

基于 Copula 函数的飞来峡水库坝址 洪水峰量联合分布研究*

王亚雄¹, 刘祖发², 陈俊合²

(1. 广东省水利电力勘测设计研究院, 广东 广州 510635;

2. 中山大学水资源与环境系, 广东 广州 510275)

摘要: 基于 Copula 函数分析了飞来峡水库坝址洪水洪峰和最大 7 d 洪量联合概率分布特征, 获得如下结论: 拟合优度检验指标表明 4 种 Archimedean Copula 函数中 Gumbel-Hougaard Copula 函数拟合两者的联合概率分布效果较好; 洪峰与最大 7 d 洪量的联合重现期小于相应边缘分布的重现期, 而同现重现期则大于相应边缘分布的重现期, 以单变量计算推算的设计值实际上达不到所要求的设计标准; 基于两变量联合分布得到的洪水频率分析计算结果更合理。

关键词: Copula 函数; 洪水频率分析; 联合分布; 飞来峡水库

中图分类号: P333 **文献标志码:** A **文章编号:** 1008-0112(2014)05-0018-05

1 研究背景

洪水是由洪峰、洪量、历时等多特征属性组成的极端水文事件, 且各属性之间存在一定的相互关系。目前, 在我国洪水频率计算分析中, 包括现行的水利水电工程设计洪水计算规范^[1], 基本上都是局限于基于单变量概率分布的统计计算, 而忽略洪水各特征属性之间的相关关系, 一些研究尝试采用多变量联合分布, 都对边缘分布有特定的要求, 存在各自的适用条件和应用局限^[2]。近年来, 随着 Copula 函数的在水文气象都变量联合分布中的深入研究, 为洪水多变量联合分布提供了新的方法。

飞来峡水利枢纽位于北江干流中游, 控制着北江流域面积的 73%, 是调蓄北江洪水的关键性工程, 保护着下游广州市、佛山市、清远市等重要城市和珠江三角洲地区的防洪安全, 其洪水调度一直以来都备受关注 and 重视, 而对洪水准确认识和掌握是洪水调度的关键因素。因此, 笔者采用 Copula 函数对飞来峡水库坝址洪水的洪峰和最大 7 d 洪量进行联合分布研究, 推求洪峰与最大 7 d 洪量的联合分布函数及联合重现期, 并与单变量洪水频率计算进行比较分析, 以期为对飞来峡水库的洪水认识提供新的思路和信息, 这对其防洪调度管理、洪水资源的规划利用等有重要的

意义。

2 研究方法

2.1 Copula 函数理论

Skar's 定理是 Copula 函数理论的核心, 也是其在统计学领域应用的基础。Skar's 定理(Sklar, 1959)如下^[3]: 令 F 为一个 n 维变量的联合累积分布函数, 其中各变量的边缘累积分布函数记为 F_i , 那么存在一个 n 维 Copula 函数 C , 使得

$$F(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n)) \quad (1)$$

可见, Copula 函数将联合概率分布分为变量的相关性结构和随机变量的边缘分布两个独立的部分来分别进行研究, 首先确定边缘分布, 然后选择一个能够恰当地反映变量间的相关结构的 Copula 函数。

Copula 函数有多种类型, 如椭圆 Copula 函数、Archimedean Copula 函数和 Plackett Copula 函数等, 其中 Archimedean Copula 函数是最经常用来构造联合分布函数, 也是目前在水文领域上应用最成熟的, 郭生练^[4]等较为全面的总结介绍了 Copula 函数在多变量水文分析计算中的应用。本文在此仅列出在水文及相关领域文献里经常出现的 4 种二维 Archimedean Copula (见表 1)。

收稿日期: 2014-03-24

作者简介: 王亚雄(1987), 男, 硕士, 助理工程师, 主要从事水文水资源研究。

* 基金项目: 国家自然科学基金项目(51379223)

表 1 水文领域常用的 4 种二维 Archimedean Copula 函数

Copula 函数	表达式 $C(u, v)$	特点
Gumbel - Hougaard (GH)	$C(u, v) = \exp(-((-\ln u)^\theta + (-\ln v)^\theta)^{1/\theta})$ $\theta \in [1, \infty]$	仅能够适用于变量存在正相关的情形(如洪峰和洪量、洪量与洪水历时等相互之间存在正相关性的变量联合分布)
Clayton	$C(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{1/\theta}$ $\theta \in [0, \infty]$	仅适用于描述正相关的随机变量
Ali - Mikhail - Haq (AMH)	$C(u, v) = uv / (1 - \theta(1-u)(1-v))$ $\theta \in [-1, 1)$	既能够描述正相关的, 也能够描述存在着负相关性的随机变量, 但是不适用于非常高的正相关性或负相关性
Frank	$C(u, v) = -\frac{1}{\theta} \ln \left(1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{(e^{-\theta} - 1)} \right)$ $\theta \in \mathbf{R}$	既能够描述正相关的, 也能够描述存在着负相关性的随机变量, 且对相关性的程度没有限制

