

波高与风速联合概率分布研究

陈子燊

(中山大学 水资源与环境系, 广东 广州 510275)

摘要: 基于 copula 函数论述了两变量的联合概率分布方法。此方法的主要优点是边缘分布可由不同的分布函数构成, 变量间可具相关性。以粤东汕尾海域极值波高与相应风速为研究实例, 经分析获得以下结果: (1) 优选的极值波高和风速可分别由 P-III 型和 GEV 分布表示; (2) 拟合优度检验指标表明二者的最优连接函数为 Archimedean copula 类的 Gumbel-Hougaard copula; (3) 与联合分布比较, 重现期介于 5—200 年之间的波高边缘分布设计值的相对差值大约介于 3.1% ~ 8.1% 之间, 风速设计值相对差值大约介于 2.8% ~ 6.4% 之间; (4) 特定风速设计频率条件下, 波高与风速的遭遇概率随波高设计频率的减小而减小, 特定波高设计频率随风速条件频率的减小, 二者的遭遇概率随之增大。

关键词: 极值波高和风速; 边缘分布; 联合概率分布; Copula 函数; 条件分布

中图分类号: P731.22

文献标识码: A

文章编号: 1001-6932(2011)02-0158-06

Study on joint probability distribution of wave height and wind velocity

CHEN Zi-shen

(Department of Water Resource and Environment, Zhongshan University, Guangzhou 510275, China)

Abstract: This article introduced the method of bivariate joint probability distribution based on the copula function. A major advantage of this method is that marginal distributions of individual variables can be of any form and the variables can be correlated. Some of conclusions were reached by using extreme wave height and wind speed as an example collected in Shanwei sea area as the following: (1) Optimized marginal distributions of wave height and wind velocity can be represented by the Pearson pattern three and generalized extreme value distribution, respectively; (2) Gumbel-Hougaard Copula that belongs to Archimedean copula family was the optimal copula selected by the goodness-of-fit test; (3) The relative differences of the special frequency design values between the marginal distribution of wave height and the joint distribution fall in between 3.1% and 8.1% for the return periods between 5 and 200 years, and the relative differences of wind velocity fall in between 2.8%~6.4%; (4) The encountering probabilities of wave height given wind velocity decrease along with decreasing the frequency of wave height; whereas the encountering probability increase while the specific wave height frequency is along with decrease of wind velocity frequency.

Keywords: extreme wave height and wind velocity; marginal probability distribution; joint probability distribution; copula function; conditional probability distribution

全球暖化可能导致出现极端气象事件的频率和强度加大, 甚则常态化, 此对于海岸和海洋工程安全构成极大威胁。众知, 风浪的形成与发展主要是由风直接作用下的海洋动力过程。大浪与极端风速密切相关, 也都是造成海岸工程损毁和海岸侵蚀的重要动力因子。深入分析极值波高和风速的联合概率分布对海岸工程与海岸防护的规划设计和风险控制十分重要。

至今已有一些多变量联合概率分布应用于海洋水文气象要素的研究。如, 周道成应用耿贝尔逻辑

模型描述了年极值风速和有效波高两随机变量的联合分布^[1]; 董胜采用边缘分布为皮尔逊III型的联合分布估计了年极值有效波高与风速的联合出现概率^[2]; 潘锦娥使用对数正态分布探讨了波高与周期的联合分布^[3]。然而, 采用传统的多变量联合概率分布时存在一定的约束条件, 其要求各个边缘分布函数类型与多变量分布函数类型一样, 而且各个边缘分布必须完全相同。而且, 不同的水文气象要素或不同空间位置的同一水文气象要素通常并不服从同一分布, 水文气象要素的正态分布也难以

收稿日期: 2010-08-30; 收修改稿日期: 2010-09-20

基金项目: 国家自然科学基金资助项目(40576041)。

作者简介: 陈子燊, (1952—), 男, 福建福州人, 教授, 研究方向: 河口海岸环境。电子邮箱: eesczs@mail.sysu.edu.cn。

满足。这使得传统的多变量分布函数的概率分析受到较大限制。近十年来, 解决多变量联合概率分布的一个强有力工具的 Copula 函数已被广泛应用于金融工程领域, 国内外在陆地水文气象领域也得以较多应用^[4-12], 但国内在海洋水文气象领域的研究还仅见于参考文献^[13]。本文将在论述 Copula 函数主要理论与方法基础上, 应用实例构造二变量水文气象要素的联合概率分布模式, 深入分析水文气象变量间的概率分布主要特征。

