Jasna Prpić-Oršić, University of Rijeka, Faculty of Engineering, Vukovarska 58, Rijeka, Croatia
Anton Turk, University of Rijeka, Faculty of Engineering, Vukovarska 58, Rijeka, Croatia

FITTING OF EXTREME WAVE PROBABILITY DISTRIBUTION FUNCTION USING GENETIC ALGORITHM

Summary

Two major statistical issues can be distinguished in the procedure of wave extreme prediction. The first issue is that predicted extreme values must be based on data collected in relatively short time. The second issue is extrapolation of the observed data into its extreme region, typically lying well beyond from even the most extreme available observation. The process of extrapolation plays a fundamental role in this area of analysis and therefore it is essential to fit empirically a convenient probability distribution that describes the available data as closely as possible. Optimization of extreme values probability distribution parameters by genetic algorithm is applied to improve the methodology of extreme sea state prediction. Illustrative applications of the method are given for a North Atlantic sea environment. The results are presented as crest height maximum values occurring with a given probability or in a design storm that has a specified return period.

Key words: *extreme sea state, probability density function, genetic algorithm*

ODREĐIVANJE PARAMETERA FUNKCIJE GUSTOĆE VJEROJATNOSTI EKSTREMNIH VALNIH VISINA GENETSKIM ALGORITMOM

Sažetak

U postupku predviđanja ekstremnih valnih visina pojavljuju se dva osnovna statistička problema. Prvi je činjenica da se procjena ekstremnih vrijednosti tijekom dugog perioda mora temeljiti na podacima prikupljenim tijekom relativno kratkog vremena. Drugi se problem sastoji u ekstrapolaciji ekstremnih vrijednosti u područje koje je najčešće daleko izvan najviše vrijednosti dostupnih podataka. U ovom je dijelu analize postupak ekstrapolacije od ključnog značaja te je veoma važno što bolje prilagoditi distribuciju vjerojatnosti koja opisuje dostupne podatke. Stoga je, u cilju poboljšanja postupka određivanja ekstremnog stanja mora, za procjenu parametara funkcije distribucije vjerojatnosti ekstremnih valnih vrijednosti u radu primijenjen postupak optimizacije genetskim algoritmom. Slikovit primjer primjene metode dat je za valove Sjevernog Atlantika. Rezultati su prikazani u obliku ekstremnih vrijednosti valnih amplituda za koje se procjenjuje da će se s određenom vjerojatnošću pojaviti tijekom oluje određenog povratnog perioda.

Ključne riječi: *ekstremno stanje mora, funkcija gustoće vjerojatnosti, genetski algoritam*

1. Introduction

An important [1] step often encountered in design is estimation of an extreme design wave on the basis of recorded or hindcast data. This generally involves selecting and fitting a suitable probability distribution to wave height data, and extrapolating this to locate a suitable design wave, such as the so-called "50-year wave". This is characteristically large wave height that might be expected with a certain small probability during the lifetime of the structure. In the procedure of wave extreme prediction the two major statistical issues can be distinguished. The first issue is that predicted extreme values, for let say hundred years, must be based on data collected in relatively short time, for example ten years. It is understandable that reliable data collected over hundred years will certainly not be available. The second issue is extrapolation of the observed data into its extreme region, typically lying well beyond even the most extreme of the available observation. Asymptotic theory suggests that such maxima are well modeled by generalized extreme value distributions that were originally developed by Fisher and Tippet and later systematized by Gumbel [2, 3]. An important requirement to an adequate analysis method is that the short term (conditional) exceedance probabilities are consistently accumulated into a resulting long term (marginal) exceedance probability, so it is important to find the distribution that covers the data best.

2. Wave environment prediction

2.1. Extreme wave amplitude – All Sea State approach

In its general form the All Sea State method (ASS) requires the complete probability density function of the response for each specific condition, that is, each combination of speed, heading, wave spectrum and sea state. The short term probability density function $f_{ST}(R_a)$ has to be calculated for all wave direction intervals β_i and sea state intervals defined by the significant wave height and the wave period T_k. The prediction of the characteristics of long term extreme values deals with the occurrence of rare events as opposed to the short term statistics which determine the normal deviations. The long term probability function $f_{LT}(R_a)$ can be calculated as the weighed sum of all short term results:

$$f_{LT}(R_a) = \sum_{i=1}^{N_\beta} \sum_{j=1}^{N_H} \sum_{k=1}^{N_T} f_{ST}(R_a)_{ijk} \cdot f_i \cdot f_{jk} \tag{1}$$

in which N_β , N_H , N_T are number of wave direction, wave height and wave period intervals respectively, f_i and f_{jk} are long term probabilities of wave direction interval β_i and long term probabilities of wave height and period interval (H_j i T_k) within the wave direction interval. Probabilities f_i and f_{jk} can be obtained from wave scatter diagrams [4] in a way that

$$f_i \cdot f_{jk} = \frac{\text{number of observations in interval}\,(\beta_i, H_j, T_k)}{\text{number of observations for all}\,(\beta_i, H_j, T_k)} \tag{2}$$

Cumulative distribution function (CDF) which represents the distribution of probability that the value of response amplitude will be lower then some value a can be calculated as:

$$F_{LT}(a) = \int_0^a f_{LT}(R_a) \cdot \mathrm{d}\,R_a \tag{3}$$

The zero-crossing period of the response in a certain sea state is given by the

$$T_{zR} = 2\pi\sqrt{\frac{m_{0R}}{m_{2R}}} \tag{4}$$

and the mean period over number of years T_R can be determined by weighting these periods over the wave direction and sea states by using the wave scatter diagram. Then the number of observations or cycles during T_R years can be calculated as a ratio of total time and mean period:

$$n_{T_R years} = \frac{T_R \cdot 365 \cdot 24 \cdot 60 \cdot 60}{\sum_{i=1}^{N_\beta} \sum_{j=1}^{N_H} \sum_{k=1}^{N_T} T_{zR}(i,j,k) \cdot f_i \cdot f_{jk}} \tag{5}$$

The above analysis assumes that the observer is continuously exposed to this one wave climate at the particular location.

