Discussion on analytical method of three parameters of Weibull distribution of environmental parameters

Chen Hai-long[1]，Li Lin-bin[2]，Xu Hui[1]

1.CCS Offshore Engineering Technology Center, Tianjin, 300457 China

2.CCS Offshore Engineering Department, Beijing, 100007 China

Key words: Maximum likelihood method; Statistical moment method; Least squares method; Three parameters of Weibull distribution; Optimization;

ABSTRACT

The maximum likelihood method, statistical moment method and least squares method are used to analyze the three parameters of Weibull distribution, and the three-parameter analysis of the Weibull distribution is optimized by means of element elimination and order reduction and automatic software solution.

1. Introduction

In marine engineering, the environmental parameters of the 100-year return period are usually used for the ultimate limit state (ULS) check of the structure. When the accidental limit state (ALS) is checked, the 1 000-year return period needs to be used. Environment parameters [1][2]. These environmental parameters are generally obtained by observing and recording the sea conditions of a specific sea area within a sufficiently long period of time (generally no less than several decades), and then fitting the distribution of the observed data to derive the environmental parameters of the expected return period. However, the prediction of large-value return period environmental parameters has a high degree of uncertainty, including statistical uncertainty and uncertainty of simulation fitting, these factors will lead to the deviation of the prediction results [3]. In order to reduce the uncertainty in the simulation fitting of the environmental parameter distribution, it is necessary to adopt the optimal fitting technique.

The modeling of long-term statistical data of marine environmental parameters (Significant wave height Hs, current velocity Uc) usually adopts the three parameter Weibull distribution. The three parameters in the Weibull distribution are the shape parameter (α), the scale parameter (β) and the location parameter (γ). Due to the nonlinear and multi-parameter characteristics of this distribution, it brings challenges to solving the three-parameter Weibull distribution. In engineering, people usually use methods including maximum likelihood method, statistical moment method, and least squares method to fit Weibull distribution. Among them, the curves fitted by the least squares method and the maximum likelihood method have the advantages of low deviation and low coefficient of variation [3], and the statistical moment method has high accuracy and is simple and fast. In this paper, the above three methods are studied to solve the three parameters of Weibull distribution, and the conventional solution of the three parameters of Weibull distribution is given. At the same time, the solution of the three parameters of Weibull distribution is optimized by means of element elimination, order reduction and software automatic solution, which provides a reference for fitting the marine environmental parameters and deducing the environmental parameters in the expected return period.

2. Solving Weibull three parameters by maximum likelihood method

Maximum Likelihood Estimation (MLE) is to select the parameter value that makes L(θ) reach the maximum value within the possible range of parameter θ as the estimated value of the parameter.

Likelihood function is a function about the parameters in the statistical model, which indicates the likelihood of the parameters in the model. Likelihood is similar to probability in meaning, and refers to the possibility of an event. However, in statistics, probability describes the rationality of the result after given model parameters, and does not involve any observed data. Likelihood describes whether the model parameters are reasonable after given specific observations.

2.1 Conventional Solution of Maximum Likelihood Method (MLE)

The likelihood function of three-parameter Weibull distribution is the probability density function, as shown in formula (1)：

$$F_X(x) = \frac{\alpha}{\beta^\alpha}(x-\gamma)^{\alpha-1}\exp\left(-\left(\frac{x-\gamma}{\beta}\right)^\alpha\right) \quad (1)$$

Where x is a random variable of long-term statistical data of waves.

Then, the probability density function is used to establish the likelihood function of wave samples, as shown in formula (2):

$$L(x_1,\dots,x_n,\alpha,\beta,\gamma)=\prod_{i=1}^{n}F_X(x_1,\dots,x_n,\alpha,\beta,\gamma)=\left(\frac{\alpha}{\beta^\alpha}\right)^n\prod_{i=1}^{n}(x_i-\gamma)^{\alpha-1}\exp\left(-\sum_{i=1}^{n}\left(\frac{x_i-\gamma}{\beta}\right)^\alpha\right)\quad(2)$$

