

文章编号:1004-4574(2012)05-0049-06

多变量洪水频率的计算

汪丽娜¹, 李艳², 陈晓宏^{3,4}, 顾继光⁵

(1. 华南师范大学地理科学学院, 广东广州 510631; 2. 广东商学院资源与环境学院, 广东广州 510320; 3. 中山大学水资源与环境研究中心, 广东广州 510275; 4. 华南地区水循环与水安全广东省教育厅重点实验室(中山大学), 广东广州 510275; 5. 暨南大学水生生物研究中心, 广东广州 510632)

摘要:将 Gumbel-logistic 模型和 Gumbel-mixed 模型进行了对比,并解析了不同组合情形下的洪水频率结果。研究表明:当输入的二元变量的相关性不大于 $2/3$ 时,两模型计算出的重现期值差异不大,当输入的二元变量的相关性大于 $2/3$ 时,两模型计算出的重现期值存在较大的差异,这是由于此时 Gumbel-mixed 模型已经失效的缘故,因此 Gumbel-logistic 模型的应用范围较 Gumbel-mixed 模型更宽。就 Gumbel-logistic 模型而言,对不同洪水特征量的组合情形得到的重现期差异较大,所得重现期差异幅度达 800 年。可见,二元变量的洪水频率计算,模型的选择和洪水特征量组合情形的选取是决定洪水频率结果的主要影响因素。

关键词:二元变量;洪水;频率;模型

中图分类号:X43

文献标志码:A

Calculation of multivariable flood frequency

WANG Lina¹, LI Yan², CHEN Xiaohong^{3,4}, GU Jiguang⁵

(1. School of Geography Science; South China Normal University; Guangzhou 510631; China; 2. Department of Natural Resources and Environment, Guangdong University of Business, Guangzhou 510320, China; 3. Center for Water Resources and Environment; Sun Yat-sen University; Guangzhou 510275; China, Guangzhou 510275, China; 4. Key Laboratory of Water Cycle and Water Security in Southern China of Guangdong High Education Institute, Sun Yat-sun University, Guangzhou 510275, China; 5. Institute of Hydrobiology, Jinan University, Guangzhou 510632, China)

Abstract: This paper compared the Gumbel-logistic model and Gumbel-mixed model and analyzed the flood frequencies under different combination situation. The results show that when the correlation coefficient between inputted binary variables is not greater than $2/3$, the difference of recurrence periods calculated from the two models is not obvious; but when the correlation coefficient is greater than $2/3$, there is a large difference between the two recurrence periods because of failure of the Gumbel-mixed model at that time. Therefore, the applicable range of the Gumbel-logistic model is wider than that of the Gumbel-mixed model. As for the Gumbel-logistic model, the difference of recurrence periods from different combination of flood characteristic quantities is more obvious and the recurrence period obtained in this paper reaches to 800 years. Therefore, in calculation of binary variables' flood frequencies, the selection of models and the choice of combination of flood characteristic quantities are the main influencing factors on flood frequency results.

Key words: binary variable; flood; frequency; model

收稿日期:2011-12-05; 修回日期:2012-02-10

基金项目:国家重点基础研究发展计划(973)项目(2010CB428405);国家自然科学基金重点项目(50839005,41001019);广东省自然科学基金博士启动项目(S2011040005992);广东高校优秀青年创新人才培养计划(育苗工程)项目(LYM11049);教育部高等学校博士学科点专项科研基金新教师类资助课题(20114407120006)