2.2 Copula 函数参数估计

对于二维 Copula 函数参数估计, 最常用的方法为相关指标法^[5], 它是根据 Copula 函数的参数 θ 与变量 Kendall 秩相关系数 τ 之间的关系进行参数估计, 该方法简单实用, Kendall 秩相关系数相关概念及计算方法详见文献[6]。水文领域常用的 4 种二维 Archimedean Copula 函数的参数 θ 与变量 Kendall 秩相关系数的关系式如下^[7]:

1) Gumbel - Hougaard (GH) Copula

$$\tau = 1 - \frac{1}{\theta} \quad \theta \in [1, \infty] \quad (2)$$

2) Clayton Copula

$$\tau = \frac{\theta}{2 + \theta} \quad \theta \in [0, \infty] \quad (3)$$

3) Ali - Mikhail - Haq (AMH) Copula

$$\tau = \left(\frac{3\theta - 2}{\theta} \right) - \frac{2}{3} \left(1 - \frac{1}{\theta} \right)^2 \ln(1 - \theta) \quad \theta \in [-1, 1) \quad (4)$$

4) Frank Copula

$$\tau = 1 + \frac{4}{\theta} \left(\frac{1}{\theta} \int_0^\theta \frac{t}{\exp(t) - 1} dt - 1 \right) \quad \theta \in \mathbf{R} \quad (5)$$

2.3 Copula 函数拟合优度评价

在估计各种 Copula 函数的参数后, 需要进一步进行拟合优度评价, 以选择最合适的 Copula 函数来描述变量之间的相关结构。常用的 Copula 函数拟合优度评价的方法^[8]有离差平方和准则法 (OLS)、AIC 信息准则法、Genest - Rivest 方法^[9]。

本文采用 AIC 信息准则法和离差平方和准则法进行 Copula 函数的拟合优度评价。

1) AIC 信息准则法

AIC 信息准则, 是 Akaike 于 1973 年从信息论的

角度出发提出的用于模型的定阶和选择的方法^[10], 形式如下:

$$MSE = \frac{1}{n} \sum_{i=1}^n (F_{emp}(x_{i1}, x_{i2}, \dots, x_{im}) - C(u_{i1}, u_{i2}, \dots, u_{im}))^2 \quad (6)$$

$$AIC = n \ln(MSE) + 2k \quad (7)$$

式中 $F_{emp}(x_{i1}, x_{i2}, \dots, x_{im})$ 、 $C(u_{i1}, u_{i2}, \dots, u_{im})$ 分别为经验频率和理论频率值; m 是模型维数, n 是观测样本个数, k 是模型参数个数; AIC 值越小, 表明模型拟合的越好。

2) 离差平方和最小准则(OLS)

$$OLS = \sqrt{\frac{1}{n} \sum_{i=1}^n (F_{emp}(x_{i1}, x_{i2}, \dots, x_{im}) - C(u_{i1}, u_{i2}, \dots, u_{im}))^2} \quad (8)$$

式中符号意义同上。

2.4 概率分布与重现期

根据 Copula 函数, 两变量联合概率分布可表示为:

$$F(x, y) = P(X \leq x, Y \leq y) = C(F_x(x), F_y(y)) = C(u, v) \quad (9)$$

变量 X 、 Y 的边缘分布为: $u = F_x(x)$, $v = F_y(y)$ 。变量 X 、 Y 边缘分布的重现期的计算公式如下:

$$T(x) = \frac{1}{1 - P(X \leq x)} = \frac{1}{1 - F_x(x)} \quad (10)$$

$$T(y) = \frac{1}{1 - P(Y \leq y)} = \frac{1}{1 - F_y(y)} \quad (11)$$

对于多变量来说, 就有联合重现期和同现重现期两种定义:

联合重现期是指多个变量中至少有一个超过某一特定值时, 事件发生的重现期:

$$T_o(x, y) = \frac{1}{P(X > x \cup Y > y)} = \frac{1}{1 - F(x, y)} =$$

$$\frac{1}{1 - C(u, v)} \quad (12)$$

同现重现期是指多个变量同时超过特定值时, 事件的重现期:

$$T_a(x, y) = \frac{1}{P(X > x, Y > y)} = \frac{1}{1 - u - v + C(u, v)} \quad (13)$$

