

洪水超定量模型的广义 Pareto 分布及其应用

郝金梅¹, 王冰², 李占强¹, 王志芳¹

(1.河北省石家庄水文水资源勘测局,河北 石家庄 050051;2.天津市中水科技咨询有限责任公司,天津 300170)

摘要:为了扩大洪水信息,提高洪水模拟精度,研究超定量洪水频率分析模型。介绍了洪水超定量模型的基本理论,假设超定量洪水年发生次数服从 Poisson 分布,超定量洪水系列服从广义 Pareto(GP)分布,给出年最大超定量洪水分布和超定量洪水重现期的计算方法,提出通过模型拟合优度检验来综合确定超定量系列阈值的方法。将超定量模型应用在海河流域小觉站洪峰频率分析中,结果表明:通过模型拟合优度检验确定超定量系列阈值的方法有效且可靠,洪水超定量系列年平均发生次数服从 Poisson 分布,GP 分布洪峰设计值略大于 P- 分布洪峰设计值,应用在水利工程设计及风险分析中是偏安全的。

关键词:洪水频率分析;超定量洪水模型;广义 Pareto 分布;Poisson 分布

中图分类号:P333

文献标识码:A

文章编号:1000-0852(2014)01-0078-05

1 前言

长期以来,洪水频率分析被归结为极值统计,常采用的方法年最大值(AM)模型,即每年选取一个最大的洪水特征值(如洪峰、最大 3d 洪量等)作为样本,对一定年限的样本系列进行频率计算分析,该方法被广泛应用于工程设计和实验中。只对洪水年最大值系列进行分析,利用的洪水信息有限,实际上,丰水年的第二大甚至第三大洪水特征值可能要大于枯水年的最大洪水特征值,因此,将丰水的其他较大的洪水系列作为研究对象,是扩大洪水信息的一个重要途径。超定量(POT 或 PDS)模型是对超过某一设定阈值的所有洪水进行分析的方法,与年最大值模型相比,这种方法增加了样本容量,最大限度的利用了有极值分析意义的洪水信息,从而达到提高设计洪水精度的目的。国内外相关研究较多,并取得了一系列成果^[1-13]。

本文将超定量模型应用于海河流域子牙河水系小觉站洪峰的频率分析中,并对超定量模型设计洪峰值和 P- 分布计算值进行比较,为海河流域防洪减灾提供新的科学依据。

2 超定量洪水概率模型

经典的超定量模型包含 3 个基本假定:①超定量

发生次数服从 Poisson 分布;②超定量样本独立;③超定量样本服从指数分布(EXP)。在本研究中假定超定量洪水年发生次数服从 Poisson 分布,采用广义 Pareto 分布对超定量洪水系列进行频率分析。

2.1 超定量洪水取样原则

为保证超定量模型模拟的效果,频率分析计算所用样本必须满足独立同分布的条件。超定量洪水选样无统一标准,国内外洪水选样标准有以下几种方法:

(1)美国水资源协会(USWRC,1976)提出采用 5d(至少 5d)加上流域面积(m^2)的自然对数来分离连续洪水事件,再附加另一个条件,即两个连续洪峰之间的流量 Q_{\max} 必须小于二者中较小值的 75%,则第二个洪峰 Q_2 的拒绝条件为:

$$\theta < 5 + \ln A \text{ 或 } Q_{\min} > (3/4) * \min(Q_1, Q_2) \quad (1)$$

式中: θ 为两峰值发生时间间隔(h); A 为流域面积(m^2); Q_i 为第 i 个日最大流量。

(2)Cunnane(1979)提出连续洪峰中,第二个连续洪峰 Q_2 被拒绝的条件为:

$$\theta > 3T_p \text{ 或 } Q_{\min} > (2/3) * Q_1 \quad (2)$$

式中: T_p 为洪水过程中连续 5 个洪峰发生时间间隔的平均值(h)。

(3)Miquel(1984)建议对各洪峰滞时 1 和滞时 2 的自相关系数进行检验。在给定的显著性水平下,若拒

收稿日期:2013-01-29

作者简介:郝金梅(1981-),女,河北沧州人,工程师,硕士研究生,主要从事水文测报及水文资料整编方面的研究。

E-mail:haojinmei0827@163.com

绝不相关的假设,则需要采用较大的 θ 值来重新选择洪峰。

(4)王善序(1999)认为对于连续洪峰过程,只选取其中最大一次洪峰的峰值或洪量;前后两次洪峰取样时,两个洪峰发生时间间隔必须大于流域的汇流时间;并且,在两场洪水过程之间,前场洪水应已明显退落到其起涨流量之下,或接近汛期平时的流量。认为不同年份洪水之间相互独立。

