



超定量系列波浪重现期值研究

罗耀, 朱良生

(华南理工大学 土木与交通学院, 广东 广州 510641)

摘要: 首先建立波浪超定量概率模型, 提出超定量抽样分组原则以保证数据独立性, 通过检验可知该原则下波浪超定量系列能较好地拟合GP分布; 提出根据年平均出现频次、是否服从Poisson分布和拟合GP情况综合分析判断, 寻找合理阈值的方法; 分别用GP分布和相应的复合极值分布分析超定量系列, 并比较传统的P-III分布, 分析可知, 采用GP分布和P-GP(Poisson-GP)复合分布计算超定量系列所得的重现期值基本一致, 且P-III曲线所得结果接近。

关键词: 波浪重现值; 超定量; GP分布; Poisson分布

中图分类号: P 731.22

文献标志码: A

文章编号: 1002-4972(2013)05-0049-06

PDS methods for determining extreme wave height

LUO Yao, ZHU Liang-sheng

(South China University of Technology, Guangzhou 510641, China)

Abstract: The paper sets an exceedance model and proposes the sampling principle for maintaining the independence of the sample. By fitness test, the sample is preferably fitted. Then the threshold is determined by the average number of occurrence every year and goodness of fit for GPD. The result of comparative analysis shows that the results of the GPD model and P-GP model is the same and approaches that of P-III distribution.

Key words: return wave height; PDS; GPD; Poisson distribution

通过波浪历史数据估计工程设计波浪一直被各国使用, 该方法实质是根据已有波浪数据建立波高和其出现频率之间关系, 再通过外延的方法来估计百年甚至万年一遇的波高。这种方法最大的缺陷就是已有实测波浪数据时间短暂。在我国, 各海洋水文站大多不过50 a左右数据。而要正确建立波高与频率间的关系, 首先所选取的极值样本序列必须是独立同分布的随机变量, 这是频率分析的前提。目前选取波浪极值样本序列的方法主要有传统的波浪重现值和PDS法两种。

传统的波浪重现值是基于年最大值序列AMS (Annual Maxima Series) 来求解, 即每年只挑选一个最大值进行频率分析。这种方法在水文统计很多领域应用, 因为它对数据的要求比较低, 处理数据方法和计算方法相对简单。它的最大的缺

点是浪费数据, 在小浪年份相对较小的波高被选入, 而大浪年份的“次大浪”被舍弃。

如何能从短暂的数据中获得更多的信息进行波浪重现期值的计算, 水文学家Shane等^[1-2]首先在水文计算中提出PDS(Partial Duration Series)模型, 与其类似的有使用较多的POT法和我国暴雨洪峰频率分析中使用超定量序列模型^[3]。PDS法是将数据按照某规则(同一台风过程、寒潮过程等)分组, 选取每组最大值(Block Maximum)组成区组最大值序列, 最后选取区组最大值序列中大于某一定值(阈值 μ)的值作为超定量极值样本序列(PDS)进行频率分析。其实AMS可以认为是PDS的特例: 以年为分组单位选取最值序列, 以最值序列的最小值为阈值所选取的极值样本即为AMS。PDS应用主要难点是如何分组选取最值样本以及如

收稿日期: 2012-11-16

作者简介: 罗耀(1985—), 男, 博士研究生, 从事海岸工程设计, 极值海洋环境事件统计分析研究。

何选取阈值让极值样本满足独立性。王超等^[4]采用该法进行设计波浪分析, 但并未考虑如何保持样本间的独立性, 国外对PDS研究相对较多^[5-8]。

本文主要研究PDS如何具体应用在波浪多年观测资料的分析中, 以及通过PDS和传统的P-III方法^[9]的比较, 分析利用PDS计算重现期值的可行性。首先介绍超定量概率模型, 根据极值理论中GP (Generalized Pareto) 分布的理论及其性质, 超定量系列应满足GP分布, 并用Poisson分布拟合超定量年发生次数; 以中国南海某观测站波浪数据为例, 讨论如何选取极值样本, 以及在具体应用方面进行探索。