3 飞来峡水库坝址洪水峰量联合分布

本文选取北江干流横石水文站(1953年设站, 1998年因飞来峡水库的修建而停测), 1915年、1931年、1953年至1998年洪水系列, 其中1915年、1913年为查测的历史洪水, 1953-1998年为实测洪水系列, 数据均来自广东省水文局。采用年最大值法取样, 得到年最大洪峰流量和最大7d洪量系列。

3.1 边缘分布

洪峰与最大7d洪量的边缘分布均采用P-III分布函数, 先采用线性矩法估计参数的初值, 在此基础上以离差平方和最小准则(OLS)进行优化适线, 最终确定参数的值。年最大洪峰流量的均值、变差系数、偏态系数分别为9342 m³/s、0.35、0.64; 最大7d洪量的均值、变差系数、偏态系数分别为36.09亿m³、0.42、0.6。由参数即可确定边缘分布(边缘分布函数公式略)。

3.2 Copula函数的参数估计与拟合优度评价

由Kendall秩相关系数τ计算得到的4种二维Archimedean Copula函数的参数θ见表2。采用AIC信息准则法和离差平方和准则法进行Copula函数的拟合优度评价的结果见表3, 可见对于横石站洪峰与最大7d洪量, G-H Copula拟合的最好。

表2 Copula函数的参数值

峰量组合	Kendall秩相关系数τ	Copula函数参数θ			
		G-H	Clayton	Frank	AMH
洪峰与最大7d洪量	0.755 2	4.085 6	6.171 2	14.486 8	0.880 8

表3 拟合优度评价结果

峰量组合	AIC信息准则				OLS信息准则			
	G-H	Clayton	Frank	AMH	G-H	Clayton	Frank	AMH
洪峰与最大7d洪量	-283.555 2	-280.631	-280.150 9	-230.709 4	0.039	0.040 3	0.040 5	0.071

因此, 选用二维Gumbel-Hougaard Copula函数作为横石站洪峰与最大7d洪量联合概率分布的连接函数, 将确定的洪峰、最大7d洪量的边缘分布函数的参数及GH Copula函数的参数代入其分布函数的公式

即可建立横石站洪峰与最大7d洪量联合概率分布函数, 由根据公式(12)、公式(13)分别计算洪峰、最大7d洪量的联合重现期、同现重现期, 并绘制其分布图和等值线图, 见图1。

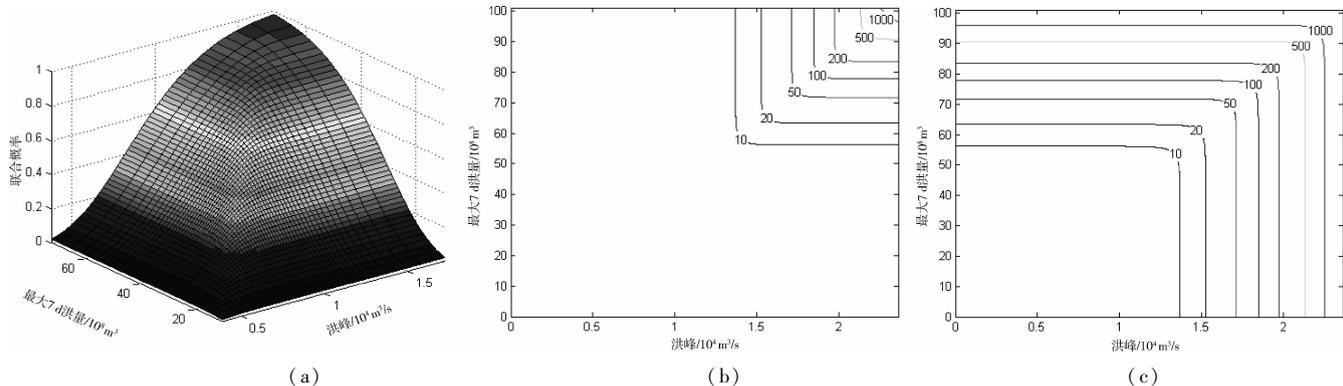


图1 洪峰与最大7d洪量联合分布图(a)、联合重现期等值线图(b)、同现重现期等值线图(c)

根据分布图, 即可查找飞来峡坝址洪水洪峰与最大7d洪量的联合分布值、联合重现期和同现期。如以1915年发生的珠江流域全流域特大洪水为例, 1915年特大洪水, 横石站洪峰、7d洪量分别为21000 m³/s、91.1亿m³, 其边缘分布重现期为417 a、526 a; 横石