作者简介:汪丽娜(1981-),女,博士,讲师,主要从事水资源与环境变异研究。E-mail:linawang2004@163.com

洪水作为一种自然现象,一旦发生大洪水,若未有及时的防范措施其灾害损失难以估量^[1]。洪水频率的计算不仅为防灾减灾工程的建立提供基础数据,也是解析洪水特性的基础工作。关于洪水频率计算的研究一直是国内外相关学者关注的焦点问题,总体上来说,主要包括以下 3 方面的内容:一是参数、非参数方法的研究,Faucher 和 Rasmussen 等^[2]将 Altman 和 Leger^[3]估计法进行改进,提出了一种新的适用于洪水及干旱非参数频率分析的内插估计法,比经典的最小二乘交叉证实法(LSCV)更能减小估计偏差。二是单元变量向二元、多元变量方向的发展,Yue^[4]采用二维正态分布描述洪峰和洪量的联合分布;我国学者熊立华和郭生练^[5]采用 Gumbel 逻辑模型的结构形式,构造边缘分布为 P-III 型分布的两变量联合分布,以年最大洪峰和年最大洪量为输入变量,应用所建立的联合分布研究长江流域某站点的洪水频率;三是 Sklar^[6]将 Copula 理论运用到统计学中,由于 Copula 函数根据不同的构造方法,将变量之间的相关性的要求从有具体的限制中解放出来,并克服了要求所有的边缘分布都服从统一分布的要求,因此 Copula 模型被国内外学者引入到洪水频率计算中,以 Favre^[7]等人于 2004 年在《Water resources research》上发表的名为《Multivariate hydrological frequency analysis using Copulas》文章为代表性,文中探讨了 Copula 函数在多维极值分布建模中的应用。

本文以武江流域犁市(二)站 1955-2007 年的 53 场洪水为例,采用二元变量的 Gumbel-mixed 模型和 Gumbel-logistic 模型,分析不同输入情况下洪水重现期值的差异。

1 方法比较

Gumbel-logistic 模型是刻画边缘分布服从 Gumbel 分布正相关的两随机变量联合概率分布的理想模型^[8]。Gumbel-logistic 模型具有参数估计方便,结构简单的优势,在计算二元变量的重现期研究中有较多的应用^[8-9]。本文选用与 Gumbel-logistic 模型的表述相似的 Gumbel-mixed 模型,将两个模型进行比较^[9],分析各种组合情形下的洪水重现期值。

表 1 Gumbel-logistic 模型和 Gumbel-mixed 模型比较

Table 1 Comparison between Gumbel-logistic model Gumbel-mixed model

| 模型比较 | Gumbel-logistic 模型 | Gumbel-mixed 模型 |
|---|---|--|
| 联合分布函数 | $F(x,y) = \exp\{-[(-\ln F(x))^m + (-\ln F(y))^m]^{\frac{1}{m}}\}$ | $F(x,y) = F(x)F(y)\exp\{-\theta[\frac{1}{\ln F(x)} + \frac{1}{\ln F(y)}]^{-1}\}$ |
| $F(x)$ 和 $F(y)$ 分别是随机变量 X 和 Y 的边缘分布函数 | $F(x) = \exp[-\exp(-\frac{x-u_x}{\alpha_x})]$ $F(y) = \exp[-\exp(-\frac{y-u_y}{\alpha_y})]$ | $F(x) = \exp[-\exp(-\frac{x-u_x}{\alpha_x})]$ $F(y) = \exp[-\exp(-\frac{y-u_y}{\alpha_y})]$ |
| 重现期 | $T_{x,y} = \frac{1}{1-F(x,y)}$ | $T_{x,y} = \frac{1}{1-F(x,y)}$ |
| 参数 | $m = \frac{1}{\sqrt{1-\rho}} \quad (0 \leq \rho \leq 1)$ $\rho = \frac{E[(X-\mu_x)(Y-\mu_y)]}{\sigma_x\sigma_y}$ | $\theta = 2[1 - \cos(\pi\sqrt{\frac{\rho}{6}})] \quad (0 \leq \rho \leq \frac{2}{3})$ $\rho = \frac{E[(X-\mu_x)(Y-\mu_y)]}{\sigma_x\sigma_y}$ |

2 统计检验

采用 K-S 检验法来检验选取的洪水特征值的边缘分布是否服从耿贝尔分布,计算结果表明在显著性水平 $\alpha = 0.05$ 时,53 场洪水的洪峰水位、洪峰流量、最大 1 d 洪量、最大 2 d 洪量、最大 3 d 洪量、洪水总量和洪水历时的柯尔莫哥洛夫的统计量(D)值如表 2 所示。

表 2 各洪水特征值的 K-S 检验结果

Table 2 K-S check-up results of flood characteristic values

| 项目 | 洪峰水位 | 洪峰流量 | 最大 1 d 洪量 | 最大 2 d 洪量 | 最大 3 d 洪量 | 洪水总量 | 洪水历时 |
|-----|---------|---------|-----------|-----------|-----------|---------|---------|
| D | 0.100 9 | 0.122 4 | 0.101 6 | 0.124 0 | 0.148 0 | 0.111 9 | 0.075 3 |