1 PDS模型基本理论

设X表示波高随机变量, 正实数μ为计算前设定的阈值, 则波浪超出量序列Y定义为:

$$Y=X-\mu \quad (1)$$

由于各国在海洋工程设计中都采用以年为单位的重现期作为标准, 在本文也采用这一重现标准。另外, 本文中PDS和超定量序列是同一定义。

1.1 超定量波浪发生次数的概率模型

从多年观测资料所选取的超定量极值序列中, 每年拥有的样本个数n组成一个离散的年出现次数序列, 由于波浪的随机特性, 可认为年内发生超定量波浪的次数N也为随机变量。如果N服从Poisson分布, 则极值样本中, 年内发生N次超定量波浪的概率 $P_\lambda(N=n)$ 为:

$$P_\lambda(N=n) = e^{-\lambda} \frac{\lambda^n}{n!} \quad n=1,2,\dots \quad (2)$$

式中: λ为Poisson分布参数, 通过最大似然法进行参数估计可知 $\lambda=\bar{N}$, 它表示在极值样本中每年波浪超定量平均发生次数。

1.2 广义Pareto分布

根据极值理论, 广义极值分布GEV (Generalised Extreme Value, 包括Gumble, Frechet和Weibull分布) 为区组最大值序列的极限理论分布, GP分布则是超定量序列的极限理论分布^[10]。假设超定量波浪序列为X, 则其服从GP分布:

$$G_{\mu,\delta,\varepsilon}(x) = 1 - \left(1 + \varepsilon \cdot \frac{x - \mu}{\sigma}\right)^{-1/\varepsilon} \quad (3)$$

式中: $x \geq \mu$, $1 + \varepsilon(x - \mu)/\sigma > 0$; G为广义Pareto分布; μ为阈值; σ为尺度参数; ε为形状参数。广义Pareto分布包含指数分布:

$$G_{\mu,\delta,\varepsilon}(x) = 1 - \exp\left(-\frac{x - \mu}{\sigma}\right) \quad \varepsilon=0 \quad (4)$$

超定量序列频率曲线的纵坐标计算式为:

$$x = \mu + \frac{\sigma}{\varepsilon} [1 - (1 - P)^{-\varepsilon}] \quad (5)$$

式中: P为累积频率值。

GP分布作为超阈值的稳定分布函数被许多学者介绍过, 有较强的极值理论背景。Pickands^[11]首先提出GP分布为超阈序列的稳定分布; Wang等^[12-13]也将GP分布运用到水文频率分析中。其参数估计方法也是各国学者研究的热点, Hosking等^[8]利用L阶矩估计参数σ和ε, 并给出概率权重矩。文献^[14-18]也分别用极大似然估计参数σ和ε, 并认为极大似然估计法具有一系列优势。本文也将利用极大似然法估计参数σ和ε。

1.3 复合极值模型

本文采用的是Liu等^[19]提出的一维复合极值分布理论。有一维离散型随机变量的概率分布律 $(0, 1, 2, \dots, k, \dots)$, 有一种连续型随机变量的概率分 $H(x)$ 布, 记

$$F_0(x) = \sum_{k=0}^{\infty} p_k [H(x)]^k \quad (6)$$

则称 $F_0(x)$ 为这两种分布构成的复合极值分布。

假设X为超定量波高序列, K为超定量波高在每年发生的频次离散序列。记 Y_{max} 为1 a内发生的若干超定量中数值最大的那个超定量, 即年最大超定量。根据复合极值理论, 可知年最大超定量 Y_{max} 的分布 $F(Y_{max})=F_0(X)$ 。p_k为Poisson分布, H(x)为Gumbel分布时就叫做Poisson-Gumbel复合极值分布:

$$F_0(x) = e^{-\lambda[1-H(x)]} \quad (7)$$

$$H(x) = \exp\{-\exp[-a(x - \mu)]\} \quad x \in R \quad (8)$$

式中: a, μ为待定参数。

Poisson-Gumbel复合极值法选取每次台风浪 (或寒潮等引起浪) 的最大值进行计算, 增加了

资料使用率, 考虑了台风频次的影响, 计算稳定性有所增加^[19-20]。但是通过调查目前文献可知, 复合极值法只用来计算台风或某一因素主导的波浪要素, 不能考虑所有因素(海啸、潮波)共同导致的极端波高进行有效分析。Poisson-Gumbel复合极值分布模型也能看作超定量取样的特例, 它巧妙地将台风过程作为分组的原则, 这样满足了取样间的独立性; 由于我国大部分海域, 台风浪是极值波浪的主要组成部分, 因此对台风浪分析基本能代表极端波浪出现的情况。而且根据台风后报波浪的方法已经较为成熟, 台风浪数据便于获取, 因此得以广泛应用。但是在小台风年, 经常出现台风浪小于大风或其他原因产生的波浪, 这时Poisson-Gumbel复合极值分布模型将不能全面考虑极端波浪出现的情况。

假设超定量序列满足 $H(x)$ 为GP分布, 则P-GP (Poisson-GP) 复合极值分布为:

$$F_{0|\mu,\varepsilon,\sigma}(x) = \exp\left[-\lambda\left(1 + \varepsilon \cdot \frac{x - \mu}{\sigma}\right)^{-1/\varepsilon}\right] \quad (9)$$

则超定量序列频率曲线的纵坐标计算式为:

$$x = \mu - \frac{\sigma}{\varepsilon} [1 - (-\lambda/\ln P)^\varepsilon] \quad (10)$$

式中: P 为累积频率。

1.4 波浪超定量序列重现期

如果 $F(x)$ 是AMS的分布函数, 则根据经典定义, 重现期 T 表示为:

$$T = \frac{1}{P(X > x)} = \frac{1}{1 - F(x)} \quad (11)$$

另外一些学者提出^[2,6,13], 假设超定量序列的年发生频次满足Poisson过程, 超定量序列的分布函数为 $G(x)$, 则重现期 T 能被定义为:

$$T = \frac{1}{\lambda[1 - G(x)]} \quad (12)$$

式中: λ 为Poisson过程强度, 在超定量序列模型中即为超定量序列的年平均发生次数。

2 波高超定量序列抽样原则

为保证超定量序列的独立性, 如何选取超定量序列样本的分组原则则很重要。海洋水文站一般每天进行多次定时观测, 得到波浪历时数据。一次大

浪过程可能有几次观测值, 且观测值都可能相对较大, 但这些数据并不相互独立。所以抽样原则如下: 一次大浪过程, 只取其中一个最大值。大浪过程应包括对波浪造成重大影响的台风过程、寒潮大风过程和地震等一次连续影响造成的过程。

3 算例分析

以南海某水文站23 a波浪资料为例, 分析超定量系列的具体取样方法和其模型计算波浪重现期值的可行性。

为满足极值样本间的独立性, 本文提出以下具体步骤进行实测波浪数据超定量取样:

1) 用统一的时间间隔50 d进行基本分组, 选取每组最大值。即每年有7~8个极值入选。

2) 逐年对每个极值进行核查、筛选, 可添加漏选的满足独立性要求的极大值和删除不满足抽样原则的两个极值波高属于一次大浪过程中的一个。对于较小的值, 可以不删除, 因为最终的波浪超定量序列必须大于阈值。

