

考虑不确定度的 POT 模型 在洪水重现期分析中的应用研究

纪忠华,王璐,路雨

(环境保护部核与辐射安全中心,北京 100082)

摘要:以淮河紫罗山子流域出口日平均流量数据为研究对象,基于超阈值(POT)模型,采用最大似然法估计广义 Pareto(GP)分布参数并计算出重现期水平和相应的置信区间范围。拟合优度检验结果显示 POT 模型在扩大洪水样本提高使用效率的同时,对样本经验点据的适线性也较好。通过对 5 种时段长度的水文实测流量数据重现期计算发现:实测数据长度对重现期计算结果不确定性有重要影响,在工程水文推荐选取恰当的置信区间上界作为设计值加以解决。

关键词:防洪安全;GP 分布;最大似然法;置信区间

中图分类号:TV697 **文献标识码:**A **文章编号:**1000-0852(2017)04-0006-06

1 前言

工程水文分析计算方法主要分为确定性方法和随机性方法两类。确定性方法主要依靠数学工具与物理模型来反演洪水灾害过程,而随机性方法则以水文历史或实测数据为基础,利用数理统计方法外推,从而得到在所求重现期的洪水量值。年最大值法(Annual maximal series,AMS)由于概念清晰、理论成熟在国内水文统计计算中得到广泛应用,但在实际工程应用中仍存有诸多限制。首先为保证计算结果可靠,一般要求数据观测年限不得少于 30 年,导致在缺资料地区应用较为困难,即使按有关技术规范可对缺测数据进行插补和延拓,但仍会对洪水分析带来不确定性。此外,年最大值方法也会丢弃丰水年许多有价值的洪水数据,造成水文信息的极大浪费,与统计分析中最大程度获取极值样本的原则相背离。

目前,国内外水文研究者提出了多种扩充样本的方法,其中较为常用的是超阈值(Peaks over threshold, POT)模型。该模型将水文事件视为离散随机过程,把超过某一洪水阈值的极值数据都作为有效样本加入到统计分析中。该方法可最大程度地挖掘实测资料中有

价值的信息,增加统计分析中的样本数量,进而提高频率计算结果精度。Pickands^[1]等人对 POT 模型研究表明:当阈值选取得足够大,则超出阈值的样本序列是符合广义帕累托(Generalized Pareto,GP)分布的稳定分布。POT 模型计算可靠性和拟合优度检验方面,Lang 等^[2]对将 GP 分布应用到 POT 模型过程中需注意的典型问题进行了总结,探讨了阈值的选取对频率计算结果的影响,认为一个合适阈值应使超阈值样本量为数据年限长度的 2~3 倍左右;王剑锋等^[3]将 GP 分布应用到 POT 模型中,提出改进的线性矩法以提高计算的精度,研究表明:在重现期较大时,三参数 GP 分布的理论频率曲线与经验频率点的拟合效果比传统 Pearson-III 型分布更好;张丽娟等^[4]对比分析了历史洪水数据对 POT 模型计算结果的影响,也认为 POT 方法的拟合优度优于 AMS 方法。

由于洪水灾害是在多种复杂环境要素作用下形成的,频率计算不仅受统计分布函数、参数估计方法等因素的影响,还受数据量多寡、观测时间长短及自然环境的变化等客观因素的影响,特别是在全球气候变化和人类活动的大背景下,常有超越预期的极端洪水事件出现,这促使工程设计中需对计算成果的不确

收稿日期:2016-03-04

基金项目:科技部国家软科学研究计划(2013GXS4B075);环保部公益性行业科研专项(201309056)