Take the natural logarithm of both sides of formula (2) at the same time to obtain formula (3):

$$\ln\big(L(x_1,\dots,x_n,\alpha,\beta,\gamma)\big)=n\ln\left(\frac{\alpha}{\beta^\alpha}\right)+(\alpha-1)\sum_{i=1}^{n}\ln(x_i-\gamma)-\frac{1}{\beta^\alpha}\sum_{i=1}^{n}(x_i-\gamma)^\alpha\qquad(3)$$

Find the partial derivatives of α, β and γ in formula (3), and get the ternary equations composed of formulas (4), (5) and (6):

$$\frac{n}{\alpha}+\sum_{i=1}^{n}\ln\left(\frac{x_i-\gamma}{\beta}\right)-\sum_{i=1}^{n}\left(\left(\frac{x_i-\gamma}{\beta}\right)^\alpha\ln\left(\frac{x_i-\gamma}{\beta}\right)\right)=0\qquad(4)$$

$$\beta^\alpha-\frac{1}{n}\sum_{i=1}^{n}(x_i-\gamma)^\alpha=0\qquad(5)$$

$$(\alpha-1)\sum_{i=1}^{n}\frac{1}{x_i-\gamma}-\frac{\alpha}{\beta^\alpha}\sum_{i=1}^{n}(x_i-\gamma)^{\alpha-1}=0\qquad(6)$$

Finally, the equations can be solved by MATLAB, Mathcad and other analysis software, and the three parameters of Weibull distribution can be obtained.

It should be noted that the uncertainty of fitting Weibull distribution includes fitting parameters and threshold. When maximum likelihood is applied to a small sample size (for example, n≤100), the uncertainty of parameters is more significant than the uncertainty of threshold. If necessary, bootstrap method can be considered to eliminate bias [3].

2.2 Optimization Solution of element elimination and order reduction

Through observation, it is found that substituting β^α in formula (5) into formula (3), then β can be eliminated, and formula (7) is obtained:

$$\ln\big(L(x_1,\dots,x_n,\alpha,\beta,\gamma)\big)=n\ln(\alpha)-n\ln\left(\sum_{i=1}^{n}(x_i-\gamma)^\alpha\right)+n\ln(n)+(\alpha-1)\sum_{i=1}^{n}\ln(x_i-\gamma)-n\qquad(7)$$

Then, according to the maximum likelihood method, the partial derivatives of the parameters (α and γ) in formula (7) are obtained, and the binary equations composed of formulas (8) and (9) are obtained:

$$\frac{n}{\alpha}-\frac{n\sum_{i=1}^{n}\left((x_i-\gamma)^\alpha\ln(x_i-\gamma)\right)}{\sum_{i=1}^{n}(x_i-\gamma)^\alpha}+\sum_{i=1}^{n}\ln(x_i-\gamma)=0\qquad(8)$$

$$\sum_{i=1}^{n}(x_i-\gamma)^{\alpha-1}-\frac{\alpha-1}{n\alpha}\sum_{i=1}^{n}(x_i-\gamma)^\alpha\sum_{i=1}^{n}\frac{1}{x_i-\gamma}=0\qquad(9)$$

The equations are solved by MATLAB, Mathcad and other analysis software, and the parameter (α and γ) of Weibull distribution is obtained, and then the parameter (β) are solved according to formula (5).

By means of element elimination and order reduction, the difficulty of solving the original ternary equations is reduced, and the solution accuracy is consistent.

2.3 Optimization Solution by using automatic solution function of software

Modern analysis software can realize the above calculation process through simple programming, such as:

Maximize function in Mathcad software is used to solve the three parameters of Weibull distribution under maximum likelihood. Set formula (3) as function (10):

$$MLE\_3P(\alpha,\beta,\gamma)=n\ln\left(\frac{\alpha}{\beta^\alpha}\right)+(\alpha-1)\sum_{i=1}^{n}\ln(x_i-\gamma)-\frac{1}{\beta^\alpha}\sum_{i=1}^{n}(x_i-\gamma)^\alpha\qquad(10)$$

Use the function $Maximize(MLE\_3P,\alpha,\beta,\gamma)$, The variables α, β and γ can be obtained.