流域汇流时间采用式(3)进行计算:

$$\tau=0.0078*\left(\frac{L}{S^{0.5}}\right)^{0.77} \quad (3)$$

式中: τ 是流域汇流时间(min); L 是流域汇流长度(英尺); S 为流域平均坡度。

2.2 超定量洪水频率曲线

根据洪水抽样原则,挑选出某一符合独立同分布的洪水随机序列 Y , μ 为已选定的阈值,则从序列 Y 中筛选出超过阈值的序列 X 即为超定量序列,该序列应服从独立且同分布的要求。假设超定量洪水随机变量 X 服从广义 Pareto 分布,则随机变量 X 的分布函数为:

$$G(x;\mu,\sigma,\xi)=1-\left(1+\xi\frac{x-\mu}{\sigma}\right)^{-1/\xi} \quad x\geq\mu, 1+\frac{\xi(x-\mu)}{\sigma}>0 \quad (4)$$

式中: $\mu\in R$ 为位置参数,即超定量系列阈值; $\sigma>0$ 为尺度参数; $\xi\in R$ 为形状参数。

其反函数,也即超定量频率曲线的纵坐标公式为:

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

洪水随机变量 X 的密度函数为

$$g(x;\mu,\sigma,\xi)=\frac{1}{\sigma}\left(1+\xi\frac{x-\mu}{\sigma}\right)^{-1/\xi-1}, \quad x\geq\mu \text{ 且 } 1+\xi(x-\mu)/\sigma>0 \quad (6)$$

2.3 超定量系列发生次数概率模型

若一年中超过阈值洪水发生的次数为 k , 则年内超定量洪水发生次数 k 是一个随机变量,假设 k 服从 Poisson 分布。则年内超出量 $X-\mu$ 发生次数的概率为:

$$P(k)=\frac{\lambda^k}{k!}e^{-\lambda}, \quad k=0,1,2,\dots \quad (7)$$

式中: k 为超定量发生次数的随机变量; λ 为 Poisson 分布参数,即年平均发生的超阈值个数。

2.4 年最大超定量洪水分布模型

根据复合极值分布模型可以得到 Poisson 分布与广义 Pareto 分布的复合极值分布函数,也即为年最大超定量洪水分布模型。由复合极值理论可知,离散型分

布为 Poisson 分布,连续分布为广义 Pareto 分布的复合极值分布属于广义极值分布(GEV),则其具体形式如下:

$$F_0(x)=e^{-\lambda[1-G(x)]} = e^{-\lambda\left[1+\xi\frac{x-\mu}{\sigma}\right]^{-1/\xi}} \quad (8)$$

式中: F_0 为年最大超定量分布函数; x 为年最大超定量洪水; $G(x)$ 为广义 Pareto 分布函数(见式(4))。由式(8)可以得到年最大超定量洪水频率曲线纵坐标公式:

$$y=\mu+\frac{\sigma}{\xi}\left[\left(-\frac{1}{\lambda}\ln F_0\right)^{-\xi}-1\right] \quad (9)$$

式中: y 为随机变量(年最大超定量洪水); μ 为 GP 分布的位置参数; σ 为 GP 分布的尺度参数; ξ 为 GP 分布的位置参数形状; λ 为 Poisson 分布的参数; F_0 为概率值(年最大超定量洪水的频率值)。

