

# 基于多元分布函数的区域洪水频率分析

卢韦伟<sup>1</sup>, 陈璐<sup>1</sup>, 周建中<sup>1</sup>, 陈华<sup>2</sup>

(1. 华中科技大学 水电与数字化工程学院, 湖北 武汉 430074;

2. 武汉大学水资源与水电工程科学国家重点实验室, 湖北 武汉 430072)

**摘要:**极端洪水事件的频率分析往往局限于单个站点,当研究区域内包含多个水文站点时,单变量频率分析方法,会导致低估或高估洪灾风险率。因此,需要进行区域频率分析。传统区域重现期计算方法,同一重现期对应多种设计洪水组合,而基于 Kendall 分布函数的重现期计算方法(KRP)有效的解决了这一问题。故本文引入三维非对称 Copula 分布函数拟合区域内各个站点年最大流量的相关关系,利用半参数法估计 Copula 函数的参数,并采用 KRP 推求区域洪水发生的重现期。结果表明:区域发生  $T$  年一遇的洪水概率远远大于单个站点发生  $T$  年一遇的洪水概率;KRP 克服了实测序列较短的问题,且能准确估算洪水重现期。本研究为防洪部门制定防洪措施提供一定的科学依据。

**关键词:**区域洪水概率;三维 Copula 函数;多维重现期计算

中图分类号:P333

文献标识码:A

文章编号:1000-0852(2015)05-0006-05

## 1 引言

我国洪水灾害发生频繁,极端洪水事件的频率分析一直是水文领域的研究重点。在水文频率分析中,常见的是用单变量的分布函数来估计某个站点年最大洪水的发生概率<sup>[1]</sup>。但对于一个流域而言,流域范围内某一区域站点流量之间具有一定的相关性,忽略这种相关性将会低估或者高估设计值及其相应的风险率,因此需评估区域的洪灾风险率。

区域洪水频率分析的核心问题是利用区域多站的水文资料,进而推求区域洪水发生的联合分布函数。Copula 函数能够将多个变量的边缘分布以特定的方式连结起来,从而得到相应的联合分布。如熊立华等<sup>[1]</sup>利用 Clayton copula 函数对同一河流上下游两个水文站建立二维最大洪水联合分布函数。张翔和冉启香<sup>[2]</sup>采用 Clayton copula 函数以及 Gumbel-Hougaard copula 函数分别构造水量、水质的二维和三维联合分布。陈璐等<sup>[3]</sup>基于多维阿基米德 Copula 函数分别建立长江上游屏山站、高场站、北碚站以及宜昌站的洪水发生时间和量级的联合分布,全面地分析长江上游的洪水遭遇。

Copula 函数分为椭圆型、阿基米德型和二次型三种类型。其中,阿基米德族 Copula 函数结构相对简单,且计算较为简便,在水文领域应用最为广泛。在区域洪水频率分析中,传统区域洪水重现期计算方法所推求的区域重现期与各站点洪水设计值无法一一对应,传统的重现期计算方法已失效<sup>[4-5]</sup>,其会导致重现期估计的较大偏差。故本文选用三种常用的非对称阿基米德 Copula 函数对区域站点年最大洪水序列进行拟合以进行区域洪水频率分析,并采用 Salvadori 提出的 KRP (Kendall return period)对区域洪水重现期进行计算。

## 2 Copula 函数理论

本文采用三维非对称阿基米德 Copula 函数,构造站点年最大洪量的联合分布。常用的几种三维非对称阿基米德 Copula 函数的表达式为:

(1) Clayton copula 函数

$$C(u_1, u_2, u_3) = C_{\theta_1}(u_3, C_{\theta_2}(u_1, u_2)) \\ = [(u_1^{-\theta_1} + u_2^{-\theta_1} - 1)^{\theta_1/\theta_2} + u_3^{-\theta_1} - 1]^{-1/\theta_1} \quad (1)$$

