

Research  
Watershed Ecology—Article

## 三峡库区香溪河流域多变量生态水文风险的不确定性分析

Yurui Fan<sup>a</sup>, Guohe Huang<sup>b,\*</sup>, Yin Zhang<sup>c</sup>, Yongping Li<sup>b</sup>

<sup>a</sup> College of Engineering, Design and Physical Sciences, Brunel University, London, Uxbridge, Middlesex, UB8 3PH, UK

<sup>b</sup> State Key Laboratory of Water Environment, School of Environment, Beijing Normal University, Beijing 100875, China

<sup>c</sup> College of Engineering and Mines, University of Alaska Fairbanks, Fairbanks, AK 99775, USA

### ARTICLE INFO

#### Article history:

Received 26 March 2018

Revised 25 May 2018

Accepted 4 June 2018

Available online 20 September 2018

#### 关键词

洪水风险

copula

多变量水文频率分析

概率分布

马尔可夫链蒙特卡罗

### 摘要

本研究基于 copula 函数开发了一种多变量生态水文风险评估框架，用于分析三峡库区香溪河流域极端生态水文事件的发生频率。通过马尔可夫链蒙特卡罗 (MCMC) 方法量化边缘分布及 copula 函数中参数的不确定性，并基于后验概率揭示联合重现期的内在不确定性，同时可进一步得到双变量及多变量风险的概率特征。研究结果显示所得概率模型的预测区间可很好地匹配观测值，尤其对洪水持续时间而言。同时，“AND”联合重现期的不确定性随着单个洪水变量重现期的增加而增加。此外，低设计流量及高服务年限可能导致高洪水风险且伴随大量不确定性。

© 2018 THE AUTHORS. Published by Elsevier LTD on behalf of Chinese Academy of Engineering and Higher Education Press Limited Company This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. 引言

由于气候变化及人为影响，极端水文事件的发生频率和强度日益增加，已经造成了巨大的经济损失，并给居民及其生活环境带来了极大的影响[1–5]。在许多生态水文问题的管理中，推断某一特定事件（如洪水、干旱和风暴）在不同发生概率下的量级至关重要[6–8]。例如，在水利工程建筑设计中，大坝泄洪道的设计需要量化洪水事件的大小及发生频率[9]。然而，许多水文事件难以通过单一的变量来描述，并且其通常显示多种特征。一个典型的例子就是，洪水通常以其峰值流量、洪量和持续时间为特征。此外，洪水的这3个属性是相

互关联的。然而，传统的单变量洪水风险分析主要集中在洪峰的概率特征上，这就难以提供洪水特征的完整描述（如对于洪峰、洪量和持续时间的依赖）。因此，多变量洪水频率分析对于深入描述极端水文事件以及推求可靠的风险预测是极其重要的[10,11]。

由于其描述相关变量间复杂相关关系的能力，copula 函数已被广泛应用于生态水文问题中。具体而言，copula 函数已被应用于许多洪水[10–13]、干旱[14,15]、暴雨[16,17]等灾害的风险分析问题中。此外，该方法还被用于径流模拟[18–21]及其他能源以及环境系统分析问题上[22–25]。与传统的多变量统计方法相比，copula 方法可以分别构建其边缘分布及相关性模型，因而可以使

\* Corresponding author.

E-mail address: [huang@iseis.org](mailto:huang@iseis.org)

用不同类型的边缘和联合概率函数[2,13,26,27]。因此,在洪水风险分析中,可以通过各种参数或非参数分布来量化洪峰、洪量和持续时间的分布。例如, Karmakar和 Simonovic[2]研究了边缘分布对基于copula函数多变量水文风险分析的影响。

水文风险分析的一个重要的问题是由于数据可用性、模型选择以及参数估计所导致的大量不确定性。不确定性有两个主要来源:①源于水文过程中固有变异性的自然不确定性;②源于系统不能完全解析所导致的认知不确定性[28]。尤其,数据可用性是洪水频率分析不确定性的主要来源。以前的一些研究已经解决了洪水风险分析中的不确定性问题。例如, Liang等[29]基于贝叶斯理论分析了模型和参数不确定性。然而,前人的大多数研究主要集中于单变量的水文风险上,很少有研究涉及多变量水文风险的不确定性分析。

因此,本研究旨在通过对洪水中多重属性的互动分析,提出一个多变量风险评估框架,其中将运用贝叶斯方法对双变量水文风险的不确定性进行评估。通过马尔可夫链蒙特卡罗(MCMC)方法,量化边缘和联合分布参数的后验概率,进而揭示首次双变量重现期的内在不确定性。

## 2. 方法

### 2.1. Copula 概念

基于copula函数,可以构造一个多变量概率分布:

$$F(x_1, x_2, \dots, x_n) = C[F_{X_1}(x_1), F_{X_2}(x_2), \dots, F_{X_n}(x_n)] \quad (1)$$

式中,  $F_{X_1}(x_1), F_{X_2}(x_2), \dots, F_{X_n}(x_n)$ 表示随机变量( $X_1, X_2, \dots, X_n$ )的边缘分布。如果这些边缘分布是连续的,那么就存在一个copula函数 $C$ ,如下式所示[13]:

$$C(u_1, u_2, \dots, u_n) = F[F_{X_1}^{-1}(u_1), F_{X_2}^{-1}(u_2), \dots, F_{X_n}^{-1}(u_n)] \quad (2)$$

式中,  $u_i = F_{X_i}(x_i); i = 1, 2, \dots, n$ 。Nelsen [30]详细描述了copula的理论背景和性质。

当前已有许多copula函数可用于实际的多变量分析问题中,主要包括阿基米德、椭圆、极值copula函数。在本研究中将会使用阿基米德copula,因为其能够描述多种相关结构[31]。阿基米德copula可以采用多样的形式,此外,它们可以有不同的上尾和下尾相关特征。表1列出了在本研究中采用的copula函数及其固有特性。

### 2.2. 条件分布

基于由方程(1)表示的copula函数,可以推出每个 $X_i$ 的条件分布。以一个双变量分布为例,假设  $U_1 = F_{X_1}(x), U_2 = F_{X_2}(x)$ , 当  $U_2 = u_2$ 时,  $U_1$ 条件分布可被表示为[12]:

$$\begin{aligned} C_{U_1|U_2=u_2}(u_1) &= C(U_1 \leq u_1 | U_2 = u_2) \\ &= \frac{\partial}{\partial u_2} C(u_1, u_2) | U_2 = u_2 \end{aligned} \quad (3)$$

类似地,当  $U_2 \leq u_2$ 时,  $U_1$ 条件分布可被表示为

$$C_{U_1|U_2 \leq u_2}(u_1) = C(U_1 \leq u_1 | U_2 \leq u_2) = \frac{C(u_1, u_2)}{u_2} \quad (4)$$

此外,一个copula的概率密度函数(PDF) $c(u_1, u_2)$ 可表示为:

$$c(u_1, u_2) = \frac{\partial^2 C(u_1, u_2)}{\partial u_1 \partial u_2} \quad (5)$$

