



Research
Watershed Ecology—Article

Uncertainty Quantification for Multivariate Eco-Hydrological Risk in the Xiangxi River within the Three Gorges Reservoir Area in China



Yurui Fan^a, Guohe Huang^{b,*}, Yin Zhang^c, Yongping Li^b

^a College of Engineering, Design and Physical Sciences, Brunel University, London, Uxbridge, Middlesex, UB8 3PH, UK

^b State Key Laboratory of Water Environment, School of Environment, Beijing Normal University, Beijing 100875, China

^c 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

Keywords:

Flood risk

Copula

Multivariate flood frequency analysis

Distribution

Markov chain Monte Carlo

ABSTRACT

This study develops a multivariate eco-hydrological risk-assessment framework based on the multivariate copula method in order to evaluate the occurrence of extreme eco-hydrological events for the Xiangxi River within the Three Gorges Reservoir (TGR) area in China. Parameter uncertainties in marginal distributions and dependence structure are quantified by a Markov chain Monte Carlo (MCMC) algorithm. Uncertainties in the joint return periods are evaluated based on the posterior distributions. The probabilistic features of bivariate and multivariate hydrological risk are also characterized. The results show that the obtained predictive intervals bracketed the observations well, especially for flood duration. The uncertainty for the joint return period in “AND” case increases with an increase in the return period for univariate flood variables. Furthermore, a low design discharge and high service time may lead to high bivariate hydrological risk with great uncertainty.

© 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. Introduction

Due to climate change and anthropogenic effects, extreme hydrological events occur with increasing frequency and intensity, resulting in great economic losses and posing intensified risks to people and the environment [1–5]. The frequency of a specific event (e.g., floods, droughts, or storms) is crucial for inferring magnitudes of the event with different occurrence probabilities, which are used for many eco-hydrological management practices [6–8]. For example, the design of hydraulic structures such as dam spillways requires the frequency of flood in the drainage to be quantified [9]. However, many hydrological events cannot easily be described by a univariate variable, and generally show multidimensional features. Floods are a typical example, as they are generally characterized by peak flow, volume, and duration. Moreover, these three attributes of a flood are correlated with each other. However, a traditional univariate flood risk analysis mainly focuses on the probabilistic feature of flood peak flow, which does not fully describe the flood characteristics (e.g., dependence among flood peak, volume, and duration). Therefore, flood frequency analysis

within a multivariate context is required in order to obtain an insightful description of extreme hydrological events and generate reliable risk inference [10,11].

The concept of the copula function has been widely used for many eco-hydrological issues due to its powerful capability to reflect complicated dependence structures among correlated variables. A number of copula-based approaches have been proposed for the multivariate risk analysis of floods [10–13], droughts [14,15], and storms [16,17]. In addition, some studies have reported the development of copula-based forecasting approaches for probabilistic streamflow simulation [18–21] and for energy and environmental systems analysis [22–25]. Compared with conventional multivariate statistical approaches, the copula-based approach allows separate quantification of the marginal distributions and the dependence modeling, thus making it possible to use different types of marginal and joint probability functions [2,13,26,27]. Thus, in flood risk analysis, the distributions of flooding peak, volume, and duration can be quantified through various parametric or non-parametric distribution formulations. For example, Karmakar and Simonovic [2] investigated the impacts of the selection of marginal distributions on the performance of copulas.

One major issue in hydrological risk analysis is the extensive uncertainty resulting from data availability, model selection, and

* Corresponding author.

E-mail address: huang@iseis.org (G. Huang).

parameter estimation. There are two primary sources of uncertainty: ① natural uncertainty stemming from the inherent variability of the underlying hydrological processes, and ② epistemic uncertainty coming from an incomplete understanding of the studied system [28]. In particular, data availability is a major source of uncertainty in flood frequency analysis. Several previous studies have addressed uncertainties in flood risk analysis. For example, Liang et al. [29] analyzed model and parameter uncertainties using a Bayesian-based approach. However, most previous research primarily focused on univariate hydrological risk. Few existing studies examine the uncertainties of hydrological risk within a multivariate context.

Therefore, the objective of this research is to propose a multivariate risk-assessment framework through an interactive analysis of the multiple attributes of a flood. The uncertainty of the bivariate hydrological risk is evaluated based on the Bayesian method, and the posterior probabilities of the parameters in the marginal and joint distribution are obtained based on the Markov chain Monte Carlo (MCMC) method. Uncertainties in the primary bivariate return periods are quantified.

2. Methodology

2.1. The concept of the copula

Based on the copula function, a multivariate probability distribution can be constructed as follows:

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

where $F_{X_1}(x_1), F_{X_2}(x_2), \dots, F_{X_n}(x_n)$ denote the marginal distributions of the random vector (X_1, X_2, \dots, X_n) . If these marginals are continuous, a copula function C exists, such that [13]:

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

where $u_i = F_{X_i}(x)$ and $i = 1, 2, \dots, n$. Nelsen [30] provides a detailed description of copulas in terms of their theoretical background and properties.

Many copula functions are available for practical multivariate analysis; these are mainly classified as Archimedean, elliptical, and extreme value copulas. This study uses Archimedean copulas due to their ability to capture a wide range of dependence structures [31]. These copulas can take a great variety of forms. Furthermore, they can have distinct upper and lower tail dependence coefficients. Thus, Archimedean copulas can capture dependence in the upper and lower tail dependences. Table 1 lists the potential copula functions used in this study and their inherent features.