作者简介:纪忠华(1985-),男,江苏扬州人,工程师,工学硕士,现主要从事核电厂防洪安全审评工作。E-mail:jizhonghua@chinansc.cn

通讯作者:王璐(1983-),女,山东青岛人,高级工程师,主要从事核电厂厂址安全审评工作。E-mail:wanglu@chinansc.cn

定性进行合理考量,确保工程设计成果具有相当的安全水平并留有足够的裕度。对此,国内外相关学者也从不同的角度,进行了多种理论方法的研究^[5-7],这些方法普遍着眼于模型参数估计中的不确定性,即根据参数估计的误差,通过设置合理的置信区间来表征不确定度的大小。本文选用淮河紫罗山流域为研究对象,根据流域出口断面流量资料,尝试采用基于 GP 分布的 POT 模型,构建并求解模型的似然函数及其置信区间范围,分析不同观测年限的样本数据对水文频率计算结果的影响。同时,样本的一致性检验、POT 模型阈值的选取原则以及适线结果的拟合优度检验也在此过程中一并加以讨论。

2 理论分析

2.1 统计分布函数

对于独立同分布的样本序列 $\{X_i\}$,当存在固定值 u ,使得 $X_i > u$ 。则称 u 为阈值, $x-u$ 为超阈值,其平均超出量函数可以表示为:

$$F_u(x-u) = \Pr\{X-u \leq x | X > u\} = \frac{F(x) - F(u)}{1 - F(u)}, x \geq 0 \quad (1)$$

对于三参数 GP 分布函数可表示如下:

$$F_{(x;\mu,\zeta,\sigma)} = 1 - (1 + \zeta \frac{x-\mu}{\sigma})^{-1/\zeta}, x \geq u, 1 + \zeta(x-\mu)/\sigma > 0 \quad (2)$$

式中: $F(x)$ 为不超越概率分布函数; μ 为位置参数; $\sigma > 0$ 是尺度参数; ζ 为形状参数。

综合(1)(2)两式,对于采用 POT 抽样方法获得的样本,当样本数量足够大时,GP 分布函数可进一步改写成以下形式:

$$F_{(y;\zeta,\sigma)} = \begin{cases} 1 - (1 + \frac{\zeta y}{\tilde{\sigma}})^{-1/\zeta}, & \zeta \neq 0 \\ 1 - e^{-y/\tilde{\sigma}}, & \zeta = 0 \end{cases} \quad (3)$$

式中: y 变量表示样本超过阈值 u 的部分 $\{y: Y = X - u\}$; $\tilde{\sigma}$ 为修正后的尺度参数,与原参数的关系为: $\tilde{\sigma} = \sigma + \zeta(u - \mu)$ 。因此在后续参数估计中只需要针对形状和修正后的尺度参数进行相应计算。

可以证明^[8]POT 模型筛选得出的时间段(w)内样本序列属于速率为 λw 的齐次泊松过程(Homogeneous Poisson Process, HPP),且相邻洪水发生的时间间隔满足指数分布。在实际应用中,一般将实测数据划分为若干固定时段(通常为 1a)。当每年超过阈值的洪水平均发生次数为 λ 时,则 T 年重现期的洪水发生频率为

$1/(\lambda T)$, POT 模型洪水重现期计算公式可表示为:

$$T = 1/\lambda(1 - F(x)) \quad (4)$$

式中: $F(x)$ 为 GP 分布函数; T 为重现期; λ 为年超越阈值速率(年平均超阈值样本数量)。类似的符合超阈值分布的经验频率则可以用经过修正的期望值公式表达:

$$P_e \approx i \cdot \lambda / (N+1) \quad (5)$$

式中: P_e 为超阈值样本经验频率; i 代表超阈值样本 $\{x_i\}$ 降序排列后的序号。

2.2 参数估计方法与分位点估计

最大似然法是利用极大似然原理来估计统计模型参数的通用点估计方法,通过构建目标函数,进而求解适用于模型最优参数估计。根据公式(4)可以看出,采用 GP 分布来计算重现期水平,除了需估计尺度(σ)和形状(ζ)参数外,还需要估计年超越阈值速率 λ 。由前文所述,超阈值的洪水年发生次数服从齐次泊松过程,因此除了采用两参数 GP 分布的密度函数乘积的形式外,还需将泊松分布函数加入到最大似然目标函数中,所构建的似然方程可表示为:

$$L(\lambda, \sigma, \zeta) = \frac{(\lambda w)^N}{N!} e^{-\lambda w} \times \prod_{i=1}^N f_X(x_i; \sigma, \zeta) \quad (6)$$