式中: $\theta_1, \theta_2$  为 Copula 函数的未知参数,满足条件  $\theta_2 \geq$

收稿日期:2014-12-28

作者简介:卢韦伟(1991-),女,湖北仙桃人,硕士生,研究方向为水文学与水资源。E-mail: luweiwei@hust.edu.cn

通讯作者:陈璐(1985-),女,山东济南人,讲师,博士,主要从事水文分析与计算方面的研究。E-mail: chl8505@126.com

$\theta_1$ ,且取值范围为 $[0, \infty)$ ;  $u_1, u_2, u_3$  分别为随机变量边缘分布函数的具体取值。

(2) Frank copula 函数

$$C(u_1, u_2, u_3) = C_{\theta_1}(u_3, C_{\theta_2}(u_1, u_2))$$

$$= -\frac{1}{\theta_1} \ln \{ [1 - (1 - e^{-\theta_1})^{-1}] (1 - [1 - (1 - e^{-\theta_2})^{-1}] (1 - e^{-\theta_2 u_1}) (1 - e^{-\theta_2 u_2}))^{\theta_1/\theta_2} (1 - e^{-\theta_1 u_3}) \} \quad (2)$$

式中:  $\theta_1, \theta_2$  为 Copula 函数的未知参数, 满足条件  $\theta_2 \geq \theta_1$ , 且取值范围为  $[-\infty, +\infty] \setminus \{0\}$ ;  $u_1, u_2, u_3$  分别为随机变量边缘分布函数的具体取值。

(3) Gumbel copula 函数

$$C(u_1, u_2, u_3) = C_{\theta_1}(u_3, C_{\theta_2}(u_1, u_2)) \quad (3)$$

$$= \exp \{ 1 - [(-\ln u_1)^{\theta_2} + (-\ln u_2)^{\theta_2}]^{\theta_1/\theta_2} + (-\ln u_3)^{\theta_1} \}^{1/\theta_1}$$

式中:  $\theta_1, \theta_2$  为 Copula 函数的未知参数, 满足条件  $\theta_2 \geq \theta_1$ , 且取值范围为  $[1, \infty)$ ,  $u_1, u_2, u_3$  分别为随机变量边缘分布函数的具体取值。

常用 Coupla 函数的参数估计方法有: 相关性指标法、极大似然法、半参数法<sup>[5]</sup>以及 IFM 估计法<sup>[6]</sup>等。对于三维及以上的 Copula 函数, 相关性指标法显然不再适用, 故本文采用半参数法估计三维非对称阿基米德 Copula 函数的参数  $\theta_1$  和  $\theta_2$ 。

本文采用 Kolmogorov-Smirnov 检验( K-S 检验)以判断 Copula 函数是否通过拟合分布假设检验, 并选用均方根误差(RMSE)评价 Copula 函数的拟合效果, 最后通过 AIC 准则选择最优的 Copula 函数。

### 3 区域洪水频率分析

#### 3.1 区域洪水概率及重现期

区域洪水概率为区域所有站点中, 至少有一个站点出现  $T$  年一遇洪水的概率<sup>[7]</sup>。其表达式为:

$$P_1(T) = 1 - P(X_1 \leq x_1(T), \dots, X_n \leq x_n(T)) \quad (4)$$

式中:  $P_1(T)$  为  $T$  年一遇洪水的区域洪水概率;  $X_i (i=1, 2, \dots, n)$  为第  $i$  个站点的年最大径流量;  $x_i(T) (i=1, 2, \dots, n)$  为第  $i$  个站点  $T$  年一遇的设计洪水值;  $P(X_1 \leq x_1(T), \dots, X_n \leq x_n(T))$  为联合分布函数计算值。

传统区域重现期遵循联合重现期<sup>[7]</sup>的定义, 其表达式为:

$$T_p = \frac{1}{1-p} = \frac{1}{1-C(\mathbf{u})} \quad (5)$$

式中:  $p=C(\mathbf{u})=P(X_1 \leq x_1, \dots, X_n \leq x_n)$ ;  $C(\mathbf{u})$  为 copula 函数;  $\mathbf{u}$  为变量  $u_1, \dots, u_n$  的集合,  $\mathbf{u}=(u_1, \dots, u_n)$ 。

与式(5)对应的  $n$  维变量取值集合可表示为:

$$R_p^{\leq} = \{ \mathbf{x} \in R^n; \mathbf{u}_x \leq \mathbf{u} \} \quad (6)$$