2.5 超定量洪水重现期

当超过阈值 μ 的洪水年发生次数服从参数为 λ 的 Poisson 分布时,超过阈值 μ 的洪水年平均发生次数即为 λ 次, T 年中发生超定量洪水的个数为 λT 个,则重现期为 T 年的洪水在超定量系列中发生概率为 $1/(\lambda T)$ 。根据指数分布特性可以得到超定量洪水重现期的计算公式:

$$T=\frac{1}{\lambda*p_e} \quad (10)$$

式中: T 为重现期(a); λ 为 Poisson 分布的参数,即超定量洪水年平均发生次数; p_e 为超定量洪水发生概率。

3 参数的确定

3.1 阈值的确定

广义 Pareto 分布可以得到应用的一个首要前提就是阈值 μ 的选取,只有确定了阈值 μ ,才能进一步估计尺度参数 σ 和形状参数 ξ 。如果 μ 值选取过大,会导致超出数据量太少,估计量的方差较大;如果 μ 值选取过小,则不能保证超定量分布的收敛性,估计量偏差较大,导致超出量分布与广义 Pareto 分布相差较大。选择阈值的方法有很多种,本文采用常用的平均剩余寿命图和判断阈值改变引起估计量的变化方法综合考虑进行阈值的选取。

3.2 参数估计

本文采用极大似然估计进行参数估计。设 $Y=(y_1, y_2, \dots, y_n)$ 是独立同分布的观测数据,总体分布函数 G 是尺度参数和形状参数分别为 σ 和 ξ 的广义 Pareto 分布,密度函数为

$$g(y)=\frac{1}{\sigma}\left(1+\xi\frac{y}{\sigma}\right)^{-1/\xi-1} \quad (11)$$

似然函数为

$$L(y_1, y_2, \dots, y_n) = \prod_{i=1}^n g(y_i; \sigma, \xi) \quad (12)$$

对数似然函数为

$$\ln L = -n \ln \sigma - \left(\frac{1}{\xi} + 1 \right) \sum_{i=1}^n \ln \left(1 + \xi \frac{y_i}{\sigma} \right) \quad (13)$$

式(13)中变量 y_i 必须满足定义域的限制, 即 $y_i \in D(\sigma, \xi)$ (如式(14)所示), 采用数值方法即可求解 σ 和 ξ 的值。

$$y_i \in D(\sigma, \xi) = \begin{cases} [0, \infty), & \xi \geq 0 \\ [0, -\sigma/\xi], & \xi < 0 \end{cases} \quad (14)$$

3.3 拟合检验与优选

由于平均剩余寿命图法和判断阈值改变引起估计量变化的方法是通过目测判断来选取阈值, 精度较低, 只能明确阈值的范围, 不能给出具体数值。因此, 需要对超定量模型进行假设检验以判断选取的阈值和估计的参数能否准确的描述变量的真实特性。对于可能存在的满足独立性、年平均超阈值个数服从 Poisson 分布的多个超定量系列需要进行拟合优度对比以选择相对最优超定量模型。本文采用 Kolmogorov-Smirnov (K-S) 检验方法检验超定量序列是否服从 GP 分布。

可以采用离差平方和最小准则 (OLS) 法和 AIC 信息准则法对超定量模型进行拟合优度评价。OLS 值和 AIC 值越小, 说明超定量模型拟合度越高。

4 实例研究

由于频率分析计算所用的洪水样本必须满足独立同分布的条件, 本文综合美国水资源协会、Cunnane、Miquel 和王善序提出的洪水选样标准, 对小觉站 1960~1991 年共 32a 的实测洪水资料进行洪水独立性选择, 共挑选出 197 场洪水。

首先根据平均剩余寿命图和判断阈值改变引起估计量变化的方法确定洪峰超定量模型阈值的范围, 然后采用极大似然估计法估计不同阈值时 GP 分布的尺度参数 σ 和形状参数 ξ 。对不同阈值的超定量洪峰系列进行假设检验以及拟合度优选, 计算出能够通过假设检验的超定量洪水模型的 OLS 和 AIC 值, 根据拟合优度最高的原则 (OLS 和 AIC 值最小), 最终确定小觉站洪峰系列的阈值 μ 为 $88\text{m}^3/\text{s}$, 洪峰超阈值个数 119 个, 年最大超定量洪峰个数为 28 个, 相应的超定量洪峰年平均发生次数 (泊松参数值) 为 3.72 次。