结合式(6),可以进一步推得其对数函数形式:

当 $\zeta \neq 0$ 时:

$$\begin{aligned} \log L(\lambda, \sigma, \zeta) &= N \log(\lambda w) - \lambda w - \log(N!) \\ &\quad - N \log \sigma - (1 + 1/\zeta) \sum_{i=1}^N \log(1 + \zeta x_i / \sigma) \end{aligned} \quad (7)$$

当 $\zeta = 0$ 时:

$$\log L(\lambda, \sigma, \zeta) = N \log(\lambda w) - \lambda w - \log(N!) - N \log \sigma - \sigma^{-1} \sum_{i=1}^N x_i \quad (8)$$

式中: N 为超阈值样本总数; w 为时段长度,本文以自然年长度为固定时段,其它参数含义同上。

求解上述似然方程可得到 GP 分布的参数估计值,可直接用于 POT 模型的在各个分位点处的计算,得到与累计概率(p)对应的分位点的极大似然估计 \hat{z}_p 为:

$$\hat{z}_p = u + \frac{\hat{\sigma}}{\hat{\zeta}} \left[\left(\frac{\hat{\lambda}}{1-p} \right)^{\hat{\zeta}} - 1 \right], \quad \zeta \neq 0 \quad (9)$$

$$\hat{z}_p = u + \hat{\sigma} \log \left(\frac{\hat{\lambda}}{1-p} \right), \quad \zeta = 0 \quad (10)$$

值得说明的是,对于年平均超越概率的估计 $\hat{\lambda}$,可

通过计算超阈值样本总数与样本统计年限的比值直接获得,即 $\hat{\lambda}=N/\omega$ 。

2.3 置信区间计算

从极大似然函数中待估计量的渐进正态性假设出发,采用 Delta 方法来计算置信区间,即通过 Taylor 展开式对分布函数相关参数置信区间求解得到,GP 分布上(下)侧置信区间边界可用下式表示:

$$z_{U,L}=\hat{z}_p\pm K_\alpha\sqrt{\text{var}(\hat{z}_p)}\approx\hat{z}_p\pm K_\alpha\sqrt{D_pVD_p^T} \quad (11)$$

式中: $z_{U,L}$ 为上、下侧置信边界; K_α 为正态分布 α 分位点处的概率值, $\alpha=(1-p)/2$; $\text{var}(\hat{z}_p)$ 为累计概率 p 水平下的误差; V 为费雪信息阵(Fisher information matrix),可由目标似然函数 $\log(L)$ 的协方差矩阵计算得到; D_p 为似然函数对各参数变量的偏导数向量;对于 GP 分布, V 、 D_p 可根据上述式(8)和式(9)进一步推算得到:

$$V=\begin{vmatrix} \text{var}(\hat{\lambda}) & 0 & 0 \\ 0 & \text{var}(\hat{\sigma}) & \text{cov}(\hat{\sigma},\hat{\xi}) \\ 0 & \text{cov}(\hat{\sigma},\hat{\xi}) & \text{var}(\hat{\xi}) \end{vmatrix} \quad (12)$$

$$D_p=\left(\frac{\partial z_p}{\partial \lambda}, \frac{\partial z_p}{\partial \sigma}, \frac{\partial z_p}{\partial \xi}\right)=(\sigma(1-p)^{-\xi}\lambda^{\xi-1}, \xi^{-1}[(1-p)^{-\xi}\lambda^{\xi}-1], -\sigma\xi^{-2}[(1-p)^{-\xi}\lambda^{\xi}-1]+(1-p)^{-\xi}\lambda^{\xi}\log[(1-p)\lambda]) \quad (13)$$

