

海洋重现期波高统一化计算方法*

尤再进

(大连海事大学港口与航运安全协创中心 辽宁大连 116085)

摘要 重现期波高是港口海岸及海洋工程设计中不可避免的一个重要设计参数,尤其对深水海港、海上平台、海底油气管道、沿海核电站等重大涉海工程设计具有巨大的经济价值和深远的社会效益。但是,现有重现期波高推算缺乏统一的计算方法,导致计算结果相差悬殊。研究重现期波高的统一化计算方法,分析重现期波高计算中存在的各种不确定因素,提出减少这些不确定因素的新方法,建立误差小、应用方便、方法统一的重现期波高计算方法。基于澳大利亚悉尼站的长期连续观测波浪数据,研究发现:广义帕累托函数(generalized Pareto distribution III, GPD-III)和威布尔(Weibull)是重现期波高计算的最佳候选极值分布函数,新推导的函数形状参数计算公式较好提高重现期波高的计算精度,极值波高数据的分析方法和样本大小是影响重现期波高计算精确度的两个重要因素,短期波浪资料和年极值法可能高估重现期波高值。逐个风暴的极值波高数据分析法及最佳候选极值分布函数 GPD-III 和 Weibull 建议应用于涉海工程设计的重现期波高推算。

关键词 重现期; 重现期波高; 极值分析; 广义极值函数; 广义帕累托函数

中图分类号 TV139.2 doi: 10.11693/hyhz20220300051

海洋波浪重现期/设计波高是港口海岸和海洋工程设计中不可避免的一个重要设计参数,尤其对深水海港、沿海核电站、海洋平台、海底油气管线等重大涉海工程设计具有巨大的经济价值和深远的社会效益(尤再进, 2016)。国际学者采纳的设计波高或重现期波高的计算方法可以归纳以下三种: (1) 年 n 个大波法(annual n -largest method, ANL; 曹兵等, 2006), 此方法是由年极值法($n=1$)延伸而来(Sobey *et al.*, 1995), 从时间序列的波浪数据中每年提出 n 个独立的大浪峰值($n \geq 1$), 组成一个极值波高数据样本, 然后应用广义极值函数(generalized extreme value, GEV)推算多年一遇的重现期波高; (2) 阈值法(peaks-over-threshold method, POT; Hosking *et al.*, 1987), 该方法是设置统一的波高阈值 H_0 , 从时间序列波浪数据中提出大于 H_0 的独立大浪峰值, 组成一个极值波高数据样本, 然后应用广义帕累托函数(generalized Pareto distribution, GPD)推算极值波高; (3) 个风暴法(storm-by-storm method, SAS; Goda, 1988), 是从时间序列的波浪数据中提取所有风暴波

浪峰值, 组成一个风暴波浪峰值数据样本, 然后应用右偏分布的概率函数(如 Weibull、Pearson-III 等)推算重现期波高。

但是, 这三种重现期波浪计算方法通常会导致不同的计算结果。首先, 这三种极值波浪数据分析方法(ANL, POT, SAS)将会导致显著差异样本的重现期波高数据。Sartini 等(2015)采用 ANL 和 POT 两种方法来计算百年一遇的极值波高, 发现这两种广义分布函数的计算结果差别很大。POT 方法的难点是如何确定一个合适的波高阈值 H_0 (Liang *et al.*, 2019)。You (2012)探讨了如何正确地选择极值波高阈值, 认为 SAS 是 POT 的一种特殊情况, 当 POT 的阈值等于 SAS 的当地最小台风/风暴波浪峰值时, 这两种方法就变成一种方法。ANL 方法每年都提取 n 个独立和随机的大浪峰值, 而不能保证每年的波高阈值相同或者认为波高阈值是时间的变量, 而 POT 方法保持每年的波高阈值相同, 而每年获取的风暴波浪数不等, 或者认为 POT 方法每年提取的独立和随机大浪数是一个变量。所以, 这两种不同方法生成不同的极

* NSFC-山东联合重点基金, U1806227 号。尤再进, 博士生导师, 教授, E-mail: b.you@dlmu.edu.cn

收稿日期: 2022-03-08, 收修改稿日期: 2022-04-05

值波高数据样本, 导致重现期波高数据样本的不确定性。但是, 如果 ANL 方法每年选取的 n 个独立的最大波浪数据点是足够多(如 $n \geq$ 每年独立台风浪个数), 这两种方法能够生成几乎相同的重现期波高数据样本(You *et al.*, 2015)。

其次, GEV 和 GPD 是由不同的极值分析法而推导出的两种广义分布函数。理论上 GEV 只能由 ANL 方法生成的极值波高数据样本计算重现期波高, 同样 GPD 只能由 POT 方法生成的极值波浪样本来推算重现期波高。ANL 或 POT 方法生成的极值波高数据样本和右偏分布的普通概率函数(如 Weibull, Pearson-III)也应用于重现期波高计算(尹宝树等, 2002)。Isaacson 等(1981)指出, 没有一个极值分布函数是最佳的, 只有与具体数据比较以后, 才能够判断出哪个函数是最佳函数。

再者, 国内外学者们应用多种不同函数参数估算法计算极值函数参数, 如矩法(method of moment, MOM), 最小二乘(least-square method, LS), 极大似然(maximum likelihood method, ML), 概率权重矩(probability weighted moments, PWM)等参数估算法, 其中 MM, ML 和 LS 是重现期波高计算中常用的三种主要参数估算法, 这些方法的物理含义比较清晰(陈子燊, 2011; You *et al.*, 2015)。MM 法应用比较简单, 从极值波高数据样本中分别估算出样本的均值、方差、偏度等特征参数。这就是概率论中 n -阶矩的定义, 也就是波高变量的 n 次方对极值概率密度的积分。所以, MM 方法通常适用于一些简单的两参数概率函数(如极值 I 型分布, FT-I), 而且极值波高数据样本量需要很大, 使得计算的极值波浪样本特征值(均值、方差、偏度)趋于稳定。ML 方法适用于大多数极值函数, 但是在估算三参数或多参数函数的参数时(如 Weibull, GEV, GPD), 其收敛速度比较慢, 甚至不收敛(Mazas *et al.*, 2011)。You (2011)推广了 LS 方法, 高精度计算 Weibull、GEV、GPD 的三函数参数。基于澳大利亚新南威士州海岸采集的长期和高质量波浪数据, You (2012)定量讨论了 ML 和 LS 两种方法的关系, 研究发现 LS 方法是 ML 方法的一种特殊形式。很多学者认为, LS 方法需要计算极值波高的经验频率分布, 而 ML 方法却不需要, 能够直接基于极值波浪数据样本确定极值函数参数。但是, 事实上所有极值波浪计算方法均需要重现期波高数据, 因为它们的计算结果需要与重现期波高数据进行比较。现有的经验频率计算公式大多是从序列函数推导出来的,

表达形式随着函数不同而改变, 导致生成不同的重现期波高数据样本(Goda, 1988)。基于波浪重现期的定义, You 等(2015)推导出唯一的极值波浪经验频率的表达式, 消除重现期波高数据样本的不确定性。

最后, 国内外学者建立了不同的评估标准, 选择最佳候选极值函数推算重现期波高。现今应用最多的评估标准是均方误差(mean squared error, MSE): $MSE = \Sigma(\text{观测的重现期波高} - \text{计算的重现期波高})^2/n$, 其中, n 是重现期波高数据样本量。但是, 也有少部分学者采用观测和计算的极值波高经验频率分布的均方误差 MSE 作为选择最佳极值函数的评估标准(Liu *et al.*, 2021), 而极值波浪经验频率分布不是直接观测的极值波浪数据, 而且计算方法也具有不确定性。You 等(2015)采用延伸的 LS 函数参数估算法, 建立了最佳候选极值函数评估标准, 其方法收敛快且精度高。

