

# 基于 Copula 函数的组合变量联合概率分布研究及应用

林 娴<sup>1,2</sup>, 欧阳昊<sup>3</sup>, 陈晓宏<sup>1</sup>, 李深林<sup>1</sup>

(1. 中山大学水资源与环境研究中心, 广东 广州 510275; 2. 广东省水利厅, 广东 广州 510000;  
3. 珠江流域水资源保护局, 广东 广州 510000)

**摘要:** 基于 Copula 函数原理, 利用武江流域实测水文资料, 以广义 GDP 为洪峰洪量边缘分布, 构建了流域组合变量 Copula 概率分布模型, 分析了洪峰与洪量、洪量与洪水历时、洪峰与洪水历时的联合概率分布, 绘制各种变量组合下的联合分布图及重现期等值线图, 并比较了同重现期条件下, 洪水单变量设计值与多维联合设计值的区别。结果表明: 广义 GDP 分布能很好的描述洪峰、洪量边缘分布, 而基于广义 GDP 分布和指数分布构建的两变量 Copula 联合概率分布模型不限定变量的边缘分布, 对各种类型的水文变量联合分布拟合效果较好; 能全面反映洪水各特征属性不同等级下的联合发生频率, 对同一频率下联合分布推求的洪水设计值比单变量设计值偏于安全。基于 Copula 函数的组合变量概率分布模型描述洪峰流量、洪量、洪水历时等特征的联合分布, 较为全面地反映组合特征的洪水发生的概率和重现期, 进一步反映洪水风险。

**关键词:** Copula 函数; 组合变量概率分布模型; GDP 分布; 频率分析; 武江流域

**中图分类号:** P333      **文献标识码:** A      **文章编号:** 1000-0852(2017)01-0001-07

## 1 引言

传统水文事件频率分析多集中在单变量水文频率研究, 洪水作为研究最广泛的水文事件之一, 具有多个特征属性(洪峰流量、洪水总量、峰现时间、洪水历时等), 且各属性间存在一定的相依性, 立足于洪水变量多属性特征的研究其联合概率分布及条件概率分布, 对水利工程的规划设计及风险评估具有重要意义<sup>[1]</sup>。Copula 函数通过任意边缘分布和相关性结构来构造多维联合概率分布, 形式灵活多样, 计算较为简单, 是求解组合变量概率问题优良的数学工具, 国内外学者已将 Copula 函数理论与方法广泛应用于组合变量水文分析计算领域<sup>[2-3]</sup>; 针对洪水事件多特征属性, 利用 Copula 函数理论与方法构建了洪水洪峰、洪量和历时等各种属性的联合分布模型, 也取得了一些成果<sup>[4-6]</sup>, 为更精确地估算极端事件发生频率, 控制工程风险提供了重要的理论基础和技术手段。

运用 Copula 函数法构建洪水特征量的组合变量联合分布时, 首先要确定各洪水特征量的边缘分布。基于极值统计、极值理论的应用和研究开展的洪水频率分析为描述洪水等极端事件提供了有效理论基础, 洪水超阈值峰量模型(POT 模型)的应用可有效获取更多的洪水极值数据, 增加资料系列长度, 基于超阈值理论的洪水频率计算与分析不断得到探索和研究<sup>[7-9]</sup>。本文以武江流域为例, 利用犁市(二)站实测水文资料, 基于洪峰和洪量均服从 GDP 型分布、洪水历时服从指数分布, 应用 Archimedena Copula 函数, 探讨了武江流域组合变量联合概率分布模型, 以期对地区水利工程规划设计和风险评估提供依据。

## 2 研究方法

### 2.1 概率分布函数

Copula 函数由 Sklar 在 1959 年提出的, 是一类将变量联合累积分布函数同变量边缘累积分布函数连接起

收稿日期: 2015-11-26

基金项目: 国家自然科学基金项目(51210013, 51479216); 国家科技支撑项目(2012BAC21B0103); 水利部公益项目(201201094, 201301002-02); 院土科技咨询项目(2015-ZD-07-04-03); 广东省水利科技创新项目(2011-11)

作者简介: 林娴(1981-), 女, 福建长乐人, 硕士研究生, 主要从事水文与水环境研究。E-mail: gqtgds1@163.com

通讯作者: 欧阳昊(1982-), 男, 湖南株洲人, 高级工程师, 主要从事水文与水环境研究。E-mail: 443016171@qq.com

来的函数,边缘分布描述的是单个变量的分布,Copula函数描述的是组合变量之间的相关性。设  $F$  为一个  $n$  维联合累积分布函数,其边缘分布为  $F_1, F_2, F_3, \dots, F_n$ , 存在  $n$  维 Copula 函数  $C$ , 使得对任意  $x \in R^n$  有:

$$F(x_1, x_2, \dots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n) \\ = C(F_1(x_1), F_2(x_2), \dots, F_n(x_n)) \quad (1)$$