式中:  $\mathbf{u}_x$  为各站设计值集合  $\mathbf{x}$  所对应的各站点频率集合。

对于单一随机水文变量  $X_1$ , 重现期  $T_p$  与  $X_1$  取值呈现一一对应的关系。但当考虑  $n (n \geq 2)$  维联合分布的重现期时, 不同的  $\mathbf{u}$  可能得到相同的重现期  $T_p (P=C(\mathbf{u}))$ , 此时各站点设计洪水集合  $R_p^{\leq}$  与  $T_p$  不再一一对应。为解决此问题, Salvadori<sup>[8]</sup>提出多维重现期的一种新定义形式——KRP。此方法引入 Kendall 分布函数  $K_C$  以计算重现期,  $K_C$  仅与 Copula 函数有关。任意给定一常数  $t (t \in [0, 1])$ ,  $K_C$  的表达式为:

$$K_C(t) = P\{ \mathbf{u} \in I^n; C(\mathbf{u}) \leq t \} \quad (7)$$

与式(7)对应的  $n$  维变量取值集合可表示为:

$$R_t^{\leq} = \{ \mathbf{x} \in R^n; C(\mathbf{u}) \leq t \} \quad (8)$$

式中:  $\mathbf{u}=(u_1, \dots, u_n)=(F_1(x_1), \dots, F_n(x_n))$ , 即 Copula 函数取值小于或等于给定  $t$  时的所有边缘分布取值集合。 $R_t^{\leq}$  为 Copula 函数取值小于或等于给定  $t$  时的所有变量  $X$  的取值集合。

$n$  维联合分布的重现期定义为:

$$T_p = \frac{1}{1-p} = \frac{1}{1-K_C(t)} \quad (9)$$

当给定重现期  $T_p$  时,  $p, K_C(t)$  的取值同样可得。  $t$  作为 Copula 函数的上边界, 表达式为:

$$t = \inf \{ s \in I; K_C(s) = p \} \quad (10)$$

由式(8)、(9)、(10)可知,  $T_p, K_C(t), t$  与  $R_t^{\leq}$  均存在一一对应的关系。

#### 3.2 区域洪水频率分析

区域洪水频率分析的具体步骤为:

(1) 采用 P- 型分布对区域各个站点的最大洪水序列分别进行频率曲线拟合, 即计算区域联合分布的边缘分布函数;

(2) 运用三维非对称 Copula 函数推求区域所有站点的联合分布函数, 并进行拟合优度检验, 选取 AIC 值较小即拟合最优的 Copula 函数;

(3) 计算三个站点同时发生  $T$  年一遇洪水时的区域洪水概率, 当重现期为  $T (T = \{1000, 500, 100, 50, 20, 10, 5\})$  年时各个站点的设计最大洪水以及对应的区域洪水概率;

(4) 计算区域洪水的重现期。给定  $t$  值, 采用 KRP 方法计算区域重现期的具体步骤为:

①随机生成均匀分布的边缘分布概率集合  $u_1, \dots, u_m, u_i=(u_1, \dots, u_n)$ 。其中,  $i=1, 2, \dots, m, m$  为生成的概率集合总数,  $n$  为 Copula 函数的维数,  $u_1, \dots, u_n$  分别为随机变量  $X_1, \dots, X_n$  所对应的边缘分布概率; ②计算联合分布函数值  $v_i=C(u_i) (i=1, 2, \dots, m)$ ; ③  $K_C(t)$  的估计值为  $K_C(t)=\frac{1}{m} \sum_{i=1}^m M$ , 其中  $M$  为  $v_i \leq t$  的个数; ④根据式(9)计算区域重现期  $T_p$ 。

#### 4 应用研究

本文选取湘江流域为研究对象, 通过流域上衡阳站、湘潭站、甘溪站等水文站 1969~2008 年日流量序列对湘江流域的区域洪水频率进行分析。图 1 为湘江流域衡阳、湘潭、甘溪水文测站大致分布图。衡阳站、湘潭站为干流站点, 甘溪站为支流站点。地理分布相对位置大致为: 衡阳站、湘潭站距离最远, 而甘溪站、衡阳站距离最近。



图 1 湘江流域三站点的地理分布

Fig.1 Location of the 3 stations in the Xiangjiang River basin

#### 4.1 联合分布函数的建立

##### 4.1.1 边缘分布函数的确定

本文采用 P- 型分布对 3 个站点的最大洪水进行拟合, 表 1 列出 3 个站点的 P- 分布的参数值及置信水平为 0.05 时 K-S 假设检验结果。由表 1 结果可知,  $p$  值均远远大于置信水平 0.05, P- 型分布的拟合