3) 确定阈值 μ , 选取超阈极值为波浪超定量序列。

以某年波高数据为例, 图1为该年波浪日最大值。通过步骤1, 入选的极值用圆圈表示, 但是在181 d时有较大值未入选, 它与前一极值相差20 d左右, 应该属于2次不同的大风过程, 所以添加入选极值样本。假设阈值 $\mu=3.0$ m, 则该年入选只有1个极值波高。根据方彬等^[21]的观点, 平均每年超阈值的个数为2~3个时, 其超阈量波高年发生次数才能较好服从Poisson分布。所以只需保持每年中较小的几个极值间独立性即可。经过下面分析可知, 以50 d为间隔进行分组, 所得的极值基本能满足抽样原则。

3.1 阈值 μ 的确定

完成步骤1和2后, 还需选取合适的阈值 μ 。阈值偏小, 则入选数据较多, λ 偏大, 数据间独立性不能保证; 阈值偏大, 则入选数据较少, λ 偏小, 年平均出现次数不能满足Poisson分布假设。本文Poisson分布用显著性水平0.05的 χ^2 拟合检验法检验, 经过试算可知, 阈值 μ 取值情况见表1。

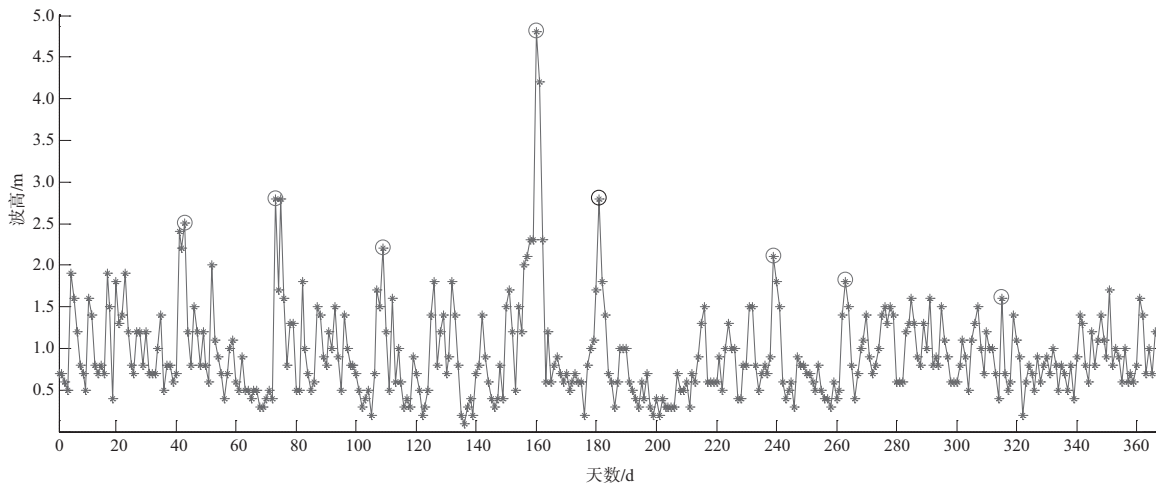
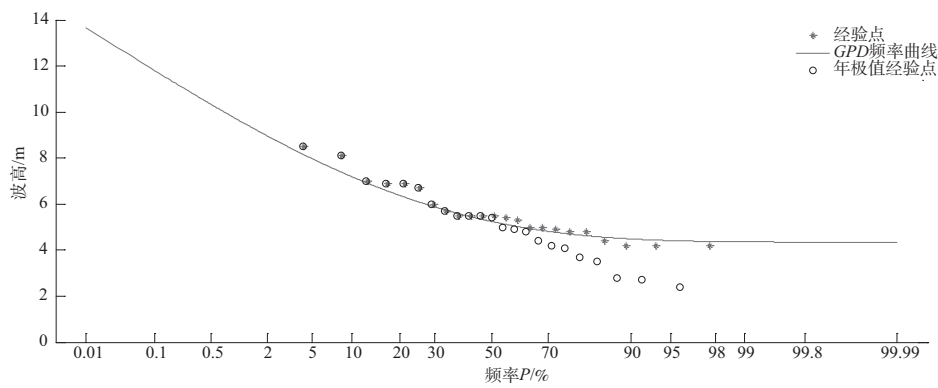


