

基于广义 Pareto 分布的洪水序列频率分析*

赵玲玲^{1,2}, 陈子燊³, 刘昌明^{2,4}, 许扬生⁵

- (1. 广州地理研究所广东省地理空间信息技术与应用公共实验室, 广东 广州 510070;
2. 中国科学院地理科学与资源研究所中国科学院陆地水循环及地表过程重点实验室, 北京 100101;
3. 中山大学地理科学与规划学院, 广东 广州 510275;
4. 北京师范大学水科学研究院, 北京 100875;
5. 广东省水文局, 广东 广州 510050)

摘要: 广义 Pareto 分布是超定量洪水频率分析的常用线型, 其阈值的合理选择是关键。对洪水序列超阈值分布的阈值选择和超定量样本的适用性检验两个方面深入讨论, 为设计洪水提供更优的频率分析方法。基于广东省曹江流域大拜水文站 1968 ~ 2013 年日流量的洪水序列数据, 采用不同的抽样方法, 检验不同洪水超定量的泊松分布、广义极值分布和超定量样本分布的拟合优度指标。将择优的 GPD、GEV 和 P-III 型模型推算的洪水重现水平做了对比分析。获得以下结论: 大拜站洪水序列 GPD 模型属于重尾分布; 洪水 GPD 阈值的选择可以参考经验平均超过函数图, 而最佳阈值应采用多种指标综合确定; 三种概率分布的拟合优度结果显示, 构建的大拜站洪水 GPD 模型精度与 GEV 和 P-III 型相比较为高; GPD 的参数估计方法对洪水重现水平的计算结果有较大影响。

关键词: 广义 Pareto 分布; P-III 型分布; 超定量洪水阈值; Anderson-Darling 检验

中图分类号: TV 122 **文献标志码:** A **文章编号:** 0529-6579 (2019) 03-0032-08

Flood sequence frequency analysis based on generalized Pareto distribution

ZHAO Lingling^{1,2}, CHEN Zishen³, LIU Changming^{2,4}, XU Yangsheng⁵

- (1. Guangzhou Institute of Geography, Guangzhou 510070, China;
2. Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China;
3. School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, China;
4. College of Water Sciences, Beijing Normal University, Beijing 100875, China;
5. Hydrological Bureau of Guangdong Province, Guangzhou 510050, China)

Abstract: The threshold selections, excess value sample fitness test of peaks over threshold model of the flood series were deeply approached to provide a better frequency analysis method for flood designing in this paper. As a flood POT model example with the daily discharges measured at Dabai hydrologic station located in Caojiang River basin of Guangdong Province. The threshold selections were used by empirical mean excess function chart, and the Poisson distribution and the fitness indicators of excess values of

* 收稿日期: 2018-08-20

基金项目: 国家自然科学基金 (41771044, 41501046); 广东省水利科技创新项目 (2016-14); 广东省科学院发展专项资金 (2019GDASYL-0104003); 广东省科技计划项目 (2018B030324002); 广东省科学院实施创新驱动发展能力建设专项 (2017GDASCX-0101, 2018GDASCX-0101)

作者简介: 赵玲玲 (1980 年生), 女; 研究方向: 中小流域暴雨洪水预警预报; E-mail: linglingzhao@gdas.ac.cn

通信作者: 陈子燊 (1952 年生), 男; 研究方向: 极端水文气象事件风险概率研究; E-mail: eesczs@mail.sysu.edu.cn

peaks above different high level were tested. And then, a contrastive analysis was made among the flood models of optimal GPD and Pearson Type III distribution. The main findings were as follows: 1) The Caojiang River flood extreme value distributions were heavy tail type; 2) The empirical mean excess function chart can be used as an important selection reference of Dabai flood threshold, but the optimal threshold determination should be investigated with multi-index; 3) The good-fitness test indicated that the GPD model was better than Pearson Type III; 4) The parameter estimate methods have higher impact on the calculation results of flood return levels.

Key words: Generalized Pareto distribution; Pearson III type distribution; peak over flood thresholds; Anderson-Darling test

变化环境下导致极端事件呈现增加趋势, 其中洪水属于典型的灾害事件。基于有限的水文观测数据中发现尽可能多的洪水的规律, 提高推算洪水重现水平的精度, 对防洪工程规划和灾害风险评估有重要作用。超限频率分析是极值统计建模理论的重要组成部分^[1], 国内外已有不少探索与研究。研究人员对超定量样本的频率分布线型的适用性做了较多研究, 如指数分布^[2-3]、Gamma 分布和 Weibull 分布^[4]。之后研究集中在广义 Pareto 分布 (GPD)^[5-8]。而 GPD 应用于实际数据集的成功与否很大程度上取决于参数估计过程。Hosking 等^[9]使用极大似然法 (ML) 和概率权重矩法 (PWM) 估计 GPD 模型的参数, Zhang^[10]提出了基于 GPD 的参数估计似然矩法。Bermudez 等^[11-12]详细介绍了多数情况下使用的最大似然 (ML)、概率加权矩 (PWM) 和矩法 (MOM) 参数估计方法及其优缺点。王剑峰等^[13]用常规线性矩法和改进线性矩法对广义 Pareto 分布参数进行估计, 并对比分析了超定量序列频率。周长让等^[14]采用高阶概率权重矩估计其分布参数, 统计试验表明该方法具更高的参数估计精度, 估参结果对应的 GPD 曲线能较好地拟合稀遇频率洪水段的经验频率分布。但在洪水超限频率分析中如何选取阈值这一关键问题、超定量数和超定量分布的拟合优度检验等还需要更多的探索与实践^[7]。