式中:  $C$  为 Copula 函数,是定义在  $[0,1]$  区间均匀分布的联合概率分布函数,对它的每一个变量都是递增的。

Copula 函数不要求具有相同的边缘分布,任意边缘分布经 Copula 连接都可构成联合分布,由于变量的所有信息都包含在边缘分布中,在转换过程中不会造成信息失真。由此可见,Copula 函数不仅为求取联合分布函数提供了一条简便途径,还由于 Copula 函数的不变性,由其导出的一致性和相关性测度的值也不会变化,应用范围更广,适用性更强。

Copula 函数的构造形式多种多样,其中 Archimedean Copula 函数结构相对简单,计算方便,可以构造出多种形式、适应性强的组合变量联合分布函数,实际中应用的较为广泛。常见的二维 Archimedean Copula 函数主要有

Gumbel-Hougaard Copula:

$$F_{(X,Y)}(x,y) = C(u,v) = \exp\{-(\ln u)^\theta + (\ln v)^\theta\}^{1/\theta}, \theta \in [1, \infty) \quad (2)$$

Clayton Copula:

$$F_{(X,Y)}(x,y) = C(u,v) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}, \theta \in (0, \infty) \quad (3)$$

Frank Copula:

$$F_{(X,Y)}(x,y) = C(u,v) = -\frac{1}{\theta} \ln\left[1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1}\right], \theta \in R \quad (4)$$

式中:  $C(u,v)$  为二维的 Copula 函数;  $u, v$  分别为边缘分布函数,  $u = F_X(x), v = F_Y(y)$ ;  $\theta$  为 Copula 函数的参数。

洪水样本的识别是洪水频率分析的基础,本文采取 POT 模型选取洪水样本,样本中洪峰和洪量边缘分布均服从广义 GDP 分布,洪水历时服从指数分布。洪峰样本的独立性的判别及阈值的确定是 POT 模型的关键问题。关于连续洪峰独立性的判别,本研究采用美国水资源协会(USWRC)推荐的方法:①两个独立洪峰的间隔时间应大于 5 日加上流域面积自然对数的和;②两个独立洪峰之间的最小流量( $Q_{\min}$ )应小于两个洪峰较小者的 75%,当两个条件同时满足时,判断这两个连续洪峰为独立洪峰。阈值的确定采用超均值函数图法确定初始阈值,并采用离差指数法检验选择阈值是否满足洪峰次数假设<sup>[7-9]</sup>,利用 K-S 拟合检验选取

最优阈值,确定 POT 洪水样本序列。

## 2.2 参数估计

常用的参数估计方法有极大似然法、相关性指标法、IFM 估计法和 MBP 估计法。本文采用极大似然估计法对洪水变量边缘分布参数进行估计,利用相关性指标法估计 Copula 联合分布函数的参数<sup>[10-11]</sup>。

相关性指标法是运用 Copula 函数的参数  $\theta$  与 Kendall 秩相关系数之间的关系来间接计算参数  $\theta$  的方法。两个独立同分布的随机变量  $\{x_i\}$  和  $\{y_i\}, i=1,2,\dots,n$ ,  $n$  为  $(x_i, y_i)$  观测数目, Kendall 秩相关系数  $\tau$  可表示为:

$$\tau_n = (C_n^2)^{-1} \sum_{i < j} \text{sign}[(x_i - x_j)(y_i - y_j)] \quad (5)$$

式中:当  $x_i \leq x_j$  且  $y_i \leq y_j$  时,  $\text{sign}=1$ , 否则,  $\text{sign}=-1$ ;  $\tau_n$  为  $\tau$  的估计值。

常见 Archimedean Copula 函数的参数与 Kendall 秩相关系数之间的关系式如表 1 所示。

表1 Copula函数参数 $\theta$ 与Kendall相关系数 $\tau$ 的关系式  
Table1 The relationship between parameter  $\theta$  of Copula and parameter  $\tau$  of Kendall

Copula 函数	$\theta$ 与 $\tau$ 的关系	$\theta$ 的取值范围
Gumbel-Hougaard	$\tau = 1 - \theta^{-1}$	$\theta \geq 1$
Frank	$\tau = 1 + \frac{4}{\theta} \left( \frac{1}{\theta} \int_0^{\theta} \frac{t}{e^t - 1} dt - 1 \right)$	$\theta \neq 0$
Clayton	$\tau = \frac{\theta}{\theta + 2}$	$\theta \geq 0$

## 2.3 拟合优度评价和检验

在运用 Archimedean Copula 研究实际问题的过程中,为选择合理的 Copula 函数来正确描述变量间相关结构,需要对各函数进行拟合度评价和检验。Copula 拟合优度评价方法主要用 Genest-Rivist 图形分析法,拟合检验主要采用离差平方和准则法、AIC 信息准则法<sup>[12]</sup>。本研究采用 Genest-Rivist 图形分析法:

图形分析法:利用图形直观的描述拟合的优劣程度。将理论联合概率值和经验联合概率值点绘成散点图,如果点距较均匀的分布在 45° 线附近,则说明建立的联合概率分布模型是合理的。其中,两变量经验联合概率分布的计算公式如下:

$$F_{emp}(x_i - y_i) = P(X \leq x_i, Y \leq y_i) = \frac{\sum_{i=1}^n \sum_{k=1}^n n_{gk} - 0.44}{N + 0.12} \quad (6)$$

式中:  $F_{emp}$  为经验联合概率分布;  $n_{gk}$  为满足  $X \leq x_i$ ,

$Y \leq y_i$  的观测数目。

### 2.4 联合分布重现期

在洪水频率分析计算时,往往更多的关注洪峰、洪水总量、历时超过某一特定值或多要素共同超过某特定值时的频率,即重现期,以及某一设计频率下对应的设计洪水各特征值。设  $F_X(x)$  和  $F_Y(y)$  分别为洪水特征变量  $X$  和  $Y$  的分布函数,两变量同时超过某一特定值的重现期如式(7)所示:

$$T_{(X>x, Y>y)} = \frac{1}{1 - F_X(x) - F_Y(y) + C(F_X(x), F_Y(y))} \quad (7)$$

两变量中任一变量超过某一特定值时的重现期如式(8)所示:

$$T_{(X>x \text{ or } Y>y)} = \frac{1}{1 - C(F_X(x), F_Y(y))} \quad (8)$$

其中用  $T_{xy}$  记  $T(X>x, Y>y)$ ;  $T_{xy}$  记  $T(X>x \text{ or } Y>y)$ 。

## 3 Copula 函数在研究区的应用

### 3.1 研究区概况

武江为珠江流域北江水系的一级支流,发源于湖南省临武县三峰岭,流经湖南省,由广东省北部地区的东南方向流入广东省境内,再经过乐昌、乳源、曲江和韶关 4 个县(市、区),于韶关市区的沙洲尾汇入北江,干流全长 260km,流域面积 7 097km<sup>2</sup>,位于东经 112°50'~113°35',北纬 24°50'~25°31',流域呈阔叶树叶状,流向从西北至东南。流域属东亚季风气候区,冬季气候寒冷干燥,夏季炎热多雨,年平均降雨量在 1 300~1 500mm 之间,汛期降雨量(4~9月)约占 72%。流域年径流总量为 60.8×10<sup>8</sup>m<sup>3</sup>,70%~80%集中于汛期。洪涝灾害严重威胁着乐昌县、韶关市等地区的生命财产安全。本文研究主要采用武江流域犁市(二)(后称犁市)水文站的 1956~2010 年日流量资料(源自广东省水文局)。

### 3.2 洪水样本选取及边缘分布参数估计

根据犁市站控制面积,得到犁市站独立洪峰峰间间隔  $\theta$  应大于 13 d;取  $\theta=14$ 。根据独立性判别条件,最终选取 640 个独立洪峰样本。根据超均值函数图和离差指数法,最终选定独立洪峰阈值为 1 700m<sup>3</sup>/s,选取独立性超阈值洪峰样本 69 个,利用 K-S 双尾检验进行 GPD 分布拟合检验,效率系数为 0.952 1,拟合效果良好。

统计分析超阈值洪水样本可知,1956~2010 年期间流域特大洪水发生次数为 69 次,年均 1.27 次,多年平均洪峰值为 24 720m<sup>3</sup>/s,多年平均洪量为 3.53×10<sup>8</sup>m<sup>3</sup>,具体特征值见表 2。

表2 场次超阈值洪水要素特性  
Table2 The features of the over-threshold flood

洪水要素	均值	标准差	$C_u$	$C_s$	BK
年洪水次数	1.27	1.13	0.90	0.62	2.49
洪峰值/m <sup>3</sup> .s <sup>-1</sup>	2 472	1 001.75	0.41	2.98	14.30
洪量/10 <sup>8</sup> m <sup>3</sup>	3.53	3.49	0.99	3.19	13.49
历时/d	1.72	1.28	0.74	3.39	17.75

运用极大似然估计对洪峰、洪量及历时边缘分布参数进行估计,可得相应分布统计参数,结果如表 3 所示。利用极大似然值确定参数后,洪峰、洪量及历时拟合效果见图 1,效率系数均在 0.9 以上,拟合效果良好。

表3 边缘分布函数的参数估计  
Table3 The parameters of marginal distribution function

参数	$u$	$\xi$	$\sigma$	$\lambda$
洪峰	1700	0.24	591.63	
洪量	1.47	0.67	0.9	
历时				1.72