图1 某年波浪日最大值历时

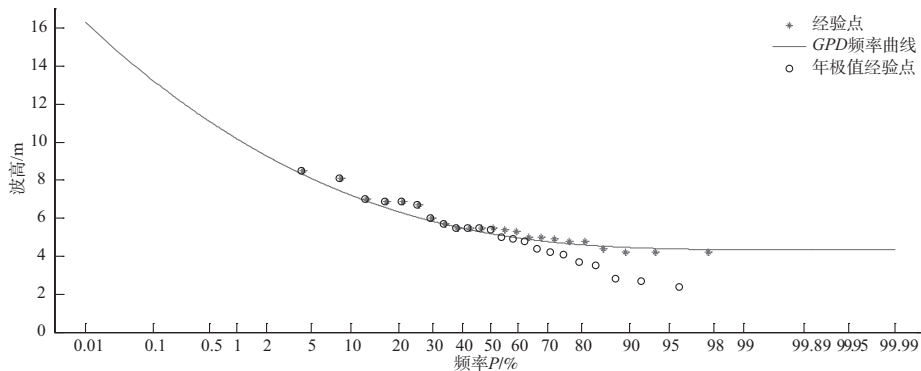
表1 阈值试算

阈值 μ/m	$\hat{\lambda}$	Possion分布拟合检验
3.0	3.70	不通过检验
3.3	2.74	不通过检验
3.4	2.52	不通过检验
3.5	2.26	通过检验
3.6	1.87	通过检验
3.7	1.65	通过检验

由表1可以看出，阈值 μ 小于3.5 m则不能满足Possion分布拟合检验， $\mu=3.7$ m时入选数据较少，平均每年仅1.65个数据入选超定量序列，所以阈值 $\mu=3.5$ m或3.55 m较好。图2分别为 $\mu=3.5$ m和3.55 m时，GP分布的拟合情况，可以看出，图中年极值点据和超定值模型所定义的点据在较大值基本重合，而且GP频率曲线也拟合较好，而在波高较小



a) $\mu=3.5$ m, $\delta_a=0.8655$



b) $\mu=3.55$ m, $\delta_b=0.9943$

图2 GP分布拟合曲线

处两种点据有较大偏差, GP拟合曲线偏差也较大。这并不影响大值区域。两图比较可以看出, 图2a) 对大值拟合更好。分别计算两图的前12个较大经验点据(大于2 a一遇值)和拟合曲线均方根误差(RMSE):

$$\delta = \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\frac{\hat{x}_i - x_i}{x_i} \right)^2} \quad (13)$$

式中: $n=12$ 。可知: $\delta_a < \delta_b$, 亦是 $\mu=3.5$ m时GP曲线拟合较好。所以本例阈值 $\mu=3.5$ m。图3为 $\mu=3.5$ m时, 超定量序列GP分布拟合的P-P和Q-Q图, 可知, 阈值 $\mu=3.5$ m时超定量序列服从GP频率曲线, 每年入选数据 $\lambda=2.26$ 。

3.2 计算结果分析

图4和图5分别是P-GP复合曲线和P-III曲线拟合图, 其计算结果列于表2。可知, 超定量模型的GP模型和P-GP复合模型计算结果一致, 仅50 a一遇以及以下值不同, 这是由于年极值点据尾部较超定量序列点据偏下, P-GP复合曲线为描述年极值点据的模型, GP曲线则直接描述的是超定量序列, 故有所不同。

