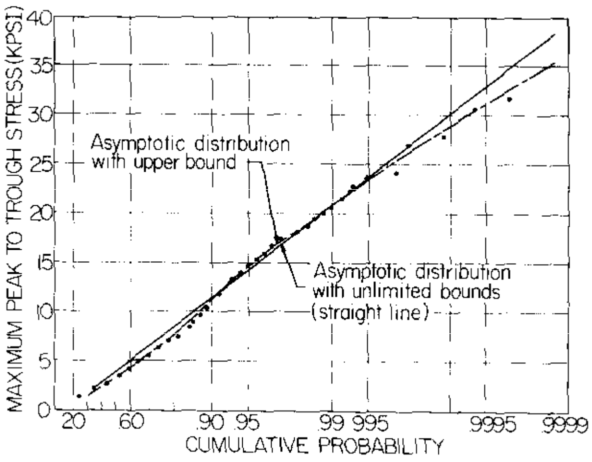

Fig. 16 Comparison between asymptotic distribution with unlimited bounds and that with upper bound using SL-7 one-year data

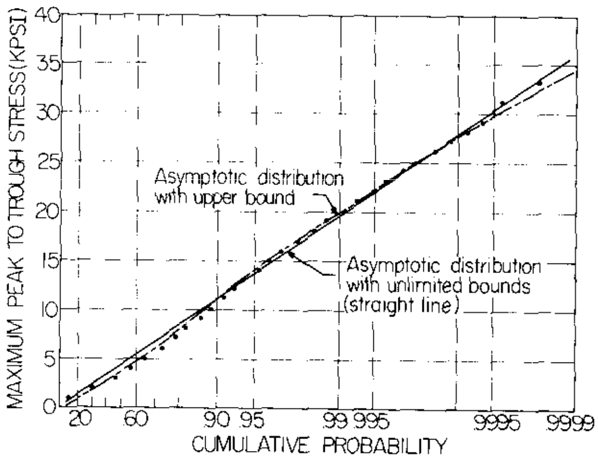

Fig. 17 Comparison between asymptotic distribution with unlimited bounds and that with upper bound for SL-7 five-data years

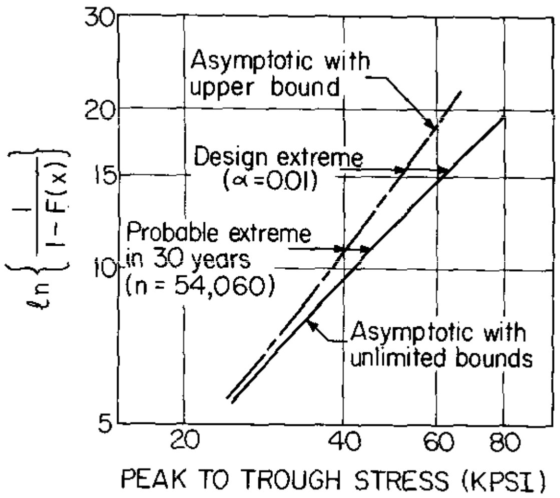

Fig. 18 Evaluation of the SL-7 lifetime (30 years) probable extreme value and design extreme value based on the one-year data

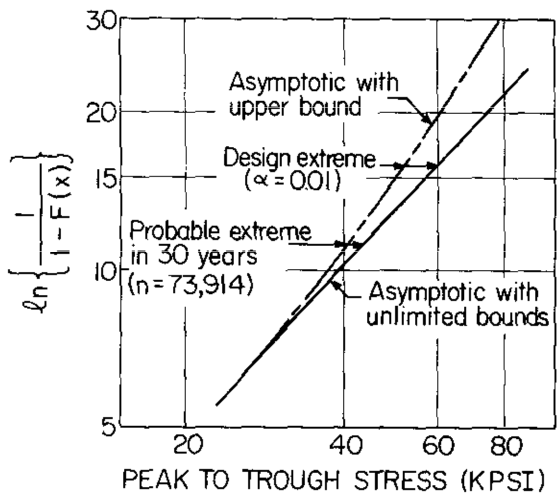

Fig. 19 Evaluation of the SL-7 lifetime (30 years) probable extreme value based on the five-data-years

less than that estimated based on the asymptotic formulation with unlimited bound.