该论文重点研究重现期波高计算的统一化方法, 分析重现期波高计算中存在影响计算结果的不确定因素, 提出减少这些不确定因素的新方法, 建立一种误差小、应用方便、方法统一的重现期波高计算方法。

1 重现期波高计算控制方程

给定一个三参数极值分布函数的一般表达式 $Q = f[(H - B)/A, \alpha]$, 其中 Q 是大于或等于一个极值波高 H 的超越概率、 B 是位置参数、 A 是尺度参数、 α 是形状参数, 极值波高 H_Q 可以从该极值分布函数中计算为

$$f^{-1}(\alpha, Q) = \frac{H - B}{A} \quad \text{或} \quad H_Q = A \times x(\alpha, Q) + B, \quad (1)$$

其中, $f^{-1}(\alpha, Q)$ 是极值分布函数的反函数, x 是函数变量。虽然方程(1)时常也应用于极值波高计算的控制方程, 但是重现期 T 常用于港口海岸和海洋工程的设计时间长度, 与超越概率 Q 的关系(You, 2012)为

$$Q = \frac{1}{T\lambda} = \frac{1}{(n/t)} \cdot \frac{1}{(m/t)} = \frac{n}{m}, \quad (2)$$

其中, $T = n/t$ 是重现期值定义(CEM, 2002), n 是 t 年观测波浪资料中共有的极值波高数据大于或等于重现期波高 H_T , m 是 t 年观测波浪资料的极值波浪总数, $\lambda = m/t$ 是 t 年极值波浪的年均个数。将方程(2)代入方程(1), 重现期波高计算控制方程推导为

$$H_T = A \times x(\alpha, T) + B, \quad (3)$$

其中, H_T 是 T 年一遇的重现期波高, $x(\alpha, T)$ 是函数变量。方程(3)是重现期波高计算的控制方程, 仅是重现期 T 的函数。

控制方程(3)中的函数参数(A, B, α)能够通过线性回归重现期波高数据(x, H_T)来确定。由于函数变量 x 是形状参数 α 的隐性函数, 函数参数(A, B)随着 α 的变化而变化。只有当 α 值满足和方差 SSE 最小要求时,

$$SSE = \sum_{i=1}^m (H_T - H_i)^2 = \sum_{i=1}^m [A(x - \bar{x}) + (H_i - \bar{H})]^2, (4)$$
 函数参数(A, B)才能够使得 SSE 值最小, 其中 $SSE = MSE \times m$, H_T 是方程(3)计算的重现期波高, H_i 是重现期波浪数据样本中的第 i 个观测波高, H 是 m 个重现期波高的平均值, x 是函数变量, \bar{x} 是极值函数变量的平均值, m 是极值波浪数据点的总数。在方程(4)的两边, 对函数形状参数 α 求导:

$$\begin{aligned} \frac{\partial SSE}{\partial \alpha} &= \sum_{i=1}^m [A(x - \bar{x}) + (H_i - \bar{H})] \left[\frac{\partial A}{\partial \alpha} (x - \bar{x}) + A \frac{\partial (x - \bar{x})}{\partial \alpha} \right] \\ &= \sum_{i=1}^m A \frac{\partial A}{\alpha} (x - \bar{x})^2 + \sum_{i=1}^m A^2 (x - \bar{x}) \frac{\partial (x - \bar{x})}{\partial \alpha} \\ &\quad - \sum_{i=1}^m (x - \bar{x}) (H_i - \bar{H}) \frac{\partial A}{\partial \alpha} - \sum_{i=1}^m A (H_i - \bar{H}) \frac{\partial (x - \bar{x})}{\partial \alpha}, (5) \end{aligned}$$

其中函数参数 A 是方程(3)中的斜率, 且当 SSE 最小且 $\partial SSE / \partial A = 0$ 时, 可由方程(6)计算

$$\sum_{i=1}^m A (x_i - \bar{x})^2 = \sum_{i=1}^m (x_i - \bar{x}) (H_i - \bar{H}), (6)$$

并将其代入方程(5), 其第一项和第三项相加等于零, 方程(5)最后简化为

$$\begin{aligned} \frac{\partial SSE}{\partial \alpha} &= \sum_{i=1}^m A^2 (x_i - \bar{x}) (x_i^* - \bar{x}^*) - \sum_{i=1}^m A (H_i - \bar{H}) (x_i^* - \bar{x}^*), (7) \end{aligned}$$

其中 $x_i^* = \partial x_i / \partial \alpha$ 和 $\bar{x}^* = \partial \bar{x} / \partial \alpha$ 当 SSE 最小时, 并令 $\partial SSE / \partial A = 0$, 形状参数 α 由方程(7)解得

$$\begin{aligned} W(\alpha) &= \left[\frac{\sum_{i=1}^m (x_i - \bar{x}) (H_i - \bar{H})}{\sum_{i=1}^m (x_i - \bar{x}) (x_i - \bar{x})} \right] \\ &\quad \times \left[\frac{\sum_{i=1}^m (x_i - \bar{x}) (x_i^* - \bar{x}^*)}{\sum_{i=1}^m (H_i - \bar{H}) (x_i^* - \bar{x}^*)} \right] - 1 = 0, (8) \end{aligned}$$

方程(8)是函数形状参数 α 计算的控制方程, 乘号的前项是方程(6)的参数 A 或斜率, 乘号的后项相当于函数参数 A 的倒数($1/A$)。Newton-Raphson 迭代法可以应用于方程(8)的求解, 迭代求解函数形状参数 α 。虽然方程(8)与曹兵等(2006, 2007)和 You 等(2006)推导结果相同, 但是本文在推导过程中有显著差异, 尤其在方程(8)的推导过程中没有给出已知极值函数的具体表达式, 该方法适用于具有极值函数普通表达式

$Q=f[(H-B)/A,\alpha]$ 的显函数形状参数估算。

2 极值波高数据分析的逐个风暴法

重现期波高数据(T, H_T)是估算方程(3)函数参数(α, A, B)的基础数据, 是由极值波高数据(Q, H)并通过方程(2)转换而成, 这两组数据(Q, H)和(T, H_T)其实是等同的。在现有的三种常用极值波高数据分析方法中, 逐个风暴法(图 1)是从长期时间序列的波高数据中逐个提取风暴/台风过程中的风暴波高峰值, 一个风暴过程仅提供一个风暴波高峰值的数据点, 提取的所有风暴波高峰值数据组成了独立和随机的极值波高数据样本。图 1 中的时间序列波高数据来源于澳大利亚新南威士州的悉尼波浪骑士站(You *et al*, 2008), 每个风暴过程中的风暴有效波高峰值均大于风暴波高阈值 3 m。逐个风暴法与波高阈值法存在一些明显区别, 尤其逐个风暴法中的风暴阈值是唯一的, 可由大气压场确定风暴刚刚开始时的初时海况(You, 2012), 而 POT 阈值法中的波高阈值至今缺乏取值的统一标准, 人为影响因素比较大。但是, 这两种方法也是相关的, 逐个风暴法可以看作是 POT 阈值法的一种特殊形式: 当逐个风暴法中的风暴波高阈值等于阈值法中的波高阈值时, 这两种方法是等同的。

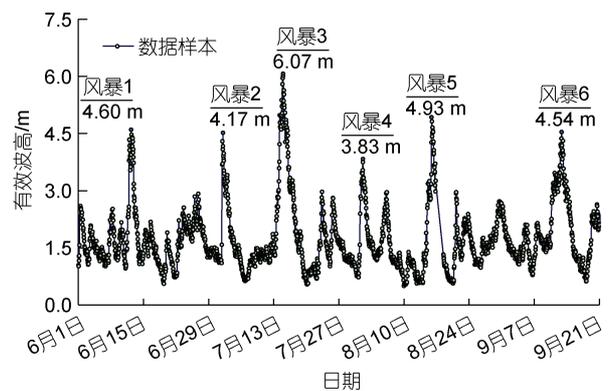


图 1 逐个风暴法应用于提取风暴过程中的波高峰值生成极值波高数据样本

Fig.1 Extreme wave height dataset analyzed with the storm-by-storm analysis method