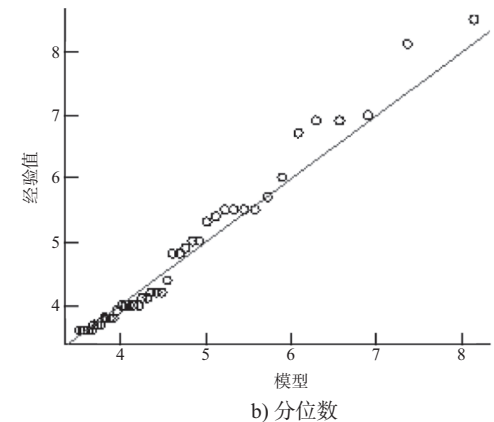
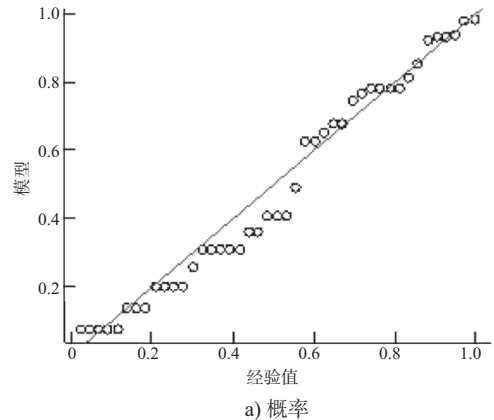


图3 GP曲线拟合检验

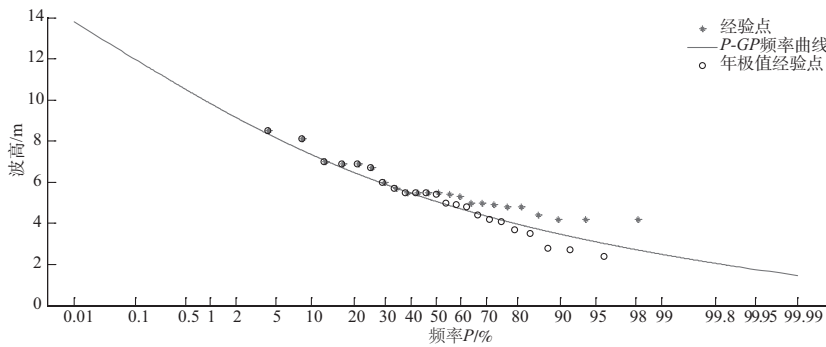


图4 P-GP复合分布曲线

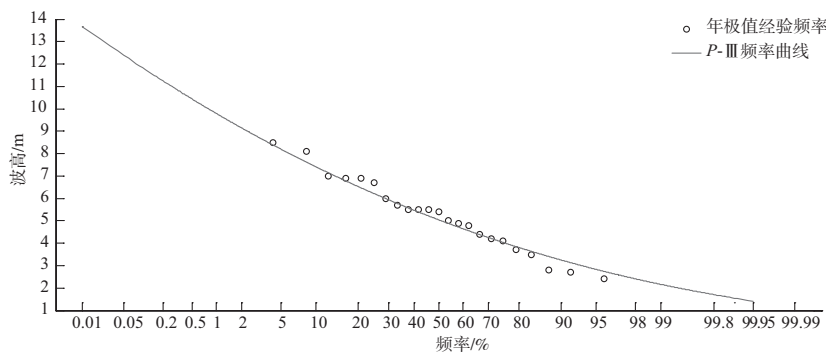


图5 P-III曲线拟合

比较波浪超定量两种模型和传统P-III曲线法的计算结果可知, 两者误差均较小, 波浪超定量模型计算设计波浪重现期值可行。

4 结论

1) 提出了波浪重现期值计算中PDS法抽样原则。按照这一原则, 提出以50 d为间隔分组取最值

表2 计算结果对比

重现期/a	GP/m	P-III/m	P-GP/m
10 000	13.65	13.67	13.65
1 000	11.80	11.83	11.80
500	11.19	11.24	11.19
200	10.34	10.44	10.34
100	9.67	9.81	9.66
50	8.96	9.14	8.95
20	7.98	8.20	7.95
10	7.20	7.42	7.14

