

广义 Pareto 分布在超定量洪水序列频率分析中的应用

王剑峰, 宋松柏

(西北农林科技大学 水利与建筑工程学院, 陕西 杨凌 712100)

[摘要] **【目的】**针对指数分布在超定量洪水序列计算时存在适线灵活性不足的缺点,应用改进线性矩法和广义 Pareto 分布研究超定量洪水序列频率的计算问题。**【方法】**以陕西省延河流域无定河赵石窑站超定量洪水序列为例,采用 Poisson 分布拟合超定量洪水发生次数,用常规线性矩法和改进线性矩法估计广义 Pareto 分布参数,并分别进行超定量序列频率分析。根据复合极值理论,结合 Poisson 分布和广义 Pareto 分布获得年最大超定量洪水分布。采用 Pearson-Ⅲ 分布,对赵石窑站年最大洪水序列频率进行分析。选用离(残)差平方和最小值(OLS)准则对上述 3 种方法计算的洪水频率进行对比分析。**【结果】**改进线性矩法估计的 OLS 值小于常规线性矩法;广义 Pareto 分布的 OLS 值小于 Pearson-Ⅲ 分布。**【结论】**改进线性矩法估计参数的精度高于常规线性矩法;建议在水利工程规划中采用广义 Pareto 分布进行超定量洪水频率分析计算。

[关键词] 超定量洪水;洪水频率分析;广义 Pareto 分布;改进线性矩法;年最大超定量洪水;Poisson 分布

[中图分类号] TV122⁺.5

[文献标识码] A

[文章编号] 1671-9387(2010)02-0191-06

Application of generalized Pareto distribution in POT flood series frequency analysis

WANG Jian-feng, SONG Song-bai

(College of Water Resources and Architecture Engineering, Northwest A&F University, Yangling, Shaanxi 712100, China)

Abstract: **【Objective】** Considering the bad agility of exponential distribution on curve fitting of Peak over threshold (POT) flood frequency analysis, the modified L-moments and generalized Pareto distribution were used to study POT flood series frequency. **【Method】** Choosing the POT flood series from Zhaoshiyao site of Wuding river of Yanhe basin of Shaanxi Province, we fitted the number of POT flood with Poisson distribution, calculated the parameters of generalized distribution by normal L-moments and modified L-moments method, and then executed the frequency analysis of POT flood series separately. Subsequently, according to multivariate compound extreme theory, the distribution of the largest POT series was obtained using Pearson-Ⅲ distribution executes frequency analysis of annual maximum flood series. In order to execute comparative analysis of the three methods, the OLS values of different frequency curve were compared. **【Result】** By comparing the OLS values of frequency curve, it shows that the OLS value of modified L-moments is less than the OLS value of normal L-moments. The OLS value of Generalized Pareto distribution is less than the OLS value of Peartson-Ⅲ distribution. **【Conclusion】** Modified L-moments is better than normal L-moments in parameters estimation, and should be used in generalized Pareto distribution to

* [收稿日期] 2009-06-30

[基金项目] 国家自然科学基金项目(50879070, 50579065); 西北农林科技大学青年学术骨干支持计划和优秀博士论文基金项目

[作者简介] 王剑峰(1984—), 男, 内蒙古察哈尔右翼前旗人, 在读硕士, 主要从事水文水资源研究。

E-mail: wjfsuperman@yahoo. com. cn

[通信作者] 宋松柏(1965—), 男, 陕西永寿人, 教授, 博士生导师, 主要从事水文水资源研究。

analyze POT flood frequency in projecting hydraulic engineering.

Key words: POT flood; flood frequency analyze; generalized Pareto distribution; modified L-moments; annual maximum POT flood; Poisson distribution

洪水频率分析是水利工程规划、设计和运行的依据。我国大多数水文站点的观测资料系列比较短缺,一般仅有几十年,在百年一遇甚至千年一遇洪水的防洪设计时误差较大。而且长期以来,水文学家通常把洪水频率分析归结为极值统计,即一年的观测资料中只选用最大一次洪水作为样本选样,摒弃年内发生的其他洪水,尽管其中有些也是大洪水。因此,从有限的水文观测资料中获得尽可能多的水文信息、增加样本系列长度,是提高设计洪水计算精度的重要途径^[1]。

