

CHAPTER 49

ESTIMATION OF EXTREME SEA SEVERITY FROM MEASURED DAILY MAXIMA

Michel K. Ochi*, David Mesa*, and De-Fu Liu**

ABSTRACT

This paper presents the results of a study to statistically estimate the most severe sea state (significant wave height) expected in 50 and 100 years from analysis of data consisting of the largest significant wave height observed each day by applying the Type III asymptotic extreme value distribution. In applying the Type III asymptotic distribution, the distribution parameters are estimated by three different methods: the maximum likelihood method, the skewness method, and a nonlinear regression method. Since none of these methods estimates values of the parameters which satisfactorily yield a distribution representing well the daily maximum data, a modified Type III asymptotic distribution is newly developed in the present study. The modified distribution yields an excellent fit over the entire range of the cumulative distribution, and the probability density function agrees well with the histogram constructed from the data.

INTRODUCTION

For the design of coastal and ocean structures, it is extremely important to statistically estimate the most severe sea state (the largest significant wave height) expected to occur over a period of time, on the order of 50 to 100 years, sufficiently long to cover the lifetime of the structure. The estimation is usually carried out based on a probability function established from analysis of data accumulated over several years. It should be noted that, in general, the probability distribution derived from an accumulation of significant wave height data is obtained empirically and that there is no theoretical basis for selecting any particular probability distribution to characterize the significant wave height.

However, if the data consist of the largest significant wave height observed every day, and if the number of observations is sufficiently large, then there is a scientific basis to choose particular probability distributions for estimating extreme significant wave

* Coastal and Oceanographic Engineering Department, University of Florida, U.S.A.

** Department of Hydraulic Engineering, Tianjin University, Peoples Republic of China

height. That is, the probability distribution applicable for the significant wave height under this situation must be one of three asymptotic distribution functions developed by Fisher, Tippett, and Frechet -- so called Type I, II, and III asymptotic extreme value distributions.

The Type I asymptotic distribution has been by far the most commonly used method for estimating extreme values from an analysis of data consisting of the daily, weekly, and monthly largest values. The results of the present analysis show that the Type I asymptotic distribution demonstrates reasonably good agreement, in general, with significant wave height data. However, the distribution deviates from the data cumulative distribution for large significant wave heights. This results in substantial overestimation of the 50 and 100 year extreme values if the estimation is made by extending the theoretical cumulative distribution function.

The Type III asymptotic distribution has a unique characteristic in that the distribution is bounded from above. This feature of the distribution appears to be pertinent for analysis of wave data, since the height of waves cannot be unlimited in reality due to breaking. For analysis of data by applying the Type III asymptotic distribution, the distribution parameters are estimated by three different methods: the method of maximum likelihood, the skewness method, and a nonlinear regression method. Unfortunately, none of these methods estimates values of the parameters which satisfactorily yield a Type III asymptotic distribution representing well the daily maximum data.

This paper presents a newly developed modified Type III asymptotic extreme value distribution which yields an excellent fit over the entire range of the cumulative distribution.

DAILY LARGEST SIGNIFICANT WAVE HEIGHT AND ASYMPTOTIC DISTRIBUTIONS

The data used in the present study were obtained by the Coastal Engineering Research Center on the eastern coast of the United States, Duck, North Carolina, at a location 450 m off the shoreline where the water depth is 8.8 m as an average. Observations of significant wave height were made four times daily during a 42 month period from 1979 to 1983. A total of 1,061 observations of largest significant wave height for each day was accumulated as shown in Table 1, and the data (for brevity, they may be called the daily maxima) were analyzed through application of Type I and Type III asymptotic extreme value distributions.

Figure 1 shows the plot of data on Type I extreme value probability paper. If the data points lie on a straight line, it may safely be concluded that the data follow the Type I asymptotic distribution, and contain all statistical characteristics thereof. This allows the extrapolation of the data to higher periods to be made very simply merely by extension of the straight line.