采用 Kolmogorov-Smirnov (K-S) 检验方法对超定量洪峰年发生次数是否服从 Poisson 分布做假设检验, 结果表明: 当阈值 $\mu=88\text{m}^3/\text{s}$ 时, 统计量 $\chi^2=3.13$, 自由度为 9, 取显著性水平 $\alpha=0.05$, $\chi^2_{1-\alpha}(v)=\chi^2_{0.95}(9)=16.92 > 3.13$, 通过检验。采用极大似然估计法得到洪峰超阈值模型的尺度参数 $\sigma=171.494$, 形状参数 $\xi=0.288$ 。

小觉站洪峰系列超定量模型的参数确定后, 可以得到 GP 分布拟合的洪峰超定量系列频率曲线如图 1 所示, 同时图 2 给出了年最大洪峰超定量频率曲线。由图 1 和图 2 可以看出, 超定量模型对小觉站洪峰频率曲线的拟合效果较好。

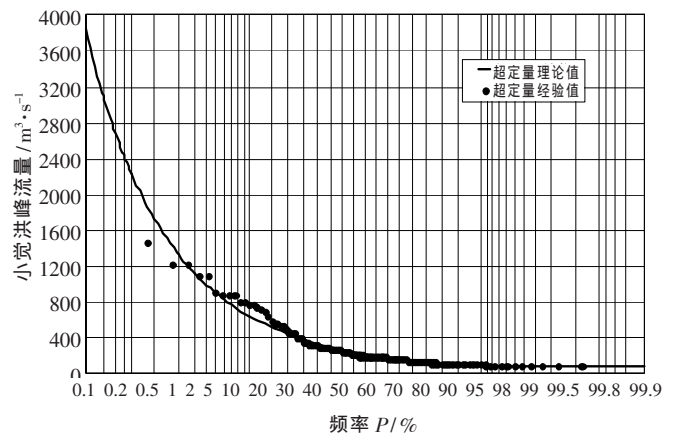


图 1 小觉超定量洪峰系列频率曲线

Fig.1 The super quantitative peak series frequency curve of the Xiaojue station

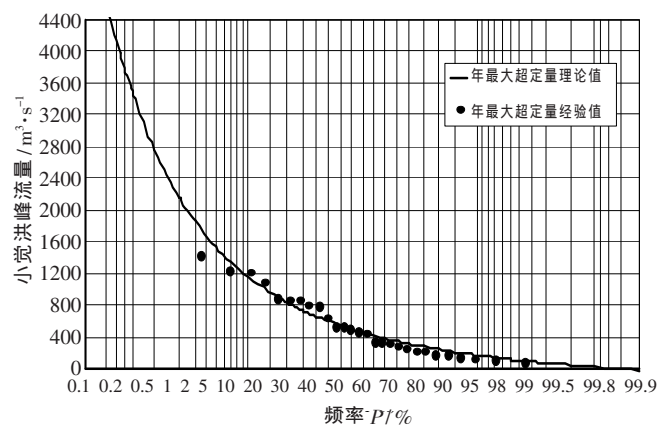


图 2 小觉年最大洪峰系列超定量频率曲线

Fig.2 The super quantitative frequency curve of the annual maximum flood peak series of the Xiaojue station

将不同重现期下小觉洪峰 GP 分布设计值和年最大洪峰超定量设计值列入表 1, 同时采用 P- 分布拟合 1960~1991 年小觉洪峰年最大值系列, 并将

表 1 不同分布模型的洪峰设计值

Table 1 The design peak values of the different distributed models

| 模型 | 不同重现期下设计洪峰值 / $m^3 \cdot s^{-1}$ | | | | | | | |
|--------|----------------------------------|-------|-------|-------|-------|-------|-------|-----|
| | 1 000a | 500a | 200a | 100a | 50a | 20a | 10a | 5a |
| GP 分布 | 5 848 | 4 697 | 3 488 | 2 763 | 2 167 | 1 537 | 1 154 | 831 |
| 年最大超定量 | 5 848 | 4 697 | 3 488 | 2 763 | 2 167 | 1 537 | 1 154 | 831 |
| P- 分布 | 4 838 | 4 244 | 3 258 | 2 714 | 2 187 | 1 523 | 1 061 | 653 |