3 统计建模

3.1 研究区概况及独立洪峰筛选

选择淮河流域紫罗山子流域为研究区,紫罗山流域位于淮河一级支流沙颍河的中上游,控制面积约1800km²,流域为半山丘陵地貌,最大海拔高度约2100m,最低约300m。平均高程约820m,地势由西向东逐渐降低。流域属大陆性季风气候,多年平均降水量约为900mm,降雨主要集中在每年6~8月份。紫罗山水文站上游未建有大中型水利工程,保持了较为原始的自然形态,因此其长期实测资料具有较好的代表性和一致性。

本文收集了紫罗山流域出口水文站1952~2010年共59a实测日流量数据。在此基础上,截取研究区1952~1991、1952~1996、1952~2001、1952~2006和1952~2010年流量数据,对比分析40、45、50、55和59a5种时长数据对计算结果不确定性的影响。

进行极值统计分析时,不仅仅要考虑量值的大小,还需保证样本的独立性,避免所使用的数据出现自相关现象。目前国内外已提出多种独立次洪过程的分割

判别标准^[9],其中较为简便的方法是限值法,即根据实测洪水数据变化情况估计出一个流量限值,提取超过这一限值的洪峰流量,该方法在忽略掉一部分小洪水事件的同时也会截断前后相继的次洪过程,本文采用该法达到获取独立洪峰数据的目的。

3.2 阈值的确定

如前文所述,利用 POT 模型对极值洪水样本进行统计计算的关键是提出一个恰当的阈值 u ,阈值选取得过高则减少了有效样本数量,太低则会导致抽样数据过多,不符合极值分布的含义。在样本选取方法上,相关学者提出了多种准则来帮助获取最优的样本数据,目前国内外还没有一个统一的标准,较为常用的有超阈值样本均值法、分散指数法^[10]、年均超阈值发生次数法等,其中超阈值样本均值法应用效果较好,该方法认为所有样本超过阈值的部分的均值(即平均剩余寿命值,Mean Residual Life)应呈稳定线性函数关系。Coles^[8]对前人的研究进行总结后进一步提出:一个最佳的阈值还应使所拟合的 GP 分布的形状及尺度参数及其置信区间也呈稳定线性状态,在稳定区间内,尽量选用较大的阈值以达到筛选极值样本的初衷。按此方法,绘出研究区 59a 独立洪峰流量数据的平均剩余寿命及 GP 分布的形状、尺度参数值随阈值变化曲线(图1)。

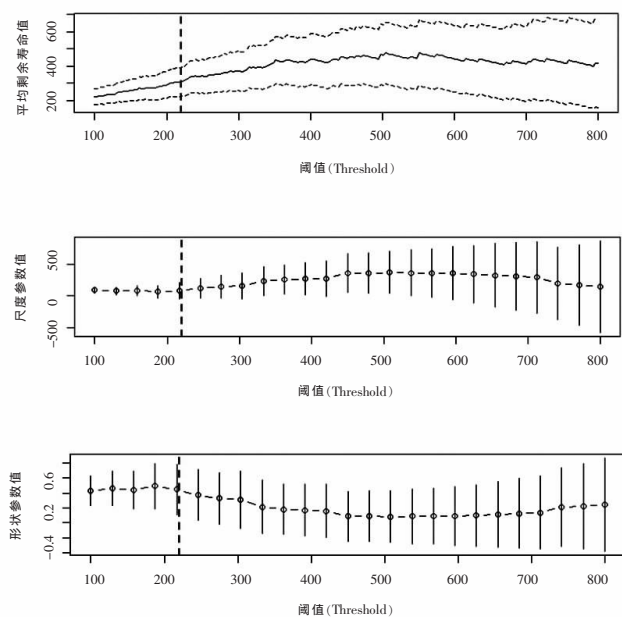


图1 剩余寿命值及 GP 分布尺度、形状参数值及其 95% 置信区间随阈值变化曲线图

Fig.1 Estimation of threshold: the GPD scale and shape parameters and mean residual life plot with 95% confidence intervals