为了解决水文信息短缺而影响设计洪水精度的问题,水文学者提出了很多方法,如考虑历史洪水资料的水文频率分析方法、洪水频率区域综合分析方法、无资料地区设计洪水的方法以及超定量洪水序列频率方法(Peak over threshold, POT)等。考虑历史洪水资料的水文频率分析方法,是通过在实测资料中加入历史洪水,扩充样本序列,从而提高设计洪水的计算精度。影响该方法设计洪水精度的关键问题是,历史洪水量级的估计精度和历史洪水重现期的确定^[2-4]。洪水频率区域综合分析方法,通过充分利用与设计站点同一水文分区的邻近站点的水文信息,扩大了水文信息量,并能够减少单站估计的任意性和误差,从而提高了设计洪水精度^[5-7]。无资料地区设计洪水的方法是针对水文资料不完整或资料序列长度不足的地区的一种方法,其应用的关键是相似流域的选择^[8-9]。POT 是对超过某一设定门限值的所有洪峰进行分析的方法,与年最大值方法(Annual maximal series, AMS)相比,该法利用了更多的洪水信息,因为其基于超定量年发生次数分布和超定量洪水系列分布模型,所以具有更多的物理相关性^[10],从而提高了设计洪水的计算精度。我国学者虽然于 20 世纪 60 年代提出了超定量抽样理论^[1],但是并未引起我国水文界的广泛注意和深入研究。经典的 POT 模型认为,超定量样本服从指数分布^[11],但在应用中发现,由于指数分布仅有 1 个参数,导致适线分析时灵活性不足。为了克服这一不足,国外许多学者研究和探讨了 POT 系列的频率分布线型,其中 Zelenhasic^[12] 建议采用 Gamma 分布;Miquel^[13] 建议采用 Weibull 分布;Van Montfort 等^[14]、Claps 等^[15] 建议采用广义 Pareto 分布。目

前,我国对超定量洪水频率计算的研究较少,尚未形成完整的理论体系。

为此,本研究针对指数分布在超定量分析中适线灵活性不足的问题,以陕西省延河流域无定河赵石窑站 1954~2003 年洪水超定量序列为例,研究广义 Pareto 分布在超定量洪水频率中的应用,分别采用常规线性矩法和 Hosking^[16] 提出的改进线性矩法估计广义的 Pareto 分布参数,进行超定量序列洪水频率分析;应用 Pearson-Ⅲ 型分布对该站年最大洪水序列进行频率分析;并按照频率曲线适线准则,分别计算出 3 种方法设计洪水频率曲线的离(残)差平方和最小值(Orthogonal least square, OLS),以期为无定河防洪抗灾提供依据。

1 超定量洪水概率模型

设 x 为洪水随机序列, q_0 为选定的门限值。根据超定量洪水定义,从洪水随机序列中筛选超定量洪水样本,样本应满足独立、同分布且保证筛选的洪峰(洪量)大于设定的门限值的要求。为表述方便,本研究用 y 表示超定量洪水。

1.1 超定量洪水发生次数的概率模型

由于洪水固有的随机特性,可以认为超定量洪水的发生也是随机的,其在单位时间内的发生次数 i 也是随机变量。假定超定量洪水发生次数 i 服从 Poisson 分布,则年内发生 i 次超定量洪水的概率 $P_r(x)$ ^[1] 为:

$$P_r(x) = \frac{r^i}{i!} e^{-r}, (x=0, 1, 2, \dots, i). \quad (1)$$

式中: x 为超定量洪水年发生次数的随机变量, r 为超定量洪水年的平均发生次数。

随机变量 x 的数学期望为:

$$E(x) = \sum_{i=0}^{\infty} i P_r\{x=i\} = r. \quad (2)$$

1.2 超定量洪水序列分布模型

假定超定量洪水 y 服从广义 Pareto 分布,则其概率密度函数(pdf)和分布函数(cdf)^[17] 分别为:

$$f_e(y) = \frac{1}{\alpha} \exp\left(1 - \frac{k(y-x_0)}{\alpha}\right)^{1/k-1}, k \neq 0; \quad (3)$$

$$F(y) = 1 - \left(1 - \frac{k(y-x_0)}{\alpha}\right)^{1/k}. \quad (4)$$

式中: $f_e(y)$ 为广义 Pareto 分布的概率密度函数;

$F(y)$ 为广义 Pareto 分布的分布函数,即超过某一超定量值的洪水发生概率,数学表达式为 $F(y) = P(Y > y | y \geq q_0)$; α 为分布尺度参数; k 为分布形状参数,与偏态系数 c_s 有关; x_0 为分布位置参数,即分布曲线的起点值。超定量频率曲线纵坐标的计算公式为:

$$y = x_0 + \frac{\alpha}{k} (1 - (1 - F(y))^k). \quad (5)$$

线性矩法最大的特点是其对洪水系列中的极大值和极小值不如常规矩法敏感,因而根据其求得的洪水频率曲线参数的估计值较常规矩法稳健得多^[18]。为了能够在参数计算时充分利用样本中的信息,本研究采用 Hosking^[16] 提出的改进线性矩法估计广义 Pareto 分布参数,其计算公式为:

$$\hat{k} = \frac{n(\lambda_1 - x_{1:n}) - 2(n-1)\lambda_2}{(n-1)\lambda_2 - (\lambda_1 - x_{1:n})}. \quad (6)$$

式中: $\hat{k}$ 为广义 Pareto 分布的形状参数估计值; n 为洪水超定量系列长度; λ_1 为一阶线性矩; λ_2 为二阶线性矩; $x_{1:n}$ 为超定量样本序列最小值。

根据线性矩定义,通过计算得到 Pareto 分布的总体参数估计 $\hat{\alpha}$, $\hat{k}$ 和 $\hat{x}_0$ 与线性矩的关系为:

$$\hat{\alpha} = (1 + \hat{k})(2 + \hat{k})\lambda_2, \quad (7)$$

$$\hat{x}_0 = x_{1:n} - \frac{\hat{\alpha}}{n + \hat{k}}. \quad (8)$$

式中: $\hat{\alpha}$ 为广义 Pareto 分布的尺度参数估计值; $\hat{x}_0$ 为广义 Pareto 分布的位置参数估计值; λ_1 , λ_2 分别为一阶线性矩和二阶线性矩,可分别由样本线性矩估计值 l_1 和 l_2 估计得到; n 为洪水超定量系列长度。

对样本线性矩估计值 l_1 , l_2 有:

$$l_1 = M_0^*, \quad (9)$$

$$l_2 = 2M_1^* - M_0^*. \quad (10)$$

式中: M_0^* , M_1^* 为样本概率权重矩,其计算公式为:

$$M_0^* = \frac{1}{n} \sum_{i=1}^n x_i, \quad (11)$$

$$M_1^* = \frac{1}{n} \sum_{i=1}^n \frac{i-1}{n-1} x_i. \quad (12)$$

式中: x_i 为超定量洪水序列样本值。

1.3 年最大超定量洪水分布模型

依据复合极值理论,设 $Y_{\max}$ 为年最大超定量洪水,则可由 Poisson 分布和广义 Pareto 分布推导年最大超定量洪水,计算公式^[19]为:

$$F_{\max}(y_{\max}) = \sum F(y)^i \cdot e^{-r} r^i / i! =$$

$$\exp(-r(1 - F(y))), \quad (13)$$

式中: $F_{\max}$ 为年最大超定量分布函数; $y_{\max}$ 为年最大超定量洪水频率曲线纵坐标; $F(y)$ 为广义 Pareto 分布函数。

由式(13)可得年最大超定量洪水频率曲线纵坐标为:

$$y_{\max} = x_0 + \frac{\alpha r^{-k}}{k} (1 - (-\ln(1 - P_{\max}))^k) + \frac{\alpha}{k} (1 - r^{-k}). \quad (14)$$

式中: $P_{\max}$ 为年最大超定量洪水频率;其他符号意义同前。

1.4 超定量洪水重现期

记 $y(p_e)$ 是频率为 p_e 的超定量洪水,根据 Poisson 过程的稀疏定理^[20],事件 $\{Y \geq y(p_e)\}$ 构成强度为 $r \cdot p_e$ 的泊松过程。