1 基于 Copula 函数的联合概率分布模型

1.1 Copula 函数、参数估计和拟合优度评价

Copula 的数学涵义是指把多个变量的联合分布与它们的边缘分布连接在一起的函数。根据奠定 Copula 函数理论基础的 Sklar 定理, 令 F 是具有单变量边缘分布函数 $F_1, \dots, F_n$ 的 n 维分布函数, 若边缘分布函数 $F_1, \dots, F_n$ 连续, 则存在一个唯一满足 $F(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n))$ 关系的连接函数 C 。相反, 如果 C 是一个 n 维 Copula 函数, $F(x_1, \dots, x_n)$ 为 n 维分布函数。作为一新的概率分布和统计推断方法, Copula 函数为描述多变量随机变量之间的相关结构提供了一个新技术手段而得到了广泛应用, Nelsen^[14]曾系统地总结了 Copula 函数的性质和这个领域的主要研究成果。

根据 Copula 函数性质, 构建两变量的联合概率分布模型可分两步进行: 首先分别确定边缘分布 $F_x(x)$ 和 $F_y(y)$, 然后选择一个能够恰当地反映变量间相关结构的 Copula 函数: $C(F_x(x), F_y(y)) = C(u, v)$ 。式中, $u = F_x(x)$, $v = F_y(y)$ 。在三大类 Copula 函数——椭圆型、阿基米德型(Archimedean)和二次型中, 阿基米德(Archimedean) Copula 是由其生成元唯一确定的单参数函数, 是当前应用于水文气象领域的一类非常重要的 Copula 函数。几种常用的二维 Archimedean Copula 函数及其适用性如下:

①Gumbel-Hougaard(GH) Copula :

$$C(u, v) = \exp\{ -[(-\ln u)^\theta + (-\ln v)^\theta]^{1/\theta} \}, \theta \in [1, \infty)$$
(1)

其生成元 $\varphi(t)$ 为: $\varphi(t) = (-\ln t)^\theta$ 。GH Copula 函数仅适用于变量存在正相关的情形, 主要刻画随机变量间的上尾相关性。

②Clayton Copula:

$$C(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}, \theta \in (0, \infty)$$
(2)

其生成元 $\varphi(t)$ 为: $\varphi(t) = (t^{-\theta} - 1)/\theta$ 。同 GH 函数, 适用于描述正相关的随机变量, 主要用来描述联合分布中随机变量间的下尾相关性。

③Ali-Mikhail-Haq (AMH) Copula:

$$C(u, v) = uv / [1 - \theta(1-u)(1-v)], \theta \in [-1, 1)$$
(3)

其生成元 $\varphi(t)$ 为: $\varphi(t) = \ln(1 - \theta(1-t)/t)$ 。AMH 函数能够描述正或负相关的随机变量, 但不适用于正相关或负相关性很高的变量, AMH Copula 结构呈对称性。

④Frank Copula 函数:

$$C(u, v) = -\ln\{1 + [(e^{-\theta u} - 1)(e^{-\theta v} - 1)/(e^{-\theta} - 1)]\} / \theta, \theta \in R$$
(4)

其生成元 $\varphi(t)$ 为: $\varphi(t) = -\ln[(e^{-\theta t} - 1)/(e^{-\theta} - 1)]$ 。

与 AMH Copula 函数类似, 但对相关性程度没有限制。Frank Copula 结构具有对称性, 即在其分布的上尾和下尾, 变量间的相关性呈对称增长。

Copula 参数 θ 的估计方法大致可分为3种: ①相关性指标法: 根据 Kendall 秩相关系数与 θ 的关系间接求得; ②适线法: 在一定适线准则下, 求解与经验点据拟合最优的频率曲线的统计参数; ③极大似然法。由于不同的 Copula 函数代表不同的相关结构, 对 Copula 函数的选择将直接影响到一些分析和统计推断的结果^[15], 因此选择合适的 Copula 函数显得十分重要。择优选用 Copula 函数的主要检验方法有由图示直观选择 Copula 函数的 Genest-Rivest 方法^[16]、均方根误差(RMSE)准则法和 AIC 信息准则法等。

1.2 联合概率分布及其重现期

根据 Copula 函数, 两变量联合分布可表示为:

$$F(x, y) = P(X \leq x, Y \leq y) = C(F_x(x), F_y(y))$$
(5)