本研究采用的重现期波高数据来源于澳大利亚新南威士州悉尼观测站的深水波浪数据(You *et al*, 2008)。该站的平均水深约 80 m, 波浪数据从 1976 年开始采集, 一直持续至今。1986 年前, 波浪数据纪录在图表纸上, 波峰值从图纸上测量。从 1986 起, 采用波浪骑士每 0.5 s 自动测量一个波高数据, 连续观测 34 min, 每小时重复一次, 数据采集的成功率大于

95%。基于悉尼站采集的 25 a 波浪数据(1987~2011 年), 图 2 给出了逐个风暴法分析的风暴波峰值数据, 其中小圆圈数据点代表每个风暴过程中的最大有效波高, 方框中的数字代表年最大有效波高(年极值法), 大圆圈中的数字代表年风暴发生次数。该论文采用的悉尼站极值波高数据样本是由图 2 中的从大到小排列的所有风暴波峰值数据组成, 并应用方程(2)将极值波高数据(Q, H)转化成重现期波高数据(T, H_T)。

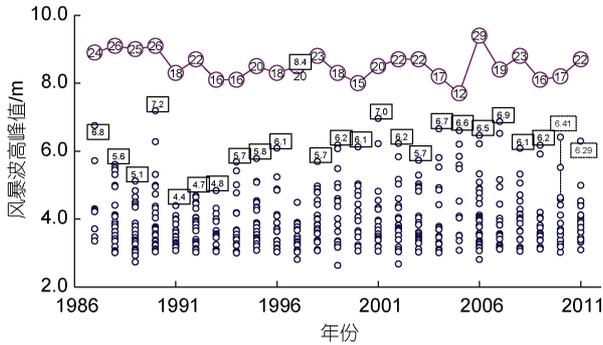


图 2 逐个风暴法分析获取的澳大利亚悉尼波浪观测站的极值波高数据

Fig.2 Extreme wave height data analyzed by the storm-by-storm analysis method from Sydney waverider buoy on the NSW coast of Australia

注: 蓝色小圆圈数据点代表每个风暴波峰值, 黑色方框中的数字代表年最大波高, 紫色大圆圈中的数字代表年风暴发生次数

3 候选极值分布函数的选取标准

重现期波高计算的候选极值函数众多, 大体可以归纳为 3 大类: 广义极值分布函数 GEV、广义帕累

托分布函数 GPD、普通概率函数(如 Weibull、Pearson-III 等), 它们的共性是右偏分布的长尾巴函数, 不同点是 GEV 应用于年 n 大波法的极值波高数据, GPD 针对大于某个波高阈值的极值波高数据, 普通概率函数仅要求随机和独立的极值波高数据。因为 GEV 和 GPD 可以简化为等同的普通概率函数 (You *et al*, 2015), 所以三类候选极值分布函数可以统一归类为右偏分布的普通概率函数。

3.1 广义极值函数

如果暂先不考虑广义分布函数 GEV 推导的假设条件, GEV 的复杂数学表达式可以直接简化为等同的指数函数表达式:

$$P = \exp \left\{ - \left[1 + \frac{1}{\sigma} \left(\frac{H - \mu}{\beta} \right) \right]^{-\sigma} \right\} = \exp \left[- \left(\frac{H - B}{A} \right)^\alpha \right], \quad (9)$$

其中, $A = \sigma\beta$, $B = (\mu - \sigma\beta)$ 和 $\alpha = -\sigma$ 。当 $\alpha = \pm\infty$, $\alpha < 0$ 和 $\alpha > 0$, 简化的 GEV 也分别称为 FT-I, FT-II 和 FT-III 函数, 是否适用于重现期波高计算将通过均方差 MSE 收敛性得以判断。

在方程(3)中, 给定的不同函数参数 α , 其他两个函数参数(A, B)由最小二乘法唯一确定。基于悉尼站的重现期波高数据, 图 3 分别给出了方程(4)计算的均方差 $MSE(\alpha)$ 和重现期波高 H_{100} 随简化的 FT-II 和 FT-III 形状参数变化的分布规律。由图 3 可见, FT-II 无最大值, 最小值收敛于 FT-I; FT-II 计算的 $MSE(\alpha)$ 存在最小值, 且当 $\alpha > 100$ 时逐渐收敛于 FT-I 计算的 $MSE(\alpha)$; FT-III 无最小值, 最大值收敛于 FT-I; 计算

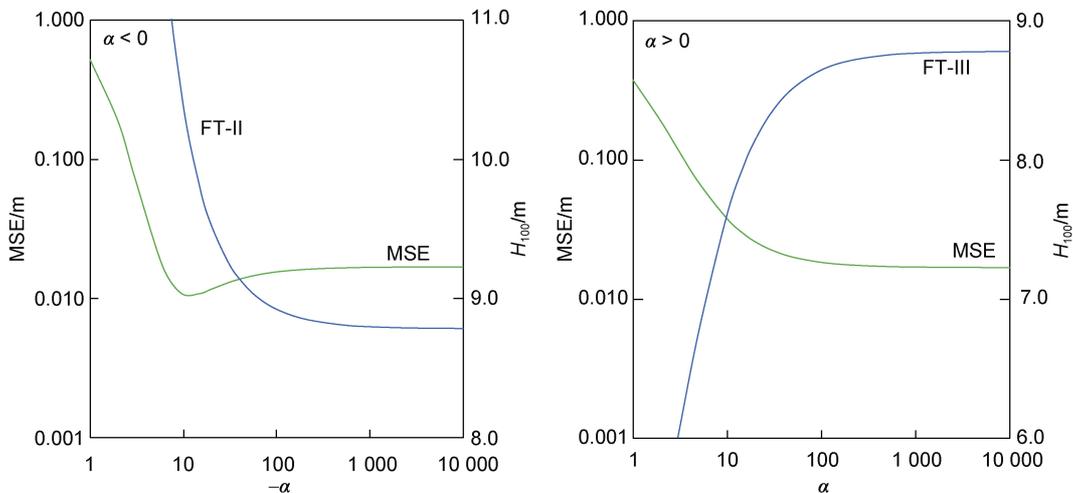


图 3 简化的广义极值分布函数(generalized extreme value, GEV)计算的均方差 MSE 和重现期波高 H_{100} 随简化的 FT-II 和 FT-III 形状参数 α 变化的分布规律

Fig.3 Variations of mean-squared error MSE and 100-year return wave height H_{100} with shape parameter α calculated by the simplified GEV

的 $MSE(\alpha < 100)$ 不存在最小值, 只有当 $\alpha > 100$ 时逐渐收敛于 FT-I 计算的 $MSE(\alpha)$ 。所以, GEV 的三种极值分布函数只有 FT-II 适用于重现期波高计算的候选函数, 三参数的 FT-II 包括二参数的 FT-I 函数。

当 FT-II 应用于重现期波高计算时, Newton-Raphson 迭代法可以应用于计算方程(8)中的函数形状参数 α , 其中 x_i^* 和 \bar{x}^* 分别计算为

$$x_i^* = \frac{\partial x_i}{\partial \alpha} = -\frac{1}{\alpha^2} (-\ln P_i) \frac{1}{\alpha} [\ln(-\ln P_i)] = -\left[\frac{\ln(-\ln P_i)}{\alpha^2} \right] x_i, \quad (10)$$

$$\bar{x}^* = \frac{\partial \bar{x}}{\partial \alpha} = \frac{1}{m} \sum_{i=1}^m \frac{\partial x_i}{\partial \alpha} = \frac{1}{m} \sum_{i=1}^m x_i^*, \quad (11)$$