在随机点过程理论中,称出现第 1 点与出现第 2 点的时间差为时间间隔(记作 T)。按 Poisson 过程理论,其点间间距服从以其强度为参数的指数分布。综上所述可知,事件过程 $\{Y \geq y(p_e)\}$ 的点间间距服从以 $r \cdot p_e$ 为参数的指数分布。另外,从指数分布的特征可知,超定量洪水重现期 $E(T)$ ^[1,19] 为:

$$E(T) = 1 / (r \cdot p_e). \quad (15)$$

2 实例计算

洪水频率分析样本应满足独立、同分布条件。依据王善序^[1]的观点,所选择的洪峰资料只要满足在前后 2 次洪峰取样时,其时差大于流域汇流时间,而且 2 次洪水之间,前次洪水已明显退落到其起涨流量之下,或者接近于平时的流量,就可认为洪峰满足独立、同分布条件。据此,本研究从《黄河流域水文资料》中摘录了陕西省延河流域无定河赵石窑站 1954~2003 年的洪水资料,选取满足独立、同分布条件的 150 个洪峰值用于计算。

超定量序列方法的关键是门限值的合理选定,如果门限值过小,则连续洪水的多次洪峰都可能入选,因而无法保障样本系列的独立性;如果门限值过大,则减少了入选的洪峰次数,不利于有效地统计分析。故门限值的选取应使超定量发生次数符合某种分布,依据方彬等^[21]的观点,通常将超定量年平均发生次数限制在 2~3 次。本研究通过编制试算程序,不断改变门限值,使超定量年平均发生次数满足条件,最终确定门限值为 120,筛选得到洪水超定量序列 104 个,年均超定量洪水发生次数约为 2 次,年

最大超定量洪水序列 50 个。按照改进线性矩法对赵石窑站超定量洪水序列参数进行估计,广义 Pareto 分布的位置参数为 120.9、尺度参数为 116.5、形状参数为 -0.156。

利用广义 Pareto 分布超定量洪水模型,拟合可得无定河赵石窑站超定量洪水频率曲线(图 1)。由 Possion 分布和广义 Pareto 分布,依据复合极值理论,推导得到年最大超定量分布模型,拟合可得无定

河赵石窑站年最大超定量洪水频率曲线(图 2)。为检验广义 Pareto 分布拟合超定量洪水序列的效果,利用传统的 Pearson-Ⅲ 分布模型对年最大洪水系列进行拟合,得到赵石窑站年最大洪水系列频率曲线(图 3)。从图 1~3 可以看出,采用广义 Pareto 分布对超定量洪水序列的拟合效果良好,尤其在大洪水部分拟合更优,其结果优于我国广泛使用的 Pearson-Ⅲ 分布模型。

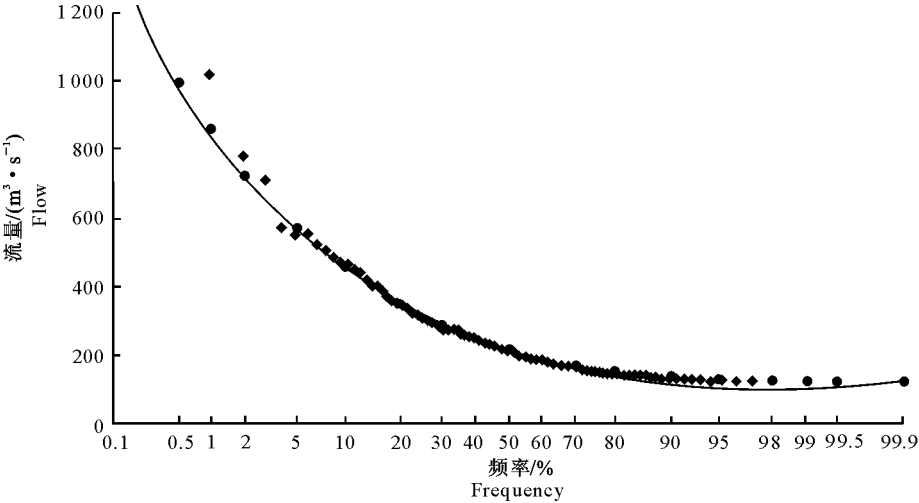


图 1 基于广义 Pareto 分布拟合的陕西省延河流域无定河赵石窑站超定量洪水频率曲线
Fig.1 POT frequency curves for Zhaoshiyao gaging station of Yanhe basin of Shaanxi Province based on generalized Pareto distribution