式中, $F_x(x), F_y(y)$ 分别代表变量 X 和 Y 的边缘分布。变量 X 和 Y 的重现期为:

$$T(x) = \frac{1}{1 - P(X \leq x)} = \frac{1}{1 - F_x(x)}, \quad T(y) = \frac{1}{1 - P(Y \leq y)} = \frac{1}{1 - F_y(y)}$$
(6)

使用“或”符号“ $\vee$ ”, 以及“和”符号“ $\wedge$ ”, 两个单变量事件理论上可以组合成八种两变量事件。

对于极端水文气象事件, 关注的主要是以下两种两变量事件:

$E_{x,y}^{\vee} = \{X > x\} \vee \{Y > y\}$, $E_{x,y}^{\wedge} = \{X > x\} \wedge \{Y > y\}$
 式中: $E_{x,y}^{\vee}$ 表示事件中水文变量 X 或 Y 超过某一特定值; $E_{x,y}^{\wedge}$ 表示事件中 X 和 Y 都超过某一特定值。当 X 或 Y 均为年最大值时, $E_{x,y}^{\vee}$ 和 $E_{x,y}^{\wedge}$ 事件的重现期分别为联合重现期 T_o 和同现重现期 T_a :

$$T_o(x, y) = \frac{I}{P[X \leq x \vee Y \leq y]} = \frac{I}{1 - F(x, y)};$$

$$T_a(x, y) = \frac{I}{P(X \leq x \wedge Y \leq y)} = \frac{I}{1 - F_x(x) - F_y(y) + F(x, y)} \quad (7)$$

由以上可得:

$$T_o(x, y) \leq \min(T(x), T(y)) \leq \max(T(x), T(y)) \leq T_a(x, y)$$

对连续性随机变量, 通常需分析随机变量落在一个区间内的概率, 由概率论可导出以下四个条件概率:

$$P(X \geq x | Y \geq y) = \frac{P(X \geq x, Y \geq y)}{P(Y \geq y)} = \frac{1 - F_x(x) - F_y(y) + F(x, y)}{1 - F_y(y)} \quad (8)$$

$$P(X \leq x | Y \geq y) = \frac{P(X \leq x, Y \geq y)}{P(Y \geq y)} = \frac{F_y(y) - F(x, y)}{1 - F_y(y)} \quad (9)$$

$$P(X \geq x | Y \leq y) = \frac{P(X \geq x, Y \leq y)}{P(Y \leq y)} = \frac{F_x(x) - F(x, y)}{F_y(y)} \quad (10)$$

$$P(X \leq x | Y \leq y) = \frac{P(X \leq x, Y \leq y)}{P(Y \leq y)} = \frac{F(x, y)}{F_y(y)} \quad (11)$$

同理, 还可定义给定随机变量 $X \geq x$ 和 $X \leq x$ 条件下的随机变量 Y 的 4 个条件概率。

表 1 边缘分布参数与优度检验值

Tab. 1 Parameters of the marginal distribution and the values of goodness of fit test

参数	分布	形态参数	尺度参数	位置参数	Cv	Cs	RMSE	Q	PPCC
波高/m	P-III	1.141	0.509	2.388	0.453	1.873	0.398	0.047	0.979
	GEV	0.181	1.225	3.667	/	/	0.260	0.033	0.977
风速/m.s ⁻¹	P-III	2.947	0.200	5.346	0.427	1.165	1.421	0.026	0.985
	GEV	0.009	6.385	16.324	/	/	0.597	0.017	0.987

2.2.2 Copula 函数的参数估计及拟合优度评价
 样本波高与风速的 Kendall 秩相关系数 (τ) 计算值为 0.629, 说明二者具有较高的相关性。采用相关性指标法计算了样本的联合概率分布的 copula 参数 θ , 并利用 AIC、RMSE 检验其拟合优度, 结果见表 2。根据拟合优度评价指标, 选择样本中 AIC 和

2 实例研究

2.1 基本数据

基本数据采用位于粤东汕尾海域的汕尾海洋观测站 1972—1992 年观测的历年热带气旋影响期间 E 向浪的最大波高与相应的极端风速。样本的基本统计特征值表明, 最大风速为 45 m/s, 达到强台风等级, 最大波高为 9.50 m。极值波高及其相应的极值风速的偏态系数分别为 1.21 和 1.10, 都属于正偏分布。