Table 1. Daily maximum significant wave height data

Significant wave height (m)	Number of observations
0.2 - 0.4	11
0.4 - 0.6	151
0.6 - 0.8	158
0.8 - 1.0	175
1.0 - 1.2	109
1.2 - 1.4	116
1.4 - 1.6	91
1.6 - 1.8	63
1.8 - 2.0	47
2.0 - 2.2	40
2.2 - 2.4	27
2.4 - 2.6	19
2.6 - 2.8	21
2.8 - 3.0	12
3.0 - 3.2	6
3.2 - 3.4	7
3.4 - 3.6	6
3.6 - 3.8	2
<hr/>	
Total	1,061
	in 42 months

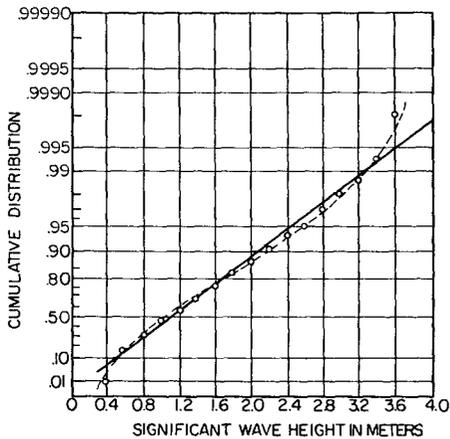


Figure 1. Cumulative distribution function of daily maximum significant wave height plotted on extreme value probability paper

It appears in Figure 1 that the data are represented reasonably well by the Type I asymptotic distribution. However, if we examine the results carefully, the cumulative distribution function of the data (the curved line given in the figure) shows a tendency to slowly deviate with different character from that of the straight line at the higher significant wave heights. Hence, by extrapolating the data and the estimated Type I distribution to comparable return periods, it is seen that the Type I distribution may yield an increasing overestimation of the extreme value prediction. This is a major shortcoming of utilizing the Type I asymptotic distribution for extreme sea state estimation.

Type III asymptotic distribution has three parameters, w , v , and k , as shown in Equations (1) and (2).

Cumulative distribution function

$$F(y) = \exp\left\{-\left(\frac{w-y}{w-v}\right)^k\right\}. \quad (1)$$

Probability density function

$$f(y) = \frac{k}{w-v} \left(\frac{w-y}{w-v}\right)^{k-1} \exp\left\{-\left(\frac{w-y}{w-v}\right)^k\right\}. \quad (2)$$

$$-\infty < y < w, \quad 0 < v, k < \infty$$

where, y = random variable, daily largest significant wave height for the present problem.

The distribution is bounded from above by the parameter w . If the parameters w and k become increasingly large, the Type III distribution tends to default to Type I distribution. The Type III distribution has not often been considered for estimating extreme values in engineering problems to the same extent as the Type I distribution. It has a feature, however, that the upper value of the distribution is bounded and thereby the extrapolation of the cumulative distribution function should not overestimate the extreme values.

In order to evaluate the parameters of the distribution from the data, the following three different methods are considered in the present study:

(1) Maximum Likelihood Method

The likelihood function of the distribution can be obtained from Eq. (1) as,

$$L(y_i | w, v, k) = \prod_{i=1}^n f(y_i) = k^n (w - v)^{-nk} \\ \times \prod_{i=1}^n (w - y_i)^{k-1} \exp \left\{ - \sum_{i=1}^n \left(\frac{w - y_i}{w - v} \right)^k \right\}. \quad (3)$$

Since $L(y_i | w, v, k)$ is a monotonic function, partial differentiation of $\ln L(y_i | w, v, k)$ with respect to the unknown parameters w , v , and k and setting each equal to zero results in

$$\frac{\partial}{\partial w} \ln L = - \frac{nk}{w - v} + \sum_{i=1}^n \left[k \left(\frac{w - y_i}{w - v} \right)^k \frac{v - y_i}{(w - y_i)(w - v)} + \frac{k - 1}{w - y_i} \right] = 0$$

$$\frac{\partial}{\partial v} \ln L = \frac{nk}{w - v} + \sum_{i=1}^n \frac{-k}{w - v} \left(\frac{w - y_i}{w - v} \right)^k = 0$$

$$\frac{\partial}{\partial k} \ln L = \frac{n}{k} - n \ln(w - v) - \sum_{i=1}^n \left[\left(\frac{w - y_i}{w - v} \right)^k \ln \left(\frac{w - y_i}{w - v} \right) - \ln \left(\frac{w - y_i}{w - v} \right) \right] = 0. \quad (4)$$

The three parameters of the distribution can be estimated from the above equations. It is found, however, that the solution cannot be obtained explicitly, but instead involves a lengthy iterative numerical procedure. Furthermore, it is found that the solution is very sensitive to the parameter v . Under these conditions a numerical iterative procedure is developed which initially fixes the parameter v , then determines the corresponding values of the parameters w and k . Since $F(y) = e^{-1}$ for $y = v$ in Eq. (2), the theoretical value of v which satisfies the relationship $F(y = v) = e^{-1}$ is used as a base from which to initiate the iterative algorithm. Graphic and regression analysis indicates a linear relationship exists between the parameters w and k for a given v .

No combination of the w , v , and k parameters determined through the maximum likelihood method, however, represents well the cumulative distribution function obtained from the data as demonstrated in Figure 2. The figure shows a comparison of the data and Type III asymptotic cumulative distribution functions obtained for three values of the parameter v (0.96, 1.00, and 1.04) and the corresponding w and k parameter sets.

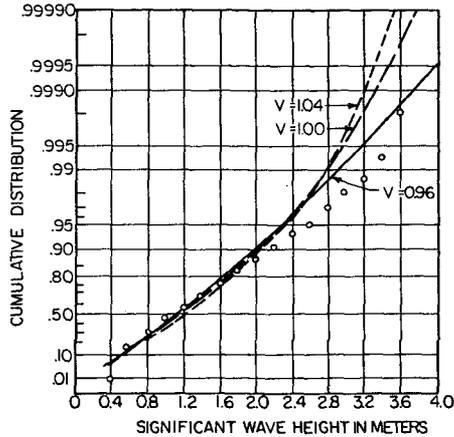


Figure 2. Daily maximum significant wave height and Type III asymptotic cumulative distributions, based on maximum likelihood method

(2) Skewness Method

The principle of the skewness method is based upon the fact that the skewness of the Type III asymptotic distribution can be theoretically expressed solely as a function of the parameter k (Gumbel, 1966). That is,

$$\text{Skewness } \gamma = \left\{ \Gamma(1 - 3/k) - 3\Gamma(1 + 2/k) \Gamma(1 + 1/k) + 2\Gamma^3(1 + 1/k) \right\} \{B(k)\}^3 \quad (5)$$

$$\text{where, } B(k) = 1/\{\Gamma(1 + 2/k) - \Gamma^2(1 + 1/k)\}^{1/2} .$$

The parameter k therefore can be evaluated from Eq. (5) by calculating the skewness γ from the observed data by

$$\gamma = \frac{E[(y - E[y])^3]}{(\text{Var}[y])^{3/2}} . \quad (6)$$

The parameters w and v are subsequently determined by evaluating the mean and variance with the aid of the following formulae:

$$E\{y\} = w - (w - v) \Gamma(1 + 1/k)$$

$$\text{Var}\{y\} = (w - v)^2 \{ \Gamma(1 + 2/k) - \Gamma^2(1 + 1/k) \} . \quad (7)$$

It should be noted that theoretical background of the skewness method is sound; however, it is sometimes difficult to determine the k -value in practice as is the case for the problem discussed below.

The theoretical skewness expression given in Eq. (5) is shown in Figure 3 as a function of k . The curve is asymptotic at $k = 0$ and monotonically increases through the zero skewness point ($k = 3.602$) until reaching its limiting value at 1.139. The skewness increases rapidly up to approximately $k = 40$, then becomes increasingly insensitive to k as k becomes large. The sample skewness as calculated from Eq. (7) is $\gamma = 1.134$ for the data given in Table 1. This value is very near the limiting value 1.139; hence, a very wide range of k values is possible in this limiting region. For this reason, the skewness method for estimating the parameter k appears to be inappropriate for the present problem.

