

# 波高的长期极值统计分布\*

王运洪

董胜

(天津市水利局 天津 300074)

(青岛海洋大学工程学院 青岛 266003)

**提要** 通过对国内外常用的 4 种极值波高分布模式的拟合与比较,得到了以下结论:(1)由于地区差异,港口工程技术规范给出的单一模式并非具有普遍性。对同一工程应该采用多种理论分布进行计算比较,从中选择最佳模式。(2)应用麦奎尔特法拟合 Weibull 分布实现了对未知参数的一举寻优,解的收敛速度快,结果稳定且精度高。同时对其它几种分布实现了资料的微机化处理。(3)本文算例用 4 种极值分布对不同重现值波高进行了比较,所得到的最小概率波高各不相同:Gumbel 分布得到的重现值最大;Log-normal 和 P-III 型分布得到的值次之;Weibull 分布得到的值最小。

**关键词** 极值分布拟合 麦奎尔特法 重现值

**学科分类号** P731.22

目前,国内外海岸及海洋工程设计波高的获取,大多采用概率统计分析的方法,即假设年极值波高符合某种理论分布,如 Gumbel 分布、P-III 型分布、Log-normal 分布或 Weibull 分布等,再推算出某一重现期的波高设计值。在工程设计规范中大多规定某一线型,如 Gumbel 分布,来求解多年一遇设计值。由于地域的差异,环境的不同,没有任何一种分布可适合于所有的工程条件,因此,对实际的观测序列应进行多种理论分布拟合的比较,以理论频率与经验频率之离差平方和最小为适线标准,确定出最佳概率分布模式。常规法拟合 Weibull 分布参数时,采用图解法或者先确定分布的位置参数,再用最小二乘法求解形状参数和尺度参数。本文则以麦奎尔特法实现了 Weibull 分布参数的寻优,提高了计算结果的精度。本文对几种极值分布型式实现了资料的微机化处理,对计算的重现值进行了大小的比较,所得结论对于海洋、海岸工程设计有实用意义。

## 1 几种极值波高理论分布模式

### 1.1 Weibull 分布

Weibull 分布是由 Weibull(1951)首先提出来,其分布函数为:

$$F(x) = 1 - \exp \left[ - \left( \frac{x-a}{b} \right)^c \right] \quad (1)$$

\* 国家科委九五攻关项目,96-922-03-03号。王运洪,男,出生于 1955 年 8 月,博士后,博士导师,研究员,E-mail: yhwang@public.tpt.tj.cn

收稿日期: 1997-04-04, 收修改稿日期: 1998-01-13

式中,  $x$  为波高;  $a$  为位置参数;  $b$  为尺度参数;  $c$  为形状参数。

由于 Weibull 分布函数有三个可调参数, 因而适用性广, 但在工程设计中, 适配曲线求解时多采用图解法, 将观测数据点绘于 Weibull 概率纸上, 若各点大致位于一条直线上, 从图上读取数值, 进而间接计算出参数  $a, b, c$  的值, 试算时需反复调整, 工作量大。近些年, 已有学者(陈上及等, 1991)采用分步最小二乘法求解  $a, b, c$  之值, 即先确定位置参数  $a$ , 再用最小二乘法推求参数  $b$  和  $c$ , 这种方法收敛的快慢取决于  $a$  的初始值离精确解的远近。本文则以麦奎尔特(Marquardt, 1963)法求解 Weibull 分布, 实现了三参数的一举寻优, 计算表明, 拟合精度较以往方法精度高, 且编程计算, 省却了大量手工劳动, 便于成批资料的处理。

麦奎尔特法微机化算法如下: 设函数  $y = f(x; P)$ , 参数  $P_k (k = 1, 2, \dots, m)$ , 已知  $n$  组观察值  $(x_i, y_i)$ ,  $(i = 1, 2, \dots, n)$ ,  $m \in n$ , 采用经验与理论频率的残差平方和最小为准则求解  $P_k, (k = 1, 2, \dots, m)$ ,

令: 
$$Q(P) = \sum_{i=1}^n [y_i - y(x_i, P)]^2 \rightarrow \min \tag{2}$$

设:  $P^{(k+1)} = P^{(k)} + \Delta P^{(k)}$ , 将  $Q(P)$  在  $P^{(k)}$  处作泰勒展开, 取前二次项。

令:  $\nabla Q(P^{(k+1)}) = 0$ , 则  $\nabla^2 Q(P^{(k)}) \Delta P^{(k)} = -\nabla Q(P^{(k)}) \quad k = 0, 1, \dots \tag{3}$