2.2. Conditional distributions and joint return period

Based on the copula function expressed in Eq. (1), the conditional distribution for each X_i can be derived. Taking a bivariate case as an example, where $U_1 = F_{X_1}(x)$ and $U_2 = F_{X_2}(x)$, the

conditional distribution for U_1 with $U_2 = u_2$ can be derived as follows [12]:

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

Similarly, the conditional distribution function of U_1 when $U_2 \leq u_2$ can be expressed as follows:

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

In addition, the probability density function (PDF) $c(u_1, u_2)$ of a copula is derived by the following:

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

Next, the joint PDF of X_1 and X_2 can easily be derived by $c(u_1, u_2)$:

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

Consequently, the conditional PDF of X_1 for a given value of X_2 , is expressed as follows:

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

The conditional PDF of X_2 , with a given X_1 , is formulated as follows:

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

Based on the copula function, the concept of return period, which is widely used in univariate hydrological frequency analysis, can be extended to the multivariate context, leading to the concept of a joint return period. To be specific, the joint (primary) return periods can be further characterized by “OR” and “AND” cases [13,32,33]:

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

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

where μ indicates the mean inter-arrival time of the two consecutive extreme events.

2.3. Bivariate hydrological risk analysis

In practical hydraulic infrastructure design, it is essential to know the random features of the floods that will flow through the infrastructure. Moreover, a flooding event is usually characterized by multiple attributes such as peak, volume, and duration. Therefore, flood risk in a multivariate context would provide more

Table 1
Basic properties of applied copulas.

Copula name	Functions $C_\theta(u_1, u_2)$	θ 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	$\exp\left\{-\left[(-\ln u_1)^\theta + (-\ln u_2)^\theta\right]^{1/\theta}\right\}$	$[1, \infty)$	$(-\ln t)^\theta$	$1 - \theta^{-1}$
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 , $D_k(x) = \frac{k}{x^k} \int_0^x \frac{t^k}{e^t - 1} dt$.

insightful information than univariate flood risk in terms of the flooding peak. For example, a flood associated with a large peak value and a long duration (e.g., more than 7 d) may lead to great economic losses, whereas a short-term high-peak flood may only be a flash flood. Thus, we introduce a bivariate risk indicator based on the joint return period in “AND,” which is used to reveal interactive effects among the multiple attributes of the flood. The bivariate risk indicator can be expressed as follows:

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

The bivariate hydrological risk R_{u_1, u_2} is introduced in this study in order to investigate the significance of effects from persistent high-risk levels due to impacts from multiple interactive flood variables. It indicates the joint risk of u_1 and u_2 respectively occurring above their thresholds. For example, for a specific hydraulic facility with a design peak flow and service time standards, if u_1 and u_2 respectively denote the flood peak and volume, then the value of R_{u_1, u_2} indicates the risk of a flood with a peak that is larger than the design peak flow and a volume that is greater than a certain threshold. Therefore, in this study, we adopt a joint return period in the “AND” case in order to express the bivariate hydrological risk.

2.4. Parameter estimation through the Bayesian model

Extensive uncertainties may be involved in the parametric estimation of a copula function due to: ① the inherent uncertainty in the flooding process, ② uncertainty in the selection of appropriate marginal functions and copulas, and ③ statistical uncertainty or parameter uncertainty within the parameter estimation process (e.g., the availability of samples). The Bayesian approach has been widely applied for uncertainty quantification, since it can incorporate various sources of information into a single analysis using Bayes’ theorem. Given the prior probability density and observations, the posterior distribution can be derived using Bayes’ theorem, which is expressed as follows:

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

where $\pi_0(\theta)$ signifies the prior parameter distribution, $L(\theta|\mathbf{X})$ denotes the likelihood function, $\int L(\theta|\mathbf{X}) \pi_0(\theta) d\theta$ is the normalization constant, $\pi(\theta|\mathbf{X})$ is the posterior probability density function, and $\mathbf{X} = (x_1, x_2, \dots, x_n)$ is the observation vector (i.e., $\mathbf{X} = (x_1, x_2)$ in this study).

For a bivariate hydrologic risk analysis using the copula, it is necessary to estimate the parameter of the copula, as well as the parameters of the two marginal probability distributions. Let θ_c , θ_1 , and θ_2 denote the parameters of the copula and of the two marginal probability distributions, respectively. The posterior distribution can then be derived as follows:

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

The term of $L(\theta_c, \theta_1, \theta_2|\mathbf{X})$ is the likelihood function of the observations. Based on the dependence structure between the copula and its marginal distributions, as expressed by Eq. (6), the likelihood function can be estimated as follows:

$$L(\theta_c, \theta_1, \theta_2|\mathbf{Y}) = c_{u_1, u_2} [F_{X_1}(x_1), F_{X_2}(x_2)|\theta_c] f_{X_1}(x_1|\theta_1) f_{X_2}(x_2|\theta_2) \quad (14)$$

where c_{u_1, u_2} is the density of the copula function, and f_{X_1} and f_{X_2} are the two marginal probability density functions, respectively.

The procedures to derive the posterior distributions are presented below:

Step 1: Apply the maximum likelihood method to estimate the initial values for the parameters in the marginal and copula models.

Step 2: Calculate the root-mean-square error (RMSE) and Akaike information criterion (AIC) values for the marginal and joint probabilities in order to choose the most appropriate marginal and joint distribution forms.