其中, $x_i = (-\ln P_i)^{1/\alpha}$ 是简化后的 GEV 函数变量, 通过将方程(9)改写成与方程(3)具有相同线性表达式而获得。当形状参数 α 确定后, 方程(3)中的尺度参数 A 由方程(6)计算, 位置参数计算为 $B = (A \times \bar{x} - \bar{H})$ 。

简化广义极值函数 GEV 有两个主要目的, 一是将复杂 GEV 表达式等地简化为指数函数表达式, 并归类于普通的概率函数类; 二是能够极大简化方程(10)~(11)的函数变量求导, 应用方程(8)精确计算函数形状参数。

3.2 广义帕累托函数

参考广义极值分布 GEV 的简化目的和方法, 广义帕累托分布 GPD 的复杂数学表达式也可以改写成简单的幂函数表达式:

$$Q = \left[1 + \frac{1}{\sigma} \left(\frac{H - \mu}{\beta} \right) \right]^{-\sigma} = \left[\frac{H - B}{A} \right]^\alpha, \quad (12)$$

式中, 简化后的 GDP 函数参数确定为 $A = \sigma\beta$, $B = (\mu - \sigma\beta)$ 和 $\alpha = -\sigma$ 。当 $\alpha = \pm\infty$, $\alpha < 0$ 和 $\alpha > 0$, 简化后的广义帕累托分布 GPD 也分别叫作 GPD-I, GPD-II 和 GPD-III, 它们是否是最佳候选极值函数可以通过均方差 MSE 收敛性得以判断。

基于悉尼站的同样重现期波浪数据, 图 4 分别给出了均方误差 MSE 随 GPD-II 和 GPD-III 形状参数变化的分布关系, 以及百年一遇的重现期波高 H_{100} 与形状参数 α 的变化规律。由该图可见, GPD-II 无最大值, 最小值收敛于 GPD-I 或指数函数 $Q = \exp[(H - B)/A]$, 计算的 $MSE(\alpha)$ 分布不存在最小值, 而 GPD-III 无最小值, 最大值收敛于 GPD-I, 计算的 $MSE(\alpha)$ 分布存在最小值。所以, GPD 的三种函数只有 GPD-III 适用于重现期波高计算。

当 GPD-III 应用于重现期波高计算时, Newton-Raphson 迭代法用于计算方程(8)的函数形状参数 α , 其中 x_i^* 计算为

$$x_i^* = \frac{\partial x_i}{\partial \alpha} = -\frac{1}{\alpha^2} Q^{\frac{1}{\alpha}} (\ln Q_i) = -\left(\frac{\ln Q_i}{\alpha^2} \right) x_i, \quad (13)$$

\bar{x}^* 由方程(11)计算, 其中 $x_i = Q_i^{1/\alpha}$ 是简化后的 GPD 函数变量。当 α 确定后, 函数尺度参数 A 由方程(6)计算, 函数位置参数为 $B = (A \times \bar{x} - \bar{H})$ 。

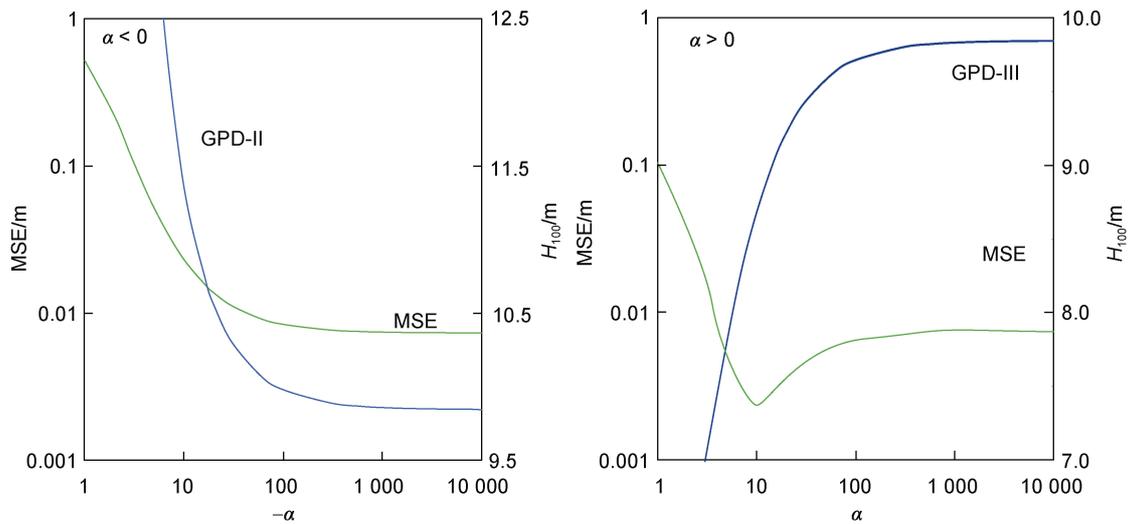


图 4 简化的广义帕累托函数(generalized pareto distribution, GPD)计算的均方误差 MSE 和重现期波高 H_{100} 随函数形状参数 α 变化的分布规律

Fig.4 Variations of mean-squared error MSE and 100-year return wave height H_{100} with shape parameter α calculated by the simplified GPD

3.3 概率分布函数

几种常采用的右偏分布概率函数(如 Weibull 和 Pearson-III)也应用于重现期波高计算。虽然简化后的 GEV 方程(9)的 FT-III($\alpha > 0$)与 Weibull 函数相似,但不适用于重现期波高的计算(图 3),且与 Weibull($\alpha > 0$)分布函数完全不同

$$Q = \exp \left[- \left(\frac{H - \mu}{\beta} \right)^\alpha \right]. \quad (14)$$

基于相同的澳洲悉尼站重现期波浪数据,图 5 分别给出了均方误差 MSE 随 Weibull 和 Pearson-III 形状参数变化的分布关系,以及百年一遇的重现期波高 H_{100} 与形状参数 α 的变化规律。由图 5 可以看出,Weibull 和 Pearson-III 函数计算的 $MSE(\alpha)$ 均存在最小值,但 Weibull 计算的 $MSE(\alpha)$ 要比 Pearson-III 收敛要快。

类似于 GEV 和 GPD 形状参数的计算方法,Newton-Raphson 迭代法也用于计算方程(8)的 Weibull 函数形状参数 α ,但方程(11)和(12)中用于计算 x_i^* 和

\bar{x}^* 的累积概率 P 由超越概率 Q 代替,函数变量为 $x = (-\ln Q)^{1/\alpha}$ 。因为 Pearson-III 函数是非显性函数,函数变量的导数比较复杂,所以其函数形状参数 α 不能由方程(8)计算,只能通过试算方法确定 α ,使得方程(4)的 SSE 值最小(You, 2012)。

4 候选极值函数的最佳拟合标准

4.1 最佳拟合度评估法

候选极值函数与重现期波高数据(T, H_T)的拟合程度可由方程(4)计算的均方误差 MSE 或者和方差 SSE 值来评价。当 $MSE(\alpha)$ 最小或方程(8) $W(\alpha) = 0$, 候选极值函数与重现期波高数据的拟合程度达到最佳。基于悉尼波浪站的重现期波高数据,两种简化的广义极值函数 FT-II 方程(9)和 GPD-III 方程(12)与重现期波高数据(x, H_T)的线性拟合度见图 6,其中 x 是极值函数变量,函数形状参数 α 由方程(8)计算,方程(1)中的函数尺度和位置参数(A, B)由最小二乘法确定。