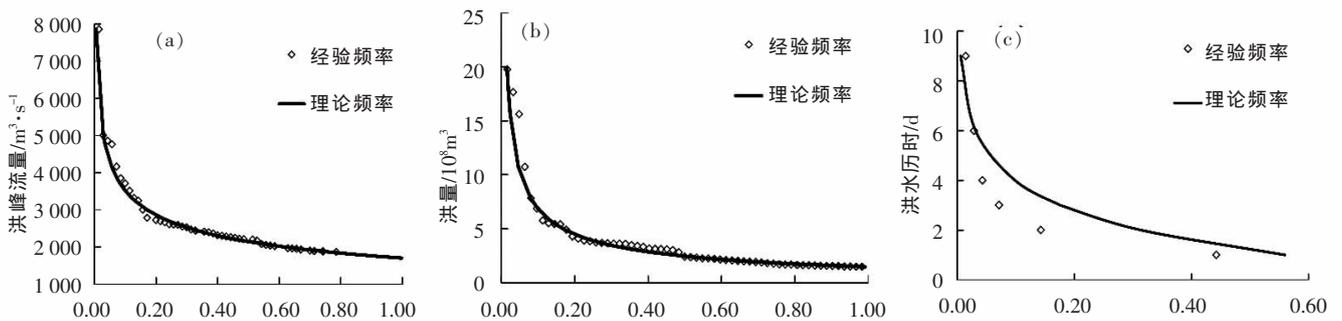


图 1 洪水特征变量拟合图  
(a.洪峰分布拟合图;b.洪量分布拟合图;c.洪水历时分布拟合图)  
Fig.1 The fitting plots of the flood characteristics variables

### 3.3 Copula 联合分布参数的确定

利用式(5)可分别计算出洪水各特征属性值之间的 Kendall 秩相关系数  $\tau$ , 结果见表 4。由  $\tau$  值可以看出, 洪量-历时及洪峰-洪量间存在较强的相依性, 而洪峰和历时相关性较弱。将  $\tau$  分别代入 Archimedean Copula 函数参数  $\theta$  与  $\tau$  的关系式(见表 1)中, 可估算其分布函数的参数  $\theta$ , 结果见表 4。

表4 Kendall秩相关系数 $\tau$ 和Archimedean Copula函数的 $\theta$ 估值  
Table4 The estimated values of  $\tau$  and  $\theta$

变量	Kendall 秩 相关系数 $\tau$	$\theta$		
		G-H	Frank	Clayton
洪量-历时	0.76	4.20	14.97	6.41
洪峰-历时	0.37	1.58	3.74	1.17
洪峰-洪量	0.67	2.99	10.01	3.99

### 3.4 拟合评价与优度检验

利用 Genest-Rivest 检验法, 分别计算各 Copula 函数的理论估计值和经验估计值, 并点绘 p-p 图(见图 2), 分析比较 p-p 图更接近于对角线的 Copula 函数, 以此来检验各种 Copula 函数对实测洪水历时、洪峰及洪量组合值的拟合程度。

由以上相关图可以得出, 三组 Copula 函数对应的理论和经验点据拟合效果总体良好。相关系数都在 0.95 以上。不同 Copula 函数下, 各变量 p-p 图拟合效果不同。统计各组 p-p 图效率系数可得: 对于洪量-历时联合分布 Frank Copula 拟合效果最好, 相关系数达 0.97, 洪峰-历时联合分布 Gumbel Copula 拟合效果最好, 相关系数为 0.98; 洪峰-洪量联合分布 Clayton Copula 拟合效果最好, 相关系数为 0.99。利用 AIC 信息准则和 OLS 准则进行拟合优度评价, 结果见表 5。

