CHAPTER 38

NEW APPROACH FOR ESTIMATING THE SEVEREST SEA STATE FROM STATISTICAL DATA

Michel K. Ochi*

Abstract

This paper discusses a feature of the generalized gamma distribution which is particularly appropriate for statistical analysis of long-term significant wave height data. The agreement between the cumulative distributions of the data and the generalized gamma distribution is shown to be satisfactory. Methods to estimate the probable extreme sea state (significant wave height) expected in a specified time period as well as the extreme sea state for design consideration of marine systems are presented.

Introduction

Probabilistic estimation of the extreme sea state expected in 50 or 100 years provides information vital for the design of offshore and nearshore structures as well as for the stochastic analysis of various coastal processes such as wave-induced sediment transport.

Sea severity as evaluated from wave height measurements depends to a great extent on the geographical location where the data are obtained, since the crucial factors for sea severity are the frequency of occurrence of storms, water depth and fetch length. In addition, sea severity depends on the growth and decay stage of a storm even though wind speed is the same. Thus, there is no scientific basis for selecting a specific probability distribution function to represent the statistical distribution of sea state (significant wave height). Because of this, various probability distribution functions have been proposed which appear to best fit particular sets of observed data. These include (a) log-normal distribution [Ochi 1978a], (b) modified

---

* Professor, Coastal & Oceanographic Engineering Department 336 Weil Hall, University of Florida, Gainesville, Florida, 32611, U.S.A.

log-normal distribution [Fang and Hogben 1982], (c) three-parameter Weibull distribution [Burrows and Salih 1986] [Mathisen and Bitner-Gregersen 1990], (d) combined exponential and power of significant wave height [Ochi and Whalen 1980] and (e) modified exponential distribution [Thompson and Harris 1972], etc.

It is highly desirable that a more rational probability distribution be developed so that sea severity evaluated from data obtained anywhere in the world can be reasonably represented and analyzed by a specific distribution thereby permitting a direct comparison of data, including extreme values. To achieve this goal, this paper introduces a probability distribution called the generalized gamma distribution, and discusses a feature of the distribution which is particularly appropriate for analysis of long-term significant wave height data.

Statistical Trend of Long-Term Significant Wave Height Data

Prior to introducing a probability distribution to represent long-term significant wave height data, it may be well to examine the general trend of the statistical distribution. For this, significant wave height data obtained at various geographical locations as well as various water depths are analyzed in the present study.

Figures 1(a) through 1(g) show the cumulative distribution functions of significant wave height plotted on log-normal probability paper. These data were obtained at locations: (a) Norwegian coast [Mathisen and Bitner-Gregersen 1990], (b) North Sea [Bouws 1978], (c) North Pacific off Japan [Tomita 1988], (d) North Pacific off Canada [National Data Buoy Center 1990], (e) Atlantic Ocean off Georgia [National Data Buoy Center 1990], (f) Florida East Coast (shallow water area) [Coastal Data Network 1990] and (g) Gulf of Mexico (shallow water area) [Work 1992], respectively. Included also in these figures is a straight line which represents the cumulative distribution function by fitting the following log-normal probability distribution:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left\{-\frac{1}{2}\left(\frac{\ell n x - \mu}{\sigma}\right)^2\right\}, \qquad 0 \le x < \infty \tag{1}$$

As can be seen in Figure 1, the cumulative distributions of all significant wave height data show a consistent trend irrespective of geographical location and water depth. That is, at least 90 to 95 percent of each set of data is well represented by the log-normal probability distribution; however, the data diverge from the log-normal distribution for large significant wave heights which are extremely critical for estimating extreme values. The divergence is always consistent in such a way that the cumulative distribution function of the data converges to unity faster than that of the log-normal distribution. This implies that the log-normal distribution will overestimate the extreme significant wave height by a substantial amount.

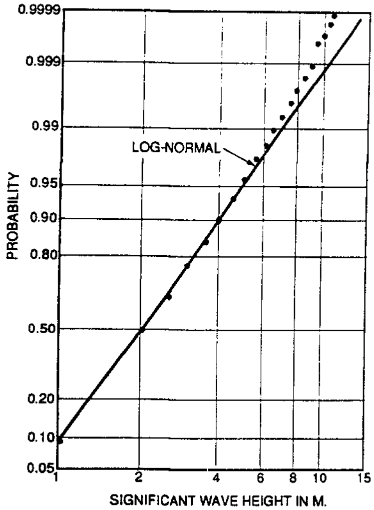

Figure 1(a): Long-term significant wave height data obtained off Norwegian coast plotted on log-normal probability paper (Data from Methisen & Bitner-Gregersen 1990)

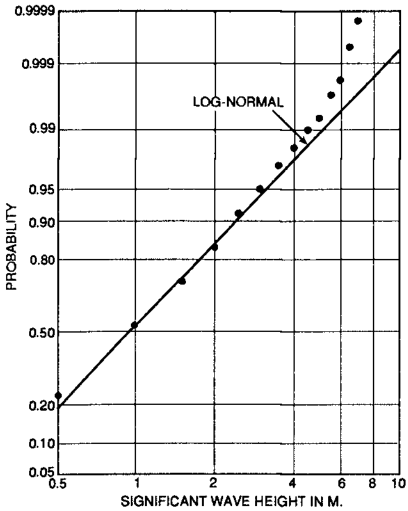

Figure 1(b): Long-term significant wave height data obtained in the North Sea plotted on log-normal probability paper (Data from Bouws 1978)

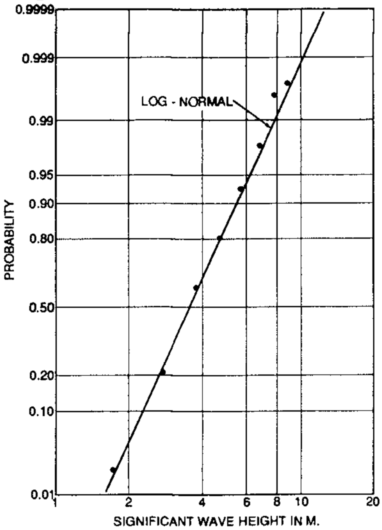

Figure 1(c): Long-term significant wave height data obtained in the North Pacific Ocean off Japan plotted on log-normal probability paper (Data from Tomita 1988)

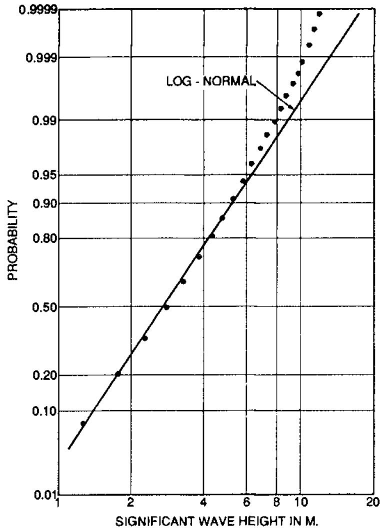

Figure 1(d): Long-term significant wave height data obtained in the North pacific Ocean off Canada plotted on log-normal probability paper (data from Nat.Data Buoy Center 1990)

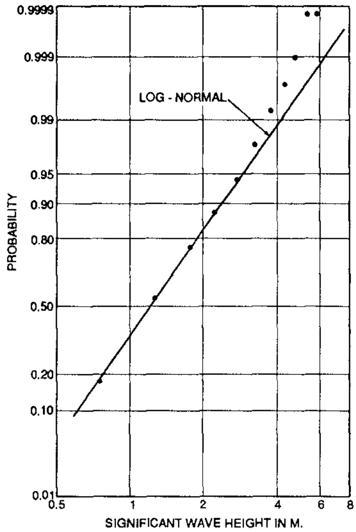

Figure 1(e): Long-term significant wave height data obtained in the Atlantic Ocean off Georgia plotted on log-normal probability paper (Data from Nat.Data Buoy Center 1990)

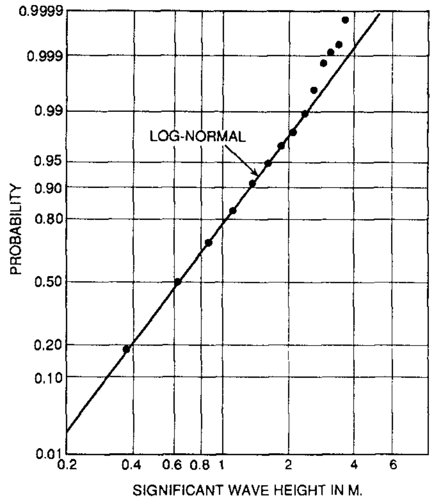

Figure 1(f): Long-term significant wave height data obtained off Florida East Coast plotted on log-normal probability paper (Data from Coastal Data Network 1990)

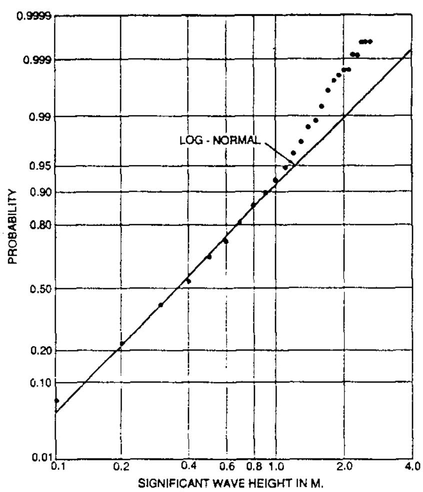

Figure 1(g): Long-term significant wave height data obtained in the Gulf of Mexico plotted on log-normal probability paper (Data from Work 1992)

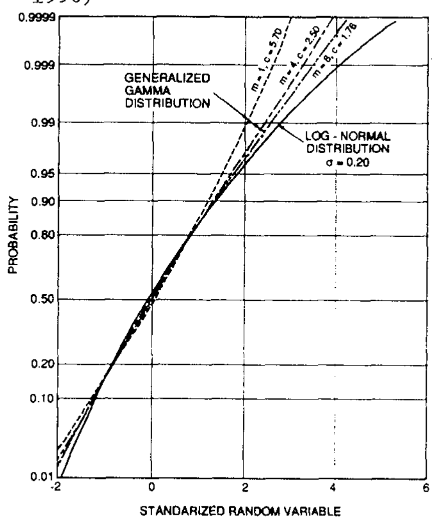

Figure 2: Comparison bewteen the standardized cumulative distributions of (a)log-normal and (b) generalized gamma distribution for σ =0.20 and for various m-values

We may conclude from the results shown in Figure 1 that the statistical characteristics of long-term significant wave height can best be represented by a probability distribution whose cumulative distribution is by and large close to that of the log-normal distribution but converges to unity much faster than the log-normal distribution at higher values, say over 0.95.

Generalized Gamma Probability Distribution

An extensive search was made to find an appropriate probability distribution which represents the cumulative distribution of significant wave height data. For this, various probability distribution functions were standardized so that a comparison of distributions could be made under the uniform condition of zero-mean and unit variance. Here, standardization was achieved through a change of random variables by subtracting the mean and dividing by the standard deviation of the original random variable. Let us define the standardized random variable as Z and its probability density function f(z). The standardized log-normal distribution is given by

$$f(z) = \frac{1}{\sqrt{2\pi}\,\sigma} \frac{\sqrt{\alpha^2-1}}{\sqrt{\alpha^2-1}\,z+1} \exp\left\{ -\frac{\left(\ell n\left\{\alpha\left(\sqrt{\alpha^2-1}\,z+1\right)\right\}\right)^2}{2\sigma^2} \right\}, \tag{2}$$

where $-1/\sqrt{\alpha^2-1} \le z < \infty, \quad \alpha = \exp\left\{\alpha^2/2\right\}.$

Note that the standardized log-normal distribution has only one parameter σ, and its lower bound is a function of σ.

It was found that the following generalized gamma distribution (in the standardized form) appears to satisfy the conditions required for analysis of significant wave height data discussed in the previous section:

$$f(z) = \frac{c\sqrt{q}}{\Gamma(m)} \left(\sqrt{q}\,z+p\right)^{cm-1} \exp\left\{-\left(\sqrt{q}\,z+p\right)^c\right\}, \tag{3}$$

where $-p/\sqrt{q} \le z < \infty, \quad p = \Gamma\left(m+\frac{1}{c}\right) / \Gamma(m),$

$$q = \left[\Gamma\left(m+\frac{2}{c}\right)\Gamma(m) - \left\{\Gamma\left(m+\frac{1}{c}\right)\right\}^2\right] / \left\{\Gamma(m)\right\}^2.$$

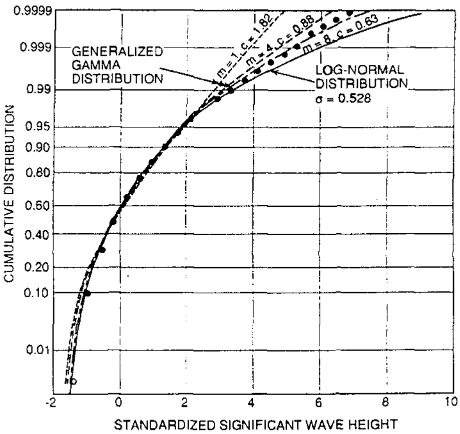

Figure 3: Comparison between the standardized cumulative distributions of (a) Norwegian data, (b) log-normal and (c) generalized gamma distribution for σ = 0.528 and for various m-values

In order to elaborate on the above statement, let us compare the two standardized probability distributions given in Eqs.(2) and (3) for an arbitrarily chosen value σ = 0.20 of the log-normal distribution and m = 1, 4 and 8 of the generalized gamma distribution. Since the lower bounds of these two probability density functions are equal, we have a functional relationship between Eqs.(2) and (3). That is,

$$\exp\left\{\sigma^2\right\} = \Gamma\left(m+\frac{2}{c}\right)\Gamma(m) \ / \ \left\{\Gamma\left(m+\frac{1}{c}\right)\right\}^2 \tag{4}$$

Hence, from Eq.(4), we can evaluate c = 5.70, 2.50 and 1.76 for each m-value with σ = 0.20. A comparison of cumulative distribution functions of the log-normal and generalized gamma distributions (in standardized form) is shown in Figure 2. As can be seen in the figure, cumulative distributions are nearly equal up to 0.90, but depending on the m and c-values in the generalized gamma distribution, the difference can become substantially large for cumulative distribution is greater than 0.90.

For further confirmation of this feature, another comparison is shown in Figure 3 for σ = 0.528 of the log-normal distribution evaluated from data obtained off Norwegian coast shown in Figure 1(a), and m = 1, 2 and 4 of the generalized gamma distribution. Included in the figure is the cumulative distribution of the Norwegian data which is also standardized by using the mean and variance evaluated from the data. It can be seen in the figure that the values of significant wave height (standardized) for a specified cumulative distribution show little difference for the two probability distributions and they both agree well with data in the range of cumulative distribution up to 0.95. However, the difference becomes substantially large for cumulative distribution greater than 0.95. The cumulative distribution of the data is very close to the generalized gamma distribution with m = 8 in this case.

This feature of the generalized gamma distribution shown in Figures 2 and 3 is considered to make its use advantageous in statistical analysis of significant wave height data. With this background, let us compare the cumulative distribution functions of measured data with the generalized gamma distribution (non-standardized) whose probability density function f(x) and cumulative distribution function F(x) are given as follows:

$$f(x) = \frac{c}{\Gamma(m)} \lambda^{cm} x^{cm-1} \exp\left\{-(\lambda x)^c\right\}, \qquad 0 \leq x < \infty. \tag{5}$$

$$F(x) = \Gamma\left\{m, (\lambda x)^c\right\} / \Gamma(m), \tag{6}$$

where the numerator is the incomplete gamma function.

Figures 4(a) through 4(g) show comparisons of cumulative distributions of data with the generalized gamma distribution. Data in each figure correspond to those shown in Figures 1(a) and 1(g), respectively. Included also in each figure are the values of the three parameters of the distribution evaluated from the data. Methods to estimate the parameter values will be discussed in the next section. As can be seen in Figure 4, the long-term significant wave height data can be well represented by the generalized gamma distribution.

Estimation of Parameters of Generalized Gamma Distribution

Methods to estimate the three parameters involved in the generalized gamma distribution are discussed by Stacy and Mihram [Stacy and Mihram 1965]. They present a procedure based on two different approaches; one being the maximum likelihood method, the other the moment method. In both methods, the logarithm of the variables (significant wave height for the present problem) is used. In particular, the second method considers the following sample mean, variance and skewness:

$$\begin{aligned} \bar{u} &= \frac{1}{n}\sum_{i=1}^{n} u_i, \quad \text{where } u_i = \ln x_i \\ s_u^2 &= \frac{1}{n-1}\sum_{i=1}^{n} (u_i - \bar{u})^2 \\ g_u &= \frac{n}{(n-1)(n-2)} \frac{1}{s_u^3} \sum_{i=1}^{n} (u_i - \bar{u})^3 \end{aligned} \tag{7}$$

By equating Eq.(7) to the theoretical mean, variance and skewness (in logarithmic form), respectively, we may estimate the three

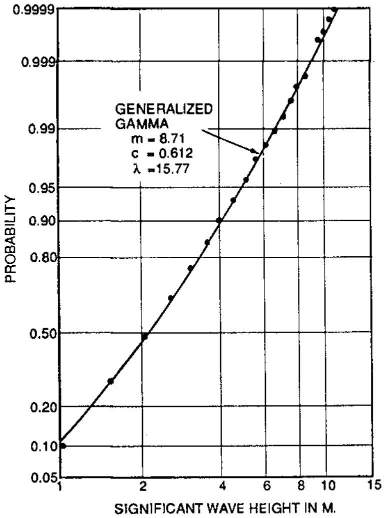

Figure 4(a): Comparison of cumulative distribution functions of data and generalized gamma distribution (Data from Mathisen and Bitner-Gregersen 1990)

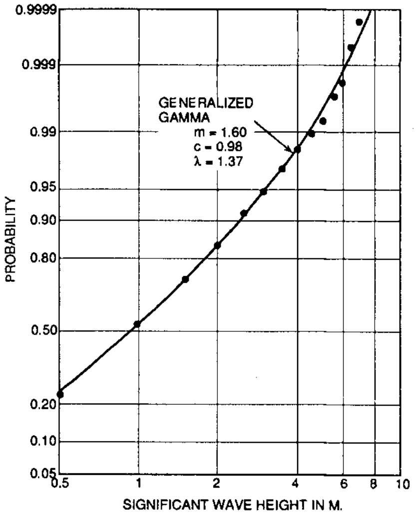

Figure 4(b): Comparison of cumulative distribution functions of data and generalized gamma distribution (Data from Bouws 1978)

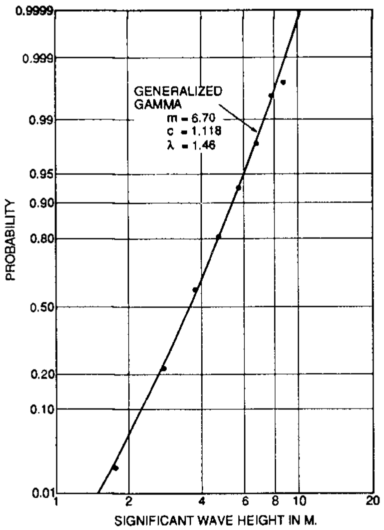

Figure 4(c): Comparison of cumulative distribution functions of data and generalized gamma distribution (Data from Tomita 1988)

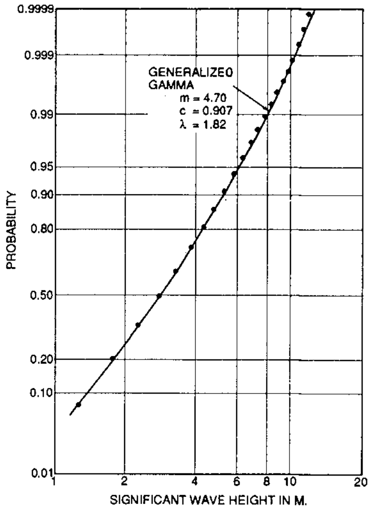

Figure 4(d): Comparison of cumulative distribution functions of data and generalized gamma distribution (Data from Nat. Data Buoy Center 1990)

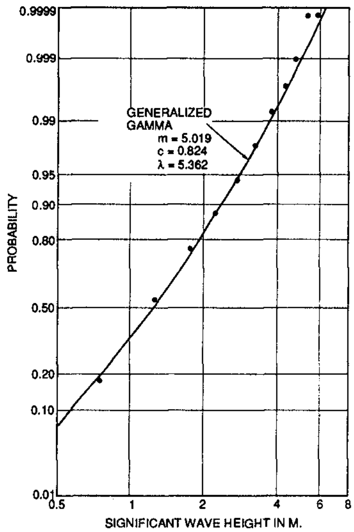

Figure 4(e): Comparison of cumulative distribution functions of data and generalized gamma distribution (Data from Nat. Data Buoy Center 1990)

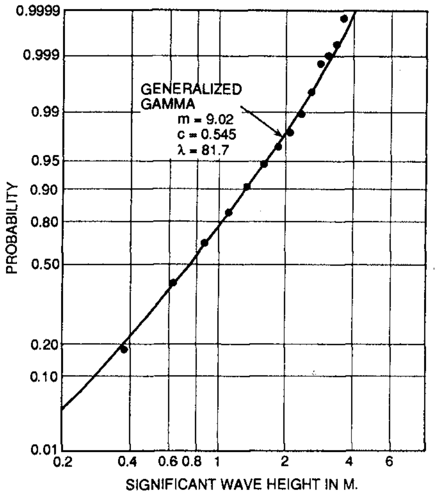

Figure 4(e): Comparison of cumulative distribution functions of data and generalized gamma distribution (Data from Coastal Data Network 1990)

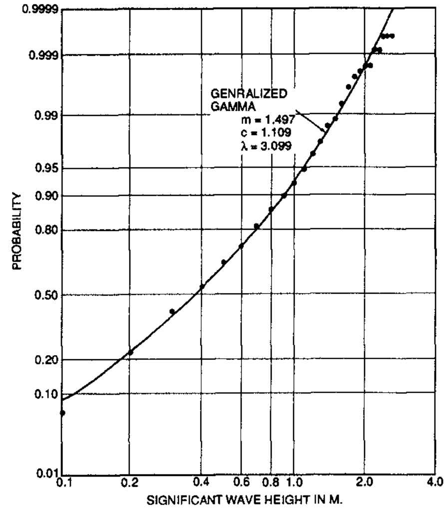

Figure 4(g): Comparison of cumulative distribution functions of data and generalized gamma distribution (Data from Work 1992)

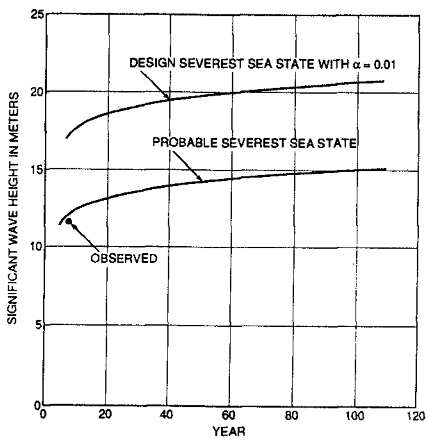

Figure 5: Probable extreme significant wave height and design extreme significant wave height with risk parameter $\alpha = 0.01$ as a function of time (Data from Mathisen & Bitner-Gregersen 1990)

parameters. For example, the parameter m can be estimated from the following equation:

$$-|g_u| = \frac{\dfrac{d^3}{dm^3}\ln\Gamma(m)}{\left\{\dfrac{d^2}{dm^2}\ln\Gamma(m)\right\}^{3/2}} \tag{8}$$

Although the Stacy and Mihram method is mathematically correct, some difficulty is often encountered in practice in solving Eq.(8) for certain set of data. It should be noted that in estimating the statistical properties of long-term significant wave height, the sample size of data is usually extremely large, on the order of several thousand. If this is the case, we may simply estimate the parameter values by equating the sample moments to theoretical moments. Here, the j-th moment of the generalized gamma distribution is given by

$$E[x^j] = \frac{1}{\lambda^j}\,\frac{\Gamma\left(m+\frac{j}{c}\right)}{\Gamma(m)} \tag{9}$$

Since the generalized gamma distribution has three unknown parameters m, c and λ, we may consider a set of equations consisting of either the first three moments or the 2nd, 3rd and 4th moments; the latter places more emphasis on the higher order moments. It is found through statistical analysis of many data that the solution of a set of three moments consisting of the 2nd, 3rd and 4th moments yields the generalized gamma distribution which well represents the cumulative distribution of the observed data. The parameter values of the generalized gamma distributions given in each example in Figure 4 are all determined by this procedure. That is, from a set of three equations for j = 2, 3 and 4 of Eq.(9), we can derive the following two equations by eliminating the parameter λ.

$$\begin{aligned} \frac{\{\Gamma(m)\}^{1/2}\,\Gamma\left(m+\frac{3}{c}\right)}{\left\{\Gamma\left(m+\frac{2}{c}\right)\right\}^{3/2}} &= \frac{E[x^3]}{\{E[x^2]\}^{3/2}} \\ \frac{\Gamma(m)\,\Gamma\left(m+\frac{4}{c}\right)}{\left\{\Gamma\left(m+\frac{2}{c}\right)\right\}^{2}} &= \frac{E[x^4]}{\{E[x^2]\}^{2}} \end{aligned} \tag{10}$$

The parameters m and c are determined from the above equations.

Estimation of Extreme Sea State

Estimation of the extreme sea state (significance wave height) expected to occur in a specified time period (50 years, for exmaple) based on the generalized gamma distribution is discussed in this section.

(a) Probable Extreme Sea State

The probable extreme sea state refers to that most likely to occur in a specified time period which is the modal value of the probability density function of extreme values. It is essentially the value of the significant wave height which satisfies the equation of return period being equal to the number of significant wave heights in a specified time. In order to avoid possible confusion, the extreme value in N-observations is denoted by Y_n-value which satisfies the following equation:

$$\ell n \left\{ 1 \,/\, \left[1 - \Gamma\left\{ m,\ (\lambda y_n)^c \right\} \,/\, \Gamma(m) \right] \right\} = \ell n\ N \tag{11}$$

where N = number of significant wave heights expected in a specified time period.

Another approach for evaluating the probable extreme sea state is through application of Cramér's asymptotic extreme value statistics. This method was used for evaluating the extreme value of the generalized gamma distribution and therefrom the following equation to estimate the probable extreme values was derived [Ochi 1978b]:

$$\frac{1}{\Gamma(m)}\, u_n^{m-1}\, e^{-u_n} = \frac{1}{N}\left\{ 1 - \frac{1}{u_n}\left(m - \frac{1}{c} \right) \right\}, \tag{12}$$

where $u_n = (\lambda y_n)^c$.

The left-side of Eq.(12) is the gamma probability distribution, and hence solution of the equation with respect to u_n can easily be obtained for a given m, c and λ. The asymptotic probable extreme value y_n can be evaluated from the extreme value of u_n. It is noted that the asymptotic probable extreme value thus obtained is very close to the value obtained as the solution of Eq.(11) when N is large.

(b) Extreme Sea State for Design Consideration

The probable extreme sea state discussed in the previous section is the modal value of the probability density function of y_n. However, the probability that the extreme value exceeds the probable extreme value is theoretically $1 - e^{-1} = 0.632$. Since this probability is very large, the probable extreme value should not be used for design of marine systems. For the design of marine systems, it is necessary to consider an extreme value for which the probability of exceedance is a very small specified value, α, called the risk parameter. It can be evaluated from the cumulative distribution function given in Eq.(6) as

$$F(x) = \frac{\Gamma\left\{m, (\lambda x)^c\right\}}{\Gamma(m)} = (1-\alpha)^{1/N} \tag{13}$$

Although Eq.(13) yields the exact solution numerically, we can derive the following equation for large N and small α:

$$1 - \frac{\Gamma\left\{m, (\lambda x)^c\right\}}{\Gamma(m)} = 1-(1-\alpha)^{1/N} \sim \frac{\alpha}{N} \tag{14}$$

Eq.(14) yields an equation in a form similar to that given in Eq.(11); and hence, the design extreme value with risk parameter, α, can be obtained by finding the y_n-value which satisfies the following equation:

$$\ln\left\{1 \,/\, \left[1-\Gamma\left\{m, (\lambda y_n)^c\right\} \,/\, \Gamma(m)\right]\right\} = \ln(N/\alpha), \tag{15}$$

where α = risk paramater; namely, the probability that the extreme value exceeds the design extreme value.

The design extreme value can also be obtained as the solution of the following equation which has a form similar to that given in Eq.(12):

$$\frac{1}{\Gamma(m)} u_n^{m-1} e^{-u_n} = \frac{\alpha}{N}\left\{1 - \frac{1}{u_n}\left(m - \frac{1}{c}\right)\right\} \tag{16}$$

As an example of the estimation of extreme sea states, Figure 5 shows the probable and design severest sea states as a function of time using the Norwegian data shown in Figure 4(a). The estimations are made by Eqs. (12) and (16). As can be seen, the magnitude of the probable as well as that of the design extreme sea state with risk parameter $\alpha = 0.01$ do not increase significantly with increase in time. However, the design extreme sea state with the risk parameter $\alpha = 0.01$ is substantially larger (approximately 40 percent) than the probable extreme sea state in this example. The severest sea state observed in 7 years at this location agrees well with the estimated probable extreme sea state for this period.

Conclusions

Results of statistical analysis of long-term significant wave height data obtained at various geographical locations and at various water depths show a consistent trend. That is, at least 90 to 95 percent of the cumulative distribution of each set of data is well represented by the log-normal probability distribution; however, the data diverge from the log-normal distribution for large significant wave heights. The divergence is consistent in such a way that the cumulative distribution of the data converges to unity faster than that of the log-normal distribution.

From comparisons of many probability distributions in standardized form, it is found that the generalized gamma distribution satisfies the condition that its cumulative distribution is by and large close to that of the log-normal distribution but it converges to unity much faster than the log-normal distribution at higher cumulative distribution values, say over 0.95. Comparisons of cumulative distribution functions of data and the generalized gamma distribution show satisfactory agreement. Methods for estimating the probable extreme sea state expected in a specified time as well as the extreme sea state for design consideration of marine systems are presented.

References

Bouws, E. (1978), "Wind and Wave Climate in the Netherlands Sector of the North Sea between 53° and 54° North Latitude", The Netherlands Meteorological Inst. Sci. Report WR78-9.

Burrows, R. and Salih, B.A. (1986), "Statistical Modelling of Long-Term Wave Climates", Proc 20th Int. Conf. on Coastal Eng., Vol. 1, pp 42-56.

Coastal Data Network (1990), "Wave Data Report 1984-1990", Univ. of Florida, CDN Report.

Fang, Z.S. and Hogben, N. (1982), "Analysis and Prediction of Long Term Probability Distributions of Wave Heights and Periods", Nat. Mar. Inst. Report R146, Feltham, England.

Mathisen, J. and Bitner-Cregersen, E. (1990), "Joint Distribution for Significant Wave Height and Zero-Up-Crossing Period", App. Ocean Res., Vol. 12, No. 2, pp 93-103.

National Data Buoy Center (1990), "Climatic Summaries for NDBC Buoys and Stations, Update I".

Ochi, M.K. (1978a), "On Long-Term Statistics for Ocean and Coastal Waves", Proc 16th Int. Conf. on Coastal Eng., Vol. 1, pp 59-75.

Ochi, M.K. (1978b), "Ceneralization of Rayleigh Probability Distribution and its Application", J. Ship Res., Vol. 22, No. 4, pp 259-265.

Ochi, M.K. and Whalen, J.E. (1980), "Prediction of the Severest Significant Wave Height", Proc. 17th Int. Conf. on Coastal Eng., Vol. 1, pp 587-599.

Stacy, E.W. and Mihram, G.A. (1965), "Parameter Estimation for a Generalized Gamma Distribution", Technometrics, Vol. 7, No. 3, pp 349-358.

Thompson, E.F. and Harris, D.L. (1972), "A Wave Climatology for U.S. Coastal Waters", Proc. Offshore Tech. Conf., Vol. 2, pp 675-688.

Tomita, H. (1988), "Wave Characteristics of North West Pacific Ocean, Part 1. Analysis of Buoy Data", Ship Research Institute, Vol. 25, No. 5, pp 21-72. (In Japanese).

Work, P. (1992), "Sediment Transport Processes at a Nourished Beach", Ph.D. Dissertation, Univ. of Florida, Coastal and Oceano. Department.