Step 3: Set the prior distributions based on the parameter values obtained in Step 1.

Step 4: Use the MCMC with the Metropolis–Hastings algorithm to derive the posterior probabilities for the parameters in the marginal and joint distributions.

Step 5: Evaluate the uncertainties in the primary and joint return periods.

Step 6: Reveal the probabilistic features of the bivariate hydrological risk.

3. Application

3.1. Catchment characterization and data collection

The proposed approach is applied to determine the hydrological risk for the Xiangxi River within the Three Gorges Reservoir (TGR) area in China. The Xiangxi River is located in Hubei Province. As shown in Fig. 1, the Xiangxi River is one of the main tributaries of the Yangtze River, and originates from the Shennongjia Nature Reserve area. It has a main stream length of 94 km and a catchment area of 3099 km² [34]. This area has a northern subtropical climate with a mean annual precipitation of 1100 mm and a fluctuation range of 670–1700 mm [35]. The flooding season in the Xiangxi River Basin is from July to August, and the main rainfall season is from May to September. In this study, the daily discharge data (1961–2010) observed at the Xingshan Hydrologic Station (110°45′0″ E, 31°13′0″ N) is used for the probabilistic assessment of flood risks for the Xiangxi River.

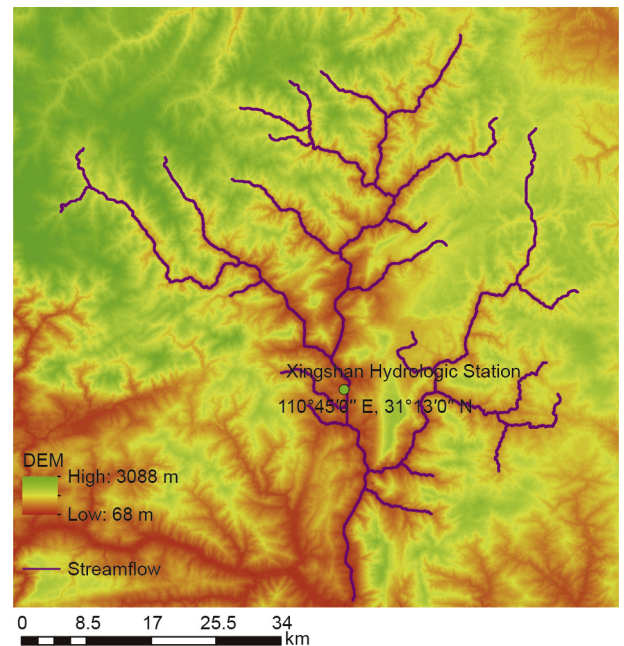


Fig. 1. The location of the studied watershed. DEM: digital elevation model.

3.2. Flooding characteristics in the Xiangxi River

In this study, the flood characteristics are identified as the peak, volume, and duration; the flood peak is the maximum peak discharge, while the volume and duration are obtained from the hydrograph of the maximum discharge. According to Yue [36,37], the flood duration (D_i) of a single peaked hydrograph in year i (Fig. 2 [31]) can be obtained by characterizing the time moment of the rise (D_{si} in Fig. 2) and fall (D_{ei} in Fig. 2) of the hydrograph. As shown in Fig. 2, the starting time of a flood is marked as a sharp rise on the hydrograph, and the end of the flood runoff is identified by an inflection point on the receding limb of the hydrograph [10]. If the rise moment of a flood is denoted by D_s (d) and the end time is denoted by D_e (d), then the flood volume (denoted as V_i , for this flood event) can be obtained as follows:

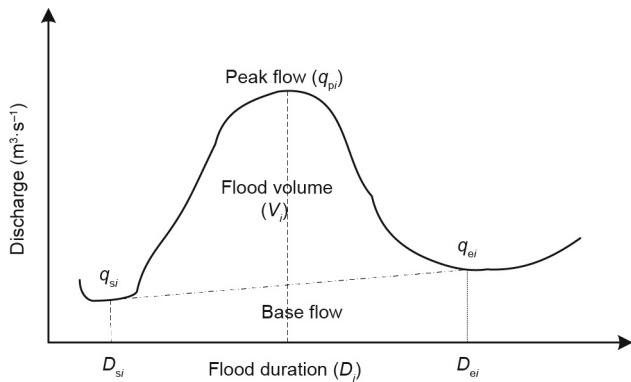


Fig. 2. Typical flood hydrograph showing flood flow characteristics (adapted from Ref. [31]).

Table 2
Statistical characteristics of flood variables.

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	

Table 3
Parameters of marginal distribution functions of flood variables.

Name	Probability density function	Variables	Parameter values		
			Peak	Volume	Duration
Gamma	$x^{a-1}e^{-x/b}/(b^a\Gamma(a)), \Gamma(a) = \int_0^\infty u^{a-1}e^{-u}du$	a	4.50	3.06	8.62
		b	113.26	301.24	0.76
GEV	$\frac{1}{\sigma} \exp\{-[1+k(x-\mu)/\sigma]^{-1/k} [1+k(x-\mu)/\sigma]^{-1-1/k}\}$	k	0.032	0.099	0.073
		μ	185.00	373.16	1.73
		σ	396.15	664.96	5.43
Lognormal	$\exp\left[-\frac{(y-\mu_y)^2}{2\sigma_y^2}\right]/(x\sqrt{2\pi\sigma_y})$ $y = \log(x), x > 0, -\infty < \mu_y < \infty, \sigma_y > 0$	μ_y	6.12	6.65	1.82
		σ_y	0.50	0.63	0.34

