

RESEARCH

Open Access



# Bivariate hydrologic risk analysis for the Xiangxi River in Three Gorges Reservoir Area, China

Y. R. Fan<sup>1\*</sup>

## Abstract

**Background** Hydrological extremes such as floods generally have multidimensional attributes with complex dependence structures. This leads to the urgent demand of hydrological risk analysis within a multivariate context. In this study, the bivariate hydrologic risk framework is proposed based on the bivariate copula method. In the proposed risk analysis framework, bivariate flood frequency would be analyzed for different flood variable pairs (i.e., flood peak-volume, flood peak-duration, flood volume-duration), and the bivariate hydrologic risk is then derived based on the joint return period of a flood variable pair. The distribution of one flood variable conditional on another flood variable can also be obtained through the copula method.

**Results** The proposed method is applied to the risk analysis for the Xiangxi River in the Three Gorges Reservoir area, China, based on 50 years streamflow measurements. The results indicate that the bivariate risk for flood peak flow-duration would keep constant for some time and then decrease as the increase of the flood duration. The bivariate risk for flood peak-volume holds a similar trend with the bivariate risk of flood peak-duration. The probability density functions (PDFs) of the flood volume and duration conditional on flood peak can also be generated through the best fitted copula function.

**Conclusion** The results indicate that the distributions of flood volume would be highly influenced by the flood peak flows, in which the flood volume would be expected to increase as the increase of flood return period. Conversely, the distribution of the flood duration would not change significantly with the variation in the flood peak return period. The obtained conclusions from the bivariate hydrologic analysis can provide decision support for flood control and mitigation.

**Keywords** Flood risk, Copula, Flood frequency analysis, Distribution, Conditional distribution

## Introduction

Extreme hydrological events such as floods, droughts and storms may lead to significant economic and social consequences and pose increasing risks on human beings and environment (Fan et al. 2012; Hu et al. 2021; Lyu

and Fan 2021). Hydrological frequency analysis procedures are essential and widely applied for analyzing and predicting such extreme hydrologic events, which pose direct impact on reservoir management and dam design (Li et al. 2008; Chebana et al. 2012; Huang and Fan 2021). At the drainage basin scale, consideration of flood risk plays a necessary role in planning of water infrastructure projects, for example in design of hydraulic structures (e.g., dam spillways, diversion canals, dikes and river channels), urban drainage systems, cross drainage structures (e.g., culverts and bridges), reservoir management, flood hazard mapping etc., (Ganguli and Reddy 2013;

\*Correspondence:

Y. R. Fan

yurui.fan@brunel.ac.uk

<sup>1</sup> Department of Civil and Environmental Engineering, Brunel University, Uxbridge, London UB8 3PH, Middlesex, UK

Mei et al. 2021). However, hydrological processes often involve multidimensional characteristics. Consequently, flood frequency analysis with considering more than one flood variable would be more effective in investigating the probabilistic flooding risk.

In recent years, copulas have been widely used for multivariate hydrologic studies, such as multivariate flood frequency analysis (Zhang and Singh 2006; Sraj et al. 2015; Fan et al. 2020a, b; 2021a), compound floods consisting of river discharges and sea level rise (e.g., Moftakhari et al. 2017; Sadegh et al. 2018; Fan et al. 2021b), drought assessment (Song and Singh 2010; Kao and Govindaraju 2010; Ma et al. 2013; Zhang et al. 2013), storm or rainfall dependence analysis (Zhang and Singh 2007; Vandenberghe et al. 2010), and streamflow simulation (Lee and Salas 2011; Kong et al. 2015). The main advantage of copula functions over classical bivariate frequency analyses is that the selection of marginal distributions and multivariate dependence modelling are two separate processes, giving additional flexibility to the practitioner in choosing different marginal and joint probability functions (Zhang and Singh 2006; Genest and Favre 2007; Karmakar and Simonovic 2009; Huang et al. 2017). Thus, in flood frequency 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 (2009) investigated the impacts of selection of marginal distributions on the performance of copulas. Besides, many other papers were proposed for copula analysis, such as unknown parameter estimation (Salvadoriet al. 2007), goodness-of-fit testing (Genest et al. 2009) and multivariate return period calculation (Salvadoriet al. 2011; Vandenberghe et al. 2010; Graler et al. 2013).

The Three Gorges Dam (TGM) is the largest hydraulic project in terms of design capacity over the world. The TGM project has produced dramatic benefits in flood control, power generation and navigation. Recently, the impacts of the TGM project on hydrology and environment have been attracting the world's attention. The Xiangxi River is the largest tributary of Yangtze River in the Hubei part of the Three Gorges Reservoir (TGR) area. Amounts of research studies have been conducted in this area, mainly focusing on hydrological modelling, water quality management and ecological studies (Ye et al. 2009; Xu et al. 2010; Han et al. 2014a, b). For instance, Han et al. (2014b) analyzed heterogeneous precipitation and streamflow trends in the Xiangxi River watershed, 1961–2010. Li et al. (2015) revealed hydrologic risk analysis for nonstationary streamflow records under uncertainty. However, extreme hydrologic events, especially flooding are one of the major natural disasters encountered by local people, due to the temporal-spatial