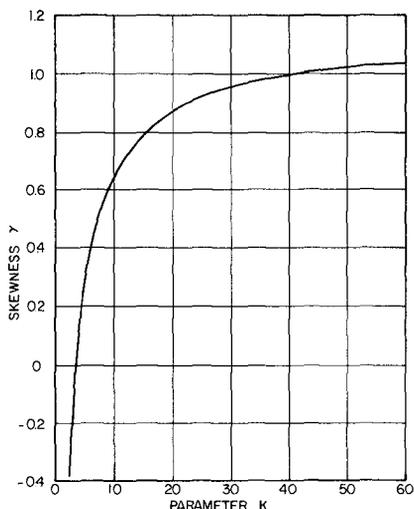


Figure 3. Skewness γ as a function of k

To overcome this difficulty, an alternative method is developed using the skewness approach as a foundation. That is, the parameter v is estimated from the cumulative distribution function constructed from the data such that the parameter v satisfies the condition that the cumulative distribution function $F(y) = e^{-1}$ for $y = v$. The other two parameters are evaluated by using the v -value thusly determined. However, the derived distribution does not well represent the observed data as shown in Figure 4.

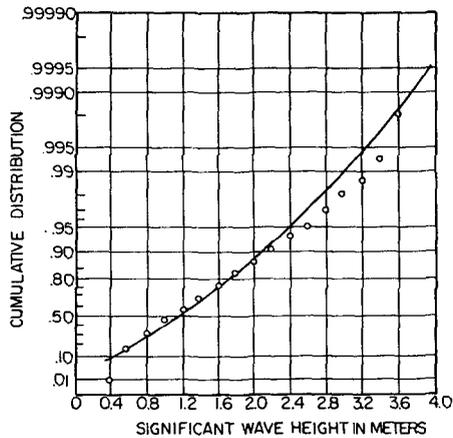


Figure 4. Daily maximum significant wave height and Type III asymptotic cumulative distributions, based on skewness alternative method

(3) Nonlinear Multiple Regression Method

Another method to establish the parameters of the Type III asymptotic extreme value distribution is to apply a nonlinear multiple regression analysis. For this, taking the logarithm of Eq. (1) twice results in

$$\ln [-\ln F(y)] = k \ln \left(\frac{w - y}{w - v} \right) . \quad (8)$$

The left-hand side of Eq. (8) is determined from the data. The right-hand side is now linear in the parameter k , and monotonic in the parameters w and v . Linearization of the exponent k substantially increases the computational stability in determining the parameter values.

A comparison between the cumulative distribution function obtained by the nonlinear regression method and observed data is shown in Figure 5. Figure 6 shows a comparison between the probability density function and histogram. Although the cumulative distribution function in Figure 5 appears to display a reasonable fit, it is apparent from the probability density function in Figure 6 that the distribution does not well represent the data. The peak does not possess the sharp rapid increase, but instead is much flatter and shifted to the right. Thus, it may be concluded that the Type III asymptotic distribution with the parameters determined by the nonlinear regression method does not well represent.

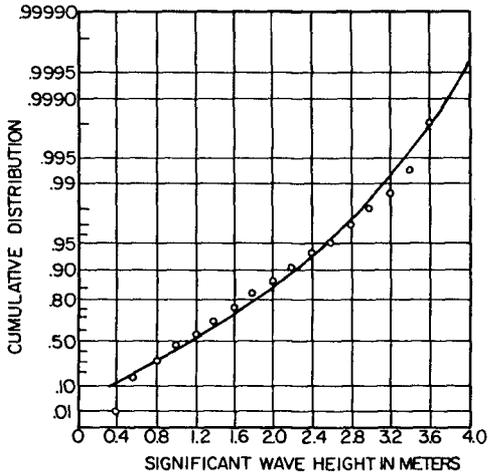


Figure 5. Daily maximum significant wave height and Type III asymptotic cumulative distributions, based on nonlinear regression method

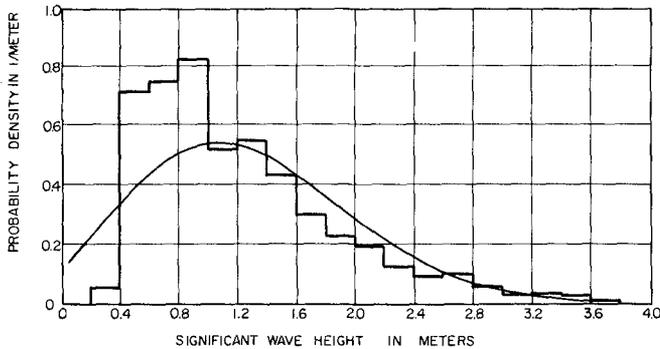


Figure 6. Comparison of daily maximum significant wave height histogram and Type III asymptotic probability density function, based on nonlinear regression method