2.1. Extreme wave height of an extreme sea state - Design Sea State method

The local storm conditions can be described by the return periods associated with a particular environmental condition. The return period T_R is defined as the average interval of time, normally in years, in which the condition is exceeded [5]. According to Design Sea State method (DSS), adequate significant wave height and zero crossing period define an extreme sea state. The probabilities for the significant wave height are known from the wave scatter diagram and the total number of observation n during return period can be obtained from the total number of observation during observation time. The most probable extreme value expected to occur in n observations x_n, can be evaluated from the initial cumulative distribution function $F(x_n)$

$$F(x_n) \approx 1 - \frac{1}{N_{50 years}} \tag{6}$$

and the given return period (for example 50 years), as the extreme value that is most likely to occur during return period. This value is typically lying well beyond even the most extreme of the available observation. The problem is thus one of extrapolation of the observed distribution of data (collected over a relatively short period of, say, ten years) into its extreme region.

In offshore calculations is generally considered an extreme sea state with a return period of 50 years – lasting 3 hours. The resulting extreme wave amplitude in this sea state can be calculated from Rayleigh distribution,

$$\zeta_{a,\max(50\ \text{years storm})} = \sqrt{2m_{0\zeta} \cdot \ln(N)} \tag{7}$$

The number of cycles N during 3 hours storm can be determined by using an average wave period belonging to the extreme significant wave height [6, 7, 8]. However, the probability that the largest peak value may exceed the probable extreme value is quite large (≈0.632), and hence it is not appropriate to use this value for engineering design. For the purposes of structural design an extreme value for which the probability of being exceeded is some acceptably small value must be obtained. This small probability value, α, is called risk parameter. Although the decision regarding which probability value should be taken for design purposes depends on the designer's own judgment, a risk parameter value of 0.01 to 0.05 is often considered as a standard design practice. If S is the storm duration in hours the maximum amplitude, with probability of exceedance α, can be expressed as

$$\zeta_{a,\max(50\text{ years storm})} = \sqrt{2m_{0\zeta} \cdot \ln\left(\frac{N}{\alpha}\right)} = \sqrt{2m_{0\zeta} \cdot \ln\left(\frac{3600\ S}{\alpha T_z}\right)} \tag{8}$$

There are various approaches of estimating the extreme wave period associated with the design wave height which has been obtained [3]. The first is to repeat the entire procedure using wave periods instead of wave heights as statistics. An alternative approach involves using the predicted wave height to set a lower limit T_L to the wave period (due to wave breaking),

$$T_{zL} = \sqrt{\frac{32\pi H_{1/3}}{g}} \tag{9}$$

and then using a series of values of wave period above this lower limit to find the worst possible effect on the structure.

2.2. Extreme value probability distribution

The long term probability could be obtained by the above method of order statistics. This method, however, is limited by the lack of long term wave data. Often, one takes the help of a theoretical probability distribution function. The process of extrapolation plays a fundamental role in this area of analysis and therefore it is essential to empirically fit a convenient probability distribution that describes the available data as closely possible. There are several such formulas available [1]. These include the log-normal distribution and the Extremal Types I, II and III known also as Gumbel (Fisher-Tippett type I), Frechet (Fisher-Tippett type II), and Weibull (Fisher-Tippett type III) probability distribution. Although these all have a theoretical base, they are used here essentially as an empirical fit to the data. The lognormal and Weibull distribution seems to fit best the empirical wave data.

The log-normal distribution has been fitted to many wave data with a varying degree of success. The cumulative probability distribution is given as

$$F(H_{1/3}) = \Phi\left(\frac{\ln H_{1/3} - \mu_{\ln H_{1/3}}}{\sigma_{\ln H_{1/3}}}\right) = \Phi(u) \tag{10}$$

where Gamma function is given as

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{x^2}{2}\right) \mathrm{d}x \tag{11}$$

$\mu_{\ln H_{1/3}}$ and $\sigma_{\ln H_{1/3}}$ are the mean value and the standard deviation of significant wave heights natural logarithm. The line u can be expressed as

$$\mathrm{u} = \frac{1}{\sigma_{\ln H_{1/3}}} \ln H_{1/3} - \frac{\mu_{\ln H_{1/3}}}{\sigma_{\ln H_{1/3}}} \tag{12}$$

where $-\frac{\mu_{\ln H_{1/3}}}{\sigma_{\ln H_{1/3}}}$ is the intercept and $\frac{1}{\sigma_{\ln H_{1/3}}}$ is the slope.

The 3-parametar Weibull probability distribution can be expressed as

$$F(H_{1/3}) = 1 - \exp\left\{ -\left(\frac{H_{1/3} - \varepsilon}{\theta} \right)^{\alpha} \right\} \tag{13}$$

where ε represents location parameter, θ is scaling parameter and α is shape parameter. The accuracy of extreme value prediction is significantly affected by the choice of these parameters. If above expression is rearranged, it may be written as a linear equation

$$\ln\left(-\ln\left(1 - F(H_{1/3})\right)\right) = \alpha \cdot \ln\left(H_{1/3} - \varepsilon\right) - \alpha \cdot \ln\theta \tag{14}$$

where $-\alpha \cdot \ln\theta$ is the intercept and α is the slope.

One way to represents data points precisely is to express the cumulative distribution with 4-parameters function developed by Ochi [8, 9] in the form:

$$F(x) = 1 - e^{-q(x)} \tag{15}$$

$$q(x) = ax^{m} e^{-px^{k}} \tag{16}$$

where *a*, *m*, *p* and *k* are distribution parameters which have to be estimated according to empirical data.