然后,  $X_1$ 和 $X_2$ 的联合PDF可以由 $c(u_1, u_2)$ 得出:

$$\begin{aligned} f(x_1, x_2) &= \frac{\partial^2 C(u_1, u_2)}{\partial x_1 \partial x_2} = \frac{\partial^2 C(u_1, u_2)}{\partial u_1 \partial u_2} \frac{\partial u_1}{\partial x_1} \frac{\partial u_2}{\partial x_2} \\ &= f_{X_1}(x_1) f_{X_2}(x_2) c(u_1, u_2) \end{aligned} \quad (6)$$

因此,对于给定的 $X_2$ ,  $X_1$ 的条件PDF表示为:

$$f(x_1|x_2) = \frac{f(x_1, x_2)}{f_{X_2}(x_2)} = f_{X_1}(x_1) c(u_1, u_2) \quad (7)$$

$$f(x_2|x_1) = \frac{f(x_1, x_2)}{f_{X_1}(x_1)} = f_{X_2}(x_2) c(u_1, u_2) \quad (8)$$

重现期已被广泛应用于单变量水文频率分析问题中。同理,可基于copula函数将重现期的概念扩展到多变量环境,从而形成了联合重现期的概念。具体地说,联合(首次)重现期可以进一步描述为“OR”和“AND”两种情况[13,32,33]:

$$T_{u_1, u_2}^{\text{OR}} = \frac{\mu}{1 - C_{U_1, U_2}(u_1, u_2)} \quad (9)$$

$$T_{u_1, u_2}^{\text{AND}} = \frac{\mu}{1 - u_1 - u_2 + C_{U_1, U_2}(u_1, u_2)} \quad (10)$$

式中,  $\mu$ 表明了两个连续的极端事件的平均到达时间。

### 2.3. 双变量水文风险分析

在实际的水利设施设计中,需要了解流经基础设施

表1 Copula的基本性质

Copula name	Functions $C_\theta(u_1, u_2)$	$\theta$ range	Generating functions $\phi(t)$	$\tau = 1 + 4\int_0^1(\phi(t)/\phi'(t))dt$
Ali-Mikhail-Haq	$u_1 u_2 / [1 - \theta(1 - u_1)(1 - u_2)]$	$[-1, 1)$	$\ln\{[1 - \theta(1 - t)]/t\}$	$(3\theta - 2)/\theta - 2(1 - \theta^{-1})^2 \ln(1 - \theta)/3$
Cook-Johnson	$[u_1^{-\theta} + u_2^{-\theta} - 1]^{-1/\theta}$	$[-1, \infty) \setminus \{0\}$	$t^{-\theta} - 1$	$\theta/(\theta + 2)$
Gumbel-Hougaard	$\exp\{-[\ln u_1]^\theta + [\ln u_2]^\theta\}$	$[1, \infty)$	$(-\ln t)^\theta$	$1 - \theta^{-1}$
Frank	$-\ln\{1 + (e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)/(e^{-\theta} - 1)\}/\theta$	$(-\infty, \infty) \setminus \{0\}$	$\ln[(e^{-\theta t} - 1)/(e^{-\theta} - 1)]$	$1 - 4[D_1(-\theta) - 1]/\theta$

$D_1$  is the first-order Debye function, and for any positive integer  $k$ ,  $D_k(x) = \frac{k}{x^k} \int_0^x \frac{t^k}{e^t - 1} dt$ .

的洪水的随机特征。此外，一个洪水事件通常具有多重属性，如洪峰、洪量和持续时间。因此，多变量洪水风险比单变量洪水风险能够提供更多、更有用的信息。例如，一场具有高洪峰值和较长持续时间（如超过7 d）的洪水可能会导致巨大的经济损失，而短期的高峰值洪水只会导致一场突如其来的山洪。因此，本研究基于“AND”联合重现期的构建了双变量风险指标以揭示洪水的多重属性之间的交互作用。该双变量风险指标可以表示为：

$$R_{u_1, u_2} = 1 - \left(1 - \frac{1}{T_{u_1, u_2}^{\text{AND}}}\right)^n \quad (11)$$

该双变量水文风险 $R_{u_1, u_2}$ 可研究多个交互式洪水变量影响下的洪水风险水平。例如，对于一个具特定设计流量和服务年限的水利设施，如果 $u_1$ 和 $u_2$ 表示洪峰和洪量， $R_{u_1, u_2}$ 表示洪峰值大于设计流量时，其洪量大于阈值风险。因此，在本研究中，我们采用“AND”联合重现期来表示双变量水文风险。

#### 2.4. 基于贝叶斯模型的参数估计

基于copula函数风险模型中蕴含大量的不确定性，其原因包括：①洪水过程中固有的不确定性；②边缘分布和copula函数选择的不确定性以及③参数估计过程中的统计不确定性或参数不确定性（如样本的可得性）。贝叶斯方法已被广泛应用于不确定性量化问题。其可通过贝叶斯定理将各种信息来源整合到一个单独的分析框架中。根据先验概率密度和观察结果，其参数后验分布可以通过贝叶斯定理得到：

$$\pi(\theta|\mathbf{X}) = \frac{L(\theta|\mathbf{X}) \pi_0(\theta)}{\int L(\theta|\mathbf{X}) \pi_0(\theta) d\theta} \quad (12)$$

式中， $\pi_0(\theta)$ 表示先验概率分布； $L(\theta|\mathbf{X})$ 表示似然函数； $\int L(\theta|\mathbf{X}) \pi_0(\theta) d\theta$ 为归一化常数， $\pi(\theta|\mathbf{X})$ 是后验概率密度函数。 $\mathbf{X} = (x_1, x_2, \dots, x_n)$  观察向量[在本研究中 $\mathbf{X} = (x_1, x_2)$ ]。

对于双变量水文风险分析，有必要对copula函数及

两个边缘概率分布的参数进行估计。假设 $\theta_c$ 、 $\theta_1$ 、 $\theta_2$ 分别表示copula的参数及两个边缘概率分布的未知分布。基于贝叶斯定理，其后验分布可表示为：

$$\begin{aligned} \pi(\theta_c, \theta_1, \theta_2|\mathbf{X}) &\propto L(\theta_c, \theta_1, \theta_2|\mathbf{X}) \pi_0(\theta_c, \theta_1, \theta_2) \\ &= L(\theta_c, \theta_1, \theta_2|\mathbf{X}) \pi_0(\theta_c) \pi_0(\theta_1) \pi_0(\theta_2) \end{aligned} \quad (13)$$

基于copula与其边缘分布之间的依赖结构， $L(\theta_c, \theta_1, \theta_2|\mathbf{X})$ 为似然函数，可以表示为：

$$\begin{aligned} L(\theta_c, \theta_1, \theta_2|\mathbf{Y}) &= c_{U_1, U_2}[F_{X_1}(x_1), F_{X_2}(x_2)|\theta_c] \\ & f_{X_1}(x_1|\theta_1) f_{X_2}(x_2|\theta_2) \end{aligned} \quad (14)$$

式中， $c_{U_1, U_2}$ 是copula函数的密度； $f_{X_1}$ 和 $f_{X_2}$ 分别是两个边缘PDF。

基于上述方法，本研究的多变量生态水文风险的具体计算过程如下。

**步骤1：**应用最大似然法估计边缘和copula中参数的初始值。

**步骤2：**计算边缘和联合概率的均方根误差（RMSE）和赤池信息量准则（AIC），以选择最合适的边缘和联合分布形式。

**步骤3：**根据步骤1中获得的参数值设置先验分布。

**步骤4：**使用MCMC与Metropolis-Hastings算法，推导出边缘和联合分布中参数的后验概率。

**步骤5：**评估首次和联合重现期的不确定性。

**步骤6：**揭示双变量水利风险的概率特征。

### 3. 应用

#### 3.1. 流域特征与数据

该方法被应用于中国三峡库区香溪河流域的水文风险分析。香溪河位于三峡库区湖北段。如图1所示，香溪河起源于神农架自然保护区，是长江的主要支流之一，干流长度为94 km，流域面积为3099 km<sup>2</sup> [34]。该区域为北亚热带气候，年平均降水量为1100 mm，波动范围为670~1700 mm[35]。香溪河流域汛期为7~8月，

主要降雨时节为5~9月。本研究利用兴山水文站（东经110°45'0"，北纬31°13'0"）观测到的日流量数据（1961—2010年），对香溪河洪水风险进行了概率评估。

### 3.2. 香溪河洪水特征

在本研究中，洪水特征是通过峰值、洪量和持续时间来确定的，其中洪峰是最大洪峰流量，而洪水量和持续时间是通过对最大流量的水文图得到的。根据Yue[36,37]的研究，可以通过洪水过程线的上升（图2中的 $D_{si}$ ）和下降（ $D_{ei}$ ）时刻来获得第 $i$ 年中单峰洪水过程（图2[31]）的持续时间（ $D_i$ ）。如图2所示，一次洪水的起始时间由水文图的急剧上升为标记，而洪水的终点由曲线[10]的后退支点上的拐点确定。如果洪水的上升时刻用 $D_s$ （d）表示，结束时间用 $D_e$ （d）表示，那么洪水事件的洪水量可用下式表示：

$$V_i = (V_i^{\text{total}} - V_i^{\text{baseflow}}) = \sum_{j=D_{si}}^{D_{ei}} q_{ij} - \frac{1}{2}(q_{si} + q_{ei})(1 + D_i) \quad (15)$$

式中， $q_{ij}$ 是第 $i$ 年第 $j$ 天的观测径流； $q_{si}$ 是第 $i$ 年洪水开始日的洪水流量； $q_{ei}$ 是洪水结束日的径流值。 $D_{si}$ 和 $D_{ei}$ 分别表示第 $i$ 年洪水的开始和结束日，能够由下式得到：

$$D_i = D_{ei} - D_{si}, \quad i = 1, 2, \dots, n \quad (16)$$

年洪水极值通过下式得到：

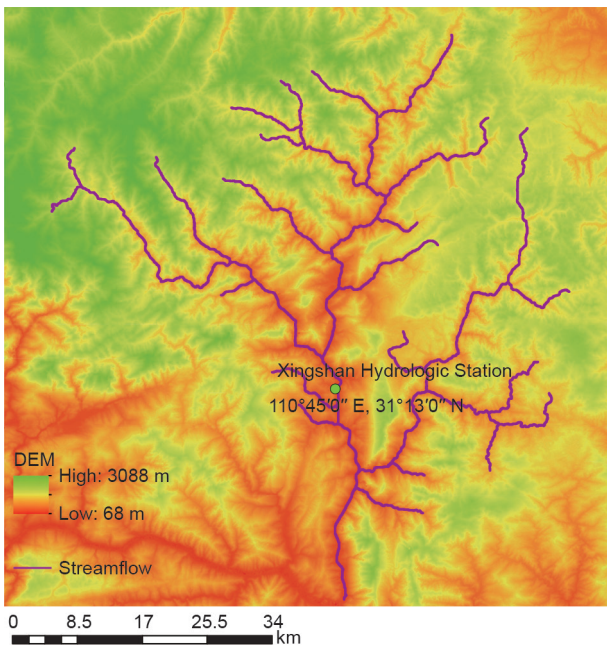


图1. 所研究流域的位置。DEM：数字高程模型。

$$q_i = \max\{q_{ij}, j = D_{si}, D_{si+1}, \dots, D_{ei}\}, \quad i = 1, 2, \dots, n \quad (17)$$

基于方程（15）~（17），洪水特征可以通过其峰值、洪量和持续时间来识别。表2显示了兴山站的洪峰、洪量和持续时间的统计特征。峰度和偏度都是正值，表明洪水变量可以用尖峰和右尾分布来模拟。

### 4. 边缘分布和联合分布的初始参数值

对于洪峰等单变量，许多参数化的分布函数可用于量化其概率特征，如广义极值分布（GEV）、对数Pearson Type III（LP3）和Pearson Type III（P3）[38,39]。在本研究中，Gamma、GEV和对数正态分布将被用来量化洪水洪峰、洪量和持续时间的概率特征，并基于最大似然估计（MLE）方法估计这些分布中的未知参数。表3给出了由MLE获得的分布函数的具体公式及未知参数的估计值。图3给出了理论分布和观测值之间的比较。

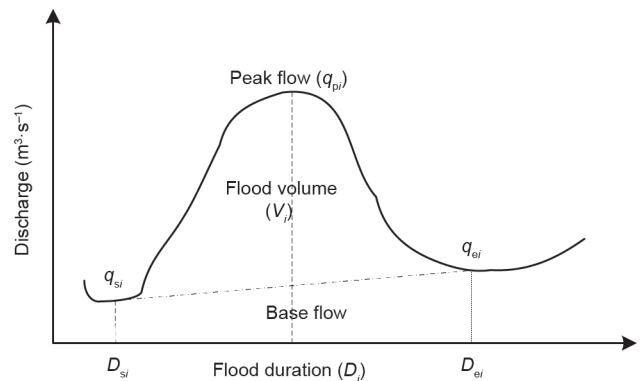


图2. 典型的洪水过程曲线图（改编自文献[31]）。

表2 洪水变量的统计特征

No.	Characteristics	Flood characteristics			
		Peak ( $\text{m}^3 \cdot \text{s}^{-1}$ )	Volume ( $\text{m}^3 \cdot \text{s}^{-1}$ )	Duration (d)	
1	Percentile	0	91	72	3
		25%	324	530.8	5
		50%	451.5	713.3	6
		75%	684	1189.5	8
		100%	1050	2430.8	13
2	Range	959	2358.8	10	
3	Mean	510	920.5	6.56	
4	Std	243.8	531.9	2.31	
5	Skewness	0.74	0.959	2.72	
6	Kurtosis	2.61	3.20	0.68	

应用RMSE和AIC准则来评价每个边缘分布的表现，如表4所示。结果表明，对数正态分布在香溪河3个洪水

变量中表现最好。

表5给出了3个洪水变量（即洪峰值、洪量和持续

表3 洪水变量的边缘分布函数

Name	Probability density function	Variables	Parameters values		
			Peak	Volume	Duration
Gamma	$x^{a-1} e^{-x/b} / (b^a \Gamma(a)), \Gamma(a) = \int_0^x u^{a-1} e^{-u} du$	$a$	4.50	3.06	8.62
		$b$	113.26	301.24	0.76
GEV	$\frac{1}{\sigma} \exp\left\{-\left[1+k(x-\mu)/\sigma\right]^{-1/k} \left[1+k(x-\mu)/\sigma\right]^{-1-1/k}\right\}$	$k$	0.032	0.099	0.073
		$\mu$	185.0	373.16	1.73
		$\sigma$	396.15	664.96	5.43
Lognormal	$\exp\left[-(y-\mu_y)/2\sigma_y^2\right] / (x\sqrt{2\pi\sigma_y})$ $y = \log(x), x > 0, -\infty < \mu_y < \infty, \sigma_y > 0$	$\mu_y$	6.12	6.65	1.82
		$\sigma_y$	0.50	0.63	0.34

表4 洪水变量不同分布函数的RMSE及AIC值的对比

Flooding variables	Marginal distribution	K-S test		RMSE	AIC
		$T$	$P$ -value		
Peak	Gamma	0.0745	0.5471	0.0378	-323.5512
	GEV	0.0741	0.5510	0.0340	-332.2144
	Lognormal	0.0548	0.7146	0.0265	-358.8192
	P 3	0.0824	0.8868	0.0361	-326.0848
	LP 3	0.0915	0.7967	0.0324	-336.9444
Volume	Gamma	0.1126	0.2615	0.0445	-307.1904
	GEV	0.1026	0.3268	0.0428	-309.1165
	Lognormal	0.0749	0.5435	0.0361	-328.2335
	P 3	0.1226	0.4077	0.0436	-307.3700
	LP 3	0.1556	0.1594	0.0706	-259.0655
Duration	Gamma	0.0745	0.5471	0.0378	-323.5512
	GEV	0.0724	0.9151	0.0275	-389.3956
	Lognormal	0.0805	0.8403	0.0283	-388.3297
	P 3	0.0893	0.7386	0.0395	-349.3274
	LP 3	0.0795	0.8508	0.0324	-371.3614

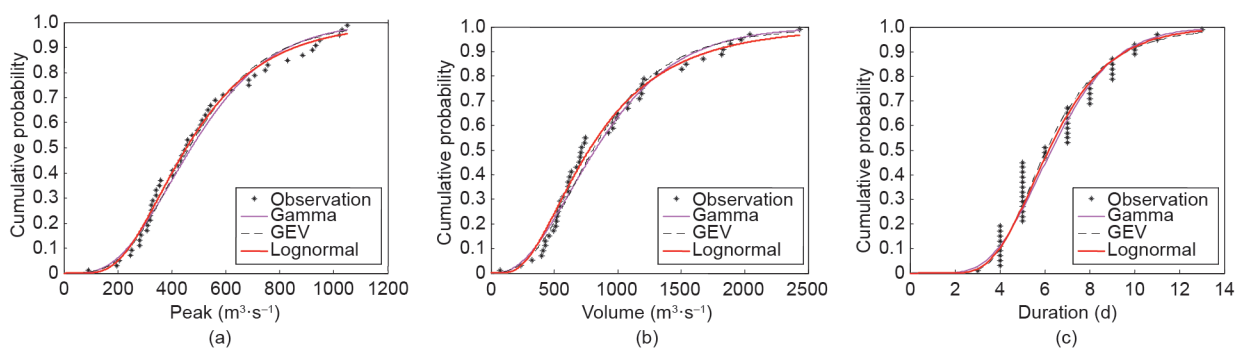


图3. 观测概率与不同概率密度模型的观测值对比。

时间)之间线性相关系数和Kendall ( $\tau$ )系数的值。结果表明,洪峰与洪量的相关性最高,其次是洪量和持续时间以及洪峰和持续时间。为了量化洪峰-洪量、洪量-持续时间和洪峰-持续时间的相关特征,采用了Ali-Mikhail-Haq (AMH)、Cook-Johnson (Clayton)、Gumbel-Hougaard和Frank copula这4个阿基米德copula函数,基于Kendall系数的反推通过矩量法(MOM)样估计器来估计它们的未知参数。用copula函数来量化洪水变量之间的相关性,利用RMSE和AIC对copula函数表现进行评价。表6显示了不同的copula函数用于模拟洪峰-洪量、洪量-持续时间和洪峰-持续时间的联合分布的RMSE和AIC结果。结果表明,通过Frank copula可以较好地量化洪峰-洪量、洪量-持续时间的联合分布,而洪量-持续时间的联合分布可以通过Ali-Mikhail-Haq得到较好的结果。

## 5. 双变量水文风险不确定性分析

为了进一步分析基于copula函数的风险分析模型的不确定性,本研究将假定模型参数的先验分布为正态

表5 洪水变量间的相关系数的取值

No.	Flood characteristics	Kendall's $\tau$	Pearson's $r$
1	Peak-volume	0.63	0.75
2	Volume-duration	0.52	0.46
3	Peak-duration	0.15	-0.06

表6 不同copula函数的RMSE和AIC值的对比

Copula	RSME			AIC		
	Peak-volume	Peak-duration	Volume-duration	Peak-volume	Peak-duration	Volume-duration
Ali-Mikhail-Haq	0.0610	0.0514	0.0707	-277.6384	-294.7895	-262.9122
Cook- Johnson	0.0665	0.2907	0.0873	-269.0263	-121.5381	-241.8286
Gumbel-Hougaard	0.0312	0.0550	0.0589	-344.8576	-288.0178	-281.1828
Frank	0.0304	0.0541	0.0559	-347.4879	-289.7032	-286.3915

表7 贝叶斯模型中各参数的先验分布

Flood variables	Prior distributions	
Peak (lognormal)	Norm (6, 4)	Norm (0.5, 0.1)
Volume (lognormal)	Norm (6, 4)	Norm (0.6, 0.1)
Duration (lognormal)	Norm (2, 1)	Norm (0.4, 0.1)
Peak-volume (Frank)	Norm (8, 4)	
Peak-duration (AMH)	Norm (1, 0.4)	
Volume-duration (Frank)	Norm (6, 4)	

分布,其均值为MLE及MOM得到的值,方差则人为假定比较大的值。表7给出具体的先验分布信息。基于MCMC计算5000次,并将最后500次设定为copula模型参数的后验分布。图4展示了边缘分布及copula函数中参数的演进过程。结果显示贝叶斯方法得到的参数最后趋于稳定并集中于某个固定值附近。图5给出了参数的后验分布信息。基于MCMC所得到的样本,可推求设计值的概率特征,其不仅可给出具体的设计值,还能得到相关的预测区间。

在本研究中,基于贝叶斯方法推求的边缘分布及copula函数中参数的值表现为一定的概率分布,进而导致了洪水变量边缘及联合分布的不确定性。图6给出了预测均值与观测值之间的对比,此外还给出了95%置信度水平下的预测区间,预测值分别为2.5%和97.5%分位数。该图显示所得到的预测区间可很好地覆盖对应观测概率。此外,对洪水持续时间而言,图3显示MLE所得到的概率分布于观测概率存在明显偏差。这可能是由于洪水持续时间的数据所导致。在本研究中,所有数据均为日尺度,因此洪水持续时间也表示为天数,进而导致洪水持续时间的概率分布蕴含大量不确定性。与之相比,图6结果显示基于贝叶斯方法所得到的95%预测区间能够覆盖大部分的洪水持续时间的观测概率值。

在贝叶斯的方法框架中,位置参数被设置为随机变量,并且设定相关先验分布;而其后的后验分布则通过MCMC来推求。同时,边缘分布及联合分布中参数的随

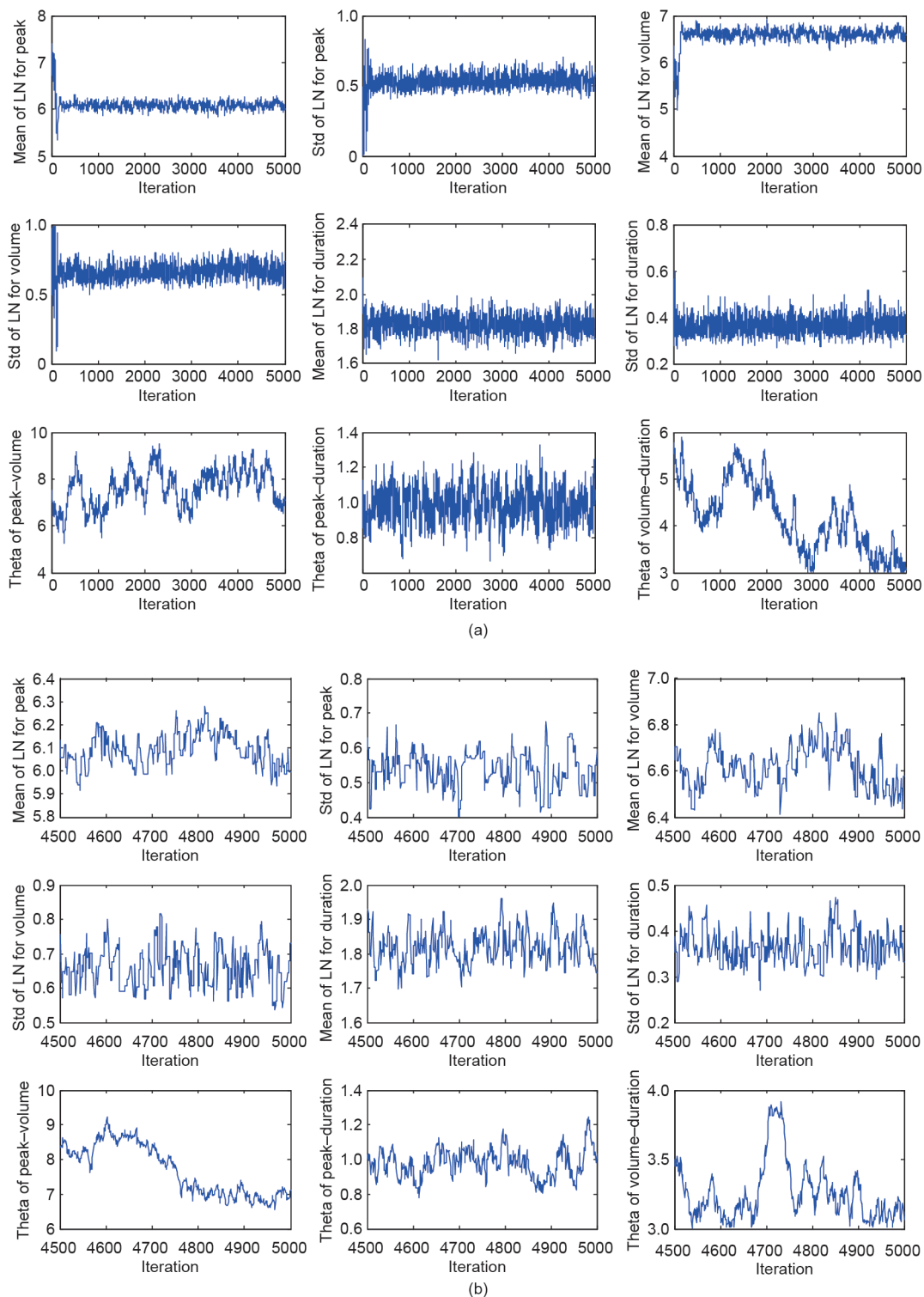


图4. MCMC过程中未知参数的演变。(a) 洪峰；(b) 燃烧阶段后。

机性又导致相关条件累积分布函数 (CDF) 及条件PDF的不确定性。图7给出了给定洪峰为95%分位数情景下的洪量及洪水持续时间的条件CDF分布；基于参数后验分布，可进一步推求所得条件CDF的预测区间。此外，通过公式 (7) 和公式 (8)，可推导不同重现期洪峰情

景下的洪量及洪水持续时间的条件PDF。图8显示了不同洪峰重现期情景下的条件PDF。假设考虑两种洪峰情景 (10年和100年重现期)，所得洪量及洪水持续时间条件概率密度分布显示随着洪峰值的增加，洪量的值也会增加，而与此同时，洪水持续时间可能不会有很大变化。

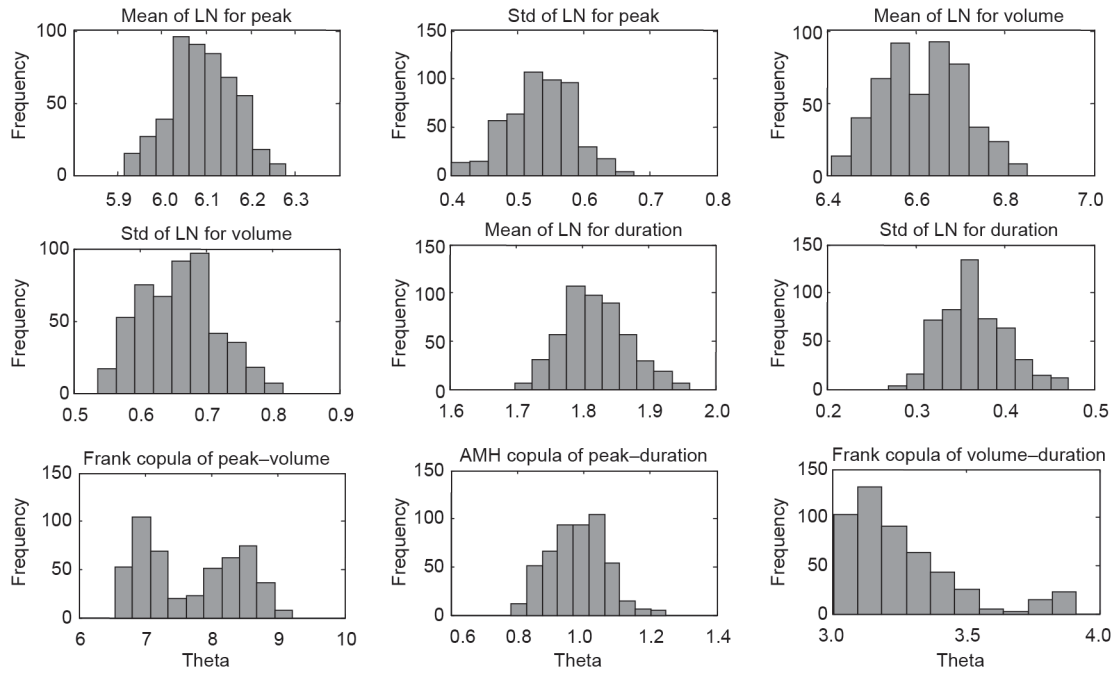


图5. 边缘分布及copula函数中未知参数的后验概率。LN对数正态分布。

基于公式(9)~(12)所得联合重现期可很好地反映洪峰-洪量、洪峰-持续时间及洪量-持续时间共同发生的概率水平。此外,模型参数的随机特征也同时导致了上述联合重现期的不确定性。表8给出了洪水的单变量及联合重现期的95%预测区间。此外,随着单变量重现期的增加,所对应的洪峰、洪量及持续时间预测值的不确定性也会增加。例如,一次5年重现期洪水事件的洪峰的95%变化区间为 $[594.1, 813.9] \text{ m}^3 \cdot \text{s}^{-1}$ ,而如果重现期为100年,其95%变化区间为 $[1156.3, 2007.1] \text{ m}^3 \cdot \text{s}^{-1}$ 。对于联合重现期,如表8所示,一个明显特征是“OR”联合重现期的不确定性不会有很大变化。与之相比,“AND”联合重现期和二次重现期的不确定性特征与单个洪水变量的重现期类似,如表8所示。例如,初始重现期为100年所对应的联合重现期 $T_{pv}^{\text{AND}}$ 的95%预测区间为 $[1229.4, 1593.8]$ 年;而相应的联合重现期 $T_{pv}^{\text{OR}}$ 的95%预测区间为 $[51.62, 52.12]$ 年。