设  $\mu$  为常数, 以最大负梯度方向作为参数迭代的改进方向, 则:

$$\Delta P^{(k)} = \mu [-\nabla Q(P^{(k)})] \tag{4}$$

记  $[\alpha] = \nabla^2 Q(P^{(k)}) \tag{5}$

$$\beta = -\nabla Q(P^{(k)}) \tag{6}$$

令:  $\Delta P = \Delta P^{(k)}$ , 则式(3)和式(4)变为

$$[\alpha] \Delta P = \beta \tag{7}$$

$$\Delta P = \mu \beta \tag{8}$$

令:  $\lambda = \frac{1}{\mu}$ , 则:  $\lambda \Delta P = \beta$ 。设  $m$  阶单位阵为  $I$ , 则:  $\lambda I \Delta P = \beta \tag{9}$

合并式(7)和式(9)可得用于微机迭代的计算式为

$$([\alpha] + \lambda I) \Delta P = \beta \tag{10}$$

若  $\lambda$  为 0, 表示前后两次迭代的  $\Delta P$  相等, 即求出的参数为最优。编程计算时, 可给定初始参数  $P^{(0)}$ , 计算  $Q(P^{(0)})$ , 取  $\lambda = 0.1$  解式(10)得  $\Delta P$ , 令  $P^{(1)} = P^{(0)} + \Delta P$ , 计算  $Q(P^{(1)})$ 。若  $|\lambda| < 10^{-6}$ , 则迭代结束; 若  $Q(P^{(1)}) > Q(P^{(0)})$  则  $P^{(0)} = P^{(1)}$ ,  $\lambda = 10\lambda$ , 再求解式(10); 若  $Q(P^{(1)}) < Q(P^{(0)})$  则  $P^{(0)} = P^{(1)}$ ,  $\lambda = 0.1\lambda$ , 再求解式(10)。

对于 Weibull 分布函数,  $P = [a, b, c]^T$ 。令  $a = 0$ , 用最小二乘法可求得  $b, c$  初始值, 则初始参数  $P^{(0)}$  可定。将分布的初始值代入式(10), 可求分布  $a, b, c$  的最佳值。

### 1.2 Gumbel 分布

Gumbel 分布又称为极值 I 型分布, 由 Fisher 首先导出, Gumbel(1958)首次把它用在洪水分析计算中。我国《港口工程技术规范》<sup>1)</sup>将其用于校核潮位的计算, 由于该分布具有

1) 中华人民共和国交通部, 1987。港口工程技术规范。312—315

充分的理论根据, 美国和日本等国已将其用于重现期波高的计算。其概率分布函数为:

$$F(x) = \exp\{-\exp[-\alpha(x-\beta)]\} \quad (11)$$

式中,  $\alpha$ ,  $\beta$  为参数;  $x$  为极值波高观测值。

本文计算时, 采用最小二乘法拟定分布参数(刘德辅, 1983)。

### 1.3 P-III型分布

P-III型分布是 Pearson 分布曲线族中的一种, 由英国学者 Pearson(1903) 首先提出, 目前, 它是水利工程界应用最为广泛的一种曲线, 其密度函数为:

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} (x - a_0)^{\alpha-1} e^{-\beta(x-a_0)} \quad (12)$$

式中,  $x$  为极值波高观测值;  $\alpha$  为形状参数;  $\beta$  为尺度参数;  $a_0$  为位置参数。

设总体的数学期望为  $\bar{x}$ 、变差系数为  $C_v$ 、偏态系数为  $C_s$ , 则可得  $\alpha$ ,  $\beta$ ,  $a_0$  与  $\bar{x}$ ,  $C_v$ ,  $C_s$  的关系如下(金光炎, 1964)。

$$\begin{cases} \alpha = \frac{4}{C_s^2} \\ \beta = \frac{2}{x C_v C_s} \\ a_0 = \bar{x} \left(1 - \frac{2C_v}{C_s}\right) \end{cases} \quad (13)$$

工程水文计算中, 对某一个频率  $p$ , 则有一个重现值  $x_p$  与之对应, 即:

$$p(x > x_p) = \int_{x_p}^{+\infty} f(x) dx \quad (14)$$

$$p(x > x_p) = \frac{\beta^\alpha}{\Gamma(\alpha)} \int_{x_p}^{+\infty} (x - a_0)^{\alpha-1} e^{-\beta(x-a_0)} dx \quad (15)$$

$$\text{令 } t = \beta(x - a_0), \text{ 则得: } p(x > x_p) = \frac{1}{\Gamma(\alpha)} \int_{t_p}^{+\infty} t^{\alpha-1} e^{-t} dt \quad (16)$$