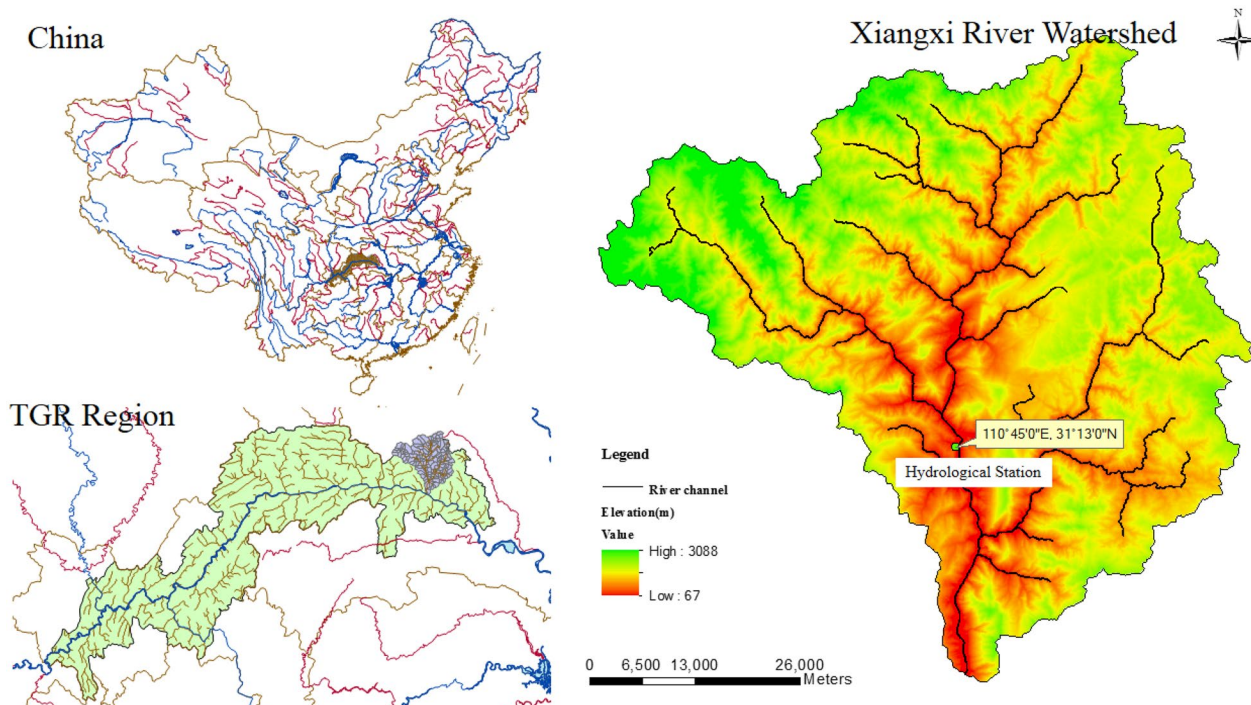
variations of precipitation and the complex terrain and geographical conditions in this area. For example, the peak flow of Xiangxi River reached 1590 m<sup>3</sup>/s on July 2, 1998; the mountain flash flood happened on August 9, 2000, leading to property losses more than three million RMB (Water Conservancy Bureau of Xingshan 2004). Consequently, robust approaches are desired for evaluating the flooding risk in Xiangxi River. Such approaches are expected to be able to reflect interactions among flood peak, volume and duration. Nevertheless, few research was conducted on flooding risk analysis in the Xiangxi River.

The objective of this research is to develop an integrated risk indicator based on interactive analysis of multiple flood variables in the Xiangxi River, China. Such an analysis will be based on provision of bivariate copulas. Notably, systematic evaluation of bivariate hydrologic risks will be undertaken, aiming at revealing significance of effects from persisting high risk levels due to impacts from multiple interactive flood variables. Moreover, the conditional probability density distributions (PDFs) under peak flows with different return periods will be characterized, intending to explore potential control and management practices once a flood has occurred. The aims of this paper are as follows: (i) establishing the bivariate copulas for the three pairs of flooding variables [i.e., flooding peak and volume (P–V), flooding peak and duration (P–D), flooding volume and duration (V–D)] in the Xiangxi River; (ii) choosing the most appropriate copulas for the three pairs of flooding variables based on the RMSE and AIC criteria; (iii) comparing primary, bivariate and secondary return periods; (iv) evaluating the bivariate hydrologic risks; and (v) characterizing the PDFs of flood volume and duration conditional on flood peak flows with different return periods.

## Overview of the studied watershed

### Site description and data collection

The Xiangxi River is located between 30.96–31.67° N and 110.47–111.13° E in Hubei part of China TGR region, draining an area of about 3200 km<sup>2</sup>, as shown in Fig. 1. The Xiangxi River originates in the Shennongjia Nature Reserve with a main stream length of 94 km, and a catchment area of 3099 km<sup>2</sup>, which is one of the main tributaries of the Yangtze River (Han et al. 2014a). The river experiences a northern subtropics climate. