◉ Technical Paper

A Statistical Method of Estimation of Extreme Sea Level

Sun H. Kwon*, Taeil Lee** and Young K. Chon***
(*Received Aug.* 31, 1991)

극한 파고 추정의 통계적 방법

권 순 홍·이 태 일·전 영기

Key Words : Extreme Sea Level(극한 파고), Type III Asymtotic Distribution(제3종 점근 분포), Nonlinear Regression Method(비선형 회귀법), Skewness Method(사행 방법), Multinomial Discretization Method(다항식 이산화방법)

초 록

본 연구에서는 극한 파고를 추정하는 방법을 제시하였다. Type III분포에 근거해서 4가지의 방법들에 의해 분포 함수의 파라미터들을 추정하였다. 실제 자료와 추정된 분포 함수 값의 차이를 다항식을 도입하여 함으로써 그 오차를 줄였다. 이 방법들의 타당성을 보이기 위해 실제 해상의 자료들을 이용하여 분포 함수를 구하고 조우 주기들에 해당하는 극한 파고를 계산하여 보았다.

1. Introduction

Extreme values from observational data are the information of special importance in several areas of engineering application. In the field of ocean engneering, wave height is main factor to be considered for various design purposes. This paper discusses the methods of statistical estimation of extreme significant wave height which may be encountered for a certain return period. Gumbel[2] had classified the asymtotic distributions of the extremes as the type I, II and III asymtotic forms. Ochi[3] has shown that the type I distribution may yield and increasingly overestimation of the extreme value with increasing variate values. In this study Type III asymtotic distribution is intensively applied to observational data.

There are many avilable methods for estimating the parameters of the distribution. Ochi[3] illustrated 3 methods of solutions. These are maximum likelihood method, skewness method, and a nonlinear regression method. The calculations using these methods do not give satisfactory re-

* Member, Dept. of Naval Architecture, Pusan National University, Pusan, Korea
** Hyundai Maritime Research Institute, Ulsan, Korea
*** Korean Register of Shipping, Seoul, Korea

sults. Ochi[3] proposed a newly developed modified type III asymtotic extreme value distribution which yields an excellent fit over the entire range of the cumulative distribution. What he did was introduction of polynomials which fit the difference between the results of the nonlinear regression method and the observed data. This newly introduced fittting function yielded an excellent fit over the entire range of cumulative distribution, and the probability density function. The authors of this paper attempted to apply that fitting function not only to the nonlinear regression method but also to the rest of the estimation methods. In this study, multinomial discretization method is newly employed in the field of extreme sea level estimation. Test of the four methods with fitting functions provide excellent results. The skewness method combined with fitting polynomials can be considered to be very practical because of its explicit form.

Table 1 Data A

Significant wave height¼M¾	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
total :	1,061

Table 2 Data B

Significant wave height¼M¾	number of observations
0.0–0.5	144
0.5–1.0	63
1.0–1.5	48
1.5–2.0	27
2.0–2.5	15
2.5–3.0	10
3.0–3.5	5
3.5–4.0	3
4.0–4.5	1
total :	316

2. Data for the Study

The data base used in this study is consisted of two types. Data A is adopted from Ochi[3] for mutual comparison purposes. Data A represents significant wave heights which were observed four times a day during a 42 month period from 1979 to 1983. Data B represents daily significant wave heights which were recorded in wave measurement station near Cheju Island in Korea during 1975 to 1985. Those data for February were compiled in data B. Those two sets of data are shown in Table 1 and 2. The probability density function of data B shows a monotonically decreasing feature which is different from most of the other propbability density functions.

3. Asymptotic Distributions

Gumbel[2] had classified the asymptotic distributions of the extremes as the type I, II, III asymptotic forms. Those relevant mathematical formulae are shown in Appendix 1.

The extreme value from an initial distribution with an exponentially decaying tail will converge asymptotically to the type I limiting form, whereas, for an initial variate that decays with a polinomial tail, its extreme value will converge to the type II asymptotic from. If the extreme is limited, the corresponding extremal distribution will converge to the type III asymptotic form. Ochi[3] has shown that the type I distribution may yield an increasingly overestimation of the extreme value with increasing variate values. Wave height can be considered to be bounded. This leads us to use the type III asymptotic distribution for extreme sea level computation. The cumulative distribution function and the probability density function of type III asymptotic distribution can be represented as

$$F(y)=\exp[-(\frac{w-y}{w-y})^{k}] \quad \cdots\cdots\cdots\cdots\cdots\cdots \quad (1)$$

$$f(y)=\frac{k}{w-v}(\frac{w-y}{w-v})^{k-1}\exp[-(\frac{w-y}{w-v})^{k}] \quad \cdots\cdots\cdots\cdots \quad (2)$$

4. Determination of Parameters

Determination of parameters are the main subject of extreme value theory. Ochi[3] illustrated 3 methods of solutions. These are nonlinear multiple regression method, maximum likelihood method, and skewness method. In this study multinomial discretization method is newly employed. Therefore four different methods are used.