2.3. Probability distribution parameters estimation

Having selected one distribution as a likely model, it remains to estimate the parameter values that will provide the best empirical fit between the distribution and the data. The most straightforward approach is to plot the individual data points on the selected probability paper and then draw a straight line through these by eye. The paper is usually constructed by taking the logarithm of the cumulative distribution function twice. Therefore, a small difference between data and cumulative distribution function drawn on the probability paper may result in a substantial difference between the histogram and probability density function. Another way to plot the data is to take logarithm of the return period. In the case of Weibull distributions, this would apply only if α or ε were chosen in advance. Since Weibull distribution includes all three empirical coefficients this increases the adaptability of this probability distribution to fitting empirical data. But it also increases the effort required to fit the data, as one of the three parameters must be estimated before the data can be plotted [5, 9,10]. One of the possibilities is to presume location parameter $\varepsilon = 0$ which corresponds to 2-parameter Weibull distribution and find a best fit line by one of three possible methods: the method of moments, the method of least squares and the method of maximum likelihood [8, 11]. The other way is to assume the value of shape parameter α. This parameter is then reestimeted until the best fit is obtained (for example α=0.75, 1.0, 1.4 and 2.0). All of these methods presume that one parameter is chosen in advance and the other two are then estimated in a way that obtained cumulative distribution displays the reasonably fit to majority of the data. The parameters involved in 4-parameters Ochi distribution can be determined numerically by a nonlinear minimization procedure. The extreme value can then be estimated based on this probability function. A genetic algorithm (GA) [12, 13] can be also used to solve the estimation problem of an extreme design wave. The objective is to minimize the discrepancies between the empirical probability distribution and the probability distribution with random parameters. So it is necessary to determine the range of possible values for all parameters. The GA searches the prospective solution space using an adaptive penalty to consider both feasible and infeasible solutions until converging to a feasible recommended design. The first step is to define GA control parameters. GA involves the evaluation of a population of solutions, which are revised over successive generations. Each solution is represented in the population by the vector. The crossover and mutation operators are used to introduce new prospective design solutions for each generation. Crossover involves the selection of parent solution vectors and the recombination of those vectors to produce new prospective solutions. Parent selection is random, but biased by the ordinal objective function ranking within a current population. Solutions that have been observed to be superior are more likely to be chosen. Mutation involves the addition or removal of components in accordance with a preselected mutation rate. This prevents premature convergence to local optima. The culling operator involves the selection of the certain number of solutions with the highest penalized objective function from among the prior population and the newly formed solutions. If satisfied, best feasible solution is selected in final population as recommended design.

After estimating parameter boundaries the genetic algorithm optimizes the solution by minimizing the objective function

$$Z(H_S) = \sum_{H_{1/3}} \left[F\left(H_{1/3}\right) - \hat{F}\left(H_{1/3}\right)\right]^2 \tag{17}$$

(measure for the accuracy) based on selected distribution function $F\left(H_{1/3}\right)$ and empirical distribution function $\hat{F}\left(H_{1/3}\right)$ for parameter tuple (significant wave height range) in order to fit empirical data best using values within parameter space. After a specified number of generations, the individual with the minimum of objective function value is the result. The advantage of application of GA method in the procedure of extreme wave estimation is that can be applied for the selection of a theoretical distribution (also multi-mode distributions, mixed different distributions, automatic selection of a theoretical distribution), it is uniform for all theoretical distributions, and there is no difficulty dealing with many parameters.

3. Predicted extreme values in the North Atlantic sea

The application of the computational methods is given for the SHIPREL [1, 4] wave scatter diagram that describes the North Atlantic sea environment given in Table 1. Data in the table represents the collected observations distribution of the significant wave height. The CFD curves obtained by Lognormal and Weibull model are compared with empirical data in Figure 1. The extreme wave amplitudes of 3-hours storm obtained by Lognormal and Weibull model are compared on Figure 2. The line fitting and the resulting CFD curves for Weibull model derived from GA compared with traditional Weibull model line and empirical data are presented Figure 3 and 4. The extreme wave amplitudes of 3-hours storm obtained by Weibull model derived from GA and traditional Weibull model are compared on Figure 5. The amplitudes at extreme sea state are computed for different return periods and different risk parameters. The extreme wave amplitudes obtained by All Sea State (ASS) and Design Sea State (DSS) method using Weibull distribution with GA optimized parameters are shown in Figure 6 as a function of return period and risk parameter. The similar analyzes have been done for the 4-parameter Ochi distribution. Comparison has been made between two approaches of obtaining extreme wave amplitudes: 3-parameter Weibull model and 4-parameter Ochi model. In both cases parameters are obtained by genetic algorithm approach. At Figure 7 and 8 the resulting CFD curves obtained by Ochi and Weibull model are compared with empirical data. The extreme wave amplitudes of 3-hours storm, for different return periods and different risk parameters, obtained by Weibull and Ochi model are compared on Figure 9.