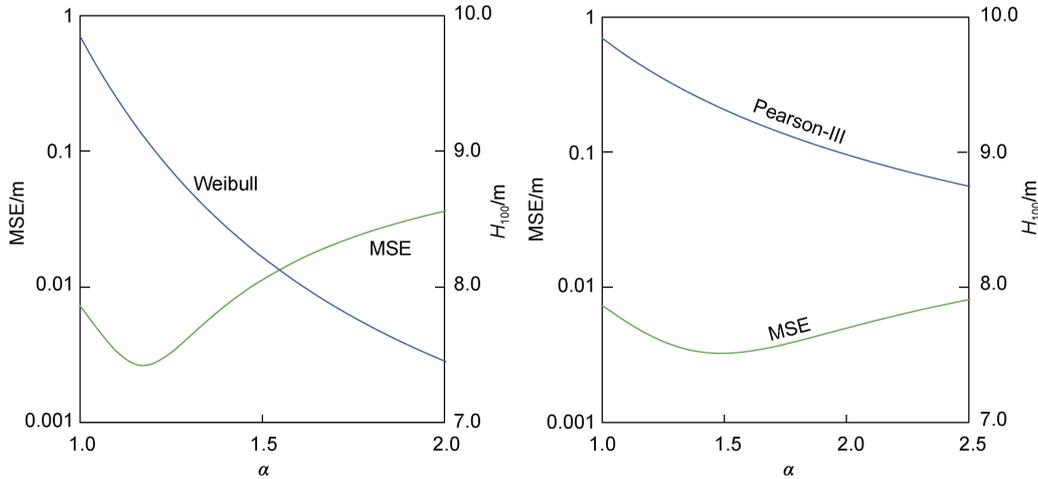


图 5 威尔布(Weibull)和皮尔逊 3 型(Pearson-III)函数计算的均方误差 MSE 和重现期波高 H_{100} 随函数形状参数 α 变化的分布规律

Fig.5 Variations of mean-squared error MSE and 100-year return wave height H_{100} with shape parameter α calculated by Weibull and Pearson-III

图 6 中的两种重现期波高数据(x, H_T)和(T, H_T)是等同的,并具有相同的分布形态。这是因为当形状参数 α 确定后, x 只是重现期 T 的唯一函数。当函数参数确定后,重现期 T 是方程(1)中的重现期波高计算的唯一函数变量,给定重现期 T 的波高 H_T 能够从方程(1)获解。图 6 也给出了计算的和观测的重现期波高的定量比较, FT-II 计算的重现期波高曲线具有上凸特点,计算的重现期波高随 T 增加而较快增加,而且计算红

线均高于重现期大于 4 a 的波高,高估了重现期波高。相比较而言, GPD-III 的重现期波高曲线具有上凹特点,计算随 T 增加而较缓慢增加,函数与数据的拟合度 R^2 要比 FT-II 的高。

图 7 给出了 Weibull 和 Pearson-III 概率函数与重现期波高数据(x, H_T)的拟合度,其中形状参数 α 先由方程(8)计算,函数尺度和位置参数(A, B)再由最小二乘法计算。这两种概率函数与重现期波高的拟合度比

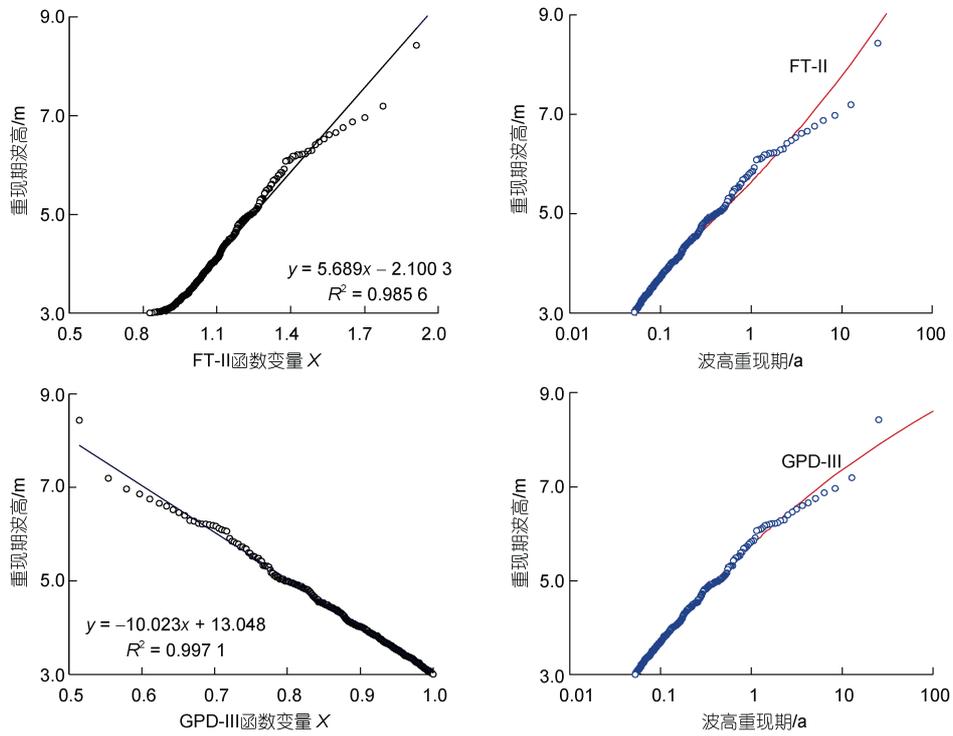


图 6 极值函数形状参数估算法方程(8)应用于简化的 GEV-II 和 GPD-III 参数估算和澳洲悉尼波浪站的重现期波高推算
 Fig.6 The shape parameters of simplified GEV-II and GPD-III are estimated by Eq.(8) and used to calculate return wave heights with different return periods at Sydney waverider buoy location on the NSW coast of Australia
 注: 图中表达式为拟合直线, R 为相关系数, 图 7 同

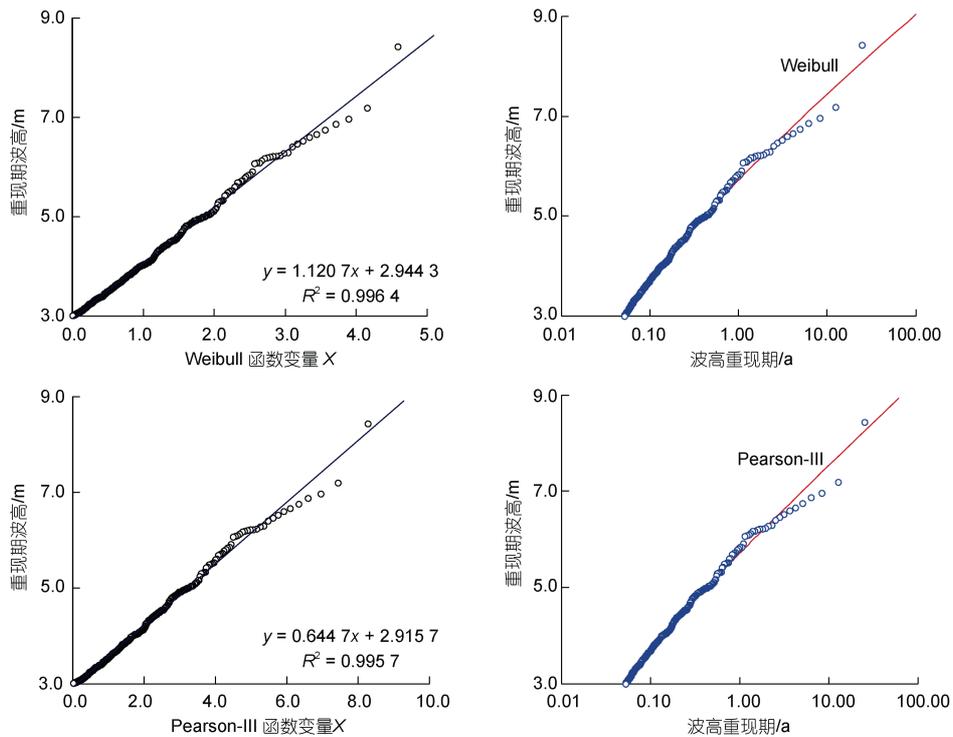


图 7 极值函数形状参数估算法方程(8)应用于概率函数 Weibull 和 Pearson-III 参数估算和澳洲悉尼波浪站的重现期波高推算

Fig.7 The shape parameters of Weibull and Pearson-III are estimated by Eq.(8) and used to calculate return wave heights with different return periods at Sydney waverider buoy location on the NSW coast of Australia