CONCLUSION

This paper presents the theoretical background of recent techniques to predict extreme values, specifically extreme loads and responses of marine systems in a seaway. There are three different approaches to evaluate extreme values depending on the information (data) available in the prediction. These are (a) the exact evaluation, (b) the approximate estimation, and (c) the estimation by using the asymptotic extreme value formulae. The underlying principle and practical application of each approach are discussed. From the results of numerical examples, the following conclusions are drawn:

One way to evaluate the design extreme value of the responses of a marine system is to calculate the extreme response through the short-term approach by using the family of wave spectra in each sea. The computations should be made for the severest loading and heading of the response, at the attainable highest speed of the system, if applicable, in each sea. From the results of the largest response in each sea severity, the design extreme value can be obtained with a specified risk parameter, α.

In the short-term prediction, the effect of the spectral band-width parameter on the magnitude of extreme values appears to be negligibly small irrespective of the number of observations.

The extreme values of the long-term response of a marine system can be evaluated through the approximate estimation method. In this case, the estimation can be made based on the cumulative distribution function which represents the data over the entire range of values. Emphasis should not be given to the representation of data points higher than the cumulative distribution of 0.99.

The extreme value through the long-term and short-term approaches are nearly equal if the difference in the number of encounters with waves for each case is taken into consideration. It is noted that the estimation procedure of the extreme values through the short-term prediction approach is extremely simple in comparison with that through the long-term approach.

The lifetime extreme response of a marine system can be estimated from data of the observed maxima by applying the asymptotic extreme value formulae. The lifetime (30 years) extreme response of the SL-7 estimated from the one-year data is very close to that estimated from the five-data-years (in which many sister ships' data are involved), if the difference in the number of observations in each case is taken into consideration.

REFERENCES

Bouws, E. (1978): "Wind and Wave Climate in the Netherlands Sector of the North Sea Between 53 and 54 North Latitude," Koninklijk Netherlands Meterologisch Institute, Report W.R. 78-9.

Gumbel, E.J. (1958): "Statistics of Extremes," Columbia University Press, New York.

Fain, R.A. and Booth, E.T. (1979): "Results of the First Five Data-Years of Extreme Stress Scratch Gauge Data Collected Abroad SEA-LAND'S SL-7". Ship Structure Committee, SSC-286.

Ochi, M.K. (1973): "On Prediction of Extreme Values," Journal of Ship Res., Vol 17, No. 1, pp 29-37.

Ochi, M.K. and Wang, S. (1976): "Prediction of Extreme Wave-Induced Loads on Ocean Structures," Proc. Sympo. Behavior of Offshore Structures, Vol 1, pp 170-186.

Ochi, M.K. and Hubble, E.N. (1976): "On Six-Parameter Wave Spectra," Proc. 15th Conf. Coastal Eng., Vol 1, pp 301-328.

Ochi, M.K. (1977): "Concept of Probabilistic Values and its Application to Ship Design," Proc. Symp. Practical Design in Shipbuilding, pp 51-57.

Ochi, M.K. and Wang, S. (1979): "The Probabilistic Approach for the Design of Ocean Platforms," Proc. Conf. Probabilistic Mechanics and Structural Reliability, Ame. Soc. Civil Eng. pp 208-213.

Ochi, M.K. and Whalen, J.E. (1980): "Prediction of the Severest Significant Wave Height," 17th Conf. Coastal Eng., Sydney, Australia (in Printing).

Russo, V.L. and Sullivan, E.K. (1953): "Design of the MARINER-Type Ship," Trans. Soc. Nav. Arch. & Mar. Eng., Vol 61.