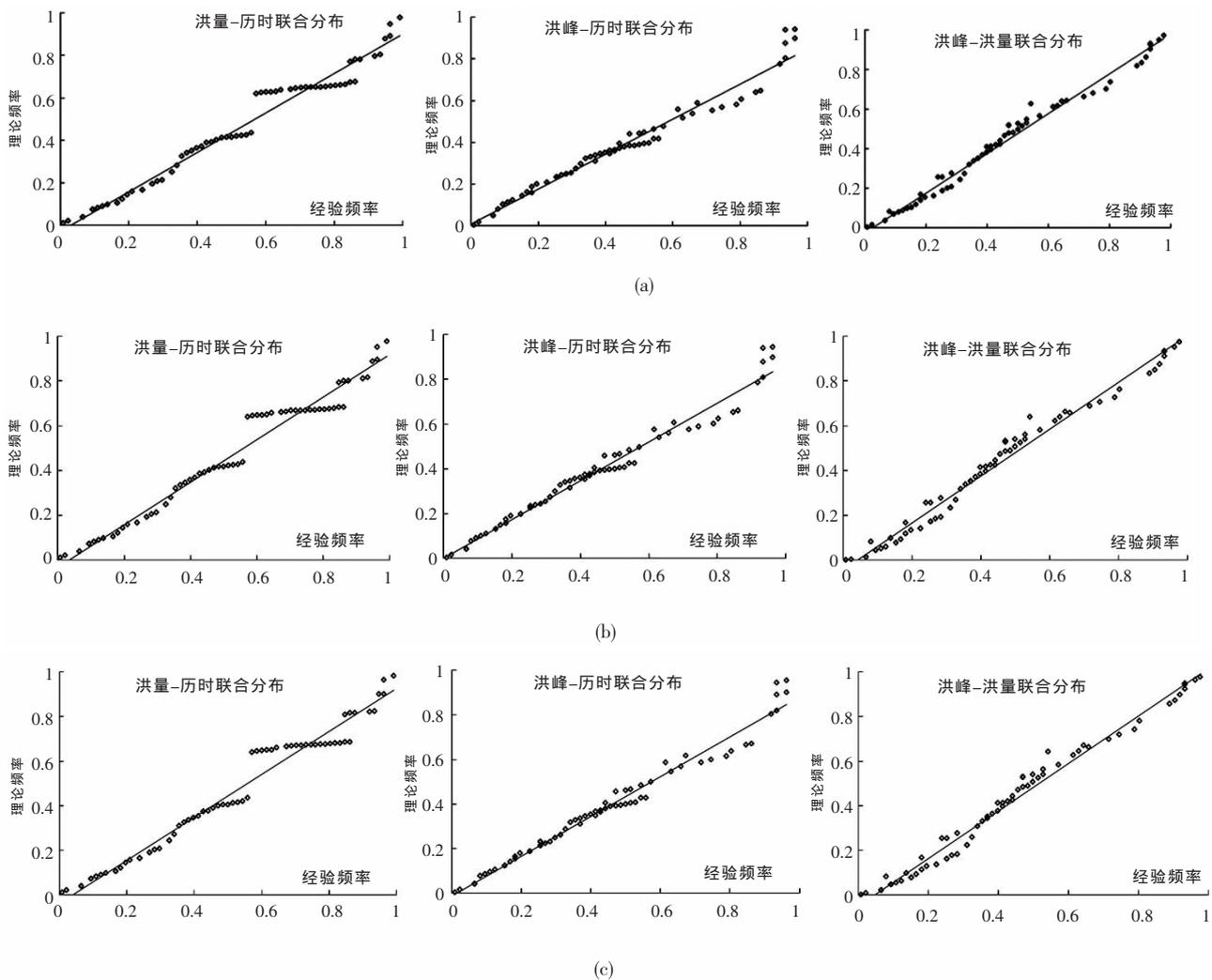


图2 Copula 函数理论与经验点据拟合图 (a.Clayton Copula ;b.Frank Copula; c.Gumbel Copula)

Fig.2 The comparison of the theoretical results and experiential results

表5 AIC信息准则和OLS准则拟合优度评价结果  
Table5 The evaluation results using AIC or OLS

Copula 类型	洪量-历时			洪峰-历时			洪峰-洪量		
	OSL	MSE	AIC	OSL	MSE	AIC	OSL	MSE	AIC
Clayton	0.08	0.01	-347.96	0.08	0.01	-339.55	0.04	0.01	-441.90
Frank	0.07	0.01	-356.95	0.08	0.01	-354.11	0.05	0.01	-421.04
Gumbel	0.07	0.01	-355.60	0.07	0.01	-357.30	0.05	0.01	-419.22

由表 5 可以看出:Frank Copula 函数拟合历时与洪量效果最好,Gumbel Copula 函数拟合洪峰与历时效果最好,Clayton Copula 拟合洪峰和洪量效果最好,与 Genest-Rivist 检验法判断结果一致。

3.5 组合变量分布函数与单变量分布函数对比分析

3.5.1 同场洪水重现期结果对比

利用拟合优度判别标准可得到拟合洪水各个特征

变量联合分布最优的 copula 概率分布函数,由拟合最优的 Copula 联合概率分布函数可推算出不同频率下洪水特征属性值,从而可绘制相应的联合概率分布图(见图 3),由式(7)和式(8)可求得洪峰、洪量及洪水历时任意两变量组合下的两种重现期,绘出其等值线图(见图 4),根据等值线图可得到任意两变量组合下的重现期。