式中,  $t_p = \beta(x_p - a_0)$ 。

由于目前所观测到的样本容量较小, 因此  $C_v$  值特别是  $C_s$  值往往有较大偏差。此处计算时, 由样本求出  $C_v$  值,  $C_s$  值随  $C_v$  值变化, 每次求出一组理论频率, 从中选出与经验频率的离差平方和最小者作为最佳 P-III 型曲线。

### 1.4 Log-normal 分布

Log-normal 分布假设变量取对数后变成正态分布。目前国内外学者已将此分布用于年极值波浪的研究(潘锦嫦, 1988)。该分布计算简单, 不必适线。设变量为  $x$ , 令  $\ln x = y$ , 则  $y$  服从正态分布,  $x$  的概率密度函数为:

$$f(x) = \frac{1}{x\sigma_{\ln x}\sqrt{2\pi}} \exp\left[-\frac{(\ln x - \overline{\ln x})^2}{2\sigma_{\ln x}^2}\right] \quad (17)$$

令  $t = \frac{\ln x - \overline{\ln x}}{\sigma_{\ln x}}$ , 则分布函数为:

$$F(x) = \int_{-\infty}^t \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{t^2}{2}\right] dt \quad (18)$$

应用最小二乘法即可求出  $\sigma_{\ln x}$  和  $\overline{\ln x}$ , 对不同的重现期  $T$ , 随机变量的出现频率  $p = \frac{1}{T}$ , 累积频率  $F = 1 - p$ , 代入式(18)可求得  $x_p$ 。

## 2 拟合优度检验

由于总体的分布函数表达式是未知的, 因而对所得理论分布  $F_0(x)$  进行柯尔莫哥洛夫检验, 验证它是否符合水文极值的总体概率分布。具体的检验步骤如下:

设原假设  $H_0$  与被选假设  $H_1$  分别为:  $H_0: F(x) = F_0(x)$ ;  $H_1: F(x) \neq F_0(x)$ 。取统计量:  $D_n = \sup_{-\infty < x < +\infty} |F_n(x) - F_0(x)|$ , 式中,  $F_n(x)$  为经验累积概率分布函数。令:  $d_i = |F_n(x) - F_0(x)|$ , 则统计量  $D_n$  的观测值为:  $\hat{D}_n = \max_{1 \leq i \leq 15} \{d_i\}$

对于显著水平  $\alpha$ , 对不同的样本容量  $n$ , 柯氏检验的临界值  $D_n(\alpha)$ 。若  $\hat{D}_n < D_n(\alpha)$ , 则接受原假设  $H_0$ , 拒绝被选假设  $H_1$ ; 否则拒绝原假设  $H_0$ 。

## 3 算例

### 3.1 原始数据

已知渤海湾中某区块连续 21 年的极大值波高序列<sup>1)</sup>, 按照降序列于表 1, 要求对不同重现值波高进行计算。

表1 年极值波高观测值

Tab.1 Observed values of annual extreme wave height

序号	1	2	3	4	5	6	7	8	9	10	11
极值波高(m)	4.35	4.29	4.03	4.02	3.96	3.88	3.87	3.73	3.64	3.60	3.57
序号	12	13	14	15	16	17	18	19	20	21	
极值波高(m)	3.54	3.50	3.42	3.36	3.28	3.21	3.21	2.99	2.92	2.69	

### 3.2 几种极值分布模式对重现值的计算

对图 1 所示极值波高序列进行 Gumbel、P-III、Weibull、Log-normal 分布拟合, 所得理论分布曲线分别绘在 Powell, Normal, Weibull 和 Log-normal 概率纸上(见图 1)。为了使图 1(a)成一直线, 对波高  $x$  作如下变换:  $t = x - a$ , 式中  $a$  为 Weibull 分布的位置参数。

1) 中国海洋石油生产研究中心, 1995. 渤海SZ36-1油田二期开发工程海洋环境条件分析研究. 124-129

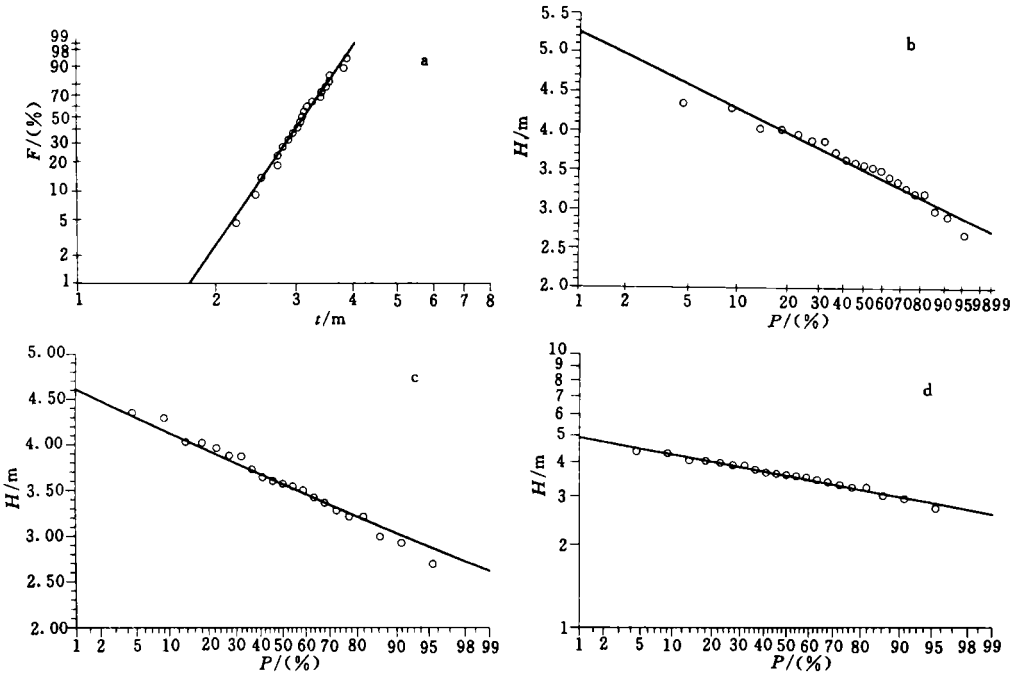


图 1 年极值波高的 Weibull(a)、Gumbel(b)、P-III(c) 和 Log-normal(d) 分布

Fig.1 Weibull(a), Gumbel(b), P-III(c) and Log-normal (d) distribution of annual extreme wave height

○观测值；——理论分布曲线

### 3.3 分布拟合检验

对极值波高序列求得 Gumbel、P-III、Weibull 和 Log-normal 等 4 种理论线型各自的分  
布参数后,进行拟合优度检验,结果见表 2。

表2 4种极值分布柯氏检验统计量

Tab.2 Kolmogorov test statistics of 4 extreme distributions

分布类型	Gumbel	P-III	Weibull	Log-normal
$\hat{D}_{21}$	0.129	0.122	0.096	0.099

当显著水平  $\alpha$  为 0.05, 样本容量为 21 时,柯氏检验的临界值  $D_{21}(0.05)$  为 0.286。由于  
本例中 4 种分布的柯氏检验统计量皆小于 0.286, 因此,波高序列的 4 种分布均不拒绝原假  
设,都可用于重现值计算。

### 3.4 最佳分布模式

将 4 种分布的理论频率与经验频率的离差平方和  $\Sigma$  列入表 3。由表 3 看出 Weibull 分  
布的  $\Sigma$  值最小, 则 Weibull 分布为最佳线型, 工程设计中以此理论分布计算重现值。

表3 理论与经验频率的离差平方和

Tab.3 Deviation square sum of theoretical and empirical probability

分布类型	Gumbel	P-III	Weibull	Log-normal
$\Sigma (10^{-2})$	4.101	2.043	1.386	1.689

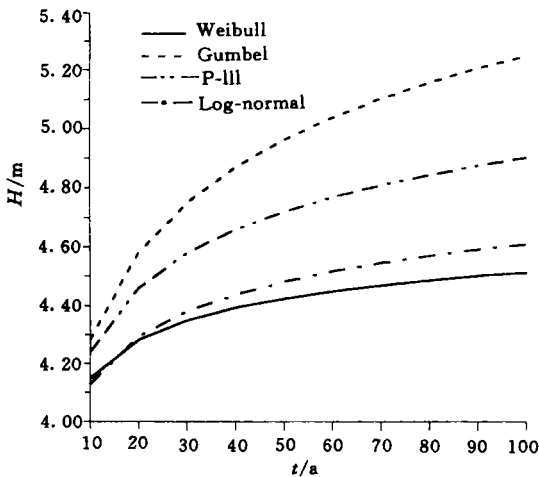


图2 4种极值分布模式的重现期波高曲线

Fig.2 Curves of return period wave heights calculated by 4 extreme value distribution models