Table 1 Significant wave height distribution according to SHIPREL data

Tablica 1. Distribucija značanih valnih visina prema SHIPREL podacima

$H_{1/3}$. m	11.5	10.5	9.5	8.5	7.5	6.5	5.5	4.5	3.5	2.5	1.5	0.5
$Q(H_{1/3})$.‰	0.5	2.328	6.494	14.62	30.58	61.02	119.83	227.77	408.11	657.65	899.78	999

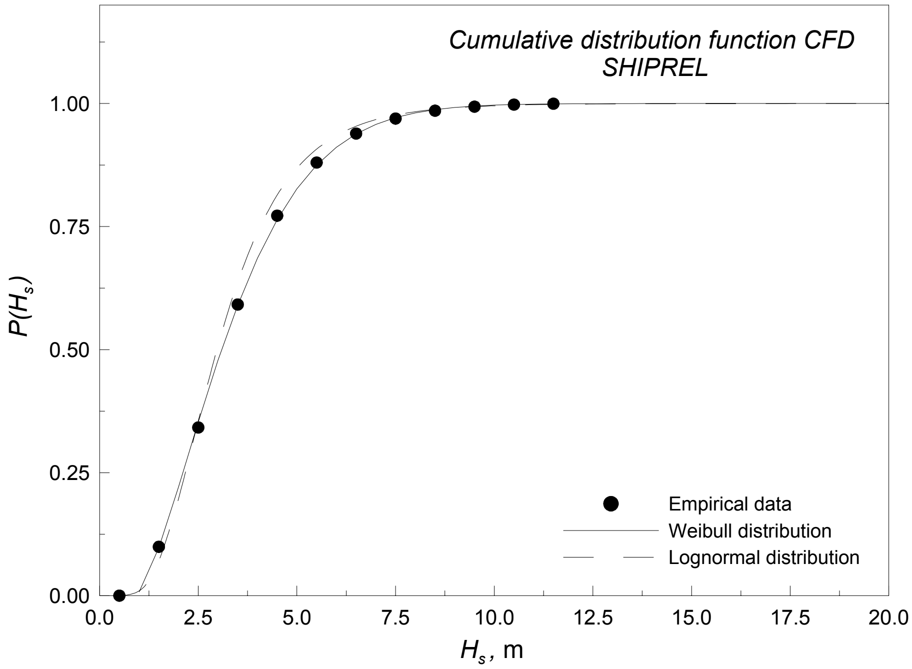

Fig. 1 Cumulative probability distribution of significant wave height (Lognormal and Weibull distribution)

Slika 1. Funkcija distribucije vjerojatnosti značajnih valnih visina (lognormalna i Weibullova distribucija)

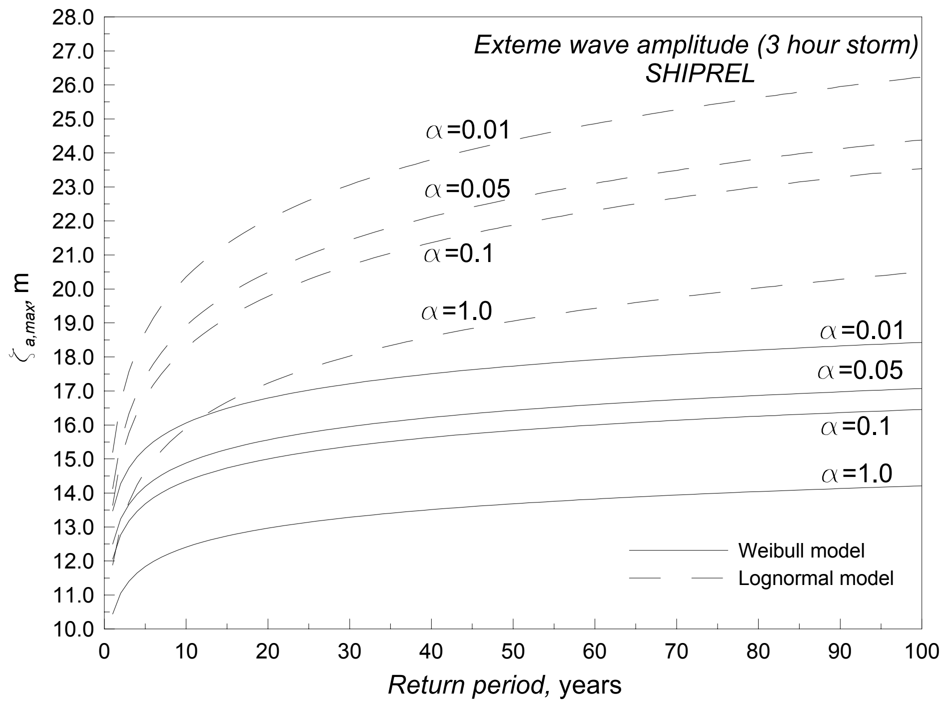

Figure 2 Extreme wave amplitude (Lognormal and Weibull distribution)

Slika 2. Ekstremne amplitude vala (lognormalna i Weibullova distribucija)