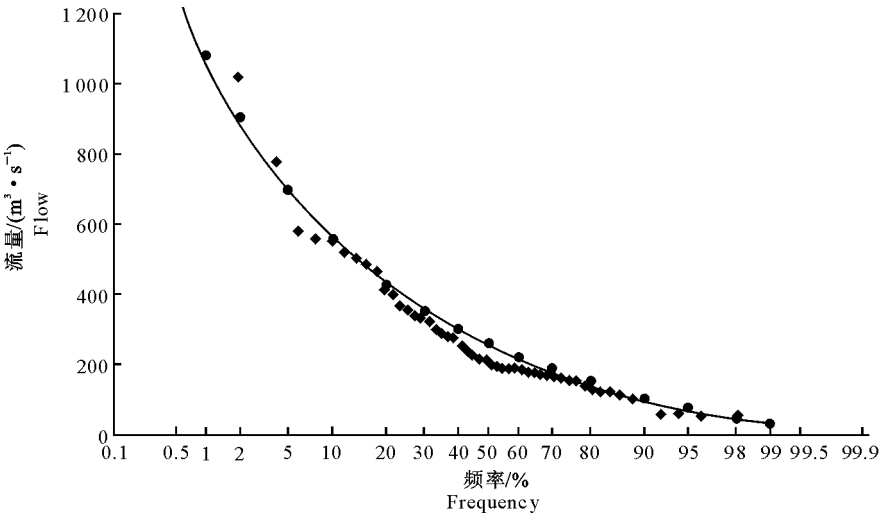


图 2 陕西省延河流域无定河赵石窑站年最大超定量洪水频率曲线

Fig.2 Annual maximal POT frequency curves for Zhaoshiyao gaging station of Yanhe basin of Shaanxi Province

表 1 列出了基于超定量洪水频率分析、年最大超定量洪水频率分析和 Pearson-Ⅲ 分布拟合年最大洪水系列的无定河赵石窑站设计洪峰流量。由表 1 可看出,超定量洪水频率分析时,采用改进线性矩法估计参数,其设计洪水值大于常规线性矩法;在计算

重现期较大(如千年一遇、500 年一遇等)的设计洪水时,其结果大于 Pearson-Ⅲ 分布的计算结果;在计算重现期较小(如百年一遇、50 年一遇等)的设计洪水时,其结果小于 Pearson-Ⅲ 分布的计算结果。年最大超定量洪水分布模型,在计算重现期较大的设