A nonlinear multiple regression analysis have been used widely in all branches of engineering. Ochi(1986) demonstrated this by taking the lograithm of equation(1) twice as follows

$$\ln[-\ln F(y)]=k\ \ln(\frac{w-yk}{w-v}) \quad \cdots\cdots\cdots\cdots \quad (3)$$

The calculated cumulative distribution functions (cdf) obtained by the nonlinear regression method have been compared against the observed data. Figures 1 and 2 show the comparison using data A and B, respectively. Figures 3 and 4 show the probabiltity denisty functions(pdf) and the

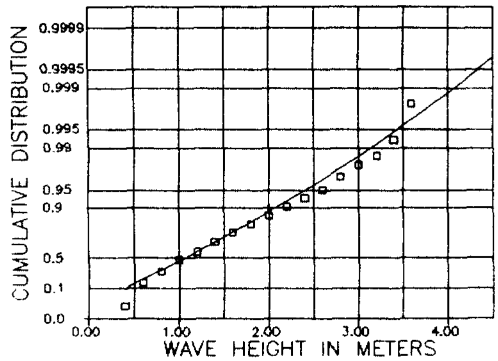

Fig. 1 Cdf of data A by nonlinear multiple regression method

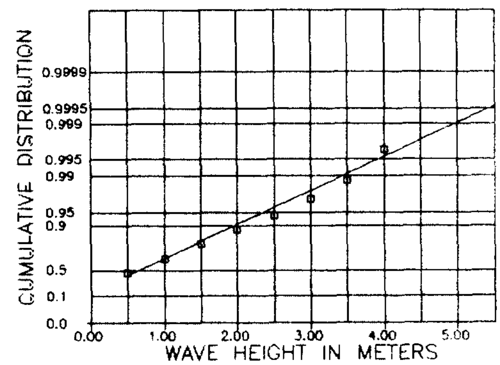

Fig. 2 Cdf of data B by nonlinear multiple regression method

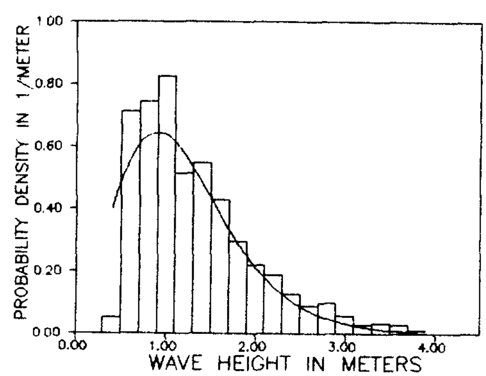

Fig. 3 Pdf of data A by nonlinear multiple regression method

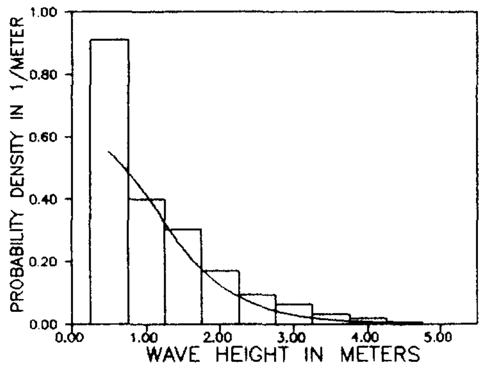

Fig. 4 Pdf of data B by nonlinear multiple regression method

histogram of observed data. A nonlinear multiple regression method does not give a good fit over the entire variate range of the cumulative distributions.

To improve the agreement between the type III asymptotic distribution and the observed data, Ochi[3] proposed modified type III asymptotic distribution. What he did is the introduction of polynomials which fit the difference between the theoretical distribution calculated from the nonlinear multiple regression mehtod and the observed data.