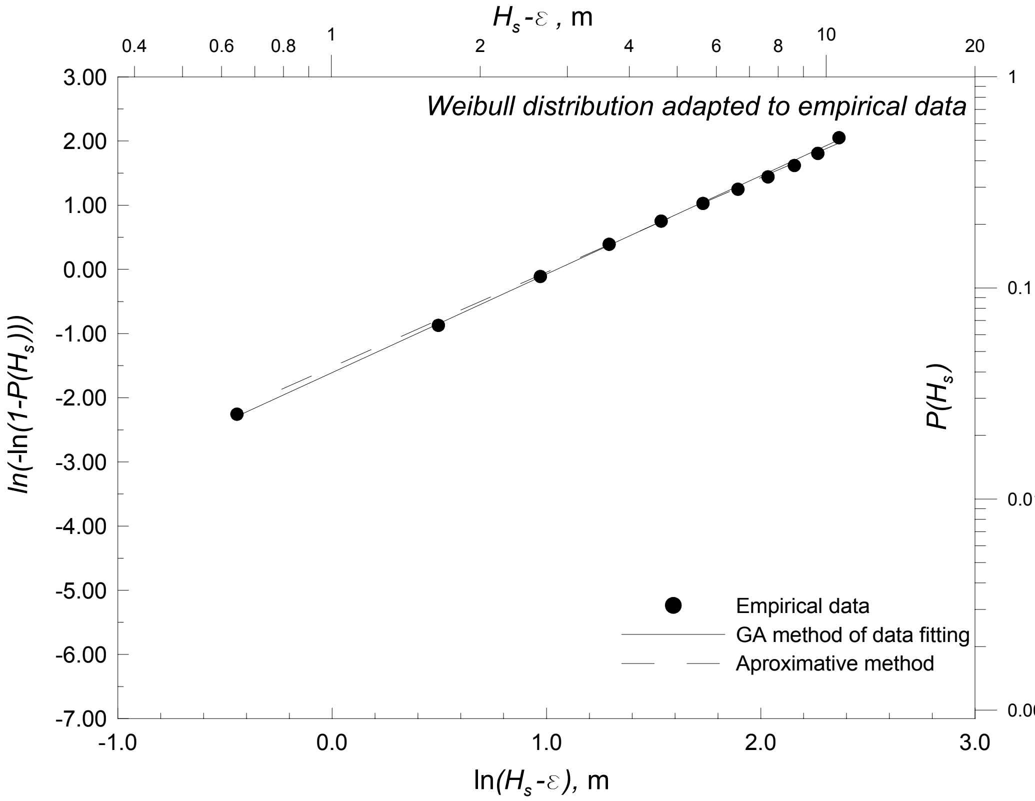

Figure 3 Weibull probability distribution of significant wave height

Slika 3. Weibullova distribucija vjerojatnosti značajnih valnih visina

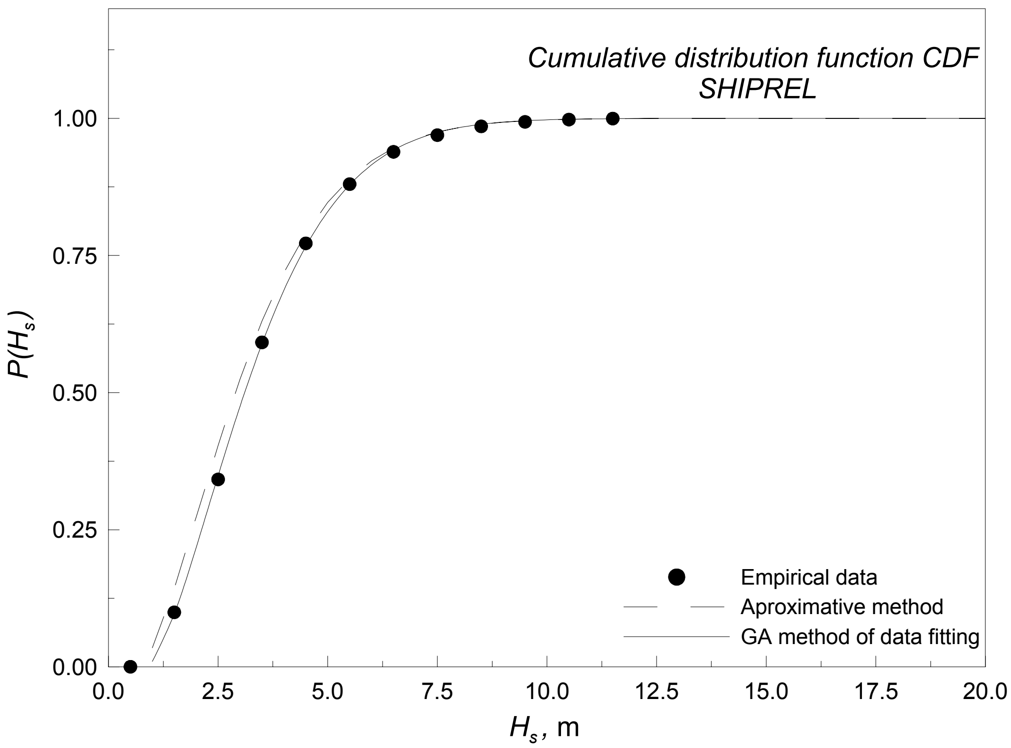

Figure 4 Weibull cumulative probability distribution of significant wave height