相应重现期下 P- 分布的设计值结果列入表 1 做对比分析。

表 1 的计算结果表明:GP 分布洪峰设计值与年最大超定量设计值无差别;除 50 年一遇 GP 洪峰设计值略小于 P- 分布设计值外,其他重现期时 GP 分布洪峰设计值均大于 P- 分布设计值,且随着重现期逐渐增大,GP 分布设计值与 P- 分布设计值的差距先减小后增大;重现期为 50~100 年时两分布设计值差距较小;重现期为 5 年和 1000 年时,两分布设计值差距较大;GP 分布洪峰设计值与 P- 分布设计值稍有差别,但均在允许误差范围内,此外,基于 GP 分布洪峰设计值略大于 P- 分布设计值,对于工程防洪设计及水库风险计算分析方面是偏安全的。因此,GP 分布适用于海河流域子牙河水系,采用该分布对子牙河水系洪水洪水频率计算分析是合理的。

5 结论

(1)采用超定量序列对洪水进行频率分析,增加了样本容量,充分的利用了洪水信息,弥补了洪水系列较短,资料短缺的不足。

(2)超定量模型阈值的选取是建立该模型十分关键的一步,本文首先确定阈值的范围,然后通过模型拟合优度检验的方法来综合确定超定量系列的阈值,阈值选取结果比较可靠。

(3)采用 GP 分布拟合洪水频率曲线,具有三个参数,在频率曲线拟合时比指数函数更为灵活,能够更好的反映洪水的特征。

(4)GP 分布和 P- 分布洪峰设计值略有差别,但均在误差范围内;GP 分布洪峰设计值略大于 P- 分布洪峰设计值,应用在水利工程设计及风险分析中是偏安全的;基于 GP 分布的超定量模型适用于海河流域小觉站洪水系列的频率计算分析。

参考文献:

- [1] Coles S. An Introduction to Statistical Modeling of Extreme Values [M]. New York: Springer Verlag, 2001:36-78.
- [2] Van Montfort M A J, Witter J V. Testing exponentiality against generalized Pareto distribution [J]. Journal of Hydrology, 1985,78: 305-315.
- [3] Rosbjerg D, Madsen H. Prediction in partial duration series with generalized Pareto distribution exceedances [J]. Water Resources Research, 1992,28(11):3001-3010.
- [4] Wang Q J. The pot model described by the generalized Pareto distribution with Poisson arrival rate [J]. Journal of Hydrology, 1991,129:263-280.
- [5] Madsen H, Pearson C P, Rosbjerg D. Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events: 2 regional modeling [J]. Water Resources Research, 1997,33(4):747-757.
- [6] Hosking J R M, Wallis J R. Regional Frequency Analysis: An Approach Based on L-Moment [M]. London: Cambridge University Press, 1997:14-43.
- [7] 丰吉闯, 李建平, 高丽君. 商业银行操作风险度量模型选择分析[J]. 国际金融研究, 2011,8:88-96. (FENG Jichuang, LI Jianping, GAO Lijun. Select analysis of commercial bank's operational risk measurement model [J]. International Financial Research, 2011,8:88-96. (in Chinese))
- [8] 陈红英. POT 模型在巨灾保险中的应用 [D]. 上海: 上海交通大学, 2010. (CHENG Hongying. POT Model and Its Application in Catastrophe Insurance [D]. Shanghai: Shanghai Jiaotong University, 2010. (in Chinese))
- [9] 冯剑丰, 王洪礼, 李胜明. 渤海湾赤潮藻类细胞密度的超阈值分布 [J]. 天津大学学报, 2007,40 (10):1194-1198. (FENG Jianfeng, WANG Hongli, LI Shengming. Super threshold distribution of harmful algal cell densities in the Bohai Bay [J]. Journal of Tianjin University, 2007,40(10):1194-1198. (in Chinese))
- [10] 张文清, 谭维翰. 激光短脉冲作用下原子吸收的超阈值离子化[J]. 物理学报, 1991,40(11):1755-1764. (ZHANG Wenqing, TAN Weihuan. Short pulse laser atomic absorption of ultra-threshold ionization [J]. Acta Physica Sinica, 1991,40(11):1755-1764. (in Chinese))
- [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 of partial duration flood series [J]. Yangtze River, 1999,30(8):23-25. (in Chinese))
- [12] 周川, 陈元芳, 魏琳, 等. 适线法在洪水超定量系列频率分析中的应用研究 [J]. 水电能源科学, 2011,29 (3):44-47. (ZHOU Chuan, CHENG Yuanfang, WEI Lin, et al.. Super application of right method in quantitative analysis of flood series [J]. Water Resources and Power, 2011,29(3):44-47. (in Chinese))
- [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, SONG Songbai. Generalized Pareto distribution applications in partial duration sequence of flood frequency analysis [J]. Journal of Northwest Sci-Tech University (Natural Science Edition), 2010,38(2):191-196. (in Chinese))