站洪峰和7d洪量的联合分布函数值为0.9974, 联合重现期为385 a, 同现重现期为588 a, 意味着发生超过1915年洪水的洪峰或者7d洪量的重现期为385 a, 发生洪峰和7d洪量均超过1915年洪水的重现期为588 a。

3.3 结果分析

飞来峡水库坝址洪水洪峰和最大 7 d 洪量同频率组合下联合、同现重现期及其设计值见表 4。从表 4 中可见,联合重现期 T_o 小于单变量边缘分布的重现期,而同现重现期 T_a 则大于单变量边缘分布的重现期;根据单变量同频率方法推算的设计值小于两变量联合分布同现重现期计算的设计值。如以 1 000 a 一遇

洪水为例,洪峰与最大 7 d 洪量的边缘分布重现期设计值分别为 22 464 m^3/s 、96.05 亿 m^3 ,联合重现期为 844.0 a,同现重现期为 1 226.7 a,同现重现期设计值分别为 22 802 m^3/s 、97.57 亿 m^3 。以上结果表明,基于单变量计算推算的设计值的实际上达不到所要求的设计标准,基于两变量联合分布得到的洪水频率分析计算结果更合理。

表 4 洪峰和最大 7 d 洪量同频率组合下的重现期及其设计值

边缘分布 设计重现期 T/a	联合分布		同现重现期设计值			
	$Q_{峰}/(\text{m}^3 \cdot \text{s}^{-1})$	$W_7/10^8 \text{m}^3$	T_o/a	T_a/a	$Q_{峰}/(\text{m}^3 \cdot \text{s}^{-1})$	$W_7/10^8 \text{m}^3$
1 000	22 464	96.05	844.0	1 226.7	22 802	97.57
500	21 299	90.79	422.1	613.3	21 646	92.36
200	19 701	83.57	168.9	245.2	20 064	85.21
100	18 439	77.85	84.5	122.5	18 816	79.55
50	17 118	71.85	42.3	61.2	17 510	73.63
20	15 246	63.33	17.0	24.4	15 665	65.24
10	13 691	56.23	8.5	12.1	14 135	58.25

4 结语

以 AIC 信息准则法和离差平方和准则法对 4 种 Archimedean Copula 函数进行拟合优度检验,结果表明 Gumbel - Hougard Copula 函数最适合作为飞来峡水库坝址洪水洪峰与最大 7 d 洪量联合概率分布的连接函数,基于此,本文建立了以 P - III 型分布为边缘分布的飞来峡水库坝址洪水洪峰和最大 7 d 洪量的联合分布。将基于 Gumbel - Hougard Copula 函数的飞来峡水库坝址洪水两变量联合分布的洪水频率计算结果与单变量洪水频率计算结果进行比较,发现两变量联合分布得到的洪水频率分析计算结果更为合理,并绘制了洪峰、最大 7 d 洪量的联合分布图和联合重现期、同现重现期等值线图,为对飞来峡水库的洪水认识提供新的信息和理论支撑。

参考文献:

- [1] SL44—2006 水利水电工程设计洪水计算规范[S]. 北京:中国水利水电出版社, 2006.
- [2] 谢华,黄介生. 两变量水文频率分布模型研究述评[J]. 水科学进展, 2008(3): 443 - 452.
- [3] Nelson RB. An Introduction to copulas[M]. New York: Springer, 1999.

- [4] 郭生练,闫宝伟,肖义,等. Copula 函数在多变量水文分析计算中的应用及研究进展[J]. 水文, 2008(3): 1 - 7.
- [5] 杜江,陈希镇,于波. Archimedean Copula 函数的参数估计[J]. 科学技术与工程, 2009(3): 637 - 640.
- [6] 严忠权. 随机变量相依关系的度量[J]. 黔南民族师范学院学报, 2008(6): 34 - 38.
- [7] Christian Genest, Anne - Catherine Favre. Everything You Always Wanted to Know about Copula Modeling but Were Afraid to Ask [J]. Journal of Hydrologic Engineering, 2007 (4): 347 - 368.
- [8] Cheng wang, Ni - Bin Chang, Gour - Tsyh Yeh. Copula - based flood frequency analysis at the confluences of confluences of river system [J]. Hydrological Processes, 2009(2): 7 273 - 7 288.
- [9] Genest C, Rivest L. Statistical inference procedures for bivariate Archimedean copulas[J]. Journal of American Statistical Association, 1993(88), 1 034 - 1 043.
- [10] Akaike H. A new look at the statistical model identification [J]. IEEE Transactions on Automatic Control, 1974, 19 (6): 716 - 723.

(本文责任编辑 马克俊)