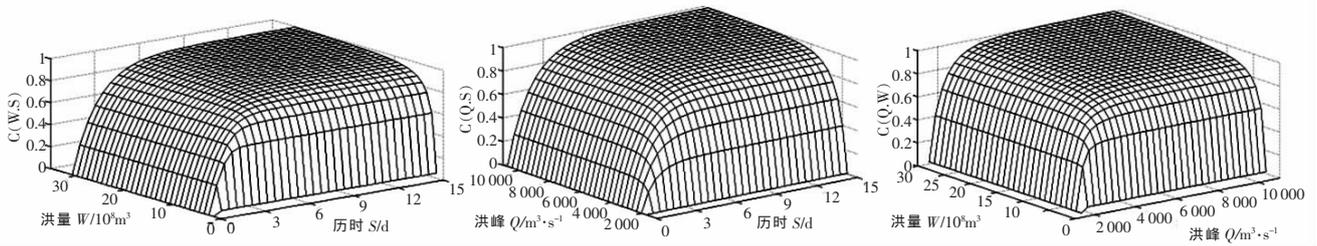


图 3 两变量联合概率分布图(洪量-历时 Frank Copula 分布;洪峰-历时 Gumbel 分布;洪峰-洪量 Clayton Copula 分布)

Fig.3 The distribution of the probability joint two elements

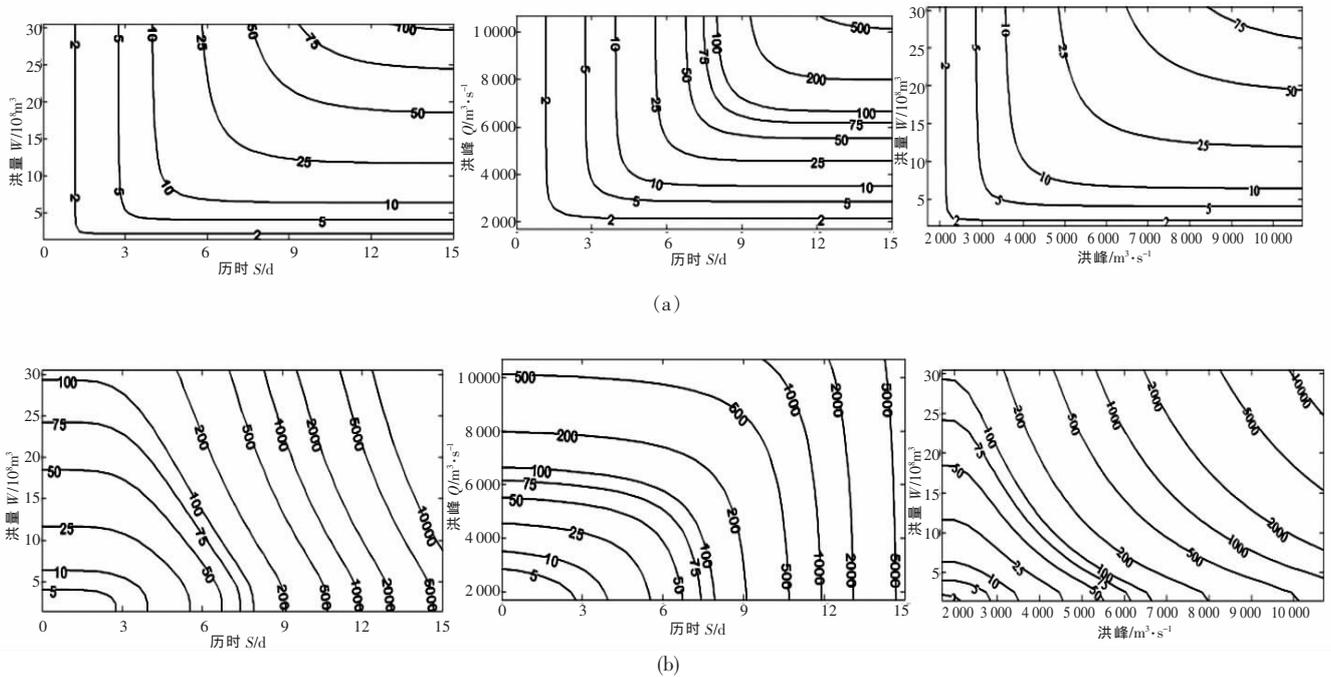


图 4 两维联合分布重现期等值线平面图(a.  $T(x|y)$ ; b.  $T(xy)$ )

Fig.4 Contour maps of the recurrence interval with combined flood elements

由于组合洪水变量与单一洪水变量采用不同的拟合分布函数,同场洪水(洪峰、洪量和洪水历时均相同)的再现期结果也不尽相同。以2006年特大洪水为例,年最大洪峰流量为 $7860\text{m}^3/\text{s}$ ,累积洪量为 $17.66\times 10^8\text{m}^3$ ,历时4d,由图3和图4可得该场次洪水各要素联合分布概率及再现期如表6所示。

表6 2006年场次洪水各特征值联合分布再现期  
Table6 The recurrence intervals of the combined flood elements from the selected floods in 2006

变量	$T(x/y)$	$T(xy)$	变量	单变量 $T$
洪量-历时	10	65	洪峰	190
洪峰-洪量	38	1884	洪量	50
洪峰-历时	11	216	历时	10