的具体抽样步骤,经计算分析可知,该原则适合设计波浪计算。超定量序列能增加样本数量,平均每年超阈个数 $\lambda=2.26$,提高数据利用率。

2) 根据年平均出现次数 λ 、年出现频次是否满足Poisson分布和拟合GP情况综合分析判断,寻找合理阈值 μ 。为阈值的确定提供新思路。

3) 计算了GPD模型、P-GP模型和P-III分布曲线的重现波高值,其中GPD模型拟合超阈数据,实现数据的充分利用;P-GP模型拟合超阈数据,但是直接计算出重现期对应的年累计频率。该模型理论性强,有良好计算稳定性,考虑了极值发生频次等因素,有较好应用前景。

参考文献:

- [1] Shane R, Lynn W R. Mathematical model for flood risk analysis[J]. J Hydraul Div: ASCE, 1964,90:1-20.
- [2] Todorovic P, Zelenhasic E. A stochastic model for flood analysis[J]. Wat Resour Res, 1970(6): 1 641-1 648.
- [3] 王善序.超定量洪水系列系列频率分析[J]. 人民长江, 1999 (8):23-25.
- [4] 王超, 刘德辅.设计波浪选取中的不确定性分析[J].海洋学报,1991,13(6):874-881.
- [5] Santiago Beguerfa. Uncertainties in partial duration series modelling of extremes related to the choice of the threshold value[J]. Journal of Hydrology, 2005(1): 215-230.
- [6] Ashkar F, Rousselle J. Partial duration series modeling under the assumption of a Poissonian flood count[J]. Journal of Hydrology,1987(3):135-144.
- [7] Rosbjerg D. Estimation in partial duration series with independent and dependent peak values[J]. Journal of

Hydrology, 1985 (1):183-196.

- [8] Hosking J R M, Wallis J R. Parameter and quantile estimation for the generalized Pareto distribution[J]. Technometrics, 1987, 29(3):339-349.
- [9] JTJ 213—1998 海港水文规范[S].
- [10] 史道济.实用极值统计方法[M].天津:天津科学技术出版社,2006.
- [11] Pickands J. Statistical inference using extreme order statistics [J]. The Annals of Statistics, 1975(3): 119-131.
- [12] Wang Q J.The POT model described by the generalized Pareto distribution with Poisson arrival rate [J]. J Hydrol, 1991,129: 263-280.
- [13] Rosbjerg D H Madsen, Rasmussen P F. Prediction in partial duration series with generalized Pareto-distributed exceedances [J]. Water Resour Res, 1992, 28: 3 001-3 010.
- [14] Martins E S, Stedinger J R. Generalized maximum-likelihood Pareto-Poisson estimators for partial duration series [J]. Water Resources Research, 2001, 37: 2 551-2 557.
- [15] DuMouchel W. Estimating the stable index α in order to measure tail thickness: A critique [J]. The Annals of Statistics, 1983(11): 1 019-1 036.
- [16] Davison A C. Modelling excesses over high thresholds, with an application[M]//Statistical Extremes and Applications. Dordrecht: Reidel, 1984: 461-482.
- [17] Smith R L. Threshold methods for sample extremes, in statistical extremes and applications [J]. Dordrecht: Reidel, 1984: 621-638.
- [18] Smith R L. Estimating tails of probability distributions [J]. The Annals of Statistics,1987, 15, 1 174-1 207.
- [19] Liu T F, Ma F S. Prediction of extreme wave heights and wind velocities[J]. J Water Port Coast Ocean Div, 1980, 106(4): 469-479.
- [20] Liu D F, Wang L P, Pang L. Theory of multivariate compound extreme value distribution and its application to extreme sea state prediction [J]. Chinese Science Bulletin, 2006, 51(23): 2 926-2 930.
- [21] 方彬, 郭生练, 柴晓玲,等. FPOT 方法在洪水频率分析中的应用研究 [J]. 水力发电, 2005, 31(2): 9-12.

(本文编辑 武亚庆)