本文以山区中小流域日流量过程为例, 深入探讨广义 Pareto 分布 (GPD) 模型的阈值选择方法, 洪水序列超定量数检验、拟合优度检验, 最后对 GPD、GEV 分布与 P-III 型分布推算的设计水平加以对比。

1 极值分布模型

单变量极值分布模型有 2 种常用方法: 一是分组区块最大值模型 (block maximum group of models, BM)。首先是对所得到的数据进行分块, 常

用年最大值方法 (annual maximal series, AMS) 采用年最大值样本作为建立模型的观测值。BM 模型的要求数据样本独立同分布 (IID)。而对洪水序列而言, 常常一年内出现多次洪水样本, 所以按年最大值抽样后会出现多个洪峰流量远大于枯水年份出现的最大值的现象。按年最大值抽样显然会造成洪水信息利用不充分; 二为峰量超限阈值模型 (peaks over threshold models, POT), 其选取超特定阈值抽取样本, 该方法可获取较多的洪水极值序列建立模型。POT 模型要求满足超定量发生的时间服从泊松分布, 且彼此相互独立服从 GPD (generalized Pareto distribution) 分布^[8]。

1.1 广义 Pareto 分布

设序列 $\{x_n\}$ 的分布函数为 $F(x)$, 定义 $F_\mu(y)$ 为随机变量 X 超过阈值 μ 的条件分布函数:

$$\begin{aligned} F_\mu(y) &= \\ P(X - \mu \leq x | X > \mu) &= \\ \frac{F(\mu + y) - F(\mu)}{1 - F(\mu)} &= \frac{F(x) - F(\mu)}{1 - F(\mu)} \\ \Rightarrow F(x) &= F_\mu(y)(1 - F(\mu)) + F(\mu) \quad (1) \end{aligned}$$

研究表明^[18-20], 当 μ 阈值足够大时, 条件超量分布函数 $F_\mu(y)$ 收敛于广义 Pareto 分布, 累积分布函数为:

$$\begin{aligned} F_\mu(y) \approx G(x, \xi, \sigma, \mu) &= \\ \begin{cases} 1 - (1 - \xi \frac{x - \mu}{\sigma})^{1/\xi}, & \xi \neq 0 \\ 1 - e^{-(x - \mu)/\sigma}, & \xi = 0 \end{cases} \quad (2) \end{aligned}$$

式中, ξ 、 σ 、 μ 分别为形态参数、尺度参数和位置参数。当 $\xi = 0$ 时, GPD 对应于指数分布, 为 Pareto I 型分布; 当 $\xi < 0$ 时为 Pareto II 型分布, $x \in [\mu, \infty)$; 当 $\xi > 0$, 为 Pareto III 型分布 (短尾型)。有关研究证明了超定量 $(X - \mu)$ 数服从泊松分布^[17]。

GPD 模型 T 年一遇的分位数 x_T 为:

$$x_T = \begin{cases} \mu + \frac{\sigma}{\xi} [1 - (\lambda T)^{-\xi}], & \xi \neq 0 \\ \mu + \sigma \ln(\lambda T), & \xi = 0 \end{cases} \quad (3)$$

1.2 超定量洪水序列阈值确定

阈值 μ 的合理确定是 GPD 模型参数 ξ 和 σ 正确估计的前提。 μ 取值过高, 超限数据量少, 使估计出来的参数方差很大; 相反 μ 取值过低, 则难以保证分布的收敛性, 估计偏差较大。阈值选取基本原则: 对于每次洪水过程, 选取洪峰流量最大值; 不同场次洪峰取样时, 各场次洪峰发生时间间隔要求大于流域的汇流时间, 以保证不同洪水之间相互独立。采用以下 4 种相结合的方法确定 μ :