本研究提出了一种两变量风险指标,如公式(12)所示,以反映洪水变量间的相互作用以及它们对具体洪水风险的潜在影响。对于一个洪水事件,水利设施的损坏主要由于高洪峰量,因此分析洪水风险时必须考虑洪峰量。所以,本研究将分析洪峰-流量及洪峰-持续时间的联合风险。图9展示了不同设计流量及服务时间情景下洪水风险随洪量的变化趋势。对于所有的设计流量及服务时间情景,洪峰和洪量的两变量风险会在一定洪量

范围内保持稳定,然后会随洪量的增加而减少。此外,对于某个洪峰-洪量组合,由于边缘分布和联合分布的参数不确定性,其具体风险水平也会是不确定的。基于参数的后验概率,可推求双变量风险的预测均值及95%预测区间,如图9所示。结果显示,低设计流量和高服务年限会导致水文风险的高不确定性。图10展示了兴山站附近堤坝在不同洪峰-持续时间情景下的水文风险。结果显示对于同样的服务时间,水文风险随设计流量的增加而降低。类似地,对于同样的设计流量,增加服务时间会同时增加水文风险。同时,洪峰-持续时间双变量风险的不确定性特征与洪峰-洪量类似。

## 6. 结论

Copula函数是水文及水资源领域应用最广泛的方法之一。许多研究基于copula函数开展多变量环境下的水文灾害(如洪水、干旱等)风险分析。然而,前人的研究在一定程度上忽略了copula模型框架中边缘分布及联合分布的参数不确定性。本研究旨在辨识多变量水文风险中的内在不确定性。本研究的创新之处在于将贝叶斯理论引入到copula模型框架中用于量化边缘分布及相关性结构中参数的不确定性。同时,进一步揭示边缘分布和联合分布参数不确定性间的互动关系。该研究通过分析香溪河流域生态水文风险来阐述其应用性。