Similarly, the formula (7) in 2.2 can be set as the function $MLE\_2P(\alpha,\gamma)$, and then the parameters (α and γ) can be obtained by using the function $Maximize(MLE\_2P,\alpha,\gamma)$, and then solved β according to the formula (5).

The thresholds of the three parameters of Weibull distribution are relatively narrow, in which α is generally between [0,10] and γ is [0,Min(x_i)]. Therefore, there is a method [4] to use the data/programming solving function of MS Excel, set the threshold of α and γ, and then to solve the bivariate equation (7). Through trial calculation, this method has a low success rate and often fails to converge, which is not as reliable and accurate as the above-mentioned method using the Maximize function of Mathcad in this paper.

3. Solving Weibull three parameters by statistic moment methods

3.1 Conventional Solution of Statistic Moment (SM)

According to the survey data of marine environment, statistical data $H_s(x_1,\dots,x_n)$ and its corresponding dispersion probability $P_{Hs}(p_1,\dots,p_n)$ are mostly given in the form of scatter diagram. When there is no cumulative probability of samples, Gringorten estimation method can be adopted, namely: $P_i = (i - 0.44)/(n + 0.12)$.Therefore, statistical moment (SM) can be used to calculate the mean, standard deviation and skewness of data distribution [5]:

$$\mu = \sum_{H_s}(H_s) \cdot P_{Hs} \tag{11}$$

$$\sigma = \sqrt{\sum_{H_s}(H_s - \mu)^2 \cdot P_{Hs}} \tag{12}$$

$$\delta = \frac{\sum_{H_s}(H_s - \mu)^2 \cdot P_{Hs}}{\sigma^3} \tag{13}$$

The long-term Weibull distribution (α, β and γ) is established by using the statistical moments obtained from scatter diagram, and the relationship between the Weibull distribution and the statistical values establishes a ternary equation system consisting of formulas (14), (15) and (16):

$$\mu = \beta\, \Gamma\left(1 + \frac{1}{\alpha}\right) + \gamma \tag{14}$$

$$\sigma = \beta\sqrt{\Gamma\left(1 + \frac{2}{\alpha}\right) - \Gamma\left(1 + \frac{1}{\alpha}\right)^2} \tag{15}$$

$$\delta = \left(\frac{\beta}{\sigma}\right)^3 \left(\Gamma\left(1 + \frac{3}{\alpha}\right) - 3\Gamma\left(1 + \frac{1}{\alpha}\right)\Gamma\left(1 + \frac{2}{\alpha}\right) - 2\Gamma\left(1 + \frac{1}{\alpha}\right)^3\right) \tag{16}$$

Then, the ternary equations are solved by MATLAB, Mathcad and other calculation software, and the three parameters (α, β and γ) of Weibull distribution are obtained.

3.2 Optimization Solution by using automatic solution function of software

Through observation, it is found that the function in formula (15) can be converted, as shown in formula (17):

$$\frac{\sigma}{\beta} = \sqrt{\Gamma\left(1 + \frac{2}{\alpha}\right) - \Gamma\left(1 + \frac{1}{\alpha}\right)^2} \tag{17}$$

Substituting equation (17) into equation (16) can obtain the unary equation of parameters:

$$\delta = \left(\Gamma\left(1 + \frac{2}{\alpha}\right) - \Gamma\left(1 + \frac{1}{\alpha}\right)^2\right)^{-\frac{3}{2}} \left(\Gamma\left(1 + \frac{3}{\alpha}\right) - 3\Gamma\left(1 + \frac{1}{\alpha}\right)\Gamma\left(1 + \frac{2}{\alpha}\right) - 2\Gamma\left(1 + \frac{1}{\alpha}\right)^3\right) \tag{18}$$

Then use MATLAB, Mathcad and other analysis software to solve the α value of formula (18), or use the function of data/simulation analysis/single variable solution (Goal seek) in Ms Excel to solve the value, and finally obtain β and γ from formulas (15) and (14).