计洪水时,其结果大于超定量洪水频率分析方法和 Pearson-Ⅲ 分布的计算结果;在计算重现期较小的

设计洪水时,其结果与 Pearson-Ⅲ 分布的计算结果接近。

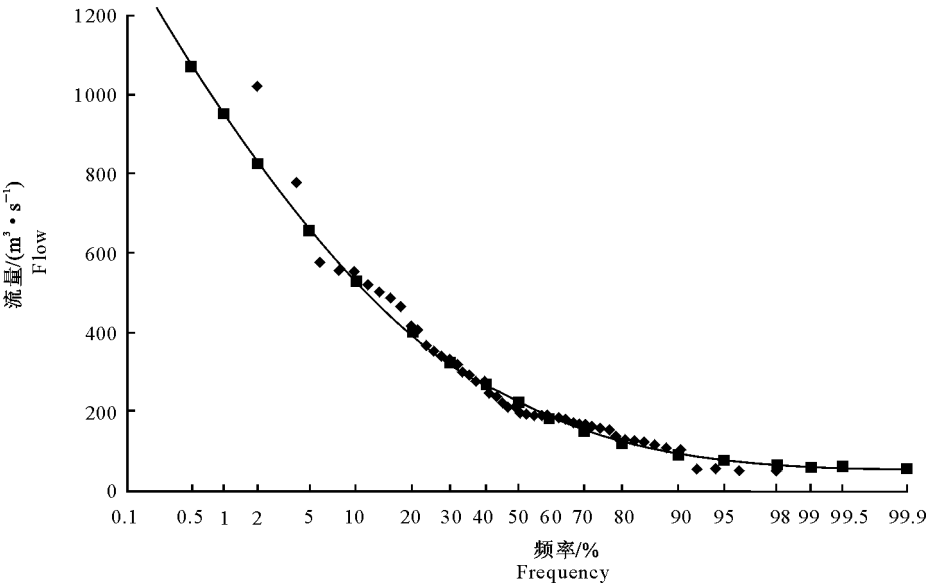


图 3 基于 Pearson-Ⅲ 分布拟合的陕西省延河流域无定河赵石窑站年最大洪水频率曲线

Fig. 3 Fitting annual maximum flood frequency curves of Zhaoshiyao gaging station of Yanhe basin of Shaanxi Province based on Pearson-Ⅲ distribution

表 1 陕西延河流域无定河赵石窑站不同方法设计洪峰流量的比较

Table 1 The comparison of designed flood peak of Zhaoshiyao gaging station of Yanhe basin of Shaanxi Province based on the different methods m³/s

重现期/年 Return period	改进线性矩法 Modified L-moments		常规线性矩法 Normal L-moments		Pearson-Ⅲ 年最大洪水 Annual maximum flood
	超定量洪水 POT flood	年最大超定量洪水 Annual maximum POT flood	超定量洪水 POT flood	年最大超定量洪水 Annual maximum POT flood	
2 000	1 818.5	2 097.5	1 510.2	1 686.8	1 471.0
1 000	1 568.0	1 818.3	1 343.5	1 510.1	1 349.7
500	1 343.1	1 567.6	1 186.0	1 343.2	1 228.1
200	1 080.8	1 275.0	991.4	1 136.7	1 066.6
100	906.0	1 079.5	853.7	990.4	943.7
50	749.0	903.5	723.7	851.7	820.2
10	443.7	556.1	449.3	554.2	528.7

按照我国水利水电工程设计洪水计算规范,计算 3 种方法设计洪水频率曲线适线准则 OLS,则采用改进线性矩法估计参数的超定量系列和年最大超定量系列 OLS 分别为22 768和 62 846,采用常规线性矩法估计参数的超定量系列和年最大超定量系列 OLS 分别为 32 511 和 86 959,采用 Pearson-Ⅲ分布的 OLS 为 63 479。可知,改进线性矩法的 OLS 小于常规线性矩法的 OLS,广义 Pareto 分布的 OLS 小于 Pearson-Ⅲ分布的 OLS。表明,采用改进线性矩法计算参数和广义 Pareto 分布拟合超定量洪水,其拟合效果优于常规线性矩法和 Pearson-Ⅲ 分布法。

3 结 论

- (1)改进线性矩法能够利用更多的样本信息,用其估计广义 Pareto 的分布参数,其精度优于常规线性矩法。
- (2)超定量洪水序列频率分析方法采用超定量序列,增加了样本长度,缓解了洪水设计过程中资料序列短缺的问题;与传统的 Pearson-Ⅲ 分布相比,其频率曲线与经验频率点的拟合效果更好,尤其在重现期较大时。
- (3)与指数函数相比,广义 Pareto 分布有 3 个参数,保证了其在频率适线时具有较好的适线弹性;

与 Pearson-Ⅲ分布相比,其函数形式为显式,可以通过反解获得设计洪水计算公式,计算过程较为简便。

(4)比较不同方法的频率曲线及设计洪水值可知,用广义 Pareto 分布拟合超定量洪水系列时,重现期较大的设计洪水的计算结果大于 Pearson-Ⅲ分布;重现期较小的设计洪水的计算结果小于 Pearson-Ⅲ分布。笔者建议,在洪水设计计算时,应综合考虑这 2 种方法的特点,以使设计更加合理。

[参考文献]