Slika 4. Weibullova funkcija distribucije vjerojatnosti značajnih valnih visina

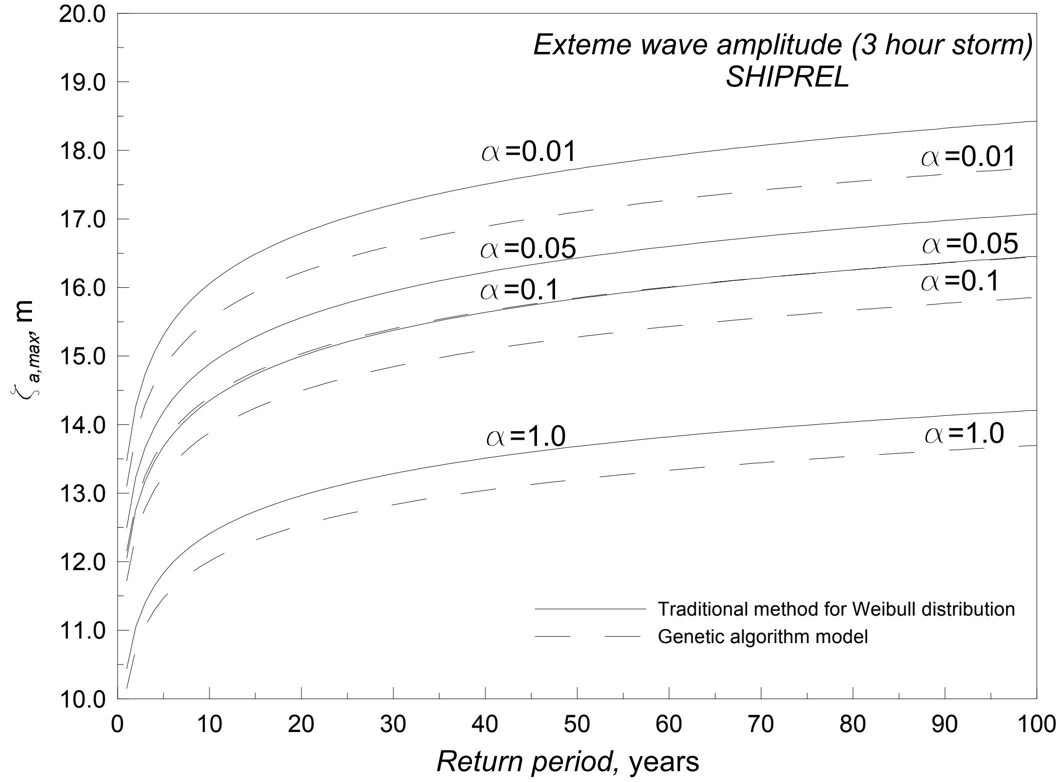

Figure 5 Extreme wave amplitude

Slika 5. Ekstremne amplitude vala

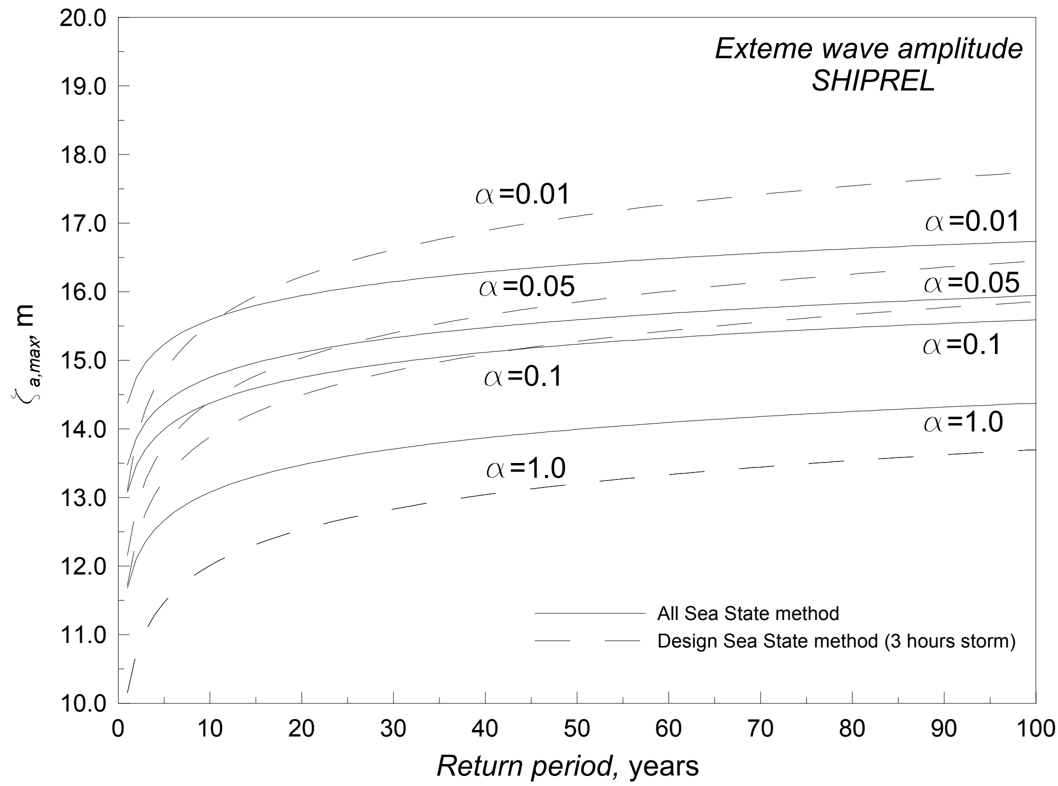

Figure 6 Extreme wave amplitude (ASS and DSS method)

Slika 6. Ekstremne valne amplitude (ASS i DSS metoda)