通过了 K-S 检验。

表1 三维边缘分布函数的参数估计及K-S检验结果  
Table1 The results of the parameters estimation and K-S test for the marginal distributions

边缘分布函数参数	$\alpha$	$\beta$	$a_0$	K-S 统计量	$p$ 值	是否通过
衡阳	7.07	0.00077	1065.07	0.068	0.987	是
湘潭	22.00	0.00124	-4914.41	0.095	0.831	是
甘溪	1.98	0.00097	988.03	0.061	0.996	是

##### 4.1.2 变量相关性分析

本文选取常用的 Kendall 秩相关系数  $\tau$  度量 3 个站点最大洪水序列的相关性。表 2 列出 3 个站点之间的  $\tau_n$  (表中下三角部分) 以及对应的非假设检验  $p$  值 (表中上三角部分)。由表 2 可看出: 湘潭站和甘溪站、衡阳站和甘溪站的  $p$  值远远小于置信水平 (一般取 0.05), 表明湘潭站和甘溪站、衡阳站和甘溪站存在较强的相关性。综上所述, 湘潭站、衡阳站和甘溪站之间存在正相关性, 适用于 Clayton copula、Frank copula、Gumbel copula 函数。

表2 站点最大洪水序列的相关性 (Kendall秩相关系数)  
Table2 The Kendal' s tau correlation coefficients

站点	衡阳	湘潭	甘溪
衡阳	1	4.44E-11	0.00734
湘潭	0.725	1	4.86E-05
甘溪	0.295	0.447	1

##### 4.1.3 Copula 参数估计及拟合优度检验

采用半参数法估计各 Copula 函数的参数  $\theta_1, \theta_2$ , 并给出 K-S 检验法、AIC 准则以及 RMSE 的计算结果, 列入表 3 中。由表 3 可看出: 在置信水平为 0.05 时, Clayton copula 函数和 Frank copula 函数通过 K-S 检验, 而 Gumbel copula 函数未通过检验; Frank copula 函数的 RMSE 值、AIC 值均最小, 拟合效果最好。故本文选取 Frank copula 作为拟合函数。

表3 Copula函数的参数估计及拟合优度检验  
Table3 The results of the parameter estimation and goodness-of-fit test of copula functions

Copula 函数	参数		RMSE	AIC	$D_{max}$
	$\theta_1$	$\theta_2$			
Clayton	0.73	2.91	0.0596	-221.62	0.1500
Frank	3.38	11.07	0.0394	-254.74	0.1073
Gumbel	0.76	0.97	0.1802	-133.11	0.3441

4.2 湘江区域洪水分析

4.2.1 区域洪水频率分析

给定单站洪水发生概率,以推求对湘江流域的区域洪水发生概率。其中,区域洪水概率指区域内至少有一个站点发生指定洪水的概率,而单站洪水发生概率指单个站点发生指定洪水的超过概率。基于式(4),计算当  $u_1=u_2=u_3$  时区域概率值,结果见表4。当各个站点相互独立时,  $C(u_1,u_2,u_3)=u_1 \cdot u_2 \cdot u_3$ , 区域洪水概率为  $1-u_1 \cdot u_2 \cdot u_3$ 。但由于区域各个站点之间存在相关性特征,因此实际求得联合分布函数值并不是简单的  $u_1, u_2, u_3$  相乘。区域洪水概率要远大于单站洪水发生概率。因区域三个站点地理位置相近,三者具有一定的相关性,且相关性越大,区域洪水概率与单站洪水发生概率差距越小。当三者完全相关时,区域洪水概率则等于单站发生洪水概率。当单站发生  $P(Q_i \geq q_i)=1-u_i=0.5 (i=1,2,3)$  的洪水时,区域内至少有一个站点发生2年一遇的洪水概率为0.69,远大于单站发生洪水概率0.5;当单站发生  $P(Q_i \geq q_i)=1-u_i=0.1 (i=1,2,3)$  的洪水时,区域内至少有一个站点发生10年一遇的洪水概率为0.21,明显大于单站洪水概率0.1。