- [1] 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 S X. The flood frequency analysis based on POT flood series [J]. Yangtze River, 1999, 30(8): 23-25. (in Chinese)
- [2] Hosking J R M, Wallis J R. The value of historical data in flood frequency analysis [J]. Water Resources Research, 1986, 22(11): 1606-1612.
- [3] 谢悦波, 刘晓风, 王 平, 等. 加入古洪水资料后设计洪水成果合理性分析 [J]. 河海大学学报, 2000, 28(4): 8-12.
Xie Y B, Liu X F, Wang P, et al. Analysis of rationality of design flood with paleofloods [J]. Journal of Hohai University, 2000, 28(4): 8-12. (in Chinese)
- [4] 黄振平, 王春霞, 马军建. 历史洪水重现期的误差对设计洪水的影响 [J]. 河海大学学报, 2002, 30(1): 79-82.
Huang Z P, Wang C X, Ma J J. Effect of estimation error of historical flood recurrence interval on design flood [J]. Journal of Hohai University, 2002, 30(1): 79-82. (in Chinese)
- [5] Hosking J R M, Wallis J R. Some statistics useful in regional frequency analysis [J]. Water Resources Research, 1993, 29: 271-281.
- [6] Hosking J R M, Wallis J R. Regional frequency analysis: an approach based on L-moments [M]. UK: Cambridge University Press, 1997.
- [7] 熊立华, 郭生练. L-矩在区域洪水频率分析中的应用 [J]. 水力发电, 2003, 29(3): 6-8.
Xiong L H, Guo S L. Application of L-moments in the regional flood frequency analysis [J]. Water Power, 2003, 29(3): 6-8. (in Chinese)
- [8] 陈守煜. 工程水文水资源系统模糊集分析理论与实践 [M]. 大连: 大连理工大学出版社, 1998.
Chen S Y. Fuzzy set analysis theory and application of engineering hydrology and water resources systems [M]. Dalian: Dalian University of Technology Press, 1998. (in Chinese)
- [9] Burn D H. Cluster analysis as applied to regional flood frequency [J]. Journal of Water Resources Planning & Management,

1989, 115: 567-582.

- [10] 郭生练. 设计洪水进展研究与评价 [M]. 北京: 中国水利水电出版社, 2005.
Guo S L. The research progress and evaluation of design flood [M]. Beijing: China Water Power Press, 2005. (in Chinese)
- [11] 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 C J, Liang Z M, Luan C M, et al. Advance in flood frequency analysis for partial duration series [J]. Advances in Water Science, 2006, 17(1): 136-140. (in Chinese)
- [12] Zelenhasic E. Theoretical probability distributions for flood peaks [R]. Fort Collins: Colorado State University, 1970.
- [13] Miquel J. Guide pratique destination des probabilités de crues [M]. Paris: Eyrolles Press, 1984.
- [14] Van Montfort M A J, Witter J V. The generalized Pareto distribution applied to rainfall depths [J]. Hydrological Sciences Journal, 1986, 31: 151-162.
- [15] Claps P, Laio E. Can continuous stream flow data support flood frequency analysis? An alternative to the partial duration series approach [J]. Water Resources Research, 2003, 39(8): 1216.
- [16] Hosking J R M. The wakeby distribution [M]. New York: IBM Research Division, 1986.
- [17] Rao A R, Khaled H H. Flood frequency analysis [M]. Florida: CRC Press LLC, 2000: 271-278.
- [18] 陈元芳, 王庆荣, 沙志贵, 等. 线性矩法在长江中下游区域水文频率计算中的应用 [J]. 河海大学学报: 自然科学版, 2003, 31(2): 207-211.
Chen Y F, Wang Q R, Sha Z G, et al. Application of L-moment based regional flood frequency analysis method to middle and lower Yangtze River [J]. Journal of Hohai University: Natural Sciences, 2003, 31(2): 207-211. (in Chinese)
- [19] 董爱红. 广义 Pareto 分布及其在水文频率分析中的应用 [D]. 南京: 河海大学, 2005.
Dong A H. Study on generalized Pareto distribution and its application on frequency analysis of hydrology [D]. Nanjing: Hohai University, 2005. (in Chinese)
- [20] 邓永录, 梁之舜. 随机点及其应用 [M]. 北京: 科学出版社, 1992.
Deng Y L, Liang Z S. Stochastic point processes and application [M]. Beijing: Science Press, 1992. (in Chinese)
- [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.
Fang B, Guo S L, Chai X L, et al. Flood frequency analysis based on FPOT method [J]. Water Power, 2005, 31(2): 9-12. (in Chinese)