从图 1 可以看出:当阈值取在 180~280m³/s 范围附近时,平均剩余寿命值、形状及尺度参数这三种指标值较为稳定,即参数在此区间内不随阈值的变化而有较大的波动,此外三种指标的置信区间在这范围内也较小,可以认为 POT 模型阈值在此区间取值时确定度较高。综合考虑,本文选择 220m³/s 作为 POT 模型的最佳阈值,由此获得的各时段超阈值极值样本数量见表 1。在此基础上利用 Mann-Kendall、Wilcoxon 以及 Wald-Wolfowitz^[12]三种非参数检验方法对超阈值样本

进行相关假设检验。其中 Mann-Kendall 检验用于时间序列的趋势性分析;Wald-Wolfowitz 检验用于检查数据的随机性,间接证明所筛选样本的独立性;采用 Wilcoxon 秩检验进行同分布检验以确保样本的一致性。从检验结果(见表 1)可以看出当阈值取为 220m³/s 时,不同时段数据时长的三种双侧检验 *P* 值均大于 0.05,不能被 95%置信水平的原假设拒绝,即所筛选超阈值样本满足 POT 模型独立同分布的前提条件。

表1 样本独立性检验结果
Table1 The independent test results for samples

| 数据时长/a | 超阈值样本量 | 超阈值样本数量/数据时长 | 检验统计量 <i>Dr/p</i> 值 | | |
|--------|--------|--------------|---------------------|--------------|----------------|
| | | | Mann-Kendall | Wilcoxon | Wald-Wolfowitz |
| 40 | 78 | 1.95 | -0.05(0.49) | 871.5(0.37) | -1.51(0.93) |
| 45 | 82 | 1.82 | -0.06(0.43) | 918.0(0.61) | -1.62(0.95) |
| 50 | 84 | 1.68 | -0.08(0.25) | 986.5(0.46) | -1.56(0.94) |
| 55 | 91 | 1.65 | -0.10(0.16) | 1178.0(0.35) | -1.40(0.91) |
| 59 | 96 | 1.63 | -0.04(0.61) | 1235.5(0.67) | -1.25(0.90) |

3.3 重现期计算结果及拟合优度检验

采用最大似然法估计 GP 分布函数中各参数,据此得到不同重现期洪量大小及相应 70% 双侧置

信区间范围(表 2),受篇幅限制本文仅给出 59 年时段数据得到 GP 分布适线图(图 2)及相应的 *Q-Q* 图(图 3)。

表2 GP分布重现期水平及70%置信区间上界计算结果
Table2 The return period level and 70% confidence upper bound of GP distribution

| 观测时段长度 | 各重现期水平及置信区间上界 | | | | | | | |
|--------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|--------------------------|--------------------------|
| | <i>Z</i> ₁₀₀ | <i>U</i> ₁₀₀ | <i>Z</i> ₂₀₀ | <i>U</i> ₂₀₀ | <i>Z</i> ₅₀₀ | <i>U</i> ₅₀₀ | <i>Z</i> ₁₀₀₀ | <i>U</i> ₁₀₀₀ |
| 40 | 3438.25 | 4891.53 | 4375.27 | 6611.32 | 5954.67 | 9737.39 | 7473.98 | 12957.63 |
| 45 | 3142.47 | 4320.19 | 3921.11 | 5702.27 | 5191.48 | 8132.77 | 6375.17 | 10557.06 |
| 50 | 2993.22 | 4051.79 | 3720.45 | 5320.97 | 4897.51 | 7536.55 | 5985.70 | 9730.31 |
| 55 | 3715.44 | 5436.45 | 5089.11 | 7958.45 | 7668.99 | 13066.18 | 10426.19 | 18908.76 |
| 59 | 2863.97 | 3820.86 | 3577.86 | 5031.10 | 4743.74 | 7153.95 | 5831.11 | 9267.04 |