2.2 联合分布计算与分析

2.2.1 边缘分布函数 对两样本的边缘分布, 都使用了以下两个三参数的概率分布函数择优:

a、广义极值分布(GEV):

$$F_x(x) = P(X < x) = \exp\left\{-\left[1 - \xi\left(\frac{x - \mu}{\sigma}\right)\right]^{1/\xi}\right\}, \xi \neq 0 \quad (11)$$

b、皮尔逊三型分布 (P-III):

$$F_x(x) = P(X < x) = \frac{\beta^\alpha}{\Gamma(\alpha)} \int_{-\infty}^x (x - \mu)^{\alpha-1} \exp[-\beta(x - \mu)] dx \quad (12)$$

式中, $\xi, \beta, \mu, \alpha, \sigma$ 为分布函数参数。参数估计使用统计稳健性良好的线性矩 (L-矩) 方法。样本的经验频率分布 P_i 使用 Gringorten 公式计算: $P_i = (i - 0.44) / (n + 0.12)$ 。拟合结果采用均方根误差 (RMSE)、经验频率和理论频率拟合误差平方和 (Q) 和概率点据相关系数 (PPCC) 检验其拟合优度。根据优度检验结果 (表 1) 综合比较, 分别选用 P-III 分布和 GEV 分布为极值波高和极值风速的较优边缘分布 (图 1)。

RMSE 最小、 K_T - K_e 关系图中点据和理论直线最接近 45° 对角线的 GH Copula 函数作为联合概率分布的连接函数 (为节省篇幅, K_T - K_e 关系图略), 由此构建的极值波高和极值风速的 GH Copula 联合分布函数如下:

$$C_I(F_H(h),F_V(v))=exp\{-[(-\ln F_H(h))^{2.692}+(-\ln F_V(v))^{2.692}]^{1/2.692}\}$$

(13)

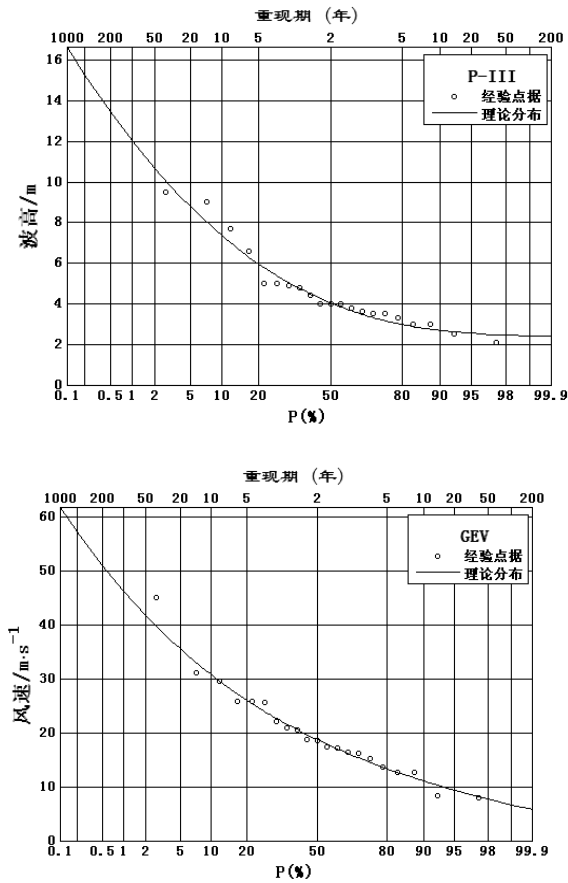


图 1 两变量边缘分布图：波高（左图），风速（右图）

Fig. 1 Diagrams of marginal distribution of wave heights (left panel) and wind velocities (right panel) in the sample one.

表 2 样本的 Copula 参数及拟合优度评价指标

Tab. 2 Parameters of Copula functions and the evaluation indices of goodness of fit for the sample

Copula 函数	AMH	GH	Clayton	Frank
θ	0.8252	2.6923	3.3846	8.7456
AIC	-49.119	-52.078	-48.142	-39.723
RMSE	0.0776	0.0669	0.0815	0.1242

2.2.3 联合概率分布与重现期 极值波高和极值风速的联合概率分布、联合重现期和同现重现期见图 2 所示，不同重现期的设计波高和设计风速见表 3。显然，波高风速联合重现期小于边缘分布重现期，边缘分布重现期则小于同现重现期。按照同频率原理，即 $T_0(h,v)=T$ ，单变量推算的设计值小于波高风速联合重现期设计值，差值随重现期减小而有所增大，重现期 5-200 年的波高设计值相对差