较接近, 线性拟合系数 R^2 几乎相等。图 7 也给出了计算的重现期波高与重现期波高数据的定量比较, 计算的重现期波高曲线均具有上凹特点, Pearson-III 计算的重现期波高稍微比 Weibull 计算的大些。这可能是因为 Weibull 函数形状参数 α 是由方程(8)精确计算的, 而 Pearson-III 的表达式是一个半隐式式, 函数形状参数只能通过试算方法来确定, 计算精度比较低。再者, 由图 5 可见, 当 $MSE(\alpha)$ 靠近其最小值时, $MSE(\alpha)$ 随形状参数 α 变化缓慢, 收敛性较差, 而 Weibull 计算的 $MSE(\alpha)$ 收敛性要比 Pearson-III 的好些。

4.2 数据量对最佳拟合度影响

基于重现期波高计算控制方程(1)的推导假设条件, 仅当 $MSE(\alpha)$ 值最小时, 候选极值函数与重现期数据的拟合度才能达到最佳。但是, 应用于计算 $MSE(\alpha)$ 的重现期波高数据量 m 可能影响 $MSE(\alpha)$ 的收敛性以及该计算方法的适用性。同样基于悉尼站的重现期波高样本数据量, 不同样本的重现期波高数

据($H_1 > H_2 > H_3, \dots, > H_k$)从悉尼站的重现期波高数据中($m=485$)提取前 k 个重现期波高数据组成, 共计生产 20 个不同样本的重现期波高数据($k=25, 50, 75, 100, \dots, 450, 475$)。其实, 这里的极值波高数据样本量是由波高阈值的取值大小决定的。

图 8 给出了 FT-II 计算的 $MSE(\alpha)$ 收敛性以及重现期波高 H_{100} 随重现期波高数据量 m 增加的变化规律。由该图可以发现, $MSE(\alpha)$ 均存在最小值, 但是当 $m < 150$ 时, 其最小值位置和收敛值随 m 增大而分别向右方移动并明显减小, MSE 最小值对应的重现期波高 H_{100} 随着 m 增加而显著减小; 当 $m \geq 150$ 时, $MSE(\alpha)$ 最小值位置和收敛值随 m 增大而趋近稳定, 对应 MSE 最小值的重现期波高 H_{100} 随着 m 增大而变化不显著。综上所述, 随着重现期波高数据 m 增加, FT-II 计算的 $MSE(\alpha)$ 始终存在最小值, 但仅当 $m \geq 150$ 时, 计算的重现期波高 H_{100} 才趋近稳定。这就要求采集的波高数据的时间长度(M 年)应该足够长或者数据点足够多, 满足生成的极值波高数据量大于 150。

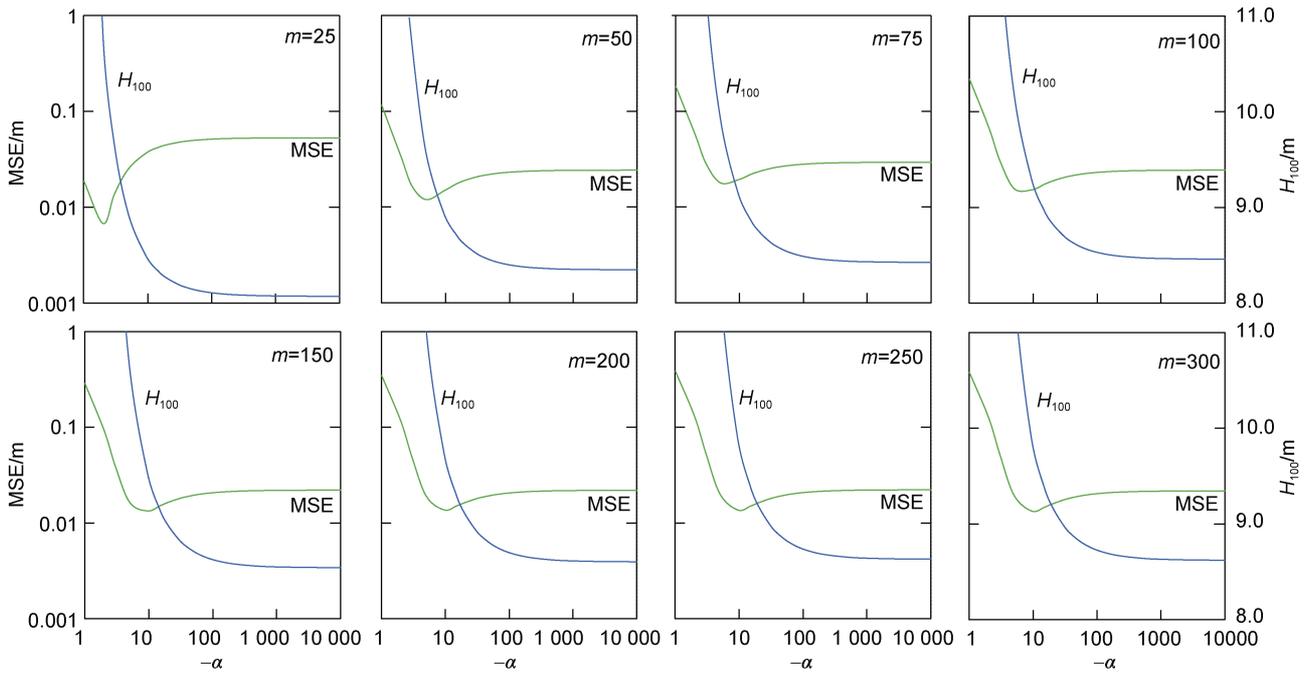


图 8 GEV-II 函数计算的 $MSE(\alpha)$ 收敛性和重现期波高 H_{100} 随重现期波高数据样本量 m 的变化规律

Fig.8 The variations of mean-squared error $MSE(\alpha)$ and 100-year return wave height, which are calculated by GEV-II, with the sample size of extreme wave height data

同样, 图 9 也给出了 GPD-III 计算的 $MSE(\alpha)$ 收敛性以及重现期波高 H_{100} 随数据量 m 增加的变化规律。由图可见, 当 $m < 50$ (相当于年均采集两个极值波高数据点), $MSE(\alpha)$ 不存在最小值, 但当 $\alpha > 100$ 时,

$MSE(\alpha)$ 开始收敛, 三参数的 GPD-III 简化为两参数的指数函数 $Q = \exp[(H - B)/A]$; 当 $m \geq 200$, $MSE(\alpha)$ 存在最小值, 其最小值位置和收敛值随 m 增大而趋近稳定, MSE 最小值对应的重现期波高 H_{100} 随着 m 增