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

where q_{ij} is the observed stream flow on the j th day of the i th year, q_{si} is the streamflow on the start day of the flood in the i th year, and q_{ei} is the streamflow value on the end day of the flood. D_{si} and D_{ei} respectively denote the start and end day for a flood occurring in the i th year. D_i denotes the flood duration, which is obtained by the following:

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

The annual peak of the flood is obtained by the following:

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

Based on Eqs. (15)–(17), the flood characteristics can be identified as the peak, volume, and duration. Table 2 provides descriptive statistics values for the flood peak, volume, and duration at Xingshan Hydrologic Station. The positive values of kurtosis and skewness suggest that the flood variables can be modeled by sharp and right tailed distributions.

4. Initial parameter values for marginal and joint distributions

For a single flood variable such as flood peak, many parametric distributions are available to quantify its probabilistic feature, such as the general extreme value (GEV), log-Pearson Type III (LP3), and Pearson Type III (P3) distributions [38,39]. In this study, the gamma, GEV, and lognormal distributions are employed to quantify the probabilistic features for flood peak, volume, and duration, and the maximum likelihood estimation (MLE) method is used to estimate the unknown parameters in these distributions. Table 3 provides detailed expressions of the potential distributions and the parameter values obtained by MLE. A comparison of the theoretical distributions and the observed values is plotted in Fig. 3. The RMSE and AIC indices are applied to evaluate the performance of each marginal distribution, as shown in Table 4. The results show that the lognormal distribution performs best for all three flood variables in the Xiangxi River.

Table 5 provides the values for the linear correlation coefficient and Kendall's tau (τ) among the three flood variables (i.e., peak, volume, and duration). The results indicate that the highest correlation is between flood peak and volume, followed by the correlation of volume–duration and peak–duration. To quantify the dependence structures of flood peak–volume, volume–duration, and peak–duration, four Archimedean copulas—that is, the Ali-Mikhail-Haq (AMH), Cook-Johnson (Clayton), Gumbel-Hougaard, and Frank copulas—are employed, and their unknown parameters are estimated by a method-of-moments (MOM)-like estimator based on the inversion of Kendall's τ . The performances of the copulas in quantifying dependence

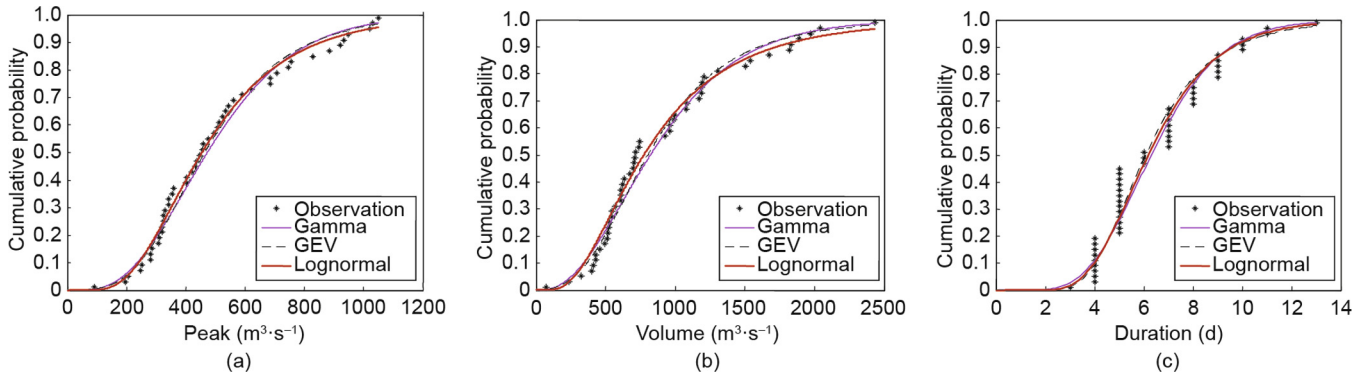


Fig. 3. Comparison of different probability density estimates with observed frequency.

Table 4
Comparison of RMSE and AIC values of flood variables for different statistical distributions.

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

among flood variables are evaluated by RMSE and AIC. Table 6 provides the results of the RMSE and AIC for different copulas for modeling the joint distribution of flood peak–volume, peak–duration, and volume–duration. The results suggest that joint distributions of flood peak–volume and volume–duration can be quantified best by the Frank copula, whereas the joint distribution of the flood peak–duration can be quantified best by the Ali–Mikhail–Haq copula.

Table 5
Values of correlation coefficients for flooding characteristics.

No.	Flood characteristics	Kendall's τ	Pearson's r
1	Peak–volume	0.63	0.75
2	Volume–duration	0.52	0.46
3	Peak–duration	0.15	–0.06

Table 6
Comparison of RMSE and AIC values for joint distributions through different copulas.

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

5. Uncertainty evaluation for bivariate hydrological risk

The prior distributions are specified for all parameters using the normal distributions, with the mean values being set as the initial values obtained through the MLE and MOM methods, and with relatively large variance values, as shown in Table 7. There are a total of 5000 iterations in the MCMC process, with the last 500 iterations being set as the posterior parameters of the parameters in the copula models. Fig. 4 presents the evolution of unknown parameters in the marginal distributions and copulas during the MCMC process. It can be seen that the parameters estimated using the Bayesian method are stationary and concentrated to stable values. The posterior distributions of the unknown parameters in the marginal distribution and copula functions are presented in Fig. 5. Based on the sampling sets of the parameters obtained through MCMC, the sampling

Table 7
Prior distribution for each parameter of Bayesian model.