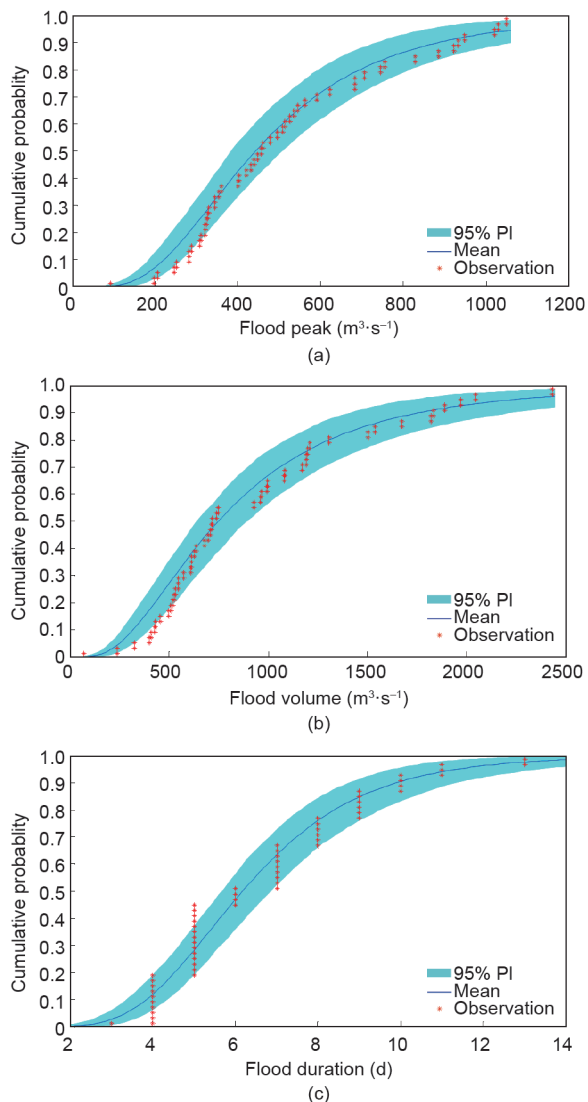


图6. 预测概率与观测概率值间的对比。(a) 洪峰; (b) 洪量; (c) 洪水持续时间。

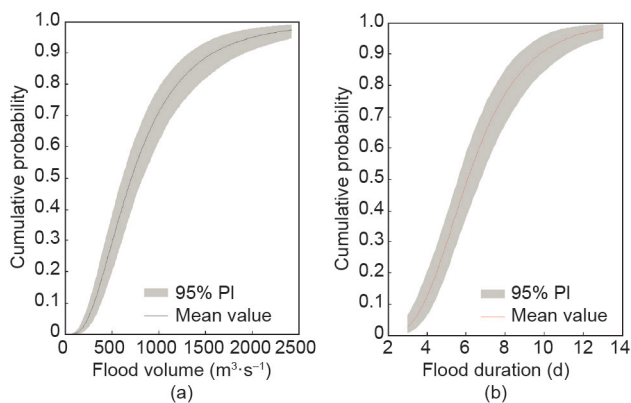


图7. 95%洪峰分位数情景下的条件CDF。(a) 洪量; (b) 持续时间。

研究结果显示, 对数正态分布比较适合量化香溪河流域洪水变量的边缘分布。就联合分布而言, Frank

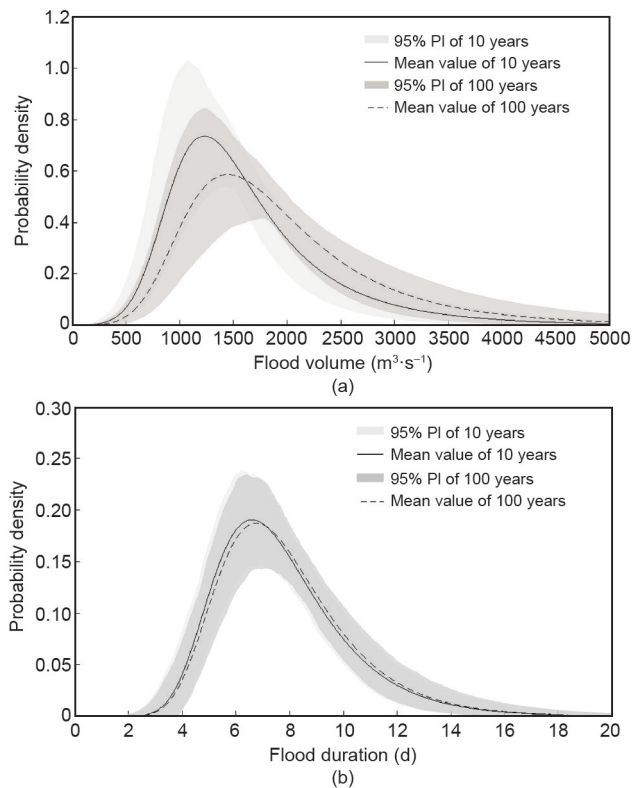


图8. 不同洪峰重现期情景下的条件PDF。(a) 洪量; (b) 持续时间。

copula适合于量化洪峰-洪量、洪量-持续时间的联合分布; Ali-Mikhail-Haq copula则适合于构建洪峰-持续时间的联合分布。MCMC方法用于量化边缘分布及联合分布的参数不确定性, 其结果显示所得预测区间能够很好地覆盖观测概率, 尤其对洪水持续时间而言。此外, 基于模型参数后验概率分布, 可推求“AND”和“OR”联合重现期的95%预测区间。所得结果显示“AND”联合重现期的不确定性将随单变量重现期的增加而增大, 但“OR”联合重现期的不确定性不会随之剧烈变化。最后, 该研究还揭示了洪峰-洪量、洪峰-持续时间两组双变量水文风险的不确定性, 以反映洪水变量间对水利设施损失风险的互动效应。

自从De Michele和Salvadori[40]将copula函数引入水文分析以来, 此类方法已被广泛应用于多变量水文风险分析问题中。本研究将贝叶斯方法引入到copula模型框架中以辨识多变量水文风险分析中的内在不确定性, 并进一步揭示洪水变量(如洪峰、洪量和持续时间)不确定性的互动效应对相应风险预测的影响。通过香溪河的实例证明该方法的可靠性。但是该方法还可用于其他水文事件(如干旱、暴雨等)在不同区域的风险分析问题中。本研究尝试评估基于copula的多变量水文风险分析

表8 单变量重现期及首次重现期的95%预测区间

$T$ (year)	Peak ( $\text{m}^3 \cdot \text{s}^{-1}$ )	Volume ( $\text{m}^3 \cdot \text{s}^{-1} \cdot \text{d}^{-1}$ )	Duration (d)	$T_{PV}^{\text{AND}}$	$T_{PD}^{\text{AND}}$	$T_{DV}^{\text{AND}}$	$T_{PD}^{\text{OR}}$	$T_{PD}^{\text{OR}}$	$T_{PV}^{\text{OR}}$
5	[594.1, 813.9]	[1087.6, 1596.3]	[7.6, 9.5]	[7.6, 8.5]	[14.1, 16.1]	[11.0, 12.4]	[2.96, 3.04]	[3.13, 3.23]	[3.54, 3.77]
10	[734.3, 1059.3]	[1413.4, 2210.7]	[8.7, 11.3]	[20.9, 24.6]	[51.6, 59.4]	[34.9, 40.7]	[5.46, 5.54]	[5.70, 5.84]	[6.28, 6.57]
20	[859.7, 1336.6]	[1734.8, 2905.5]	[9.8, 13.3]	[64.7, 79.3]	[196.6, 227.8]	[120.9, 144.6]	[10.46, 10.54]	[10.74, 10.90]	[11.44, 11.83]
50	[1029.1, 1707.4]	[2187.9, 3975.0]	[11.1, 15.7]	[331.9, 423.0]	[1191.1, 1386.9]	[683.7, 834.0]	[25.46, 25.54]	[25.77, 25.95]	[26.57, 27.04]
100	[1156.3, 2007.1]	[2552.6, 4946.1]	[12.0, 17.8]	[1229.4, 1593.8]	[4714.8, 5497.9]	[2638.1, 3241.7]	[50.46, 50.54]	[50.78, 50.97]	[51.62, 52.12]

$T$ : return period; PV: peak-volume; PD: peak-duration; DV: duration-volume.

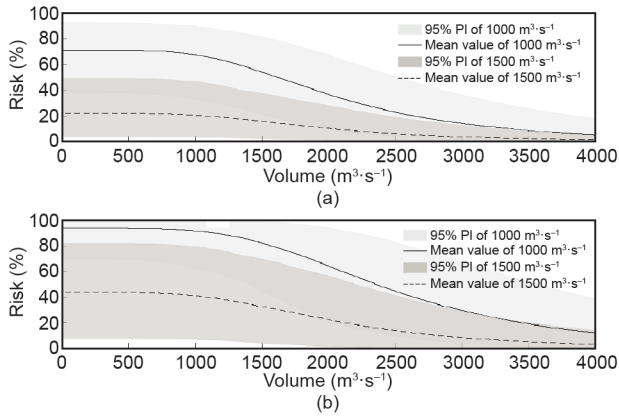


图9. 不同洪量-洪峰组合情景下的双变量洪水风险。(a) 20年服务时间; (b) 50年服务时间。

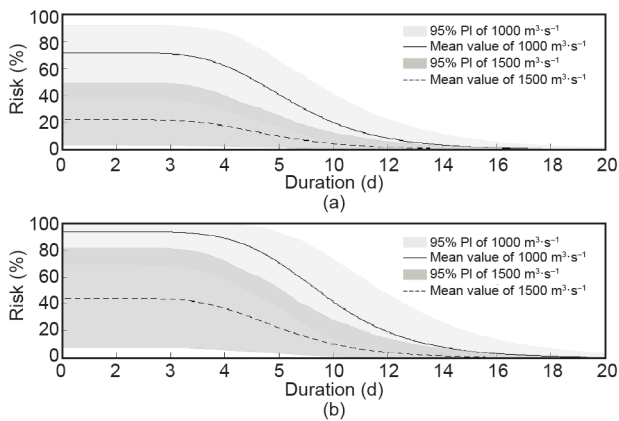


图10. 不同洪量-持续时间组合情景下的双变量洪水风险。(a) 20年服务时间; (b) 50年服务时间。

中的不确定性。然而, 该研究需要进一步拓展以分析模型选择及其他不确定来源所导致的水文风险不确定性。

## Acknowledgements

This work was jointly funded by the National Natural Science Foundation of China (51520105013 and 51679087) and the National Key Research and Development Plan of China (2016YFC0502800).

## Compliance with ethics guidelines

Yurui Fan, Guohe Huang, Yin Zhang, and Yongping Li declare that they have no conflict of interest or financial conflicts to disclose.