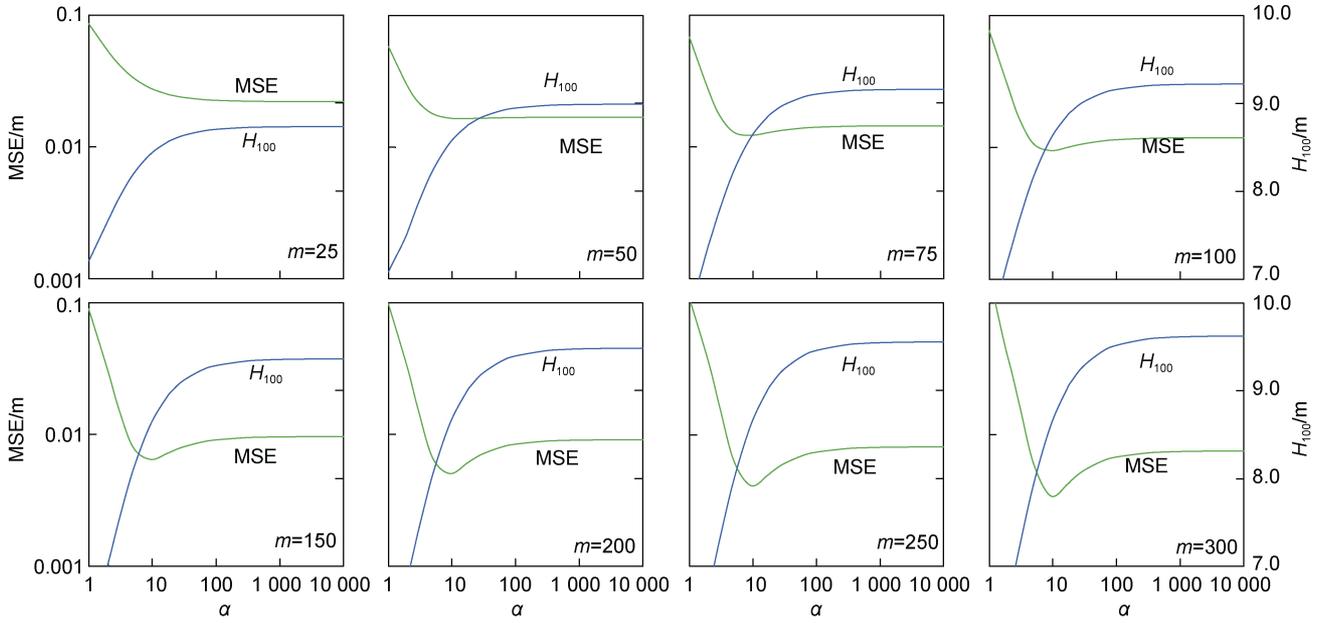


图 9 GPD-III 计算的 MSE(α)收敛性与重现期波高 H_{100} 数据随样本大小 m 的变化规律

Fig.9 The variations of mean-squared error $MSE(\alpha)$ and 100-year return wave height, which are calculated by GPD-III, with the sample size of extreme wave height data

大而没有显著变化, GPD-III 计算的重现期波高 H_{100} 随 m 增大而趋向稳定。所以, 当重现期波高数据量 m 比较小时, GPD-III 概率函数不适用于计算重现期波高, 仅当 $m > 200$ 重现期波高计算才趋近稳定。

再次, 图 10 给出了 Weibull 计算的 MSE(α)收敛性以及重现期波高 H_{100} 随数据量 m 增加的变化规律。

由图可见, 当 $m < 25$ 时(相当于年均有一个极值波高数据点), 在形状参数的适用范围内($1 < \alpha < 2$), $MSE(\alpha)$ 不满足方程(1)的推导条件 $\partial MSE / \partial \alpha = 0$, 导致 Weibull 函数不适用于重现期波高计算; 当 $50 < m < 200$ 时, $MSE(\alpha)$ 均存在最小值且 $\partial MSE / \partial \alpha = 0$, 但 MSE 收敛于最小值的速度很慢, 而靠近 MSE 最小值附近的重现

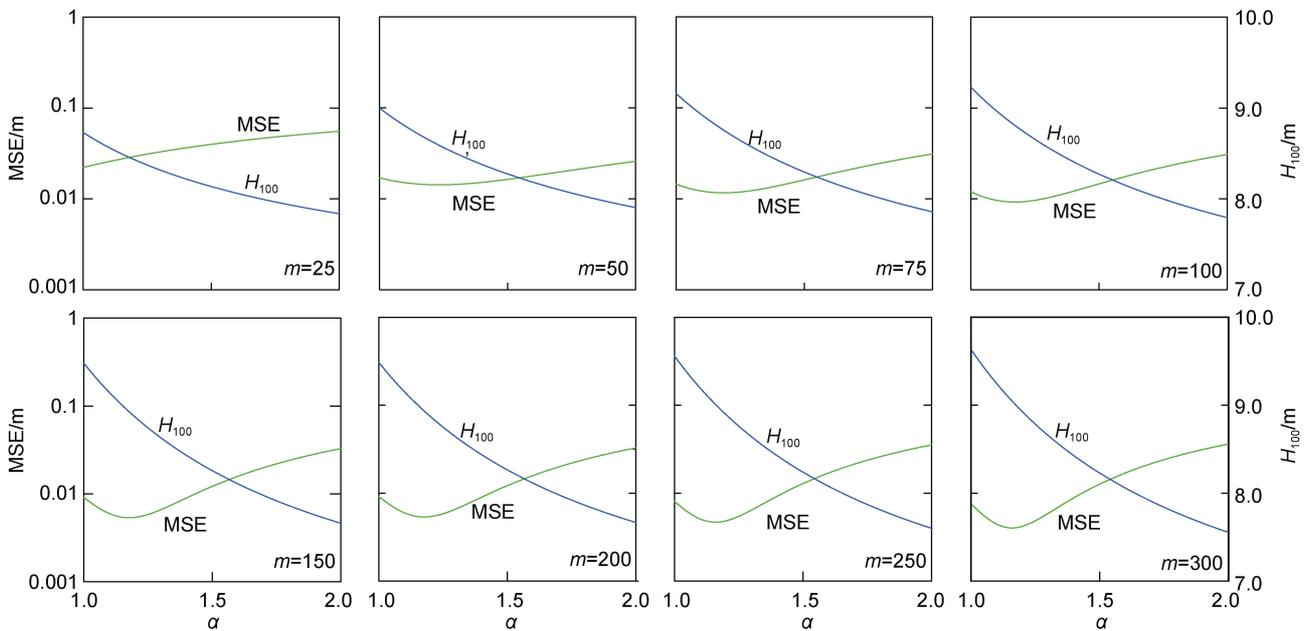


图 10 Weibull 函数计算的 MSE(α)收敛性与重现期波高 H_{100} 数据随样本大小 m 的变化规律

Fig.10 The variations of mean-squared error $MSE(\alpha)$ and 100-year return wave height, which are calculated by GPD-III, with the sample size of extreme wave height data

期波高变化较大,这意味着由 Weibull 函数计算的重现期波高受其形状参数 α 计算精度的影响比较显著;仅当 $m \geq 200$, $MSE(\alpha)$ 存在最小值,其最小值位置和收敛值随 m 增大而趋近稳定, MSE 最小值对应的重现期波高 H_{100} 随着 m 增大而没有显著变化, Weibull 计算的重现期波高 H_{100} 随 m 增大而趋向稳定。综上所述,当重现期波高数据量 m 比较小时, Weibull 概率函数不适用于重现期波高计算,但当数据量足够大时($m \geq 200$),计算的重现期波高 H_{100} 才趋近稳定,重现期波高计算受形状参数计算精度的影响较显著。

最后,图 11 给出了 Pearson-III 和 Gumbel(FT-I)计算的重现期波高与年极值法分析的重现期波高数

据的定量比较,这两函数是《港口与航道水文规范》(JTS145-2015)推荐的极值波高数据分析和候选极值分布函数。由图 7 和图 11 可见,逐个风暴法(SBS)构建的重现期波高数据样本($m=285$)远远大于由年极值法生成的极值波高数据样本量($m=25$); Pearson-III 基于年极值法计算的百年一遇重现期波高($H_{100}=9.19$ m)大于基于 SBS 法计算的重现期波高 $H_{100}=8.63$ m,两者误差约 6.5%。再比较图 8 和图 11 可知, Gumbel 基于年极值法计算的百年一遇重现期波高($H_{100}=8.94$ m)略大于基于 SBS 法计算的重现期波高 $H_{100}=8.78$ m,两者计算方法相差 1.8%,但较小于 Pearson-III 推算的 $H_{100}=9.19$ m,其中当 $\alpha \rightarrow \infty$ 时,图 8 中 FT-II 计算的重现期波高等于 FT-I 计算的。