查柯尔莫哥洛夫的统计量表得,当 $n = 53$ 时 D_n 为 0.186 6,表 2 说明 53 场洪水的洪峰水位、洪峰流量、最大 1 d 洪量、最大 2 d 洪量、最大 3 d 洪量、洪水总量和洪水历时的边缘分布都符合 Gumbel 分布。

3 结果与分析

Gumbel - logistic 模型和 Gumbel - mixed 模型均为二元变量模型,其中 Gumbel - mixed 模型要求输入的二元变量之间的相关性不大于 $2/3$,因此本文选取相关性小于 $2/3$ 的组合情形,表 3 中所列的组合为相关系数小于 $2/3$ 。

表 3 相关系数小于 $2/3$ 的组合情形表

Table 3 Combination situation for correlation coefficients less than $2/3$

| 相关系数 | 洪峰水位 | 洪峰流量 | 最大 1 d 洪量 | 最大 2 d 洪量 | 最大 3 d 洪量 | 洪水总量 |
|------|---------|---------|-----------|-----------|-----------|---------|
| 洪水历时 | 0.065 8 | 0.085 6 | 0.101 8 | 0.135 1 | 0.191 4 | 0.483 2 |

表 3 说明,武江流域犁市(二)站的洪水历时与洪峰水位、洪峰流量、最大 1 d 洪量、最大 2 d 洪量、最大 3 d 洪量、洪水总量之间的相关性较小,均小于 $2/3$ 。将上述变量的组合情形分别输入 Gumbel - logistic 和 Gumbel - mixed 模型中,比较两模型计算出的重现期值,如图 1 所示。

图 1 表明,当两变量的相关性均小于 $2/3$,且所有的变量都服从 Gumbel 分布,即输入的数据符合模型的要求时,即使输入的数据相同,采用的模型不同,得到的重现期值差异亦较小;上述 6 种组合中,武江流域犁市(二)站 1955 - 2007 年的 53 场洪水的最大重现期值小于 11 a。

由表 2 可得洪水的 7 项特征量均服从 Gumbel 分布,上述的 6 种组合情形中,两变量之间的相关性均小于 $2/3$,若两变量的相关性大于 $2/3$,此时输入的变量不符合 Gumbel - mixed 模型条件,然而 Gumbel - logistic 模型的输入条件依旧符合。研究表明,除上述的 6 种组合情形外,武江流域犁市(二)站其他两项洪水特征量组合下的相关系数均大于 $2/3$,选取 3 种组合情形,分别用 Gumbel - mixed 模型和 Gumbel - logistic 模型分析二元变量的重现期值,3 种组合情形如表 4 所示。

表 4 相关系数大于 $2/3$ 的组合情形表

Table 4 Combination situation for correlation coefficients greater than $2/3$

| 项目 | 洪峰流量 + 最大 3 d 洪量 | 最大 1 d 洪量 + 最大 2 d 洪量 | 最大 2 d 洪量 + 最大 3 d 洪量 |
|------|------------------|-----------------------|-----------------------|
| 相关系数 | 0.9551 | 0.9922 | 0.9918 |

以表 4 的 3 种组合情形作为 Gumbel - mixed 模型和 Gumbel - logistic 模型的输入,分别解析 3 种组合情形下的洪水重现期值,如图 2 所示。

图 2 表明,当所有的变量都服从 Gumbel 分布,但是输入的两变量的相关性均并不小于 $2/3$ 时,即输入的数据不完全符合模型的要求时,即使输入的数据相同,采用的模型不同,得到的重现期值差异亦较大。根据武江流域大洪水记录得,近 60 a 内武江流域 1961 年、1968 年、1994 年、2002 年和 2006 年出现了较大的洪水,进一步比较上述 5 a 的重现期值,如表 5 所示。

表 5 不同模型的重现期值比较

Table 5 Comparison of recurrence period between different models

| 项目 | | 重现期/a | | | | |
|-----------------------|----------------------|--------|--------|--------|--------|--------|
| | | 1961 年 | 1968 年 | 1994 年 | 2002 年 | 2006 年 |
| 洪峰流量 + 最大 3 d 洪量 | Gumbel - Mixed 模型 | 6.7 | 10.5 | 24.7 | 17.0 | 633.2 |
| | Gumbel - Logistic 模型 | 7.3 | 10.7 | 26.4 | 19.2 | 716.8 |
| 最大 1 d 洪量 + 最大 2 d 洪量 | Gumbel - Mixed 模型 | 8.0 | 10.4 | 28.6 | 18.6 | 671.4 |
| | Gumbel - Logistic 模型 | 9.4 | 12.0 | 32.2 | 22.4 | 809.7 |
| 最大 2 d 洪量 + 最大 3 d 洪量 | Gumbel - Mixed 模型 | 6.6 | 13.8 | 33.9 | 18.5 | 675.9 |
| | Gumbel - Logistic 模型 | 7.6 | 15.0 | 41.0 | 22.3 | 815.5 |