表4 区域洪水概率的计算结果  
Table4 The calculation results of regional frequency analysis

$u_1$	$u_2$	$u_3$	区域洪水概率
0.1	0.1	0.1	0.99
0.3	0.3	0.3	0.87
0.5	0.5	0.5	0.69
0.7	0.7	0.7	0.48
0.9	0.9	0.9	0.21

4.2.2 区域重现期分析

为检验 KRP 计算区域重现期的拟合效果,本文采用经验概率、传统重现期计算方法以及基于 Frank、Gumbel 和 Clayton Copula 的 KRP 等5种方法,计算与区域重现期具有一一对应关系的 Kendall 联合分布函数值  $K_C(t)$ 。

当  $t$  值给定时,为检验 KRP 方法的可行性,依据  $K_C(t)$  的定义式(7),采用实测序列计算  $K_C(t)$  即经验频率法的表达式为:

$$K_C(t) = \frac{k}{N+1} \tag{11}$$

式中: $N$  为实测洪水序列的长度; $k$  为由实测序列推求

的  $C(u) \leq t$  的实测点数。

当  $t$  值给定时,采用传统重现期计算方法计算  $K_C(t)$  的表达式为:

$$\overline{K}_C(t) = t \tag{12}$$

当  $t$  值给定时,采用 KRP 计算  $K_C(t)$  如区域重现期计算步骤所示。

图2描绘了随着  $t$  取值的增大,采用5种方法得到的增长趋势。由图2可得:基于 Gumbel copula 与基于 Clayton copula 的 KRP 的拟合效果低于传统重现期计算方法,而基于 Frank copula 的 KRP 则高于传统重现期计算方法,这表明对于湘江流域,并非所有 Copula 函数的 KRP 计算均优于传统重现期计算方法,但相比于传统重现期计算方法,基于 Frank copula 的 KRP 重现期计算具有合理性和优越性。

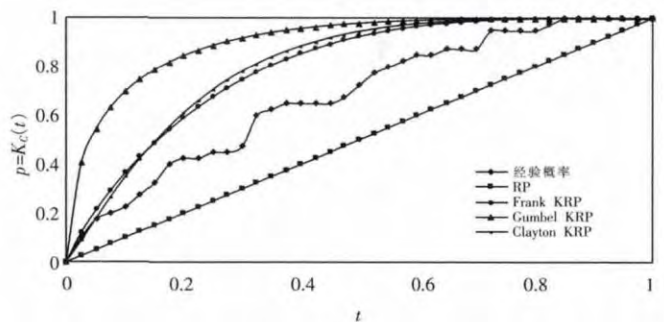


图2 区域洪水重现期的 Kendall 分布函数值的趋势图  
Fig.2 The plots of empirical and fitted Kendall's distribution functions KC for the three stations

给定各站点的重现期,计算联合分布函数值  $C(u)$ ,  $t$  表示联合分布函数的上界,故此时  $t=C(u)$ 。依照区域重现期的计算步骤,取随机生成样本数  $m=3 \times 10^7$ ,求得各个站点的年最大设计洪水值以及其对应的区域重现期,结果见表5所示。表5中,“-”表示区域重现期无穷大,可以近似认为事情发生可能性几乎为0,各站都发生重现期为50年一遇的洪水时,衡阳站、湘潭站、甘溪站的设计流量分别为  $18\ 633\text{m}^3/\text{s}$ 、 $21\ 510\text{m}^3/\text{s}$ 、 $6\ 944\text{m}^3/\text{s}$ ,而区域重现期为  $29\ 970.03\text{a}$ ,说明各站点同时发生50年一遇的洪水发生的几率极低。因衡阳站与湘潭站均为干流水文站点,且两者地理位置较近,故两者发生  $T$  年一遇洪水的设计值相差不大。

5 结论

本文利用三维非对称 Copula 函数计算了流域的区域洪水概率。以湘江流域为研究对象,基于衡阳站、湘潭站、甘溪站的年最大流量系列,建立三维联合分布



表5 区域重现期的计算结果  
Table5 The calculation results of the regional return periods

重现期/a	设计值/ $\text{m}^3\cdot\text{s}^{-1}$			区域重现期/a
	衡阳	湘潭	甘溪	
1000	24656	26936	10433	-
500	23333	25773	9641	-
100	20110	22875	7770	196078.43
50	18633	21510	6944	29970.03
20	16556	19546	5827	2160.92
10	14849	17883	4952	306.39
5	12950	15970	4035	46.68