4. Solving Weibull three parameters by least squares estimation methods

Least squares estimation (LSE) is a mathematical optimization technique. It seeks the best function matching and curve fitting of data by minimizing the sum of squares of errors. This method was invented by French scientist Legendre in 1806.

4.1 Conventional Solution of Least Squares Estimation (LSE)

The wave sample $F_{Hs}(x_1, \ldots, x_n)$ satisfies the three-parameter Weibull distribution, and the cumulative probability is the following formula (19):

$$F_{Hs}(x_i) = 1 - \exp\left(-\left(\frac{x_i - \gamma}{\beta}\right)^\alpha\right) \tag{19}$$

Take the natural logarithm of both sides of formula (19) twice at the same time, and convert it into linear function formula (20):

$$\ln\left(ln\left(\frac{1}{1 - F_{Hs}(x_i)}\right)\right) = \alpha \ln(x_i - \gamma) - \alpha \ln\ (\beta) \tag{20}$$

Let the linear function $Y_i = \mathrm{a}\, X_i + \mathrm{b}$, where $Y_i = \ln\left(ln\left(\frac{1}{1 - F_{Hs}(x_i)}\right)\right)$, $X_i(\gamma) = \ln(x_i - \gamma)$, $\mathrm{a} = \alpha$, $\mathrm{b} = -\alpha \ln\ (\beta)$, According to the principle of least square method

$$\mathrm{a}(\gamma) = \frac{n\sum_{i=1}^{n}(X_i(\gamma)Y_i) - \sum_{i=1}^{n}(X_i(\gamma))\sum_{i=1}^{n}(Y_i)}{n\sum_{i=1}^{n}(X_i(\gamma))^2 - \left(\sum_{i=1}^{n} Y_i\right)^2} = \alpha(\gamma) \tag{21}$$

$$\mathrm{b}(\gamma) = \frac{1}{n}\sum_{i=1}^{n}(Y_i) - \frac{a}{n}\sum_{i=1}^{n}(X_i(\gamma)) = \overline{Y} - b\overline{X(\gamma)} \tag{22}$$

$$\beta(\gamma) = \exp\left(-\frac{b(\gamma)}{a(\gamma)}\right) \tag{23}$$

In the above solving process, X_i, a, b, α and β are function of γ. According to the principle of least square method, the γ value can be obtained by solving the square sum formula (24) which minimizes the error.

$$\min SS(\gamma) = \sum_{i=1}^{n}\left(Y_i - \left(\mathrm{a}(\gamma)\, X_i(\gamma) + \mathrm{b}(\gamma)\right)\right)^2 \quad (24)$$

Then use MATLAB, Mathcad and other analysis software to solve the α value of formula (24), or use the function of data/simulation analysis/single variable solution (Goal seek) in Ms Excel to solve the γ value.

4.2 Simplified Solution of least squares estimation

Using the function logfit (vx,vy,vg) in the calculation software Mathcad, the vector with logarithmic curve coefficients is returned by estimating vg. The curve is $a\ln(x_i + b) + \mathrm{c}$, and it is the best approximation of the data in vx and vy. In which vx is the sample data of $F_{Hs}(x_1, \dots, x_n)$, and vy is $\ln\left(\ln\left(\frac{1}{1-F_{Hs}(x_i)}\right)\right)$, so Weibull three parameters can be obtained.

5. Engineering case analysis

Based on the samples (n = 164) observation data of typhoon effective wave heights in a certain sea area for 56 years (Table 1), the significant wave heights of typhoons with return periods of 10 years, 100 years and 1000 years are deduced.