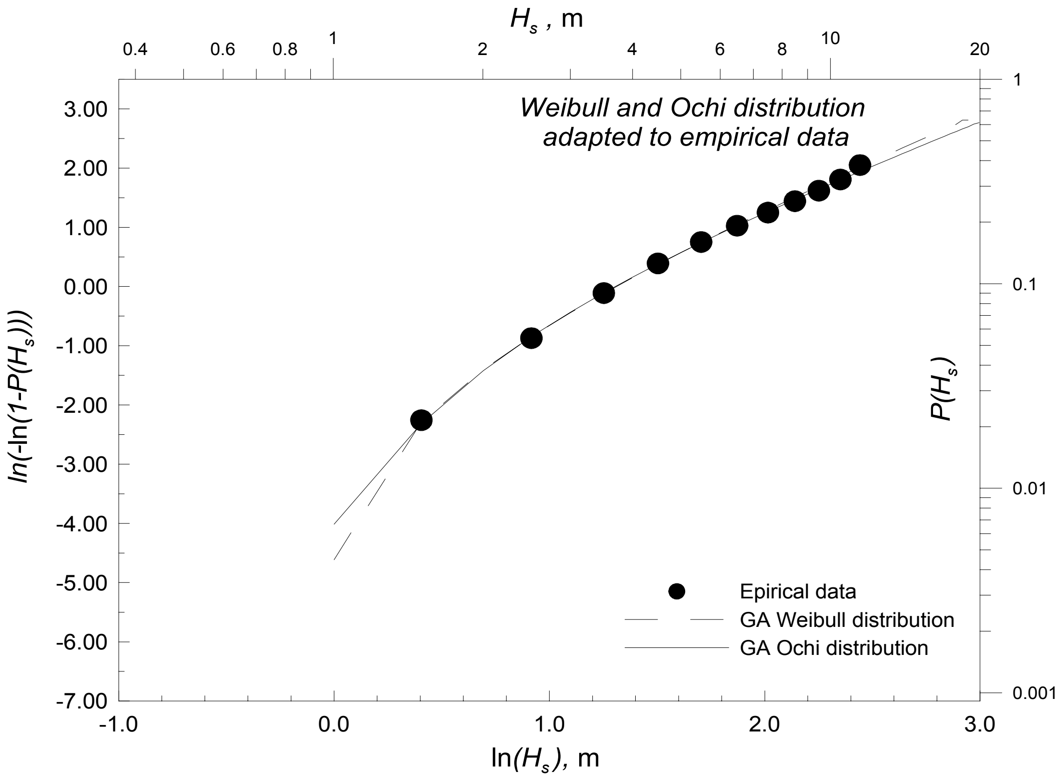

Figure 7 4-parameter probability distribution of significant wave height

Slika 7. Četiriparametarska distribucija vjerojatnosti značajnih valnih visina

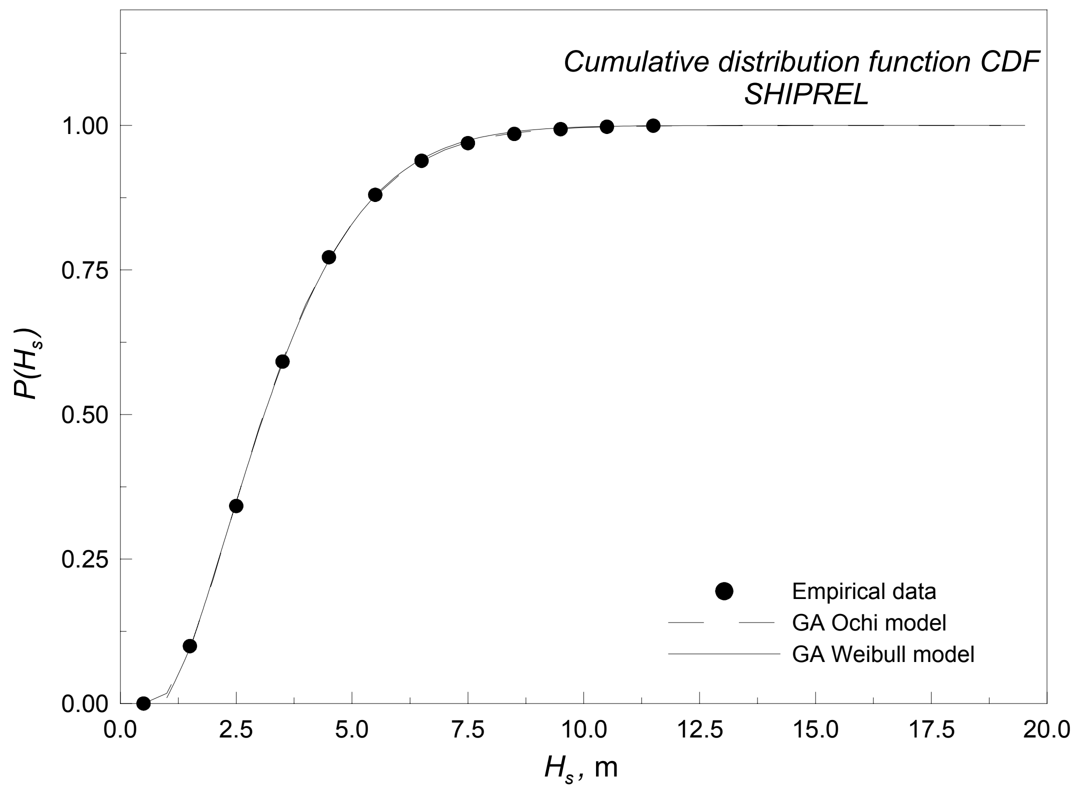

Figure 8 4-parameter probability distribution of significant wave height

Slika 8. Četiriparametarska funkcija distribucije vjerojatnosti značajnih valnih visina