The polynomial was determined by the nonlinear regression procedure. That is

$$\ln[-\ln F(y)] = k \ln\left(\frac{w-y}{w-v}\right) + \Delta y \quad \cdots\cdots \quad (4)$$

Where

$$\Delta y = a + by + cy^2 + dy^3 \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \quad (5)$$

Thus the modified type III asymptotic extreme value distribution can be wirtten as

$$F(y) = \exp\left\{-\left(\frac{w-y}{w-v}\right)^k \ e^{\Delta y}\right\} \quad \cdots\cdots\cdots\cdots\cdots \quad (6)$$

This modified cumulative distribution function still satisfies the desired condition of the cumulative distribution function. The calculated cumulative distribution functions obtained by the modified type III asymptotic distribution have been compared against the observed data in Figures 5 and 6. A significant improvement in the cumulative distribution function can be noticed immediately. The probability density functions are shown in Figures 7 and 8 which also show signifiant improvement.

The maximum likelihood method is based on the maximization of the likelyhood function with

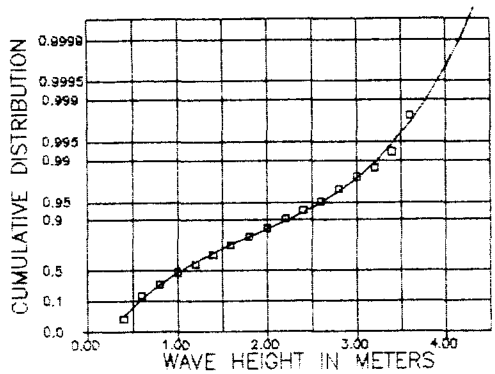

Fig. 5 Modified type III asymptotic cdf of data A based on nonlinear multiple regression method

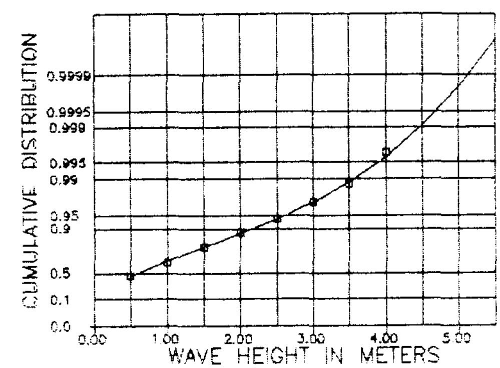

Fig. 6 Modified type III asymptotic cdf of data B based on nonlinear multiple regression method

respect to the parameters to be estimated. The likelihood fucntion of the distribution and its logarithms are shown in Appendix 2. The solutions

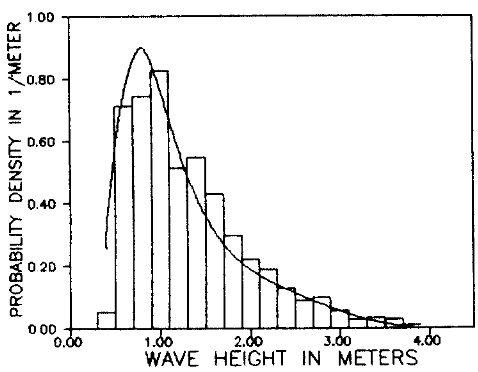

Fig. 7 Modified type III asymptotic pdf of data A based on nonlinear multiple regression method

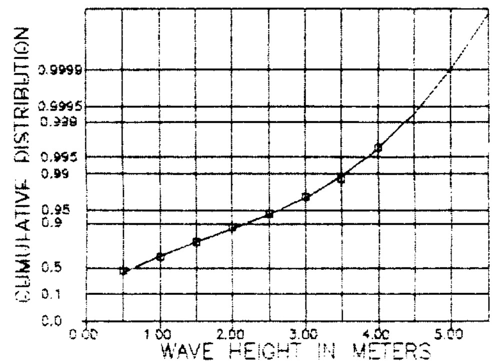

Fig. 10 Modified type III asymptotic cdf of data B based on maximum likelihood method

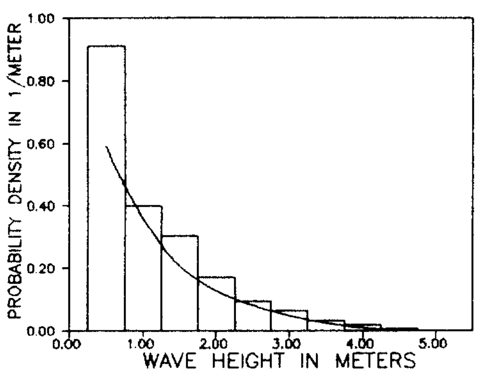

Fig. 8 Modified type III asymptotic pdf of data B based on nonlinear multiple regression method

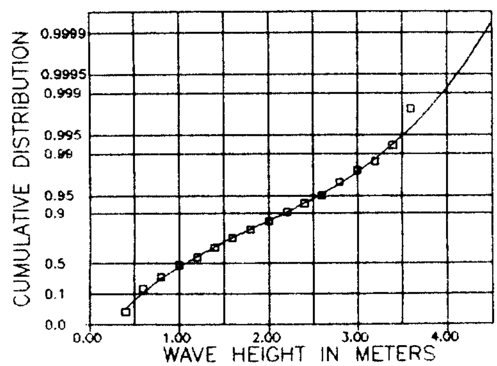

Fig. 11 Modified type III asymptotic cdf of data A based on skewness method

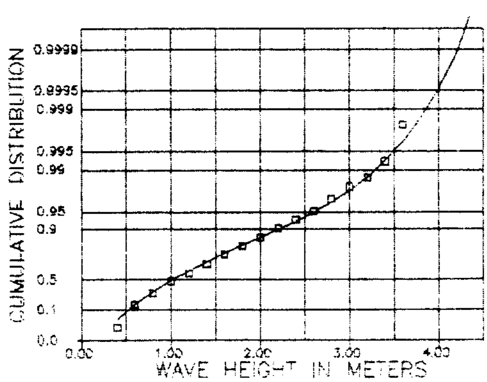

Fig. 9 Modified type III asymptotic cdf of data a based on maximum likelihood method

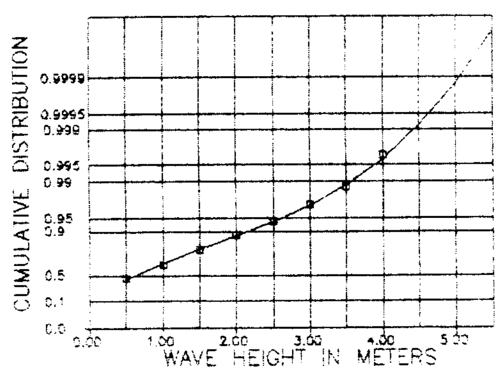

Fig. 12 Modified type III asymptotic cdf of data B based on skewness method

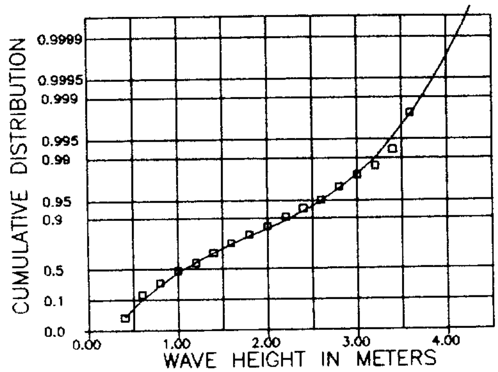

Fig. 13 Modified type III asymptotic cdf of data A based on multinomial discretization method

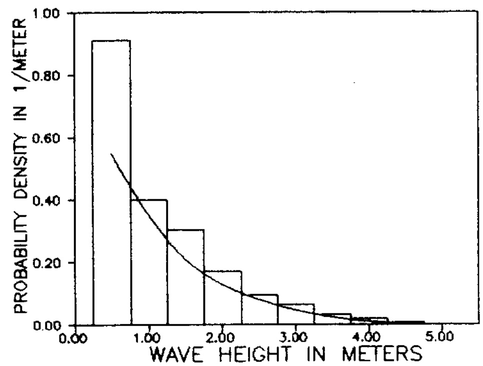

Fig. 14 Modified type III asymptotic cdf of data B based on multinomial discretization method

can be obtained by numerical iterations. The calculations are very sensitive to the initial estimation of the parameters. Maximum likelihood method does not give an excellent fit over the entire range of the cumulative distributions. Here the difference between the theoretical distribution and the observed data has been fitted again. Figures 9 and 10 show the results. Again good agreemetns are achived.

As another approach to the full specification of the 3 parameters of the type III asymptotic distribution, the skewness method is considered. The 3 parameters are related to the statistical moments. The relevant mathematics for the skewness method is summarized in Appendix 3. All 3 parameters can be obtained explicityly without iterative numerical caculation. Secant method is used as an equation solver in this study. Again the modified forms of type III asymptotic distribution function have been attempted. Improved agreements are shown in Figures 11 and 12.