如表6所示,当采用单一变量拟合分布函数时,该场次洪水洪峰流量的再现期为190a,洪量的再现期为50a,洪水历时的再现期为10a;当采用组合变量分布函数时,该场次洪水洪峰和洪量的同现再现期( $T(xy)$ )(即相同大小的洪峰和洪量同时出现)达1884a,

联合再现期( $T(x/y)$ )(即相同大小的洪峰或洪量任一出现)为38a。

整体来讲,组合变量的联合再现期( $T(x/y)$ )小于各单变量再现期,而其对应的同现再现期( $T(xy)$ )则大于各单变量再现期。该再现期结果具有合理性:以洪峰-洪量组合变量为例,该事件的联合再现期属于洪峰、洪量两个概率事件的并集,涵盖的事件数量大于任一单独事件,所以其出现的概率较单独一个事件出现概率要大,使得其再现期变小;而同现再现期为这两个事件的交集,包括的事件数量要小于任一单个事件,所以出现概率比单个事件小,即再现期更大。

### 3.5.2 同再现期洪水参数对比

进一步比较同一再现期下,单变量设计值与组合变量联合分布条件设计值的不同。以洪峰-洪量联合概率分布为例,统计分析同一再现期下,单变量设计洪水特征属性值与两维联合变量设计值的区别。由再现期计算公式可反推出某一再现期下洪峰与洪量联合分布概率如表7所示:

表7 单变量设计值与峰量联合分布下设计值  
Table7 The design values under the single element estimate or combine elements estimate

$T$	$C/\%$	单变量设计值		联合分布设计值	
		$Q/\text{m}^3\cdot\text{s}^{-1}$	$W/10^8\text{m}^3$	$Q/\text{m}^3\cdot\text{s}^{-1}$	$W/10^8\text{m}^3$
2000	99.95	14374.41	217.13	18434.69	295.38
1000	99.90	12070.27	136.57	16393.88	172.93
500	99.80	10116.26	85.92	15169.39	99.46
200	99.50	7981.24	46.58	9455.10	74.97
100	99.00	6648.56	29.34	7822.45	50.48
50	98.00	5518.39	18.49	6189.80	38.23
20	95.00	4559.95	11.67	4557.14	25.99
10	90.00	3512.72	6.38	3740.82	13.74

如表7所示,同样是百年一遇洪水,单变量分布函数给出的洪峰和洪量设计值分别为 $6648.56\text{m}^3/\text{s}$ 和 $29.34\times 10^8\text{m}^3$ ,而组合变量分布函数的设计值为 $7822.45\text{m}^3/\text{s}$ 和 $50.48\times 10^8\text{m}^3$ ,即在同一再现期下,两变量联合分布法推求的洪水各特征设计值较单变量设计值偏大。该计算结果与其他地区相关研究结果一致<sup>[13]</sup>。

因为组合变量分布函数给出的洪水设计结果较单变量结果偏大,所以就水利工程的安全性角度而言,组合变量分布函数给出的洪水设计结果更加安全,按照该设计结果进行防洪设计可以减少洪涝灾害发生风险。在实际生产应用中,则需要根据实地条件和资金状况等因素进行综合考虑。

## 4 结语

以武江流域犁市站60年的水文资料为例,利用Archimedean Copula函数中的3种函数建立了具有较高相依性关系的洪峰、洪量和洪水历时的联合分布,其中洪峰、洪量边缘分布为厚尾性的GDP分布,洪水历时为指数分布;利用指标相关性方法确定了各种类型Copula函数的参数,通过拟合优度检验各变量之间最优Copula联合分布函数,得到了洪量-历时Frank Copula拟合效果最好,洪峰-历时Gumbel Copula函数拟合效果最好,洪峰-洪量Clayton Copula函数拟合效果最好,利用最优的联合概率分布函数绘制了洪水各

特征属性组合下的联合分布图及两种重现期的等值线图,比较分析了在单变量和组合变量联合分布同频率条件下,各不同重现期时,单变量及洪量洪峰联合分特征属性组合下的联合分布图及两种重现期的等值线图,比较分析了在单变量和组合变量联合分布同频率条件下,各不同重现期时,单变量及洪量洪峰联合分布设计值,结果显示联合分布时的设计值大于单变量设计值,即利用联合分布设计值对水利工程偏于安全。由此可见,基于 Copula 函数的两变量联合分布的洪水频率分析方法能更好地描述洪水特征量之间的关系,且对边缘分布类型没有限制,为洪水频率分析计算提供了一种较优算法。

参考文献:

- [1] 冯平,崔广涛,胡明昱. 暴雨洪水共同作用下的多变量防洪计算问题[J]. 水利学报, 2000,41(4):9-53.(FENG Ping, CUI Guangtao, HU Ming-gang. A bivariate method of rainstorm and flood for design flood and its application [J]. Journal of Hydraulic Engineering, 2000,41(4):9-53. (in Chinese))
- [2] Favre A C, Adlouni S E, Perrault L, et al. Multivariate hydrological frequency analysis using copulas [J]. Water Resource Research, 2004,40:1-12.
- [3] 陆桂华,闫桂霞,吴志勇,等. 基于 Copula 函数的区域干旱分析方法[J]. 水科学进展, 2010,21(2):188-193. (LU Guihua, YAN Guixia, WU Zhiyong, et al. Regional drought analysis approach based on Copula function[J]. Advances in Water Science, 2010,21(2):188-193.(in Chinese)).
- [4] Grimaldi S, Serinresco F. Asymmetric copula in multivariate flood frequency analysis [J]. Advances in Water Resources. 2006,29(8): 150-164.
- [5] Zhang L,Singh V P. Bivariate flood frequency analysis using the copula method [J]. Journal of Hydrologic Engineering, 2006,11(2):150-164.
- [6] 方彬,郭生练,肖义,等. 年最大洪水两变量联合分布研究[J]. 水科学进展, 2008,19(4):505-511. (FANG Bin, GUO Shenglian, XIAO Yi, et al. Annual maximum flood occurrence dates and magnitudes frequency analysis based on bivariate joint distribution[J]. Advances in Water Science, 2008,19(4):505-511. (in Chinese))
- [7] 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 Changjun, LIANG Zhongmin, LUAN Chengmei, et al. Advance in flood frequency analysis for partial duration series [J]. Advances in Water Science, 2006,17(1): 136-140. (in Chinese))
- [8] 叶长青,陈晓宏,张丽娟,等. 变化环境下武江超定量洪水门限值响应规律及影响[J]. 水科学进展, 2013,24(3):392-401. (YE Changqing, CHEN Xiaohong, ZHANG Lijuan, et al. Changing rule and influence of threshold for peak-over-threshold flood series under the changing environment in Wujiang, China[J]. Advances in Water Science, 2013,24(3):392-401. (in Chinese))
- [9] 陈子燊,刘曾美,路剑飞. 基于广义 Pareto 分布的洪水频率分析[J]. 水力发电学报, 2013,32(2):68-83. (CHEN Zishen, LIU Zengmei, LU Jianfei. Analysis of flood frequency based on generalized Pareto distribution[J]. Journal of Hydroelectric Engineering, 2013,32(2):68-83. (in Chinese))
- [10] 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. (DU Jiang, CHEN Xizhen, YU Bo. Parameter estimation of archimedean Copula[J]. Science Technology and Engineering, 2009,(3):637-640. (in Chinese))
- [11] Cheng Wang, Ni-Bin Chang, Gour-Tsyh Yeh. Copula-based flood frequency analysis at the confluences of river system[J]. Hydrological Processes, 2009,(2):7273-7288.
- [12] 谢华,罗强,黄介生. 基于三维 Copula 函数的多水文区丰枯遭遇分析[J]. 水科学进展, 2012,23(2):86-73.(XIE Hua, LUO Qiang, HUANG Jiesheng. Synchronous asynchronous encounter analysis of multiple hydrologic regions based on 3D Copula function [J]. Advances in Water Science, 2012,23(2):86-73. (in Chinese))
- [13] Favre A C, Adlouni S E, Perrault L, et al. Multivariate hydrological frequency analysis using copulas [J]. Water Resources Research, 2004,40:1-12.

## Studying Probability Distribution of Flood Elements Combination Based on Copula

LIN Xian<sup>1,2</sup>, OUYANG Hao<sup>3</sup>, CHEN Xiaohong<sup>1</sup>, LI Shenlin<sup>1</sup>

(1. Center for Water Resources and Environment, Sun Yat-sen University, Guangzhou 510275, China; 2. Water Resources Department of Guangdong Province, Guangzhou 510000, China; 3. Pearl River Water Resources Protection Bureau; Guangzhou 510000, China)

**Abstract:** This study constructed a probability distribution model based on Copula to evaluate the probability of flood elements combinations, including peak and volume combination, volume and duration combination, and peak and duration combination. These combinations probability distribution maps and isograms were plotted. Comparison of the same return period flood designed from single elements and combination elements probability distribution was also made in this study. The results show that generalized GDP distribution fits flood peak and volume well, while exponent distribution fits flood duration suitably. Based on these two primary distributions, the Copula combination probability distribution fits flood elements combination appropriately, and generates a more secure design flood than traditional single element model. This study proves that the combination probability distribution model describes flood feature available, and is meaningful to evaluate flood risk accurately.

**Key words:** Copula; elements combination probability distribution model; GDP distribution; frequency analysis; Wujiang River Basin