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

distributions of the design values can be obtained; these cannot only provide the point estimation, but also give its predictive interval (PI).

The parameters in the marginal distributions and copulas, as obtained by the Bayesian method, are presented as probabilistic distributions, which lead to uncertainty in the marginal and joint distributions of the flood variables. Fig. 6 shows a comparison of the mean probabilities and the observed values. In addition, 95% predictive intervals are provided in Fig. 6, which are bracketed

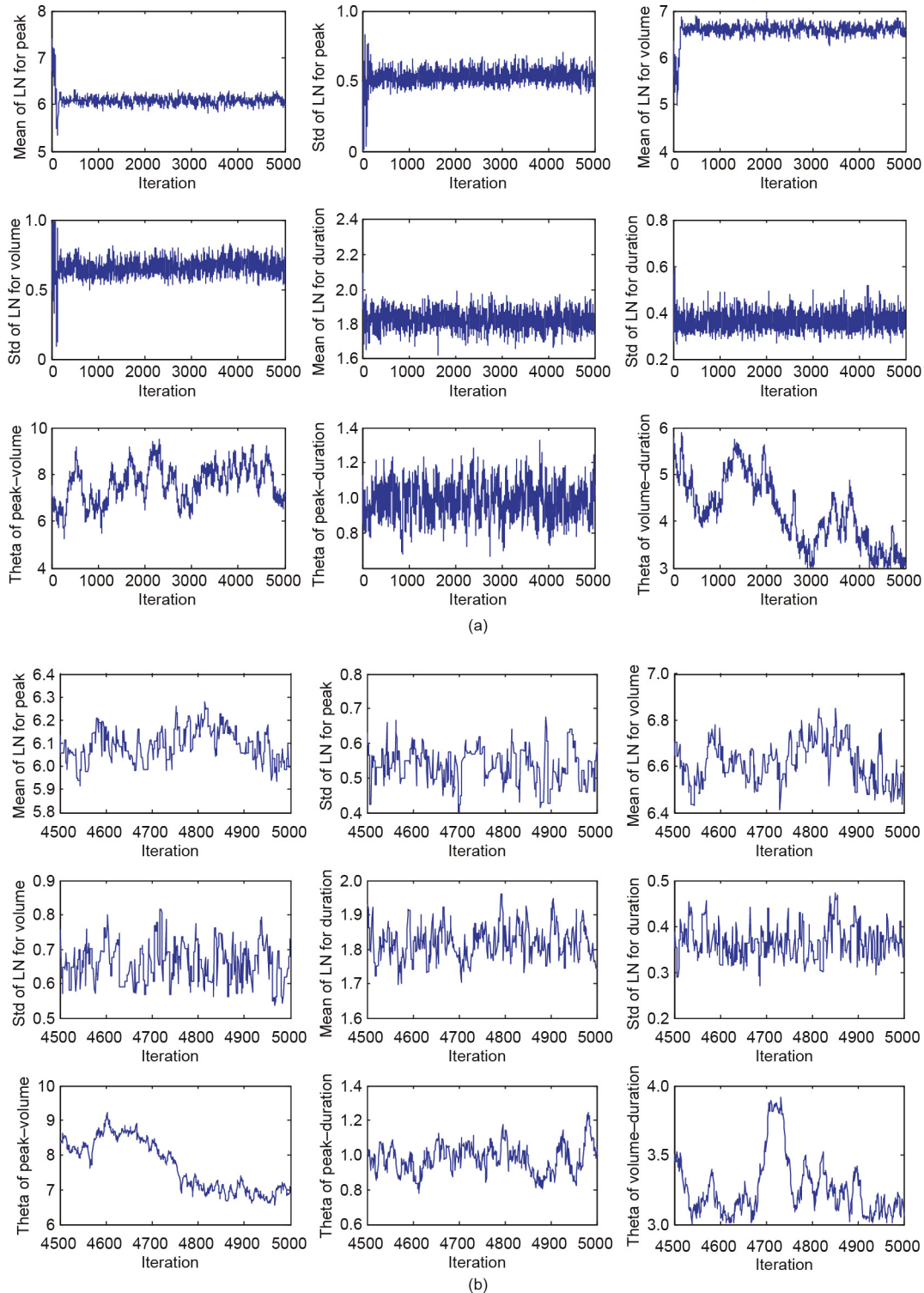


Fig. 4. Evolution of unknown parameters (a) during the MCMC process and (b) after burn-in period.