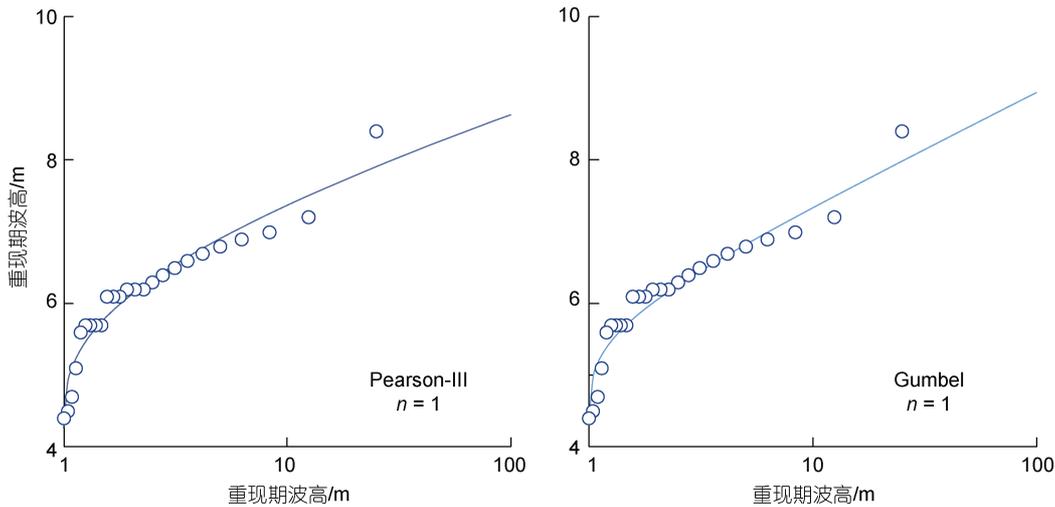


图 11 Pearson-III 和 Gumbel (FT-I)计算的重现期波高与年极大值法分析的重现期波高数据的比较

Fig.11 Comparison of return wave heights with different return periods calculated by Pearson-III and Gumbel (FT-I) with the return wave height data analyzed by the annual maximum method

注: n 表示年平均极值波高的数量

5 结论

重现期波高是港口海岸及海洋工程设计中不可避免的一个重要设计参数,但其计算方法至今缺乏统一性,计算结果存在显著不确定性。该论文建立了一种误差小、应用方便、方法统一的重现期波高计算方法。本文研究发现: GPD-III 和 Weibull 是重现期波高计算的最佳候选推算函数,新推导的它们形状参数计算公式较好提高重现期波高的计算精度,极值波高数据的样本量和分析方法是影响重现期波高计算精度的 2 个重要因素。基于上述的研究发现,我国《港口与航道水文规范》(JTS 145—2015)建议采用的年极值法和 Pearson-III 极值分布函数有可能高估推算的重现期波高,但需要中国沿海长期且高质量现

场波浪数据进一步验证。逐个风暴的极值波高数据分析法与广义帕累托函数 GPG-II 和 Weibull 建议应用于涉海工程的重现期波高计算。

参 考 文 献

- 尤再进, 2016. 中国海岸带淹没和侵蚀重大灾害及减灾策略[J]. 中国科学院院刊, 31(10): 1190-1196.
- 尹宝树, 何宜军, 侯一筠, 等, 2002. 推算波浪多年一遇波高的新方法[J]. 海洋与湖沼, 33(1): 30-35.
- 陈子燊, 2011. 波高与风速联合概率分布研究[J]. 海洋通报, 30(2): 158-163.
- 曹兵, 王义刚, YOU Z J, 2006. 三种设计波高计算方法比较[J]. 海洋工程, 24(4): 75-80.
- 曹兵, 王义刚, YOU Z J, 2007. 设计波高分布函数比较[J]. 海洋湖沼通报(4): 1-9.
- CEM, 2002. Hydrodynamic Analysis and Design Conditions [M].

- U.S Army Corps of Engineers, Coastal and Hydraulics Laboratory.
- GODA Y, 1988. On the methodology of selecting design wave height [C] // Proceedings of the 21st International Conference on Coastal Engineering. Costa del Sol-Malaga: ASCE: 899-913.
- HOSKING J R M, WALLIS J R, 1987. Parameter and quantile estimation for the generalized Pareto distribution [J]. *Technometrics*, 29(3): 339-349.
- ISAACSON M S Q, MACKENZIE N G, 1981. Long-term distributions of ocean waves: a review [J]. *Journal of the Waterway, Port, Coastal and Ocean Division*, 107(2): 93-109.
- LIANG B C, SHAO Z X, LI H J, *et al*, 2019. An automated threshold selection method based on the characteristic of extrapolated significant wave heights [J]. *Coastal Engineering*, 144: 22-32.
- LIU G L, CUI K, JIANG S, *et al*, 2021. A new empirical distribution for the design wave heights under the impact of typhoons [J]. *Applied Ocean Research*, 111: 102679.
- MAZAS F, HAMM L, 2011. A multi-distribution approach to POT methods for determining extreme wave heights [J]. *Coastal Engineering*, 58(5): 385-394.
- SARTINI L, MENTASCHI L, BESIO G, 2015. Comparing different extreme wave analysis models for wave climate assessment along the Italian coast [J]. *Coastal Engineering*, 100: 37-47.
- SOBEY R J, ORLOFF L S, 1995. Triple annual maximum series in wave climate analyses [J]. *Coastal Engineering*, 26(3/4): 135-151.
- YOU Z J, 2011. Extrapolation of historical coastal storm wave data with best-fit distribution function [J]. *Australian Journal of Civil Engineering*, 9(1): 73-82.
- YOU Z J, 2012. Discussion of “a multi-distribution approach to POT methods for determining extreme wave heights” by Mazas and Hamm, [coastal engineering, 58: 385-394] [J]. *Coastal Engineering*, 61: 49-52.
- YOU Z J, LORD D, 2008. Influence of the El Niño-southern oscillation on NSW coastal storm severity [J]. *Journal of Coastal Research*, 24(sp2): 203-207.
- YOU Z J, YIN B S, 2006. Estimation of extreme coastal wave heights from time series of wave data [J]. *China Ocean Engineering*, 20(2): 225-241.
- YOU Z J, YIN B S, JI Z Z, *et al*, 2015. Minimisation of the uncertainty in estimation of extreme coastal wave heights [J]. *Journal of Coastal Research*, 75(sp1): 1277-1281.

UNIFIED APPROACH FOR ESTIMATION OF RETURN OCEAN WAVE HEIGHT

YOU Zai-Jin

(Centre for Ports and Maritime Safety, Dalian Maritime University, Dalian 116085, China)

Abstract Return wave height is an important parameter in the design of coastal and ocean engineering. Accurate calculation of design or return wave height is of enormous economic value and social value especially for deep-water harbors, ocean platforms, subsea gas and oil pipelines, and coastal nuclear power stations. However, there is no unique approach for calculation of return wave height, and results from different methods are significantly different. The present study is undertaken to analyze uncertainty in extrapolation of return wave height, minimize errors, and develop a unified methodology for calculation of return wave height. Based on long-term wave data continuously collected at Sydney permanent wave station on the coast of New South Wales in Australia, it is found that GPD-III and Weibull are the most suitable candidate distribution functions for extrapolation of return wave heights, newly derived formulation Eq.(8) will enable us to accurately estimate the shape parameters of the two distribution functions. The sample size and analysis method of extreme wave data are two important factors affecting the accuracy of return wave height extrapolated. Short wave records and the use of annual maximum method for analysis of extreme wave data could underestimate the return wave heights. The storm-by-storm method for analysis of extreme wave data and GPD-III or Weibull for extrapolation of return wave heights are highly recommended for coastal and ocean engineering design purposes.

Key words return period; return wave height; extreme value analysis; GEV (generalized extreme value); GPD (generalized Pareto distribution)