值大约介于 3.1%~8.1%，风速设计值相对差值大约介于 2.8%~6.4%。以 100 年一遇波高为例，波高、风速边缘分布设计值分别为 12.05 m 和 46.32 m/s，同频率的联合分布设计值分别为 12.57 m 和 48.04 m/s。单变量边缘分布 50 年一遇和 100 年一遇的同现重现期分别为 70 年和 141 年。上述结果表明，如以波高风速联合概率分布为设计值，单变量无论是波高或风速推算的设计值实际上没有达到设计标准。

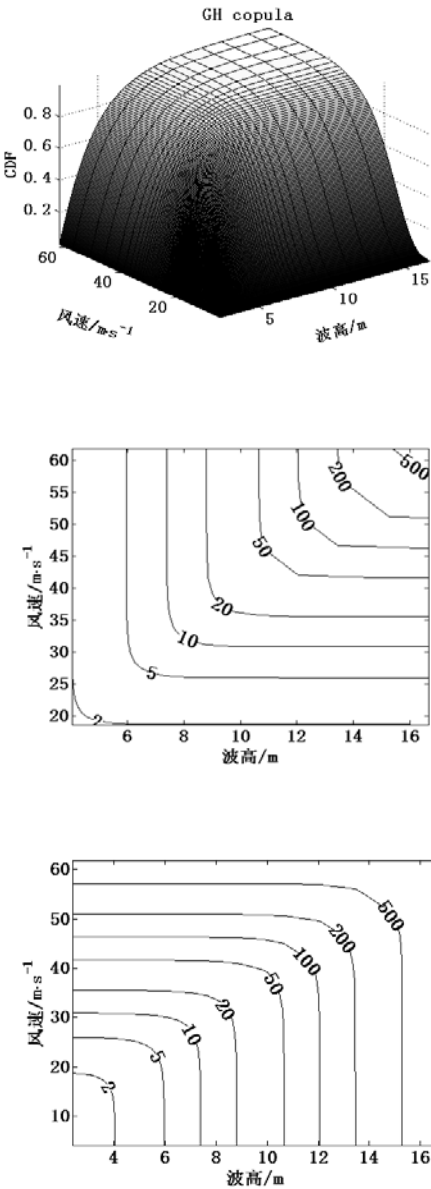


图 2 联合概率分布三维图（左图）、联合重现期(年)等值线图（中图）、同现重现期(年)等值线图（右图）

Fig. 2 Diagrams of joint probability distribution (left panel), joint return period (middle. panel) and occurrence period (right panel)

表 3 不同重现期设计波高和设计风速值
Tab.3 Designed values of different return periods of wave height and wind velocity

重现期/年			边缘分布设计值		$T_0(h,v)=T$ 设计值	
T	$T_0(h,v)$	$T_u(h,v)$	波高/m	风速/ $m.s^{-1}$	波高/m	风速/ $m.s^{-1}$
200	155	283	13.45	50.96	13.96	52.69
100	77	141	12.05	46.32	12.57	48.04
50	39	70	10.65	41.69	11.17	43.39
20	16	28	8.80	35.55	9.31	37.24
10	8	14	7.38	30.84	7.89	32.52
5	4	7	5.95	25.97	6.44	27.64

2.2.4 条件概率分布与遭遇分析 分别考虑条件概率 1： $P(H \geq h|V \geq v)$ 和条件概率 2： $P(H \geq h|V \leq v)$ 两种条件概率，其中，条件概率 1 指的是，当发生超过某一标准的风速 v 时，出现超过某一标准波高 h 的概率；而条件概率 2 的涵义则为发生不超过某一标准的风速 v 时，出现大于某一标准波高 h 的概率。二者的实质都是反映极值波高与相应风速的遭遇概率。两种不同条件概率的计算结果见表 4。表 4 显示，特定风速设计频率条件下，随波高设计频率的减小，二者的遭遇概率也随之减小；反之，特定波高设计频率随风速条件频率的减小，二者的遭遇概率随之增大。条件概率 1 表明，同频率下的极端波况和相应风况的遭遇概率不小于 70%。如：设计频率超过 20%的波高与设计频率