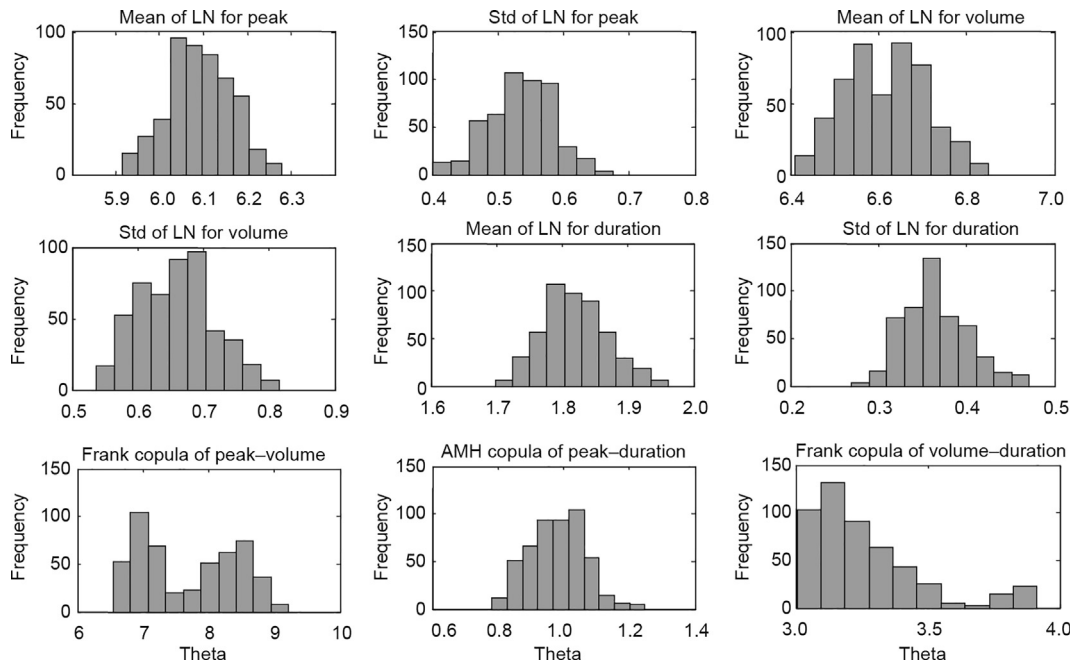


Fig. 5. Posterior distributions of the unknown parameters in marginal distributions and copulas. LN: lognormal.

by the 2.5% and 97.5% quantile values of the predictions. This figure indicates that the obtained predictive intervals can bracket the observed probabilities well. In particular, Fig. 3 shows that the estimated distributions through MLE produce an obvious discrepancy to the observed values for flood durations. Such deviation may result from data limitation regarding flood duration. Only daily streamflow records were available for this study, so the flood duration could only be calculated in days, leading to extensive uncertainty in the distribution of flood duration. In comparison, the results in Fig. 6 show that most of the observed data are bracketed within the 95% predictive intervals, which implies a good performance of the Bayesian method in parameter estimation for flood duration.

In the Bayesian method, the unknown parameters are considered as random with specific prior distributions; the posterior distributions of these parameters are then derived by the MCMC method. The randomness in the parameters of the marginal and joint probabilities lead to uncertainties in the associated conditional cumulative distribution functions (CDFs) and PDFs generated by the obtained copulas. Fig. 7 presents the conditional CDFs of flood volume and duration, with the flood peak flow specified as the 95% quantile value. Based on the posterior distribution of the parameters, the obtained CDFs and their associated 95% predictive intervals can be obtained. In addition, the conditional PDF of flood volume and duration, under a flood peak with different return periods, can be derived through Eqs. (7) and (8). Fig. 8 shows the conditional PDF of flood volume and duration. Two flood peak scenarios (10- and 100-year return periods) are assumed; the conditional PDFs of flood volume and duration indicate that as the flood peak increases, the flood volume is expected to increase as well, while the flood duration may not change significantly. Moreover, based on the posterior distributions of the parameters, the predictive intervals can also be obtained.

The joint return periods calculated by Eqs. (9) and (10) can reflect the concurrence probabilities of flood peak–volume, peak–duration, and volume–duration well. Furthermore, the random features of the model parameters leads to uncertainty in the joint return periods. Table 8 presents the 95% predictive intervals for

the univariate variable and joint return periods for the flood variables. It can be seen that the uncertainty of the flood peak, volume, and duration values increases with an increase in the return period. For example, the flood peak with a 5-year return period may vary within a 95% predictive interval $[594.1, 813.9] \text{ m}^3 \cdot \text{s}^{-1}$, while such an interval would be $[1156.3, 2007.1] \text{ m}^3 \cdot \text{s}^{-1}$ for the flood peak with a 100-year return period. For the joint return period, Table 8 shows a remarkable feature: The uncertainty of the joint return period in “OR” will not change significantly. The joint return period in “AND” has similar uncertainty characteristics as the univariate flood variables, as presented in Table 8. For example, the 95% predictive interval of T_{PV}^{AND} with a 100-year primary return period would be $[1229.4, 1593.8]$ years, while the 95% predictive interval of T_{PV}^{OR} would be $[51.62, 52.12]$.

In this study, a bivariate flood risk indicator, expressed as Eq. (12), is introduced to reflect the interactive effect among flood variables and their implications on the detailed hydrological risk. For a flooding event, the failure of hydraulic structures is mainly due to high peak flow, and the flooding peak flow is the essential factor to be considered when analyzing hydrological risks. Therefore, the bivariate flood risks for flood peak–volume and peak–duration are analyzed. Fig. 9 also shows the changing trends of the flooding risk with respect to flooding volume under different design flows and service times. The bivariate risk of flooding peak flow and volume remains constant for some time and then decreases with an increase in flooding duration for all design flows and service times. Furthermore, for a specific combination of flood peak–volume, the detailed risk may be uncertain, as a result of the parameter uncertainties in the marginal and joint distributions. Based on the posterior distributions of these parameters, the mean value and the 95% predictive interval of the bivariate risk can be derived, as presented in Fig. 9. The results in Fig. 9 suggest that a lower design discharge and longer service time may lead to greater uncertainty in hydrological risk. Fig. 10 presents the failure risk inference of the river levee around Xingshan Hydrologic Station for different flooding peak–duration extremes. It shows that for the same service time, the inferred risk decreases with an increase