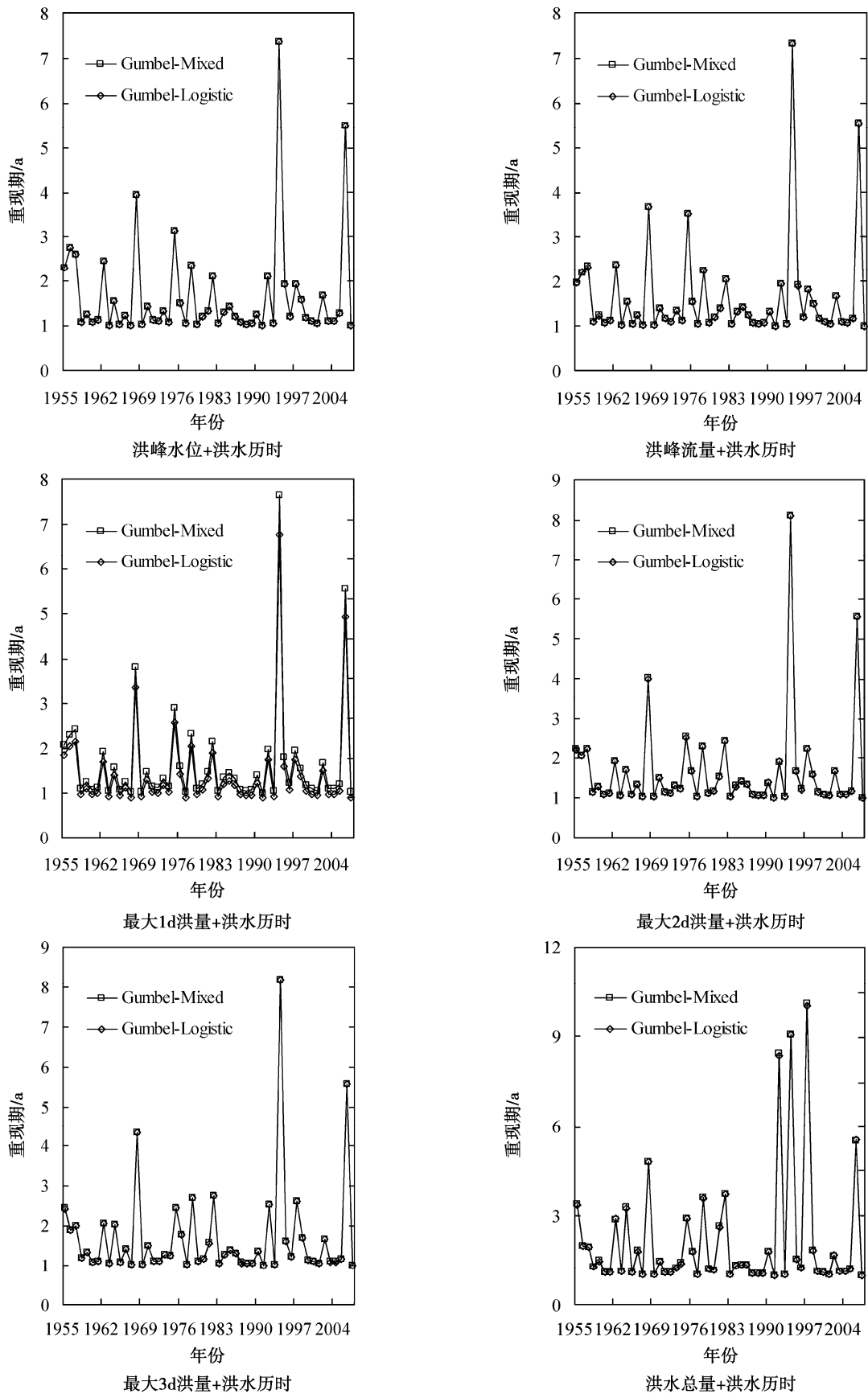


图1 Gumbel-mixed模型和Gumbel-logistic模型结果比较

Fig. 1 Comparison results from Gumbel-logistic model and Gumbel-mixed model