超过 10%风速的遭遇概率为 91.6%，而设计频率都为 10%(重现期 10 年)时，二者的遭遇概率为 72.6%。条件概率 2 则表明，当设计风速小于某一定值时，相对于条件概率 1，极端波况和相应风况的遭遇概率明显减小，同频率下极端波况和相应风况的遭遇概率都不大于 7%。如：设计频率超过 20%的波高与设计频率小于 10%风速的遭遇概率为 12.0%，而设计频率都为 10%（重现期 10 年）时，二者的遭遇概率为 3.0%。此对于如何在海岸工程选择与设计波高相匹配的设计风速，需要综合考虑遭遇概率小（风险小安全率高）和工程规模投入资金之间的平衡关系。因此需要在安全、资金投入与产出效益之间寻求平衡点，此需经过严格的经济分析与多方案综合比较才能确定。

表 4 两种条件概率计算结果*
Tab. 4 Computed results of conditional probability of two types

		V /m.s ⁻¹								V /m.s ⁻¹							
		26.0	30.8	35.5	41.7	46.3			26.0	30.8	35.5	41.7	46.3				
		H/m	P/%	20	10	5	2	1			H/m	P/%	20	10	5	2	1
条件概率 1		6.0	20	74.6	91.6	97.5	99.5	99.8	条件概率 2		6.0	20	6.3	12.0	15.9	18.4	19.2
P(H ≥ h V ≥ v)		7.4	10	45.8	72.6	90.3	97.9	99.4	P(H ≥ h V ≤ v)		7.4	10	1.1	3.0	5.8	8.2	9.1
I/%		8.8	5	24.4	45.2	71.6	92.8	97.8	I/%		8.8	5	0.2	0.5	1.5	3.2	4.1
		10.7	2	9.9	19.6	37.1	71.0	89.3			10.7	2	-	-	0.2	0.6	1.1
		12.1	1	5.0	9.9	19.5	44.6	70.8			12.1	1	-	-	-	0.1	0.3

* 表中的“-”表示不可能的遭遇概率。

3 结 语

本文以粤东汕尾海域的台风影响下的海况为实例，基于 Copula 理论与方法构建了波高和相应风速的联合概率分布模型。分析表明，两变量联合分布比单变量分布能更全面地反映极端海况条件下的重现水平，根据极端海况下的波高与风速的遭遇概率的风险分析可更好地为海岸工程规模的规划设计、资金投入和风险控制之间的均衡考虑提供科

学依据。

参考文献:

[1] 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.
[2] 董胜, 丛锦松, 余海静. 涠洲岛海域年极值风浪联合设计参数估计 [J]. 中国海洋大学学报, 2006, 36(3): 489-493.
[3] 潘锦娥. 显著(或有效)波高与平均周期的长期联合分布探讨 [J]. 海洋通报, 1989, 8(3): 1-10.
[4] 刘曾美, 陈子葵. 区间暴雨和外江洪水水位遭遇组合的风险 [J].

- 水科学进展, 2009, 20(5): 619- 625.
- [5] 闫宝伟, 郭生练, 肖义, 等. 基于两变量联合分布的干旱特征分析 [J]. 干旱区研究, 2007, 24(4): 537- 542.
- [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.
- [7] 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.
- [8] Zhang L, Singh V P. Bivariate rainfall frequency distributions using Archimedean copulas [J]. Journal of Hydrology, 2007, 332: 93-109.
- [9] Shiau J T, Song F, Nadarajah S. Assessment of hydrological droughts for the Yellow River, China, using copulas [J]. Hydrological Processes, 2007, 21: 2 157-2 163.
- [10] 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.
- [11] Joe H. Multivariate Models and Dependence Concepts [M]. Chapman & Hall, London, 1997.
- [12] 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, 28(3): 1-7.
- [13] 秦振江, 孙广华, 闫同新, 等. 基于 Copula 函数的联合概率法在海洋工程中的应用 [J]. 海洋预报, 2007, 24(2): 83-90.
- [14] Nelsen R B. An Introduction to Copulas. Springer Series in Statistics [M]. Springer, New York, 1999:216.
- [15] Embrechts P, Lindskog F, McNeil A. Modeling Dependence with Copulas and Applications to Risk Management Handbook of Heavy Tailed Distributions in Finance [M]. Elsevier Science, Publishers, New York, 2003: 329-384.
- [16] Genest C, Rivest L. Statistical inference procedures for bivariate Archimedean copulas [J]. Journal of American Statistical Association, 1993, 88:1 034-1 043.