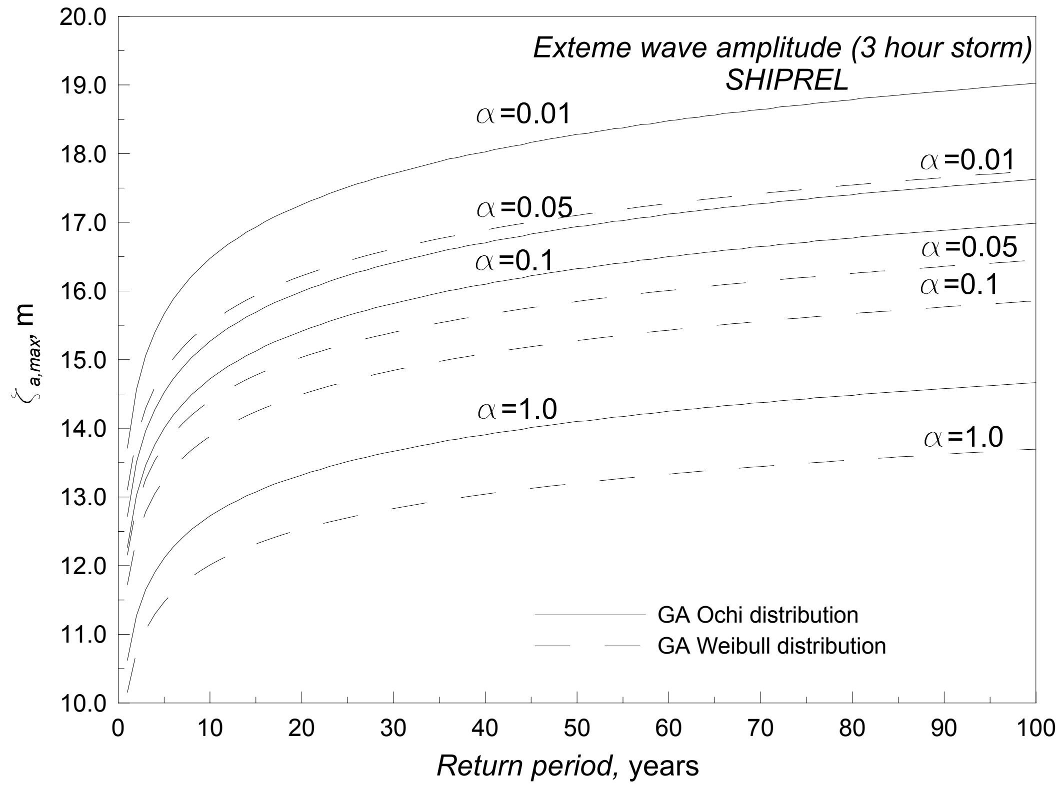

Figure 9 Extreme wave amplitude obtained by 4-parameter Ochi distribution

Slika 9. Ekstremne amplitude vala dobivene primjenom Ochijeve distribucije

4. Conclusion

The difference in extreme wave amplitudes obtained by two different methods, All Sea State and Design Sea State method, is notable especially for higher risk parameters. For SHIPREL scatter diagram difference between extreme wave amplitudes for 100-year return period and for risk parameter value of 0.01 is more then 1 m. A ship spends most of the time on small or moderate seas and therefore accuracy in estimating the characteristics of small or moderate sea states, or in calculating the ship response at these levels, could introduce an appreciable error into the estimation when using All Sea State method. For these reason, as well as for savings in computation, the Design Sea State is more appropriate for the majority of ship design application. Also, for the safety point of view the Design Sea State method can be recommended because of higher extreme values estimation.

The extreme wave amplitudes obtained by lognormal distribution are significantly different from those obtained by Weibull distribution. For example for return period of 50 years the difference in extreme wave amplitude during 3 hour storm is more then 5 m for risk parameter 1.0. The discrepancies increase as the return period increases and as risk parameter decreases. So, for the return period of 100 years and for risk parameter value of 0.01 the difference becomes even 9 m. However, the Weibull distribution seems to cover the empirical data better, which is understandable because of possibility of three-parameter tuning. The extreme wave heights obtained by application of genetic algorithm procedure of estimating Weibull parameters are slightly lower then results obtained by traditional approach. The difference in extreme wave amplitude during 3-hour storm is nearly 0.5 m for all risk parameter and for all return periods longer than 5 years. The discrepancies slightly increase as the return period increases and as risk parameter decreases. So, for the return period of 100 years and for risk parameter value of 0.01 the difference becomes 0.6 m. The Weibull distribution obtained by GA approach seems to cover the empirical data better because of simultaneous tuning of all parameters, whereas for traditional approach one of the parameters is predefined. Ochi four parameters distribution, which has the possibility to set four parameters in order to obtain the best fit curve, covers the empirical data even better. For the return period of 100 years and for risk parameter value of 0.01 the extreme wave amplitude obtained by Weibull distribution is 17.6 m while the same obtained by Ochi distribution is 19.0 m.

The use of GA optimization procedures for estimating parameters of distribution in order to predict an extreme design wave on the basis of recorded or hindcast data is well suited and can be generalized to consider a broader range of problems. Genetic algorithm is uniform for all theoretical distributions; hence it is applicable for multi-mode and mixed distributions as well as for the selection of a theoretical distribution that covers the given data best. For the estimation of extreme values, it is highly desirable to represent the data precisely by a certain probability distribution over the entire range of the cumulative distribution and, generally speaking, that can be better obtained if the distribution is multi-parameter. The advantage of GA optimization procedures application in estimation of distribution parameters is that the greater number of distribution parameters does not represent any problem.

REFERENCES

[1] Soares, C.G., Dogliani, M., Ostergaard, O., Permentier, G., Pedersen, P.T.: Reliability Based Ship Structural Design, *SNAME Transactions*, Vol. 104, 357-389, 1996.

[2] R Chakrabarti, S.K.: Hydrodynamics of Offshore Structures, Computational Mechanics Publications, WIT Press, Southampton, 2001.

[3] Sarpkaya, T., Isaacson, M.: Mechanics of Wave Forces on Offshore Structures, Van Nostrand Reinhold Company Inc., New York, 1981.

[4] Hogben, N., Dacunha, N.M.C., Olliver, G.F.: Global Wave Statistics, British Maritime Technology Ltd., Feltham, 1986.

[5] Mansour, A.E., Priston, D.B.: Return Periods and Encounter Probability, *Applied Ocean Research*, Vol. 17, 127-136, 1995.

[6] Hughes, O. F.: Ship Structural Design – A Rationally-Based, Computer Aided Optimization Approach, John Wiley & Sons, New York, 1983.

[7] Ochi, M.K.: Wave Statistics for the Design of Ships and Ocean Structures, *SNAME Transactions*, Vol. 86, 47-76, 1978.

[8] Ochi, M.K.: Ocean Waves – The Stochastic Approach, Cambridge Ocean Technology Series 6, Cambridge, 1998.

[9] Mansour, A.E.: An Introduction to Structural Reliability Theory, Ship Structure Committee, SSC-351, Washington, 1990.

[10] Parunov, J., Senjanović, I.: Methods for Long Term Prediction of Extreme Sea States (in Croatian), Journal of Naval Architecture and Shipbuilding Industry (Journal Brodogradnja), Vol. 48, No. 2, Zagreb, 2000.

[11] Oliver, J. C., Ochi, M. K.: Evaluation of SL-7 Scratch Gauge Data, Ship Structure Committee, SSC-311, Washington, 1981.

[12] Goldberg, D. E.: Genetic Algorithms in Search, Optimization, and Machine Learning, Addison-Wesley Reading, Massachusetts, USA, 1989.

[13] Coit, D. W., Smith, A. E.: Genetic algorithm to maximize a lower-bound for system time-to-failure with uncertain component Weibull parameters, Computers & Industrial Engineering 41, 2002.