函数,对区域洪水概率大小进行分析,采用 KRP 计算区域洪水重现期,并与传统重现期计算结果进行比较分析。结果表明:站点的相关性直接影响到区域洪水概率的大小,区域发生  $T$  年一遇的洪水概率远远大于单个站点发生  $T$  年一遇的洪水概率,且相比于实际洪水发生情况,传统重现期的计算结果严重偏小,因而可能对防洪减灾造成影响,而 KRP 计算重现期的方法成功解决了实测水文序列较短的问题,利于正确估计洪水发生的重现期。本研究可为区域防洪分析提供一定的科学依据。

参考文献:

[1] 熊立华,郭生练,肖义,等. Copula 联结函数在多变量水文频率分析中的应用 [J]. 武汉大学学报(工学版), 2005,38 (6):16-19. (XIONG

- Lihua, GUO Shenglian, XIAO Yi, et al. Application of copulas to multivariate hydrological frequency analysis [J]. Journal of Wuhan University (Engineering Edition), 2005,38(6):16-19. (in Chinese))
- [2] 张翔,冉启香,夏军,等. 基于 Copula 函数的水量水质联合分布函数[J]. 水利学报, 2011,42(4):483-489. (ZHANG Xiang, RAN Qixiang, XIA Jun, et al. Jointed distribution function of water quality and water quantity based on Copula [J]. Journal of Hydraulic Engineering, 2011,42(4):483-489. (in Chinese))
- [3] 陈璐,郭生练,张洪刚,等. 长江上游干支流洪水遭遇分析[J]. 水科学进展, 2011,22(3):323-330. (CHEN Lu, GUO Shenglian, ZHANG Honggang, et al. Flood coincidence probability analysis for the upstream Yangtze River and its tributaries [J]. Advances in Water Science, 2011,22(3):323-330. (in Chinese))
- [4] 郭生练. 设计洪水研究进展与评价[M]. 北京:中国水利水电出版社, 2005. (GUO Shenglian. Progress and Evaluation of Design Flood[M]. Beijing: China WaterPower Press, 2005. (in Chinese))
- [5] Genest C, Ghoudi K, Rivest L P. A semiparametric estimation procedure of dependence parameters in multivariate families of distributions [J]. Biometrika, 1995,82(3):543-552.
- [6] 杨益党,罗羨华. Copula 函数的参数估计[J]. 新疆师范大学学报(自然科学版), 2007,26(2):15-18. (YANG Yidang, LUO Xianhua. Parametric estimation of Copula function [J]. Journal of Xinjiang Normal University (Natural Sciences Edition), 2007,26(2):15-18. (in Chinese))
- [7] Shiau J T. Return period of bivariate distributed extreme hydrological events [J]. Stochastic Environmental Research and Risk Assessment, 2003,17(1-2):42-57.
- [8] Salvadori G, De Michele C. Multivariate multiparameter extreme value models and return periods: A copula approach [J]. Water resources research, 2010,46(10).

## Regional Flood Frequency Analysis Based on Multivariate Distribution Function

LU Weiwei<sup>1</sup>, CHEN Lu<sup>1</sup>, ZHOU Jianzhong<sup>1</sup>, CHEN Hua<sup>2</sup>

(1. College of Hydropower & Information Engineering, Huazhong University of Science & Technology, Wuhan 430074, China;

2. State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan, 430072, China)

**Abstract:** Traditional flood frequency method mainly focuses on one gauging station. If series gauging stations located in the same region and there are some dependence among them, at-site frequency analysis method may lead to overestimating or underestimating the regional flood risk. Therefore, it is great need to apply regional flood frequency analysis instead of at-site flood frequency analysis. In this paper, we introduced trivariate asymmetric Copula function to establish the joint distribution of annual maximum flow. The multivariate return periods were calculated based on Copulas and the Kendall's measures. Results show that the occurrence probability of  $T$ -year flood in the region is much higher than the probability of that at site.

**Key words:** regional flood frequency analysis; trivariate Copula function; multivariate return periods