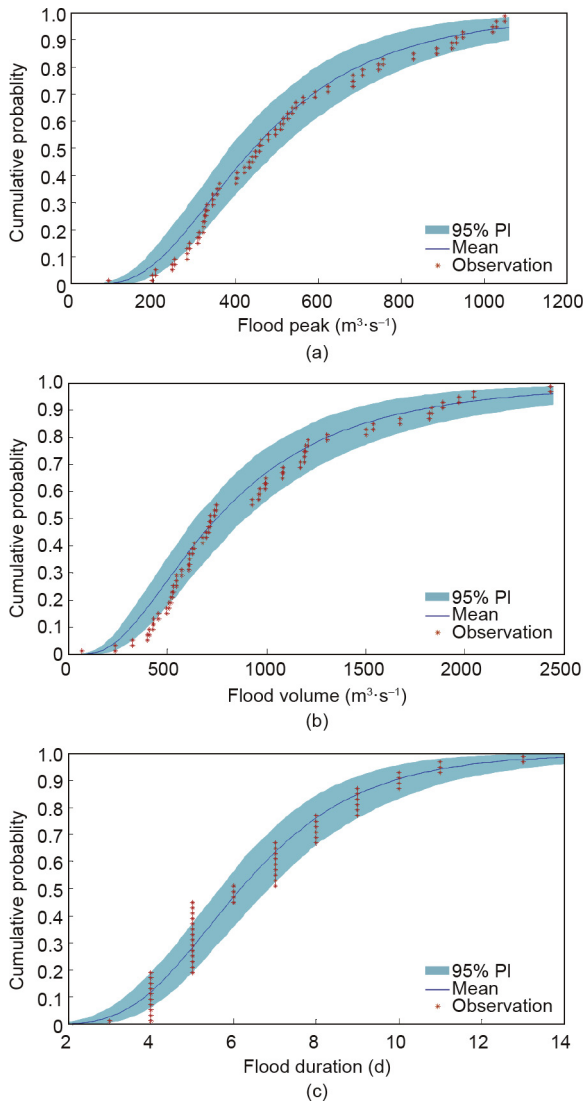


Fig. 6. Comparison of fitted and observed probabilities for flood variables: (a) flood peak, (b) flood volume, and (c) flood duration.

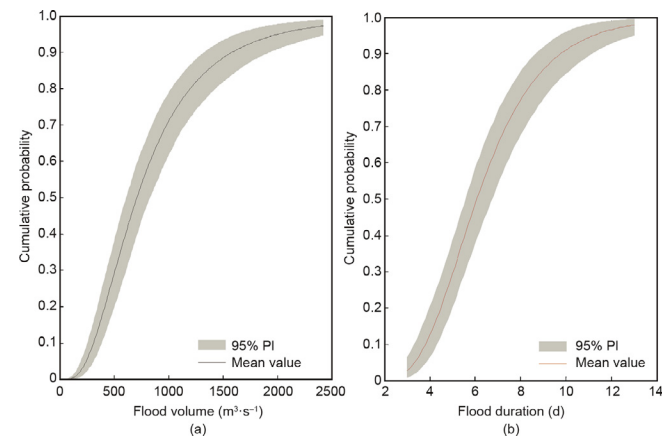


Fig. 7. Conditional CDFs of (a) flood volume and (b) duration under 95% flood peak flow.

in the designed flow. Similarly, for the same designed flow, an increase in service time leads to an increased failure risk of the river levee. Furthermore, the uncertainty of the bivariate hydrolog-

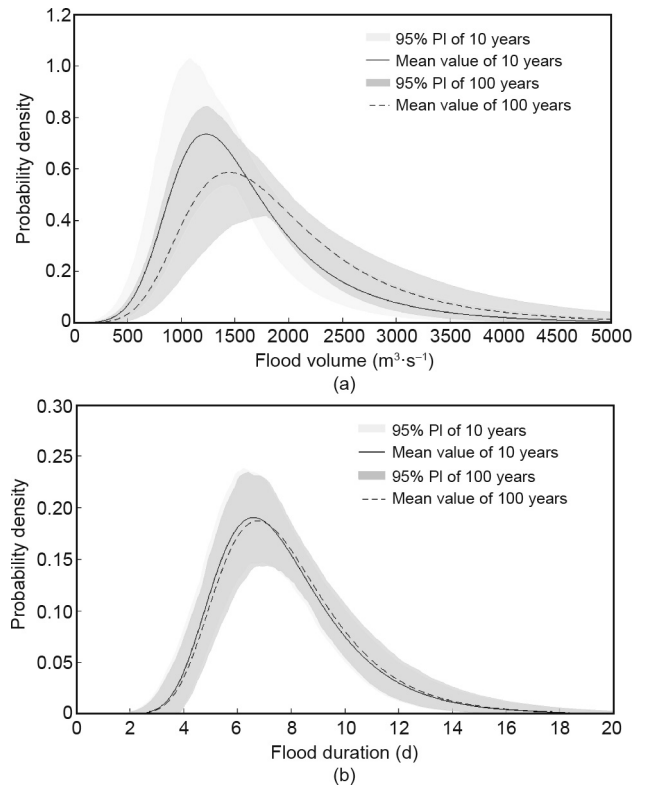


Fig. 8. Conditional PDFs of (a) volume and (b) duration under different peak flow return periods.

ical risk for flood peak–duration appears to be similar to that for flood peak–volume.