Table 1 Samples of wave height

Samples	Height (m)	Samples	Height (m)	Samples	Height (m)	Samples	Height (m)	Samples	Height (m)	Samples	Height (m)
1	4.074	29	4.555	57	5.008	85	5.748	113	6.798	141	8.408
2	4.074	30	4.570	58	5.029	86	5.788	114	6.812	142	8.503
3	4.105	31	4.628	59	5.031	87	5.801	115	6.868	143	8.529
4	4.116	32	4.632	60	5.056	88	5.821	116	6.883	144	8.622
5	4.161	33	4.668	61	5.107	89	5.851	117	6.897	145	8.968
6	4.188	34	4.676	62	5.123	90	5.880	118	6.928	146	9.036
7	4.204	35	4.686	63	5.143	91	5.941	119	6.987	147	9.239
8	4.207	36	4.697	64	5.143	92	5.947	120	7.108	148	9.319
9	4.214	37	4.707	65	5.147	93	5.957	121	7.108	149	9.327
10	4.260	38	4.712	66	5.169	94	5.959	122	7.277	150	9.364
11	4.266	39	4.737	67	5.206	95	6.042	123	7.512	151	9.491
12	4.272	40	4.803	68	5.226	96	6.063	124	7.587	152	9.979
13	4.293	41	4.807	69	5.228	97	6.071	125	7.615	153	10.007
14	4.296	42	4.832	70	5.228	98	6.136	126	7.743	154	10.237
15	4.300	43	4.835	71	5.270	99	6.183	127	7.796	155	10.242
16	4.308	44	4.838	72	5.308	100	6.241	128	7.825	156	10.321
17	4.339	45	4.850	73	5.314	101	6.249	129	7.836	157	10.385
18	4.349	46	4.861	74	5.417	102	6.372	130	7.839	158	10.689
19	4.359	47	4.868	75	5.436	103	6.392	131	7.881	159	11.026
20	4.402	48	4.875	76	5.442	104	6.443	132	7.885	160	11.273
21	4.440	49	4.881	77	5.467	105	6.493	133	7.899	161	11.629
22	4.443	50	4.894	78	5.468	106	6.540	134	7.908	162	12.834
23	4.463	51	4.911	79	5.514	107	6.545	135	7.933	163	12.881
24	4.490	52	4.944	80	5.514	108	6.604	136	7.952	164	14.878
25	4.491	53	4.946	81	5.534	109	6.697	137	7.958		
26	4.527	54	4.960	82	5.576	110	6.745	138	7.972		
27	4.531	55	4.972	83	5.650	111	6.774	139	7.985		
28	4.535	56	5.001	84	5.709	112	6.787	140	8.046		

Because we are concerned about the extreme value of the return period, we need the probability curve to better fit the large sample value, so as to extrapolate and predict the extreme value of the ultra-long return period (such as 1 000 years) more reasonably. It is necessary to filter out a certain number of samples with small values by setting a reasonable threshold. Through trial calculation, the threshold value is set at about 6.883m, and the curve has the best fitting degree, so 116~164 sample data are selected for analysis. From this, it can be estimated that the number of typhoons occurring every year is N = (164-115)/56 = 0.875 (event), and the exceeding probability of wave height in the expected return period (RP) year is equation (25):

$$P(RP) = \frac{1}{RP \cdot N} \tag{25}$$

The exceeding probability of the sample satisfying the three-parameter Weibull distribution is:

$$P_{HS}(x) = \exp\left(-\left(\frac{x-\gamma}{\beta}\right)^{\alpha}\right) \tag{26}$$

By substituting equation (25) into equation (26), the corresponding estimated wave height at the expected return period (RP) can be obtained.

Table 2 Weibull distribution parameters and estimated wave height in expected return period

Methods	Weibull distribution					wave height (m)			
	α	β	γ	μ	σ	Cov	10yr	100yr	1000yr
MLE	1.028	2.021	6.882	8.88	1.94	0.22	11.17	15.55	19.87
SM	1.196	2.162	6.769	8.80	1.71	0.19	10.90	14.34	17.48
LSE	1.190	2.260	6.795	7.74	0.80	0.10	11.13	14.75	18.07

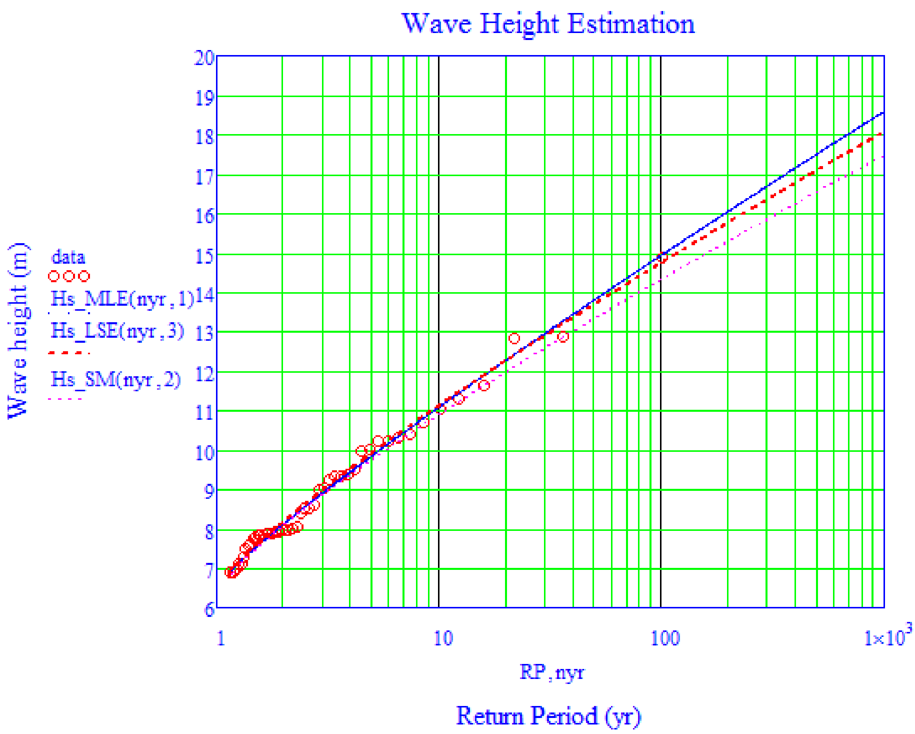

Figure 1 Return period (year) vs wave height (m)

It can be concluded from the data in Table 2 that the coefficient of variation (Cov) of curves simulated by all methods is relatively small, and the coefficient of variation of the least square method is the smallest.

From the fitting situation in Figure 1, all methods can fit the sample data well. When predicting the wave height in an ultra-long return period (such as 1 000 years), the wave height predicted by the least square method is the best, the wave height predicted by the maximum likelihood method is slightly larger, and the wave height predicted by the statistical moment method is slightly smaller.

6. Conclusion

In this paper, maximum likelihood method, statistical moment method and least square method are used to give the conventional solution of Weibull three parameters, and the solution of Weibull distribution is also optimized by means of element elimination, order reduction and automatic solution by software, which provides a reference for the fitting of marine environmental parameters and the derivation of parameters in expected return period.

The analysis shows that the above three methods can well fit the curve of marine environmental parameters, among which the curve fitted by the least square method has the smallest deviation.

In addition, when using the least square method or statistical moment method to fit the curve, there is no need for professional calculation software, and the function of data/simulation analysis/single variable solution(Goal seek) in MS Excel can be used to complete the solution of Weibull three parameters, which is relatively simpler and faster.

References

(1)王亮, 刘玉玺, 黄怀州,等. API RP 2A-WSD 22 版规范的更新概述及其工程影响[J]. 船海工程, 2018, 47(1):105.

(2)American Petroleum Institute， Recommended Practice for Planning， Designing and Constructing Fixed Offshore Platforms-Working Stress Design [S]. Twenty-Second Edition 2014.

(3)Gibson R , Grant C , Forristall G Z , et al. omae2009-79466 bias and uncertainty in the estimation of extreme wave heights and crests omae2009-79466.

(4)史景钊, 任学军, 陈新昌,等. 一种三参数 Weibull 分布极大似然估计的求解方法[J]. 河南科学, 2009, 27(7):3.

(5)DNVGL-RP-F105， Free spanning pipelines [S]. Edition June 2017.

Author introduction

Chen Hai-long, male, now working in the Offshore Technology Center of China Classification Society, with the title: surveyor, senior structural engineer, specialized in submarine pipeline. Address: Block A, Mingzhu Garden, No.11 Nanhai Road, Tianjin Economic and Technological Development Zone, China. Code 300457, tel 022-66375631-3138, and e-mail address: hlchen@ccs.org.cn.