1) 绘制样本的经验平均超过函数图

令 $X_{(1)} > X_{(2)} > \dots > X_{(n)}$, 样本的经验平均超过函数定义为^[21]:

$$e(\mu) = \frac{\sum_{i=k}^n (X_i - \mu)}{n - k - 1} \quad (4)$$

其中

$$k = \min \{i \mid X_i > \mu\}$$

绘制的平均超过函数图即为点 $(\mu, e(\mu))$ 构成的曲线, 选取较大的 μ 作为阈值, 它使得当 $x \geq \mu$ 时 $e(x)$ 为近似线性函数。当 $x \geq \mu$ 时平均超过函数图曲线向上倾斜, 表明点据服从形状参数 ξ 为负的 GPD 分布, 属于 II 型分布。当 $x \geq \mu$ 时曲线向下倾斜, 表明数据源自于尾部较短的分布, 当 $x \geq \mu$ 时曲线是水平的, 表明该数据服从指数分布。因此, 如果某个阈值 μ 后的 $e(n)$ 趋向于新的线性变化时, 可选取这个值为阈值。

2) Anderson-Darling (AD) 检验

AD 检验属于平方根类经验分布函数统计检验, 采用 AD 检验 GPD 模型的超定量样本的经验分布和理论分布的拟合优度时, AD 检验在小频率的上尾部区域赋予了更多权重。

以 F_1, F_2, K, F_n 表示各样本的累积分布函数, 假设 $H_0 = F_1 = F_2 = K = F_n$ 。样本 x_i 的个数和理论累积分布分别用 n_i 和 F_{bi} 表示, $N = \sum n_i$ 为所有样本的个数, $H_N(x)$ 为 N 个样本的累积分布函数。 k 个样本的 AD 统计量 A_{kN}^2 为:

$$A_{kN}^2 = \sum_{i=1}^k n_i \int_{B_N} \frac{[F(x) - H_N(x)]^2}{H_N(x)[1 - H_N(x)]} dH_N(x) \quad (5)$$

其中 $B_N = \{x \in R: H_N(x) < 1\}$; A_{kN}^2 经过归一化处理得到: $T_{kN} = \frac{[A_{kN}^2 - (k-1)]}{\sigma_N}$ 。 σ_N 为 A_{kN}^2 的标准差。

若 T_{kN} 小于高斯分布在置信度下的临界值 $t_{k-1}(\alpha)$,

则接受假设 H_0 。实际计算中通过插值外推得到 AD 统计量 A_{kN}^2 的 P_{AD} 值, 当 $P_{AD} > \alpha$, 接受假设; 否则, 拒绝。以逐个超定量样本的经验分布和理论累积分布为两样本进行计算, 得到统计量 T_{2N} , 然后计算 P_{AD} 值, P_{AD} 值越大表示经验分布和理论分布的拟合度越优。详细计算原理与步骤见文献^[20]。

3) 泊松分布检验

为了保证所选样本的独立性, 选择一定的阈值区间, 在显著水平 0.05 下, 分别对阈值区间内不同阈值的超限数采用 χ^2 假设检验其是否服从泊松分布:

$$P(x = k) = e^{-\lambda} \lambda^k / k!, k = 0, 1, 2, \dots$$

其中 λ 为平均发生超限的频次。对服从泊松分布样本的阈值再根据其它拟合优度检验指标作进一步的筛选。

4) 初始阈值与洪水序列的确定

以枯水年最大洪峰流量值为初始阈值, 根据流域汇流时间形成阈值区间, 采用不同的汇流时间间隔分别抽取不同间隔时段内不同序列的洪峰阈值, 在显著水平 0.05 下, 对各超定量样本数的做独立性检验。对确定的洪水序列样本估计分布参数值及其拟合优度检验值。

通过以上 4 种相结合的方法对于确定最终采用的超阈值洪水序列样本满足独立性, 阈值 μ 为优选值。

2 实例研究

2.1 研究背景与基本数据

选取广东强降水地区典型中小流域 - 曹江流域作为研究对象。曹江是粤西独流入海鉴江的一级支流, 发源于高州马贵镇山村海拔 1 141 m 的蓝蓬岭, 中上游雨量充沛, 是广东的暴雨高区之一, 流域坡降大汇流时间短, 洪水陡涨陡落, 导致洪灾频发, 对当地造成严重的生命威胁与经济损失。流域多年平均年雨量 2 160 mm, 最大平均年雨量为 3 150 mm。曹江流域出口断面大拜水文站集水面积 394 km²。

利用 1967 ~ 2013 年共 47 年大拜站逐日流量观测序列分析曹江流域洪水的超阈值频率分布特征。流量序列日最大值 778 m³/s, 最小值仅 0.46 m³/s, 日平均流量 19.4 m³/s。对大拜站流量序列的洪水过程统计表明, 通过大拜站的洪水平均传播时间绝大多数为 1 ~ 3 d, 最长历时可达 10 d (2010 年 9 月 21 ~ 30 日)。

2.2 洪水阈值

采用以下步骤确定曹江流域洪峰流量阈值:

1) 在平均超过函数图(图1)内流量序列在 $100 \text{ m}^3/\text{s}$ 左右存在线性变化折点, 随之曲线向上倾斜, 表明点据服从形状参数为负的 GPD 分布, 可考虑选取线性变化折点值为参考阈值。

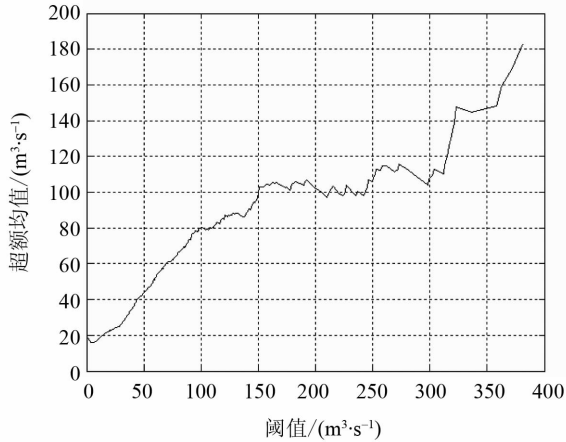


图1 超限样本的经验平均超过函数图
Fig.1 Definite mean excess function chart

2) 以枯水年(2007年)最大洪峰流量值 $74 \text{ m}^3/\text{s}$ 为初始阈值 μ_0 , 形成阈值区间: $\mu_i = \mu_0 + (i - 1) \times 10, i = 1, \dots, 15$, 阈值选择范围为 $74 \sim 214 \text{ m}^3/\text{s}$ 。

3) 按不同洪水序列之间相互独立的要求, 根据流域汇流时间短的特点, 以 3 d 时间为初始间隔时段, 以 1 d 为步长增量逐步增加至 10 d 时间间隔, 按照步骤 1) 分别抽取不同间隔时段内不同序列的洪峰阈值, 在显著水平 0.05 下, 对各超定量样本数的泊松分布加以 χ^2 检验(表 1 中原假设, $H_0 = 0$: 样本服从泊松分布); 为节省版面, 文中仅列出采用参数概率权重矩估计和极大似然估计法 4, 6, 8 和 10 d 时间间隔的洪水序列样本的参数估计值及其拟合优度检验值。

4) 超限抽样系列的 AD 检验的 P 值都远大于 $\alpha = 0.05$ 的临界值, 超定量样本的频率分布与理论 CDF 拟合良好, 表明该超限样本符合 GPD 分布。除去不符合泊松分布的超限样本和 $\lambda < 1$ 的超限样本后发现, 在 8 种采样时间间隔的洪峰序列中阈值为 $104 \text{ m}^3/\text{s}$ 的超限样本的 P_{AD} 值最大, 其中采样时间间隔中间隔 8 d 的 P_{AD} 值和 PPCC (概率图位相关系数) 值最大, RMSE (均方根误差) 最小。因此 $104 \text{ m}^3/\text{s}$ 可作为大拜站洪水序列的日洪峰流量序列的优选阈值(表 1)。

2.3 GPD 参数估计与拟合优度检验结果

极值分布模型的参数估计是统计分析的关键点

之一。不同的参数估计方法推算的分布参数直接影响极值重现水平。因此本文使用具有统计特性良好的概率权重矩法(PWM)和极大似然法(ML)^[6]估计模型的参数, 对各超限样本的 GPD 重现水平的推算结果进一步采用了 PPCC 和 RMSE 作为拟合优度检验指标, 主要结果如下:

1) PPCC 值均大于 0.984, 表明各个超定量样本点据与理论分布值的相关关系达 0.984 以上。同时 PWM 参数估计方法得到的 GPD 模型的均方误差 RMSE 小于 ML 拟合的模型误差, 其中 $\lambda = 2.0 - 3.0$ 估计的指标更可靠, 拟合优度检验结果见表 1。

2) 模型的形态参数为负值, 表明曹江洪水序列服从 Pareto II 型分布。

前述结果显示, GPD 模型确定超定量洪水是根据多个指标综合分析确定的动态过程。

综合上述结果, 确定以 8 d 间隔, 阈值取 $104 \text{ m}^3/\text{s}$, 采用 PWM 参数估计方法拟合的 GPD 模型, $G(x, \xi, \sigma, \mu) = 1 - (1 + 0.137 \times \frac{x - 104}{85.6})^{-1/0.137}$ 为最优超定量洪水的 GPD 模型。

2.4 不同概率分布模型对比分析

对比 GPD 和 GEV 两种极值分布及 P-III 型分布推算参数的分布函数拟合指标值, P-III 型分布参数估计使用常规矩法(OME)和线性矩法(L-M)。三种模型最优模型参数和拟合优度检验指标见表 2, 各分布模型推算的洪水重现水平见表 3 和图 2 ~ 4。

三种概率分布的拟合优度指标对比显示, 超限量抽样在满足超限量数服从泊松分布, 超限彼此相互独立条件下构建的 GPD 模型, 在介于 2.0 ~ 3.0 之间构建的 GPD 模型精度优于 GEV 和 P-III 型, 图 2 显示, 大拜站超定量洪水频率曲线图上的样本点与理论曲线非常吻合, 尤以 PWM 参数估计推算的 GPD 模型最佳。以 8 d 为抽样间隔的拟合优度检验指标为例, 阈值 $74 \sim 124 \text{ m}^3/\text{s}$ 范围内, 采用 PWM 参数估计方法拟合的 GPD 模型 PPCC 值大于等于 0.984, 以阈值 $104 \text{ m}^3/\text{s}$ 对应的 PAD 和 PPCC 值最大, RMSE 也明显小于 P-III 分布和 GEV 分布。

同频率设计值对比结果表明, 设计频率小于 2% (重现期 50 年) 时, GPD 模型设计值介于 GEV 和 P-III 之间, 随着设计频率增大, GPD 设计值超过 GEV 和 P-III 分布的设计值(表 3)。此或反映了自然流域洪水过程的差异性, 其原因需要通过更多的实证加以归纳说明。

表 1 GPD 模型的阈值、参数估计与拟合优度检验结果

Table 1 The selected threshold values, estimated parameters and goodness-fit test of GPD models

抽样间隔	序号	阈值 ($\text{m}^3 \cdot \text{s}^{-1}$)	H_0	POT	λ	PWM					ML				
						ζ	δ	P_{AD}	RMSE	PPCC	ζ	δ	P_{AD}	RMSE	PPCC
10 d	1	74	0	184	3.91	-0.179	76.65	0.794	0.037	0.995	-0.211	75.93	0.799	0.045	0.993
	2	84	0	162	3.45	-0.176	78.66	0.738	0.124	0.995	-0.211	77.86	0.744	0.131	0.993
	3	94	0	134	2.85	-0.050	99.16	0.739	0.043	0.993	-0.123	94.12	0.717	0.055	0.995
	4	104	0	125	2.66	-0.114	89.95	0.799	0.028	0.994	-0.169	87.23	0.797	0.029	0.994
	5	114	0	111	2.36	-0.108	92.25	0.797	0.025	0.994	-0.169	89.22	0.788	0.026	0.993
	6	124	0	101	2.15	-0.136	89.10	0.784	0.008	0.994	-0.195	86.70	0.784	0.004	0.992
	7	134	1	93	1.98	-0.184	83.04	0.739	0.027	0.992	-0.244	81.23	0.752	0.037	0.988
	8	144	1	81	1.72	-0.154	89.57	0.723	0.110	0.992	-0.219	87.39	0.721	0.118	0.989
	9	154	1	68	1.45	-0.024	112.54	0.794	0.039	0.989	-0.127	105.36	0.772	0.024	0.991
	10	164	0	60	1.28	0.064	127.71	0.723	0.050	0.980	-0.090	114.28	0.687	0.075	0.990
	11	174	1	57	1.21	-0.007	115.09	0.770	0.029	0.986	-0.135	105.72	0.743	0.045	0.990
	12	184	1	52	1.11	-0.003	116.11	0.752	0.022	0.984	-0.141	105.85	0.723	0.039	0.989
	13	194	1	47	1.00	0.033	122.10	0.661	0.063	0.980	-0.133	108.85	0.610	0.083	0.988
	14	204	1	45	0.96	-0.057	106.69	0.704	0.064	0.986	-0.194	98.00	0.675	0.071	0.987
	15	214	1	44	0.94	-0.193	85.19	0.803	0.020	0.987	-0.313	80.60	0.798	0.029	0.980
8 d	1	74	1	195	4.15	-0.206	71.16	0.789	0.035	0.994	-0.240	70.53	0.799	0.044	0.991
	2	84	0	170	3.62	-0.202	73.58	0.721	0.131	0.994	-0.238	72.91	0.732	0.141	0.991
	3	94	0	139	2.96	-0.074	94.17	0.758	0.042	0.994	-0.139	90.08	0.741	0.050	0.995
	4	104	0	129	2.74	-0.137	85.62	0.817	0.018	0.995	-0.187	83.49	0.814	0.016	0.994
	5	114	0	113	2.40	-0.114	90.65	0.802	0.024	0.994	-0.173	87.74	0.799	0.025	0.993
	6	124	0	102	2.17	-0.130	89.38	0.777	0.022	0.994	-0.190	86.75	0.774	0.019	0.992
	7	134	1	94	2.00	-0.182	82.68	0.759	0.022	0.992	-0.243	80.85	0.763	0.031	0.988
	8	144	1	82	1.74	-0.157	88.58	0.737	0.101	0.992	-0.221	86.40	0.739	0.108	0.989
	9	154	1	69	1.47	-0.035	110.04	0.801	0.032	0.989	-0.134	103.34	0.782	0.019	0.991
	10	164	0	61	1.30	0.043	123.35	0.721	0.048	0.982	-0.101	111.36	0.690	0.071	0.990
	11	174	1	58	1.23	-0.035	110.08	0.778	0.020	0.987	-0.151	102.19	0.752	0.031	0.990
	12	184	1	52	1.11	-0.003	116.11	0.752	0.022	0.984	-0.141	105.85	0.723	0.039	0.989
	13	194	1	47	1.00	0.033	122.10	0.661	0.063	0.980	-0.133	108.85	0.610	0.083	0.988
	14	204	1	45	0.96	-0.057	106.69	0.704	0.064	0.986	-0.194	98.00	0.675	0.071	0.987
	15	214	1	44	0.94	-0.193	85.19	0.803	0.020	0.987	-0.313	80.60	0.798	0.029	0.980
6 d	1	74	0	203	4.32	-0.205	69.98	0.793	0.028	0.994	-0.238	69.32	0.798	0.036	0.992
	2	84	0	175	3.72	-0.189	74.14	0.766	0.074	0.995	-0.225	73.31	0.774	0.081	0.993
	3	94	0	144	3.06	-0.073	92.60	0.755	0.040	0.993	-0.140	88.35	0.733	0.051	0.995
	4	104	0	133	2.83	-0.130	85.06	0.811	0.020	0.995	-0.183	82.60	0.801	0.021	0.994
	5	114	0	117	2.49	-0.116	88.68	0.786	0.031	0.994	-0.177	85.60	0.775	0.033	0.994
	6	124	0	106	2.26	-0.144	85.76	0.753	0.029	0.994	-0.203	83.27	0.760	0.026	0.992
	7	134	0	98	2.09	-0.210	77.32	0.740	0.032	0.991	-0.273	75.55	0.747	0.043	0.986
	8	144	1	84	1.79	-0.168	85.95	0.742	0.092	0.992	-0.232	83.88	0.750	0.100	0.989
	9	154	1	70	1.49	-0.038	108.68	0.801	0.025	0.989	-0.137	102.04	0.788	0.012	0.992
	10	164	0	62	1.32	0.032	120.64	0.734	0.051	0.983	-0.108	109.26	0.704	0.072	0.991
	11	174	0	59	1.26	-0.050	106.85	0.785	0.024	0.988	-0.161	99.66	0.773	0.033	0.990
	12	184	0	53	1.13	-0.029	111.13	0.777	0.023	0.986	-0.157	102.32	0.747	0.036	0.989
	13	194	1	48	1.02	-0.008	114.88	0.724	0.053	0.983	-0.155	104.21	0.681	0.067	0.989
	14	204	1	45	0.96	-0.057	106.69	0.704	0.064	0.986	-0.194	98.00	0.675	0.071	0.987
	15	214	1	44	0.94	-0.193	85.19	0.803	0.020	0.987	-0.313	80.60	0.798	0.029	0.980
4 d	1	74	0	218	4.64	-0.219	66.08	0.768	0.034	0.994	-0.252	65.44	0.779	0.042	0.991
	2	84	0	185	3.94	-0.190	71.99	0.758	0.072	0.995	-0.226	71.09	0.760	0.078	0.993
	3	94	0	152	3.23	-0.077	89.59	0.735	0.041	0.993	-0.145	85.27	0.706	0.053	0.995
	4	104	0	140	2.98	-0.135	82.14	0.799	0.024	0.995	-0.189	79.60	0.795	0.026	0.994
	5	114	0	122	2.60	-0.114	86.92	0.755	0.038	0.994	-0.178	83.48	0.738	0.043	0.994
	6	124	0	111	2.36	-0.154	82.29	0.730	0.036	0.994	-0.214	79.77	0.728	0.034	0.992
	7	134	0	103	2.19	-0.233	72.35	0.733	0.025	0.991	-0.301	70.48	0.753	0.035	0.984
	8	144	1	88	1.87	-0.202	79.45	0.715	0.088	0.991	-0.268	77.54	0.716	0.098	0.987
	9	154	1	72	1.53	-0.061	103.76	0.813	0.018	0.990	-0.153	98.09	0.802	0.009	0.992
	10	164	1	64	1.36	-0.011	112.44	0.791	0.036	0.986	-0.132	103.62	0.752	0.051	0.991
	11	174	1	60	1.28	-0.067	103.48	0.795	0.026	0.989	-0.173	97.01	0.779	0.033	0.990
	12	184	1	54	1.15	-0.057	106.09	0.788	0.016	0.988	-0.174	98.67	0.772	0.025	0.989
	13	194	1	48	1.02	-0.008	114.88	0.724	0.053	0.983	-0.155	104.21	0.681	0.067	0.989
	14	204	1	45	0.96	-0.057	106.69	0.704	0.064	0.986	-0.194	98.00	0.675	0.071	0.987
	15	214	1	44	0.94	-0.193	85.19	0.803	0.020	0.987	-0.313	80.60	0.798	0.029	0.980

表 2 最优 GPD、GEV、P-III 分布参数与拟合优度检验对比
Table 2 Comparison of parameters and indicators of goodness-fit test among the optimal GPD, GEV and Pearson Type III models