## References

- [1] Kidson R, Richards KS. Flood frequency analysis: assumption and alternatives. *Prog Phys Geogr* 2005;29(3):392–410.
- [2] Karmakar S, Simonovic SP. Bivariate flood frequency analysis. Part 2: a copulabased approach with mixed marginal distributions. *J Flood Risk Manag* 2009;2 (1):32–44.
- [3] Fan Y, Huang G, Huang K, Baetz BW. Planning water resources allocation under multiple uncertainties through a generalized fuzzy two-stage stochastic programming method. *IEEE Trans Fuzzy Syst* 2015;23(5):1488–504.
- [4] Fan Y, Huang W, Li Y, Huang G, Huang K. A coupled ensemble filtering and probabilistic collocation approach for uncertainty quantification of hydrological models. *J Hydrol* 2015;530:255–72.
- [5] Li Y, Nie S, Huang Z, McBean EA, Fan Y, Huang G. An integrated risk analysis method for planning water resource systems to support sustainable development of an arid region. *J Environ Inform* 2017;29(1):1–15.
- [6] Chebana F, Dabo-Niang S, Ouarda TBMJ. Exploratory functional flood frequency analysis and outlier detection. *Water Resour Res* 2012;48(4):W04514.
- [7] Guo Y, Baetz BW. Probabilistic description of runoff and leachate volumes from open windrow composting sites. *J Environ Inform* 2017;30(2):137–48.
- [8] Chen B, Li P, Wu H, Husain T, Khan F. MCFP: a Monte Carlo simulation-based fuzzy programming approach for optimization under dual uncertainties of possibility and continuous probability. *J Environ Inform* 2017;29(2):88–97.
- [9] Reddy JM, Ganguli P. Bivariate flood frequency analysis of upper Godavari River flows using Archimedean copulas. *Water Resour Manage* 2012;26 (14):3995–4018.
- [10] Fan Y, Huang W, Huang G, Huang K, Li Y, Kong X. Bivariate hydrologic risk analysis based on a coupled entropy-copula method for the Xiangxi River in the Three Gorges Reservoir area, China. *Theor Appl Climatol* 2016; 125(1–2):381–97.
- [11] Fan Y, Huang W, Huang G, Li Y, Huang K, Li Z. Hydrologic risk analysis in the Yangtze River Basin through coupling Gaussian mixtures into copulas. *Adv Water Resour* 2016;88:170–85.
- [12] Zhang L, Singh VP. Bivariate flood frequency analysis using the copula method. *J Hydrol Eng* 2006;11(2):150–64.
- [13] Srjai M, Bezak N, Brilly M. Bivariate flood frequency analysis using the copula function: a case study of the Litija Station on the Sava River. *Hydrol Processes* 2015;29(2):225–38.
- [14] Kao SC, Govindaraju RS. A copula-based joint deficit index for droughts. *J Hydrol* 2010;380(1–2):121–34.
- [15] Ma M, Song S, Ren L, Jiang S, Song J. Multivariate drought characteristics using trivariate Gaussian and Student t copulas. *Hydrol Processes* 2013;27(8):1175–90.
- [16] Zhang L, Singh VP. Bivariate rainfall frequency distributions using Archimedean copulas. *J Hydrol* 2007;332(1–2):93–109.
- [17] Vandenberghe S, Verhoest NEC, De Baets B. Fitting bivariate copulas to the dependence structure between storm characteristics: a detailed analysis based on 105 year 10 min rainfall. *Water Resour Res* 2010;46(1):W01512.
- [18] Lee T, Salas JD. Copula-based stochastic simulation of hydrological data applied to Nile River flows. *Hydrol Res* 2011;42(4):318–30.
- [19] Kong X, Huang G, Fan Y, Li Y. Maximum entropy-Gumbel-Hougaard copula method for simulation of monthly streamflow in Xiangxi River, China. *Stochastic Environ Res Risk Assess* 2015;29(3):833–46.

- [20] Fan Y, Huang G, Baetz BW, Li Y, Huang K. Development of a copula-based particle filter (CopPF) approach for hydrologic data assimilation under consideration of parameter interdependence. *Water Resour Res* 2017;53(6):4850–75.
- [21] Fan Y, Huang G, Baetz BW, Li Y, Huang K, Chen X, et al. Development of integrated approaches for hydrological data assimilation through combination of ensemble Kalman filter and particle filter methods. *J Hydrol* 2017;550:412–26.
- [22] Chen F, Huang G, Fan Y, Chen J. A copula-based fuzzy chance-constrained programming model and its application to electric power generation systems planning. *Appl Energy* 2017;187:291–309.
- [23] Huang C, Nie S, Guo L, Fan Y. Inexact fuzzy stochastic chance constraint programming for emergency evacuation in Qinshan Nuclear Power Plant under uncertainty. *J Environ Inform* 2017;30(1):63–78.
- [24] Huang K, Dai L, Yao M, Fan Y, Kong X. Modelling dependence between traffic noise and traffic flow through an entropy-copula method. *J Environ Inform* 2017;29(2):134–51.
- [25] Yu L, Li Y, Huang G, Fan Y, Nie S. A copula-based flexible-stochastic programming method for planning regional energy system under multiple uncertainties: a case study of the urban agglomeration of Beijing and Tianjin. *Appl Energy* 2018;210:60–74.
- [26] Guo A, Chang J, Wang Y, Huang Q, Zhou S. Flood risk analysis for flood control and sediment transportation in sandy regions: a case study in the Loess Plateau, China. *J Hydrol* 2018;560:39–55.
- [27] Guo A, Chang J, Wang Y, Huang Q, Guo Z, Zhou S. Bivariate frequency analysis of flood and extreme precipitation under changing environment: case study in catchments of the Loess Plateau, China. *Stochastic Environ Res Risk Assess* 2018;32(7):2057–74.
- [28] Merz B, Thielen AH. Separating natural and epistemic uncertainty in flood frequency analysis. *J Hydrol* 2005;309(1–4):114–32.
- [29] Liang Z, Chang W, Li B. Bayesian flood frequency analysis in the light of model and parameter uncertainties. *Stochastic Environ Res Risk Assess* 2012;26(5):721–30.
- [30] Nelsen RB. *An introduction to copulas*. 2nd ed. New York: Springer; 2006.
- [31] Ganguli P, Reddy MJ. Probabilistic assessment of flood risks using trivariate copulas. *Theor Appl Climatol* 2013;111(1–2):341–60.
- [32] Salvadori G, De Michele C, Kottegoda NT, Rosso R. *Extremes in nature: an approach using copula*. Dordrecht: Springer; 2007. p. 292.
- [33] Salvadori G, De Michele C, Durante F. On the return period and design in a multivariate framework. *Hydrol Earth Syst Sci* 2011;15(11):3293–305.
- [34] Han JC, Huang GH, Zhang H, Li Z, Li YP. Bayesian uncertainty analysis in hydrological modeling associated with watershed subdivision level: a case study of SLURP model applied to the Xiangxi River watershed, China. *Stochastic Environ Res Risk Assess* 2014;28(4):973–89.
- [35] Xu H, Taylor RG, Kingston DG, Jiang T, Thompson JR, Todd MC. Hydrological modeling of River Xiangxi using SWAT2005: a comparison of model parameterizations using station and gridded meteorological observations. *Quat Int* 2010;226(1–2):54–9.
- [36] Yue S. The bivariate lognormal distribution to model a multivariate flood episode. *Hydrol Processes* 2000;14(14):2575–88.
- [37] Yue S. A bivariate gamma distribution for use in multivariate flood frequency analysis. *Hydrol Processes* 2001;15(6):1033–45.
- [38] Adamowski K. A Monte Carlo comparison of parametric and nonparametric estimation of flood frequencies. *J Hydrol* 1989;108:295–308.
- [39] Kite GW. *Frequency and risk analysis in Hydrology*. Littleton: Water Resources Publications; 1988.
- [40] De Michele C, Salvadori G. A generalized Pareto intensity-duration model of storm rainfall exploiting 2-copulas. *J Geophys Res* 2003;108(D2):4067