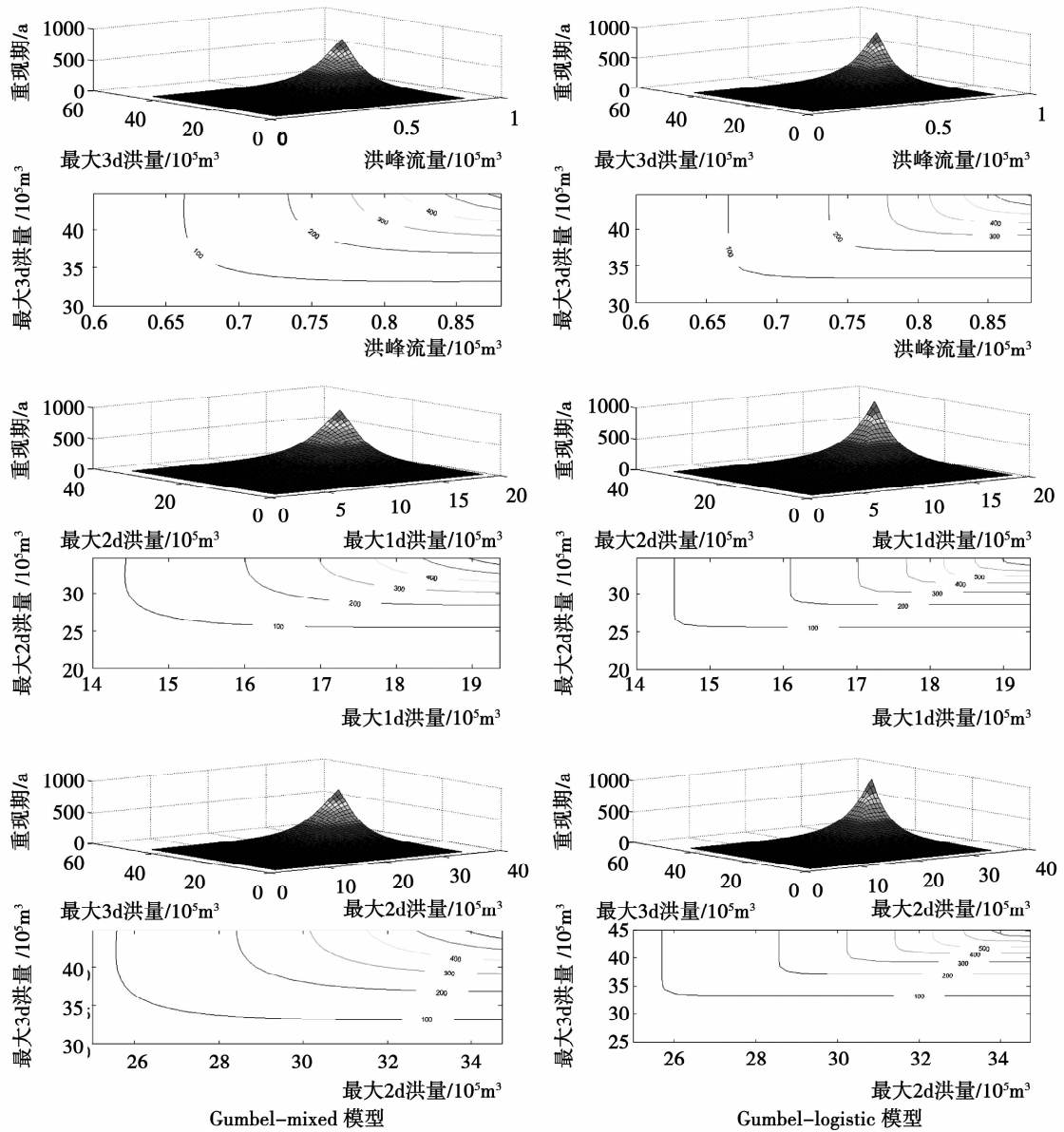


图 2 Gumbel - mixed 模型和 Gumbel - logistic 模型的联合重现期比较

Fig. 2 Comparison of joint recurrence period between Gumbel - logistic model and Gumbel - mixed model