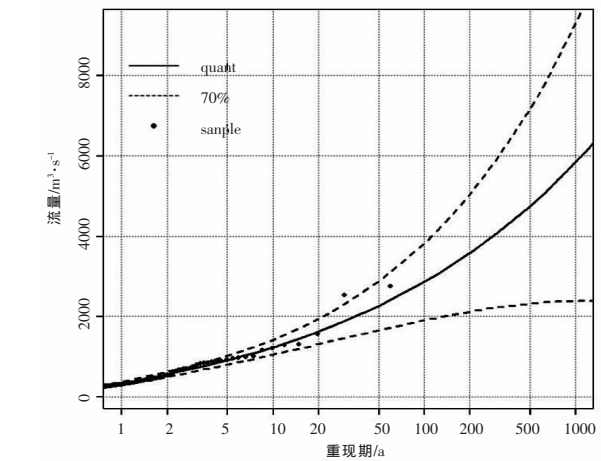


图 2 59 年时段流量数据重现期计算结果
Fig.2 Calculation of recurrence interval of flow during 59 years

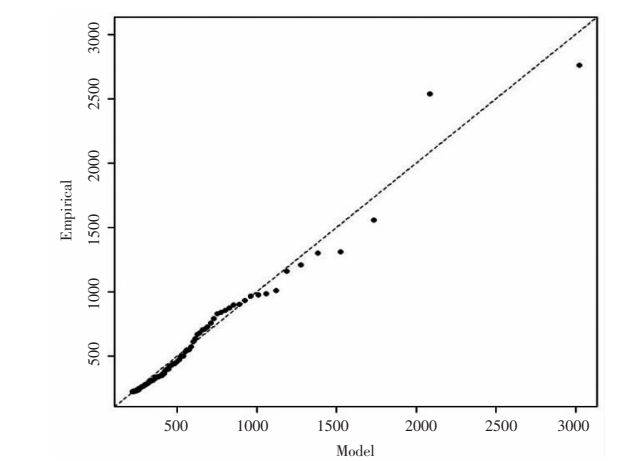


图 3 59 年时段流量 GP 分布 *Q-Q* 图
Fig.3 *Q-Q* plot based on 59-year discharge data

从图3看出该时段样本点经验累计分布与理论分布曲线基本沿对角线分布,初步说明样本满足适用GP分布的前提假设,其他时段数据的 $Q-Q$ 图形验证结果也类似,不再赘述。进一步选择较为常用的Kolmogorov-Smirnov(K-S)、Chi-square(χ^2)统计检验法对各时段样本经验点据与理论曲线吻合程度进行拟合优度检验(表3)。检验结果显示:除个别时段(40年时段数据的卡方检验结果)外,其余各时段样本理论与经验频率值拟合检验的 P 值均大于0.05,通过置信度水平为95%的K-S及卡方检验,证明样本与理论频率曲线已达到较好的适线效果。

表3 拟合优度检验
Table3 The goodness-of-fit test results

| 数据时长/a | K-S 检验 | | χ^2 检验 | |
|--------|--------|-------|-------------|-------|
| | D_n | P 值 | D_n | P 值 |
| 40 | 0.11 | 0.27 | 6.03 | 0.02 |
| 45 | 0.14 | 0.08 | 5.85 | 0.21 |
| 50 | 0.15 | 0.04 | 9.14 | 0.06 |
| 55 | 0.07 | 0.79 | 9.99 | 0.07 |
| 59 | 0.11 | 0.21 | 8.91 | 0.05 |

3.4 不确定性讨论

为分析数据量多寡及时间长度对计算结果不确定性的影响,依据前述步骤及表3,以折线的形式绘制5种时段数据计算得到的不同重现期水平值,并将相应的70%上侧置信区间范围在图中以灰色区域显示(图4)。