6. Conclusions and remarks

The copula approach is one of the most-used approaches in hydrology and water resources. Many studies have analyzed the risk of water hazards (e.g., flood and drought) using copula approaches to reflect probabilistic features in a multivariate context. However, the uncertainties that exist in both the marginal distributions and the dependence structures of the copula framework are somewhat overlooked. This study aimed to characterize the inherent uncertainties in multivariate hydrological risk analysis. The major innovation of this study is to introduce Bayesian inference into the copula framework in order to quantify uncertainties in both the marginal distributions and dependence structures; the interactions among uncertainties in marginal and joint distributions are also characterized. The approach developed in this study was applied to eco-hydrological risk analysis for the Xiangxi River in order to demonstrate the applicability of this method.

The results indicate that the lognormal distributions may be appropriate for quantifying the marginal distributions of flood variables in the Xiangxi River. The Frank copula performed better than the other copulas when quantifying the dependence of peak–volume and volume–duration, whereas the Ali–Mikhail–Haq copula performed best when quantifying the joint distribution of peak–duration. The MCMC approach was adopted to quantify the uncertainty in the marginal and joint distributions. The results showed that the obtained predictive intervals could bracket the observations well, especially for flood duration. Moreover, the 95% predictive intervals of the joint return period in “AND” and

Table 8
95% predictive intervals for the univariate and primary return periods for flood characteristics.

T (year)	Peak (m ³ ·s ⁻¹)	Volume (m ³ ·s ⁻¹ ·d ⁻¹)	Duration (d)	T _{PV} ^{AND}	T _{PD} ^{AND}	T _{DV} ^{AND}	T _{PD} ^{OR}	T _{DV} ^{OR}	T _{PV} ^{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.

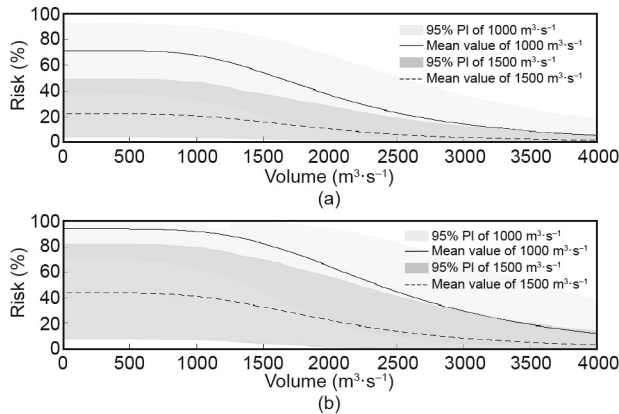


Fig. 9. Bivariate flooding risk under different flooding peak-volume scenarios. (a) 20-year service time; (b) 50-year service time.

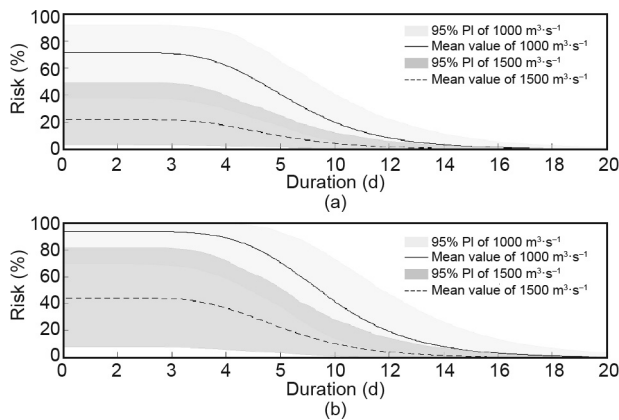


Fig. 10. Bivariate flooding risk under different flooding peak-duration scenarios. (a) 20-year service time; (b) 50-year service time.

“OR” cases could be derived based on the posterior probabilities of the model parameters. The results showed that the uncertainty of the joint return period in the “AND” case increases as the return period of the univariate flood variable increases. However, the uncertainty of the return period in the “OR” case does not change significantly. In addition, the uncertainty of bivariate risks for flood peak-volume and flood peak-duration was characterized to reflect the interactive effects among flood variables on the actual failure risk of hydraulic facilities.

Since the introduction of the copula to hydrological analysis by De Michele and Salvadori [40], these kinds of multivariate analysis approaches have been widely used for hydrological risk inference within a multivariate context. This study introduced Bayesian inference into the copula framework in order to characterize the inherent uncertainties in multivariate flood risk analysis, and to further

reveal the impacts of interactive uncertainties among flood attributes (i.e., peak, volume, and duration) on the resulting risk inferences. The approach developed in this study was demonstrated on a case study in the Xiangxi River Basin. However, this approach is transferable and can easily be applied to other risk-assessment problems (e.g., drought, storm, etc.) in different locations.

This study attempts to evaluate the uncertainty of copula-based multivariate hydrological risk. The initial idea of the research is to deal with the uncertainty in hydrological risk analysis that stems from data limitation. However, further studies are still required to characterize the uncertainty of hydrological risks that results from model uncertainty and other sources of uncertainty.

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 copula-based 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] Sraj 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] Vandenbergh 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.