表 5 说明, 3 种组合情形中, Gumbel - logistic 模型得到的重现期值, 相比 Gumbel - mixed 模型得到的重现期值大, 且 Gumbel - logistic 模型得到的最大重现期值与 Gumbel - mixed 模型相应的重现期大 80 a 以上。可见, 当输入的二元变量之间的相关性不在区间 $[0, 2/3]$ 内时, 两个模型得到的重现期值存在较大的差异, Gumbel - mixed 模型所得到的重现期值远小于 Gumbel - logistic 模型, 是由于此时的 Gumbel - mixed 模型失效^[8]。与此同时, 针对 Gumbel - logistic 模型, 当输入的变量不同时, 得到的重现期值亦存在较大的差异, 当输入的二元变量之间的相关性不大于 $2/3$ 时, 得到的最大重现期不大于 11 a; 而输入的二元变量之间相关性大于 $2/3$ 时, 本文所选的 3 种组合情形, 采用 Gumbel - logistic 模型得到的最大重现期为 815.5 a。综上所述, 进行洪水频率计算时, 不仅模型的选择是频率计算的关键问题, 洪水特征量的选取亦是决定洪水频率结果的主要因素。

4 结论

本文采用 Gumbel - logistic 模型和 Gumbel - mixed 模型分别分析不同组合情形下的洪水重现期值。研究表明: 当输入的二元变量的边缘分布服从 Gumbel 分布, 且二元变量之间的相关性不大于 $2/3$ 时,

Gumbel – logistic模型和 Gumbel – mixed 模型所计算出的重现期结果相近;即使输入的二元变量的边缘分布服从 Gumbel 分布,但二元变量之间的相关性小于 2/3 时,Gumbel – logistic 模型和 Gumbel – mixed 模型所计算出的重现期结果差异较大,是由于此时 Gumbel – mixed 模型失效的原因。当采用 Gumbel – logistic 模型计算洪水频率,结果表明:不同洪水特征量的组合情形得到的重现期差异较大,本文所得重现期差异幅度达 800 a。可见,二元变量的洪水频率计算,模型的选择和洪水特征量组合情形的选取是决定洪水频率结果的主要影响因素。

参考文献:

- [1] 杜鹃,何飞,史培军. 湘江流域洪水灾害综合风险评价[J]. 自然灾害学报,2006,15(6):38–44.
DU Juan, HE Fei, SHI Peijun. Integrated flood risk assessment of Xiangjiang River basin in China[J]. Journal of Natural Disasters, 2006, 15(6):38–44. (in Chinese)
- [2] Faucher D, Rasmussen P E, Bobée B. A distribution function based bandwidth selection method kernel quantile estimation [J]. Journal of Hydrology, 2001, 250(1/4): 1–11.
- [3] Altman N, Léger C. Bandwidth selection for kernel distribution function estimation [J]. Journal of Statistical Planning and Inference, 1995, 46(2): 195–214.
- [4] Yue S. Applying the bivariate normal distribution to flood Frequency analysis [J]. Water International, 1999, 24(3): 248–252.
- [5] 熊立华,郭生练. 两变量极值分布在洪水频率分析中的应用研究[J]. 长江科学院院报,2004,21(2):35–37.
XIONG Lihua, GUO Shenglian. Application study of a bivariate extremal distribution in flood frequency analysis[J]. Journal of Yangtze River Scientific Research Institute, 2004, 21(2): 35–37. (in Chinese)
- [6] Sklar A. Fonctions de repartition à n dimensions et leurs marges [J]. Publication de l' Institut de Statistique de l' Université de Paris, 1959, 8: 229–231.
- [7] Favre A C, Adlouni S E, Perrault L, et al. Multivariate hydrological frequency analysis using copulas [J]. Water Resources Research, 2004, 40(11): 1–12.
- [8] 周道成,段忠东. 耿贝尔逻辑模型在极值风速和有效波高联合概率分布中的应用[J]. 海洋工程,2003,21(2):45–51.
ZHOU Daocheng, DUAN Zhongdong. The Gumbel – logistic model for joint peobability distribution of extreme – value wind speeds and effective wave heights[J]. The Ocean Engineering, 2003, 21(2):45–51. (in Chinese)
- [9] Yue S, Wang C Y. A comparison of two bivariate extreme value distributions[J]. Stochastic Environmental Resea, 2004,18(2): 61–66.