模型	ζ	σ	μ	RMSE	PPCC
PWM-GPD	-0.137	85.62	104	0.018	0.995
PWM-GEV	0.134	100	197	0.061	0.991
L-M-P-III	1.645	0.008	74	0.068	0.991
MLE-GPD	-0.187	83.49	104	0.016	0.994
MLE-GEV	0.163	97	197	0.067	0.992
OME-P-III	1.977	0.009	58	0.062	0.991

表 3 三种概率分布函数洪水重现水平值
Table 3 Return levels of flood discharge of three probability distributions m³/s

T/a	GPD		GEV		P-III	
	PWM	ML	PWM	ML	L-M	MOM
200	961	1110	971	1010	875	852
100	827	934	836	860	785	767
50	705	778	712	724	693	681
20	561	602	563	567	570	564
10	463	487	461	460	474	473
5	374	386	364	361	375	377
2	268	271	235	234	232	236

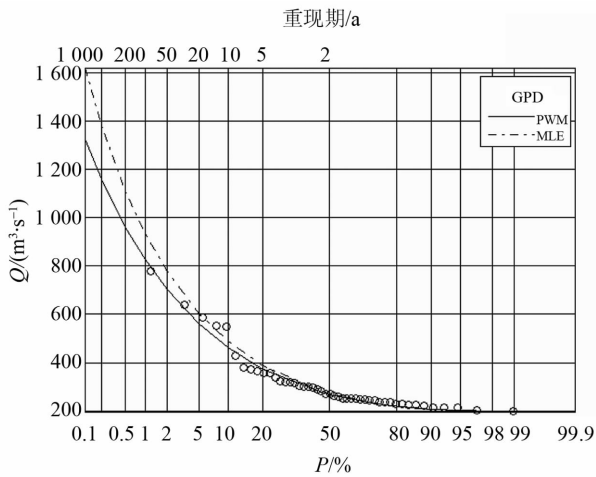


图 2 大拜站超定量洪水频率曲线图
Fig. 2 POT flood frequency curves for Daba station

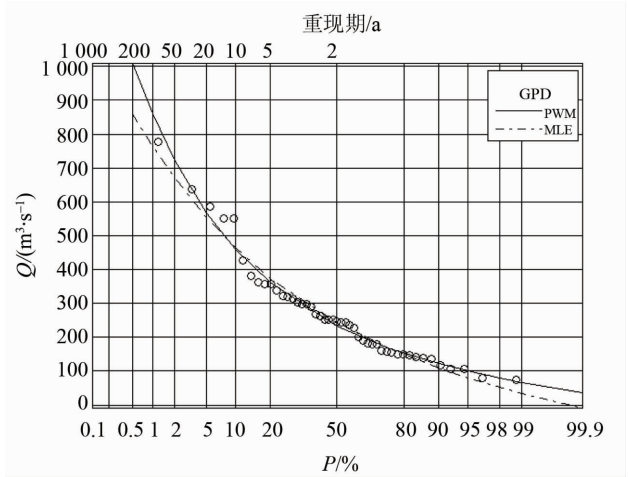


图 3 大拜站年最大洪水 GEV 分布频率曲线
Fig. 3 Annual maximal flood frequency curves based on GEV distribution for Daba station

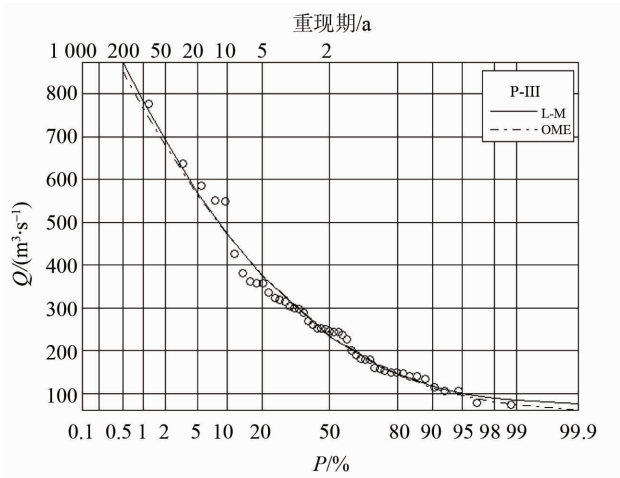


图 4 大拜站年最大洪水 P - III 分布频率曲线

Fig. 4 Annual maximal flood frequency curves based Pearson Type III distribution for Dabai station

3 结 论

1) GPD 模型的形态参数表明洪水频率分布属于 II 型, 与短尾分布 III 型不同, II 型表明密度分布函数峰值右侧分布曲线与横坐标之间的渐进性而无切点, 难于确定洪水的上限值。此是否反映了山地流域不同时段不同区域土壤含水量差异大和洪水的产汇流过程不确定性的自然属性, 有待更多的实例研究。

2) 确定超定量洪水 GPD 模型是动态择优过程, 对多个满足 GPD 模型要求的阈值, 需要通过超定量数的泊松分布检验和超限样本的拟合优度等综合评判后构建相对最优 GPD 模型。