遇值( $x_{1\%}$ )比 Weibull 分布算出的值分别大 16%, 2%, 9%, 因此, 对不同的工程, 哪一种分布更适合, 应逐一计算比较, 这也是地区、环境差异所导致的结果。本例中的最佳分布模式为 Weibull 分布。

**4.2** 本文以麦奎尔特法拟合三参数 Weibull 分布, 不仅省却了繁冗的手算工作量, 避免了图解法结果的因人而异, 同时克服了分步最小二乘法初值的选取问题, 而且计算精度高, 收敛速度快。

**4.3** 对重现期大于 20 年的波高重现值, 4 种分布的大小排列顺序为 Gumbel, Log-normal, P-III, Weibull 分布。这个结论与 Muir(1986)所得结论是一致的, 总的趋势为: 对 50 年一遇、100 年一遇的设计波高, Gumbel 所得值最大, Log-normal 次之, Weibull 最小。从本文计算的 P-III 分布计算看, 它所求出的设计值介于 Log-normal 与 Weibull 之间, 因此, 设计值的大小受分布类型的影响很大, 在不同的工程水文计算中, 计算多种分布, 从中选优, 具有现实意义。

## 参 考 文 献

- 刘德辅, 1983. 海洋石油工程环境. 北京: 石油工业出版社. 288—294
- 陈上及 马继瑞, 1991. 海洋数据处理分析方法及其应用. 北京: 海洋出版社. 366—376
- 金光炎, 1964. 水文统计原理与方法. 北京: 中国工业出版社. 83—98
- 潘锦嫦, 1988. 年极值波浪长期分布的研究. 海岸工程, 7(4): 21—29
- Gumbel E J, 1958. Statistics of Extremes. New York: Columbia University Press. 161—164
- Marquardt D W, 1963. An Algorithm for Least-squares Estimation of Nonlinear Parameters. JSIAM, 11(2): 431—441
- Muir L R, EL-Shaarawi A H, 1986. On the calculation of Extreme Wave Heights: A Review. Ocean Engineering, 13(1): 93—118
- Pearson K, Lee A, 1903. On the Law of Inheritance in Man. Biometrika, 2: 357—462
- Weibull W, 1951. A Statistical Distribution of Wide Applicability. J Applied Mechanics, 18: 293—297

## 3.5 几种分布重现值的计算与比较

将 4 种分布所得重现值作成图 2。对于同一观测序列, 不同理论分布所得重现值不同, 从图 2 可以看出, 由 Gumbel 分布所得之小概率设计值最高, 由 weibull 分布得到的值偏小, Log-normal 所得的相同重现期的设计波高与 Gumbel 分布接近, P-III 所得的结果稍大于 Weibull 分布所得的值。

## 4 结论

经过本文的计算比较, 得到以下几点结论。

**4.1** 对同一观测序列, 不同分布求得的重现值是不相同的。本例中由 Gumbel, P-III, Log-normal 分布所得到的百年一

## STATISTICAL DISTRIBUTIONS OF LONG-TERM EXTREME WAVE HEIGHT

WANG Yun-hong, DONG Sheng<sup>†</sup>

(*Tianjin Water Conservancy Bureau, Tianjin, 300074*)

<sup>†</sup> (*College of Engineering, Ocean University of Qingdao, Qingdao, 266003*)

**Abstract** Calculation of the predicted wave height against various return periods plays an important role in planning, design of coastal construction and ocean engineering. The determination of the design values will greatly influence the level of protection standard and the scale of investment. Comparison of several extreme distribution models used in China and abroad shows that because of the difference of the areas there is no popular trend to use only one distribution as a recommended model. The optimum model for every engineering design can be obtained only when much more distributions are fitted and compared. So far, 3-parameter Weibull distribution is widely used. The initially unknown values of the parameters are determined by probability paper or linear least square curve fitting at present. The former method is simple but the results may not be the same. The latter method requires estimation of an initial location parameter. The speed of solution depends to a large extent on the interval between original value and actual value of location parameter. In this paper a new approach is proposed based on the Marquardt method. Three unknown parameters can be calculated simultaneously. The speed of the iterative loop is high and the accuracy is good. Fitting was carried out by computer. Example shows that different distributions gave very different estimation of long return period wave heights. The value provided by Gumbel distribution is largest, those obtained by Log-normal and P-III distributions are less, and that by Weibull distribution is smallest.

**Key words** Extreme distribution fitting Marquardt method return period

**Subject classification number** P731.22