DERIVATION OF MODIFIED TYPE III ASYMPTOTIC DISTRIBUTION

In order to improve the agreement between the Type III asymptotic extreme value distribution and the observed data, we first examine Eq. (8). Figure 7 shows a comparison between the left-hand side of Eq. (8) presented by using data points, and the right-hand side of the equation expressed using the results of the nonlinear regression method. Writing the difference between the theoretical distribution and the observed data as $\Delta(y)$ results in

$$\ln [-\ln F(y)] = k \ln \left(\frac{w-y}{w-v} \right) + \Delta(y) . \quad (9)$$

Here, the functional form of $\Delta(y)$ may be expressed in the form of polynomials given by

$$\Delta(y) = a + b(y - y_0) + c(y - y_0)^2 + d(y - y_0)^3 . \quad (10)$$

The values of a , b , c , and y_0 are determined again by employing the nonlinear regression procedure. Thus, from Eq. (9), the modified Type III asymptotic extreme value distribution can be written as

$$F(y) = \exp \left\{ - \left(\frac{w-y}{w-v} \right)^k e^{\Delta(y)} \right\} . \quad (11)$$

Note that $F(y)$ given in Eq. (11) satisfies the conditions required of the cumulative distribution function. Hence, the addition of $\Delta(y)$ does not affect the basic characteristics of the original Type III asymptotic distribution. If $\Delta(y)$ is zero over the entire variate range, then the modified Type III asymptotic distribution reduces to the original Type III asymptotic distribution.

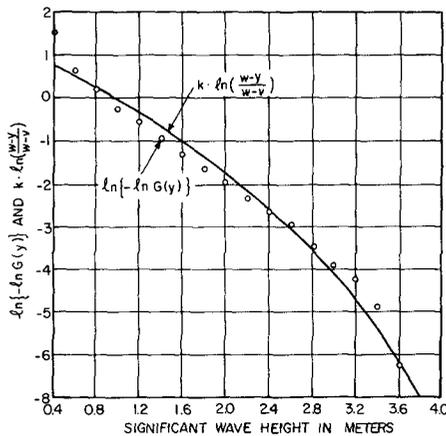


Figure 7. Comparison of cumulative distribution functions of daily maximum significant wave height and Type III asymptotic distribution, based on nonlinear regression method

Figure 8 shows the results of computations using Eq. (10) as applied to the difference $\Delta(y)$ obtained from Figure 6 for which we have $a = -0.314$, $b = 0.513$, $c = 0.464$, $d = -0.171$, and $y_0 = 1.687$. The circles in the figure are the differences between the data and the Type III asymptotic distribution obtained from Figure 7.

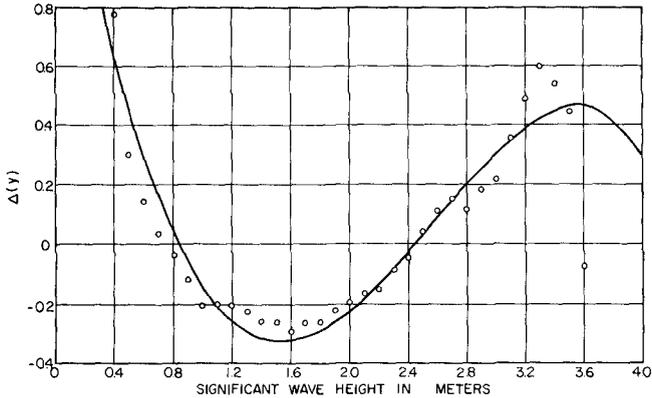


Figure 8. Comparison of $\Delta(y)$ and difference between cumulative distributions of daily maximum significant wave height and Type III asymptotic distribution, based on nonlinear regression method

A comparison between the cumulative distribution function of the modified Type III asymptotic distribution given in Eq. (11) and that obtained from data is shown in Figure 9. Good agreement can be seen between them over the entire data range. The cumulative distribution