3) GPD 模型推算的设计洪水精度普遍优于 GEV 分布和 P - III 型分布的推算成果。

4) 不同参数估计方法对于极值模型参数推算精度有较大影响。POT 样本由 PWM 法估计参数的极值模型精度高于 ML 法推算的结果。

参考文献:

- [1] JENKINSON A F. The frequency distribution of the annual maximum (or minimum) values of meteorological elements[J]. Q J R Meteorol Soc, 1955, 81: 158 - 171.
- [2] Van MONTFORT M A J, WITTER J V. Testing exponentiality against generalized Pareto distribution[J]. Journal of Hydrology, 1985, 78:305 - 315.
- [3] 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
FANG Bin, GUO Shenglian, CHAI Xiaoling, et al. Flood frequency analysis based on FPOT method[J]. Hydro-power, 2005, 31(2): 9 - 12.
- [4] ROSBJERG D, MADSEN H. Prediction in partial duration series with generalized Pareto distribution exceedances[J]. Water Resources Research, 1992, 28(11): 3001 - 3010.
- [5] Van MONTFORT M A J, WITTER J V. The generalized Pareto distribution applied to rainfall depths[J]. Hydrological Sciences Journal, 1986, 31: 151 - 162.
- [6] CLPAS P, LAIO F. Can continuous streamflow data support flood frequency analysis? An alternative to the partial duration series approach [J]. Water resources Research, 2003, 39(8): 375 - 384.
- [7] 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(自然科学版), 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(Nat Sci Ed), 2010, 38(2): 191 - 196.
- [8] 戴昌军, 梁忠民, 栾承梅, 等. 洪水频率分析中 PDS 模型研究进展[J]. 水科学进展, 2006, 17(1): 136 - 140.
DAI Changjun, LIANG Zhongmin, LUAN Chengmei, et al. Advance in flood frequency analysis for partial duration series [J]. Advances in Water Science, 2006, 17(1): 136 - 140.
- [9] HOSKING J R M & WALLIS J R. Parameter and quantile estimation for the generalized Pareto distribution [J]. Technometrics, 1987, 29: 339 - 349.
- [10] ZHANG J. Likelihood moment estimation for the generalized Pareto distribution [J]. Australian & New Zealand Journal of Statistics, 2007, 49(1): 69 - 77.
- [11] BERMUDEZ P D Z, SAMUEL K. Parameter estimation of the generalized Pareto distribution-Part I [J]. Journal of Statistical Planning & Inference, 2010, 140(6): 1353 - 1373.
- [12] BERMUDEZ P D Z, SAMUEL K. Parameter estimation of the generalized Pareto distribution-Part II [J]. Journal of Statistical Planning & Inference, 2010, 140(6): 1374 - 1388.
- [13] 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(自然科学版), 2010, 38(2): 191 - 196.

- WANG Jianfeng ,SONGSongbai. Application of generalized Pareto distribution in POT flood series frequency analysis [J]. Journal of Northwest A & F University(Nat Sci Ed),2010, 38(2):191-196.
- [14] 周长让,陈元芳,顾圣华,等. 高阶概率权重矩法在广义 Pareto 分布参数估计中的应用[J]. 水力发电学报, 2016,35(6):30-38.
- ZHOU Changrang, CHEN Yuanfang, GUSHenghua, et al, Parameter estimation of generalized Pareto distribution using high-order probability weighted moment method [J]. Journal of Hydroelectric Engineering, 2016, 35(6):30-38.
- [15] 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. Analysis on over threshold flood series [J]. Yangtze River, 1999, 30(8):23-25.
- [16] LEADBETTER M R. On a basis for peaks over threshold modeling[J]. Statistics and Probability Letters,1991,12(4):357-362.
- [17] BHUNYAA P K, BERNDTSSON R, JAIN S K, et al. Flood analysis using negative binomial and Generalized Pareto models in partial duration series (PDS)[J]. Journal of Hydrology, 2013, 497(7):121-132.
- [18] 陈海清,程维虎. 广义 Pareto 分布参数的最小二乘估计[J]. 应用概率统计,2013,29(2):121-135.
- CHEN Haiqing, CHENG Weihu. Estimation of parameters of the generalized pareto distribution by the least squares[J]. Chinese Journal of Applied Probability and Statistics, 2013,29(2):121-135.
- [19] LEADBETTER M R. On a basis for peaks over threshold modeling[J]. Statistics and Probability Letters,1991,12(4):357-362.
- [20] BHUNYAA P K, BERNDTSSON R, JAIN S K, et al. Flood analysis using negative binomial and Generalized Pareto models in partial duration series (PDS)[J]. Journal of Hydrology, 2013, 497(7):121-132.
- [21] LI Y,CAI W, CAMPBELL E P. Statistical modelling of extreme rainfall in southwest Western Australia [J]. Journal of Climate,2005,18(6):852-63.

(责任编辑 秦社彩)