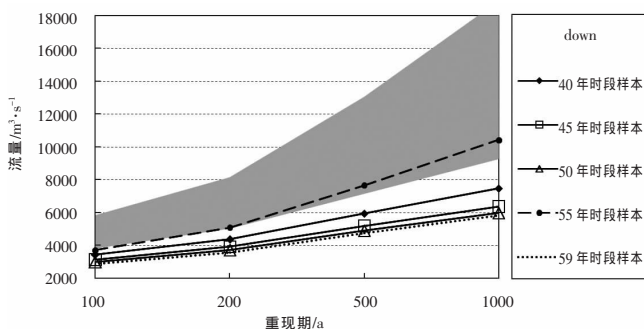


图4 各时段流量资料的重现期水平及70%置信区间上界范围
Fig.4 The recurrence interval and 70% confidence upper bound of the discharge data

从图4可以看出紫罗山流域100a、200a、500a及1000a日平均洪峰流量重现期水平分别在3000~4000m³/s、3500~5000m³/s、4600~7600m³/s和5600~10000m³/s范围左右。从纵向上来看,随着所用数据时

间长度增加,各重现期计算结果差异范围随样本数据时段的延长都呈显著增长的趋势,但对于同一重现期上的洪水流量大小与所用数据的时段长度无确定比例关系。如各重现期上最大流量值是由55年时段样本计算得出,而最小值却是由更长时段(59年)的样本算得,其余时段样本的重现期计算结果居于这两者之间。这一方面反映观测年限长度及有效样本数量的多少对重现期计算有较大的影响,另一方面则说明通过延长实测数据长度来增加样本数量时,新补充的水文数据仍受环境条件作用,当补充观测的年份属于丰水年时会得到一些偏高的离群值(Outlier)样本,而长时间的枯水年实测数据的加入则会导致代表性样本不足,即使改进统计方法(如本文所选用的POT模型),也不能完全消除此影响。通过观察图4中各重现期水平值与灰色区域位置关系可以发现:各时段70%上侧置信区间范围基本位于各重现期水平之上,且计算得到各时段不同重现期的最大流量值也仅仅与相应上侧置信区间的底部持平。可以认为,在工程设计中,如果通过设置并选取恰当置信区间,并将其上侧作为设计值时,一定程度上可以消除数据资料年限长短所带来的不确定性,使设计成果更为可靠。

4 结论

本文基于淮河紫罗山流域出口流量数据,阐述了基于GP分布的POT模型进行洪水重现期计算的理论和方法,在此基础上计算了研究区各重现期流量大小及相应的置信区间范围,最后利用K-S检验、 χ^2 检验对经验分布点据和理论分布曲线进行拟合优度检验,通过对比分析5种时段流量数据的重现期计算结果,可得出以下结论:

(1) POT模型抽样法相较于年最大值取样方法可使数据增加1.5~2倍左右,在数据量短缺情况下能有效扩大样本数量,满足统计分析中对大样本量的要求。

(2) 从拟合优度检验结果看,GP分布能较好的对洪水流量数据进行适线,在传统水文统计分析方法无法适用的情况下,POT模型可以作为替代手段在洪水统计分析中加以使用。

(3) 通过分析不同时段长度流量资料GP分布重现期计算结果,显示出资料长度对计算结果的不确定性有重要影响,在工程设计中可以通过选取一定的置信度水平的上侧置信区间值来确保设计成果的保守性。

参考文献:

- [1] James PIII. Statistical inference using extreme order statistics[J].The Annals of Statistics, 1975,3(1):119 - 131.
- [2] Lang M, Ouarda T B M J, Bobée B. Towards operational guidelines for over-threshold modeling[J].Journal of Hydrology, 1999,225(3-4):103-117.
- [3] 王剑峰,宋松柏. 广义 Pareto 分布在超定量洪水序列频率分析中的应用[J]. 西北农林科技大学学报, 2010,38(2):191-196.(WANG Jianfeng, SONG Songbai. Application of generalized Pareto distribution in POT flood series frequency analysis [J]. Journal of Northwest A & F University, 2010,38(2):191-196.(in Chinese))
- [4] 张丽娟,陈晓宏,叶长青,等.考虑历史洪水的武江超定量洪水频率分析[J].水利学报, 2013,44(3):268-275.(ZHANG Lijuan, CHEN Xiaohong, YE Changqing, et al.POT flood frequency analysis with historical floods in Wujiang River [J].Journal of Hydraulic Engineering, 2013,44(3):268-275.(in Chinese))
- [5] 尚晓三,王振龙,王栋. 基于贝叶斯理论的水文频率参数估计不确定性分析—以 P- 型分布为例[J].应用基础与工程科学学报, 2011,19(4):554 -563. (SHANG Xiaosan, WANG Zhenlong, WANG Dong. Uncertainty analysis of parameters estimation in hydrologic frequency analysis based on Bayesian method: a case study of P-distribution[J].Journal of Basic Science and Engineering, 2011,19(4): 554-563.(in Chinese))
- [6] 冯平, 黄凯. 水文序列非一致性对其参数估计不确定性影响研究[J]. 水利学报, 2015,46(10):1145-1153.(FENG Ping, HUANG Kai. A study about the impacts of non-stationary characteristic on the uncertainties of parameters estimation in hydrological series [J]. Journal of Hydraulic Engineering, 2015,46 (10):1145 -1153. (in Chinese))
- [7] Hall MJ, Van den boogaard RC, Fernando AEM. The construction of confidence intervals for frequency analysis using resampling techniques [J].Hydrology and Earth System Sciences,2004,8 (2):235-246.
- [8] Coles S. An Introduction to Statistical Modeling of Extreme Values[M]. Springer,2001.
- [9] Cunnane C. A particular comparison of annual maxima and partial duration series methods of flood frequency prediction [J]. Journal of Hydrology, 1973,18(2):257-271.
- [10] Ashkar F, Rousselle J. Partial duration series modeling under the assumption of a Poissonian flood count [J]. Journal of Hydrology, 1987,90(1-2):135-144.
- [11] 王善序.洪水超定量系列频率分析[J].人民长江, 1999,30(8):23-25. (WANG Shanxu. Frequency analysis in POT flood series [J]. Yangtze River, 1999,30(8):23-25.(in Chinese))
- [12] Wald A, Wolfowitz J. An exact test for randomness in the non-parametric case based on serial correlation[J]. The Annals of Mathematical Statistics,1943,14(4):378 - 388.
- [13] Hamdi Y, Bardet L, Duluc CM. Extreme storm surges: a comparative study of frequency analysis approaches[J]. Natural Hazards Earth System Sciences, 2014,14(1-4):2053-2067.
- [14] Begueria S. Uncertainties in partial duration series modeling of extremes related to the choice of the threshold value[J]. Journal of Hydrology, 2005,303(1-4):215-230.

Floods Return Period Calculation with Uncertainty Analysis Via POT Model

JI Zhonghua, WANG Lu, LU Yu

(Nuclear and Radiation Safety Center, MEP, Beijing 100082, China)

Abstract: Choosing the observed daily data from the Ziluoshan Station in the Huaihe River Basin, this paper used the maximum-likelihood method to estimate Generalized Pareto (GP) distribution parameters for calculating the recurrence interval level and confidence interval with the Peak of Threshold (POT) approach. The goodness-of-fitting test shows that the POT model can get a better empirical sample point while enlarge the flood sample. The sensitive analyses of different time duration flood data show that the time duration has a significant impact on the recurrence interval calculation.

Key words: flood control safety; GP distribution; maximum likelihood method; confidence interval