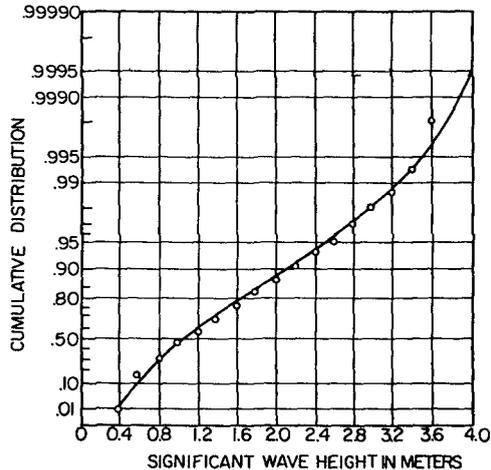


Figure 9. Daily maximum significant wave height and modified Type III asymptotic cumulative distributions

function for large significant wave heights (which is the area of interest) shows asymptotic characteristics as it approaches its limiting value. A comparison between the probability density function of the modified Type III and the histogram is shown in Figure 10 in which a good agreement can be seen over the entire variate range.

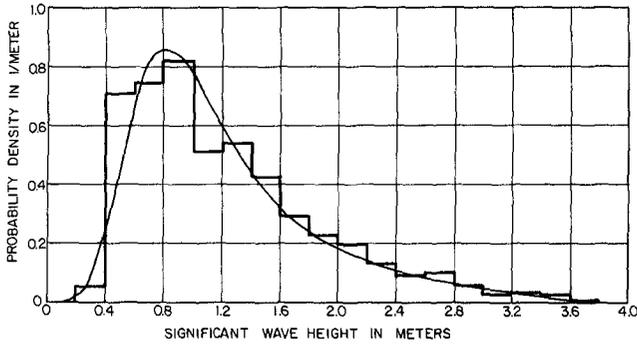


Figure 10. Comparison of daily maximum significant wave height histogram and modified Type III asymptotic probability density function

ESTIMATION OF EXTREME VALUES

The extreme sea severities most likely to occur (modal values of the extreme value distributions) in 50 and 100 years are evaluated from the data given in Table 1 by applying the Type I and the modified Type III asymptotic distributions. The estimated extreme values are tabulated in Table 2.

As can be seen in the table, the extreme significant wave heights estimated by applying the Type I asymptotic distribution are substantially greater than those estimated by applying the modified Type III asymptotic distribution. Since the Type I asymptotic distribution deviates from the data cumulative distribution for large significant wave heights as shown in Figure 1, the distribution yields an increasingly overestimation of the extreme value with increasing variate values.

Table 2 Estimation of 50 and 100 year extreme significant wave heights

Distribution	50 year (m)	100 year (m)
Type I asymptotic distribution	5.8	6.3
Modified Type III asymptotic distribution	4.2	4.3

CONCLUSIONS

This paper presents the results of a study to statistically estimate the extreme sea severity (significant wave height) expected in 50 and 100 years from analysis of data consisting of the largest significant wave heights observed each day. From results of analysis by applying the Type I and Type III asymptotic extreme value distributions, the following conclusions are drawn:

1. The cumulative distribution function of the data shows a tendency to slowly deviate with different character from that of the Type I asymptotic at the higher significant wave heights. Hence, by extrapolating the data and the estimated Type I distribution to comparable return periods, it is expected that the Type I distribution may yield an increasingly overestimation of the extreme value with increasing variate values.
2. The parameters of the Type III asymptotic distribution are estimated by three different methods: the maximum likelihood method, the skewness method, and a nonlinear regression method. None of these methods, however, estimates values of the parameters which satisfactorily yields a distribution representing well the cumulative distribution of the data.
3. The modified Type III asymptotic extreme value distribution newly developed in the present study (Equation 11) yields an excellent fit over the entire range of the cumulative distribution and the probability density function agrees well with the histogram constructed from the data.
4. The extreme significant wave heights in 50 and 100 years estimated by the Type I asymptotic distribution are substantially greater than those estimated by the modified Type III asymptotic distribution.

REFERENCES

- Gumbel, E. J.: "Statistics of Extremes", Columbia Univ. Press, 1966.