Generalized Pareto Distribution of Partial Duration Flood Model and Its Application

HAO Jinmei¹, WANG Bing², LI Zhanqiang¹, WANG Zhifang¹

(1. *Shijiazhuang Hydrology and Water Resources Bureau of Hebei Province, Shijiazhuang 050051, China;*

2. *Tianjin Zhongshui Science and Technology Consulting Co., LTD, Tianjin 300170, China)*

Abstract: Partial duration flood (PDF) frequency analysis model was studied for expanding the flood information and enhancing the simulation accuracy of the flood. The basic theory of the PDF was introduced and the hypothesis was given that the annual occurrence number of the partial duration flood series is estimated by the Poisson distribution, the partial duration flood series are estimated by the generalized Pareto (GP) distribution, then the distribution of the annual maximum partial duration flood series and the calculation formula of the PDF return period could be obtained. A model goodness of fit test has been proposed for determining the threshold of the partial duration flood. The PDF model was applied in the flood peak frequency calculation for the Xiaoju Station in the Haihe River Basin. The results show that the goodness of fit test is an efficient and reliable method for determining threshold, and the annual occurrence number of the partial duration flood series follow the Poisson distribution. The design value of the flood peak based on the GP distribution is slightly larger than that on the P- distribution, which will be more secure when used in the design and the risk analysis of the hydraulic engineering.

Key words: flood frequency analysis; partial duration flood model (PDF); generalized Pareto distribution (GP); Poisson distribution



(上接第 25 页)

Study on Spatial Pattern of Land-using Types and Hydrologic Droughts for Typical Karst Basin of Guizhou Province

HE Zhonghua^{1,2}, CHEN Xiaoxiang¹, LIANG Hong², HUANG Fasu³, ZHAO Fang⁴

(1. *School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, China;* 2. *School of Geographic and Environmental Science, Guizhou Normal University, Guiyang 550001, China;* 3. *Hydrology and Water Resources Bureau of Guizhou Province, Guiyang 550001, China;* 4. *Vocational and Technical Schools of Baiyun District of Guiyang City, Guiyang 550014, China)*

Abstract: Drought is a natural phenomenon of the global widespread occurrence, while Karst drought can not simply be attributed to the climate anomalies, reduced-rainfall. In this paper, 40 typical watersheds in Guizhou Province were selected as the sample areas, the soil water body index (SWBI) was built by using object-oriented classification techniques, to extract the remote sensing information of land-using types. To calculate the landscape density index (PD), largest patch index (LPI), shape index (SI), and landscape fractal dimension index (FDI) of the land-using types by using the calculation software of landscape indices (Fragstats4.0), and analyze the spatial distribution pattern of land-using types. To explore the variation of watershed storage-space under the different land-using patterns, the sizes of watershed water-storing capacity, and study the affect of land-using types and its spatial change on hydrologic droughts. The study has shown that: (1) The distributions of woodland, dry and shrubbery are relatively concentrated, the larger the spatial variation, and that of water and urban is relatively dispersed, the smaller the spatial variation; (2) The affects of the plaque density and largest patch distributions of land-using types on hydrologic droughts are larger, and the smaller the patch edge and patch shape distributions of land-using types; (3) The larger the patch density index, largest patch index, patch shape index and plaque fractal dimension index of land-using types, the smaller the affects of land-using types on hydrologic droughts.

Key words: typical Karst basin; land use; spatial pattern; hydrological drought; Guizhou Province