Lastly the multinomial discretization method (Castillo)[1] is employed. Let $F(y ; \phi)$ be the cumulative distribution function of a random vriables y that belongs to a ϕ−parameter family. Then, the experiment consisting of counting the number the elements, in a multinomial experiment, and its associated random variable has probability mass function given by

$$p(n_1,n_2,\ldots,n_k)=\frac{n!\ p_1^{n_1}(\phi)\ p_2^{n_2}(\phi)\cdots p_k^{n_k}(\phi)}{[n_1!\ n_2!\ldots n_k!]} \quad \cdots \quad (7)$$

where

$$p_i(\phi)=F(y_{i+1}\ ;\ \phi)-F(y_i\ ;\ \phi) \quad \cdots\cdots\cdots\cdots\cdots \quad (8)$$

The maximization of eqation(7) with respect to ϕ leads to a maximum likelihood multinomial approximation of the family $F(y ; \phi)$ of cumulative distribution function. The modified form of Type III asymptotic distribution fucntions based on the multinomial discretization method has been tired again. The modifided cumulative distribution fucntion are presented in Figures 13 and 14.

The significant wave heights which have 50 and 100 year return period have been evaluated using illustrated four modified methods. The results are shown in Table 3 with Ochi's calcuation for comparison. The illustrated four modified methods predict almost the same extreme wave heights.

5. Conclusions

1) The modified form of type III asymptotic di-

Table 3 Comparison of extreme wave heights

Method	extreme wave heights (*M*)	
	50year return period	100year return period
Ochi	4.2	4.3
Nonlinear Multiple Regression Method	4.23	4.33
Maximum Likelihood Method	4.26	4.34
Skewness Method	4.5	4.61
Multinomial Discretization Method	4.17	4.27

stribution can be applied not only to multiple regression method but also the the rest of the four methods, that is, maximum likelihood method, skewness method, and multinomial discretization method.

2) The modified type III asymptotic distribution function based on multiple regression method shows the most excellent fit over the entire variate range.

3) The application of the modifided type III asymptotic distribution to the skewness method is strongly recommended because of its sound theoretical background and its explicit forms of calculation.

4) The illustrated four methods predict almost the same extreme wave heights except that the skewness method shows a relatively higher extreme value.

Reference

1) Castillo, E. "Extreme value theory in engineeering," Academic Press INC, 1988

2) Gumbel, E. J., "Statistics of Extreme," Columbia Univ. Press, 1966

3) Ochi, M. K., "Estimation of Extreme Sea Levelity From Measured Daily Maxima," 20th International Conference on Coastal Engineering in Taipei, Taiwan, 1986

Appendix 1 ; the asymptotix distributions

Type I ; the double exponential form,

$$F(y) = \exp[-\exp(-\alpha(y-u))]$$

u ; the characteristic largest value of the initial variately

α ; an inverse measure of dispersion

y ; the random variable

Type II ; the exponential form,

$$F(y) = \exp[-(\frac{q}{y})^k]$$

q ; the characteristic largest value of the initial variately

k ; the shape parameter

Type III ; the exponential form with upper bound w,

$$F(y) ; \exp[-(\frac{w-y}{w-v})^k]$$

v ; the characteristic largest value of y

Appendix 2 ; likelihood function and its derivatives

$$L(y_i,k,w,v) = \prod_{i=1}^{n} f(y_i) = k^n(w-v)^{-nk}$$

$$\times \prod_{i=1}^{n} (w-y_i \exp\{-\sum_{i=1}^{n} (\frac{w-y_i}{w-v})^k\}$$

$$\frac{\partial}{\partial w} \ln L = -\frac{nk}{w-v} + \sum_{i=1}^{n} [k$$

$$(\frac{w-y_i}{w-v})^k \frac{v-y_i}{(w-y_i)(w-v)} + \frac{k-1}{w-y_i}] = 0$$

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

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

Appendix 3 ; mathematical relations between parameters and moments

$$\phi = \frac{E[(y-\mu)^3]}{\sigma^{3/2}}$$

$$\phi = -\{\Gamma(1+3/k) - 3\Gamma(1+2/k)\Gamma(1+1/k) + 2\Gamma^3(1+1/k)\} \times \{1/(\Gamma(1+2/k) - \Gamma^2(1+1/k))^{1/2}\}$$

$$\mu = w - (w-v)\Gamma(1+1/k)$$

$$\sigma^2 = (w-v)^2\{\Gamma(1+2/k) - \Gamma^2(1+1/k)\}$$

where μ ; mean
σ ; variance
ϕ ; skewness coefficient
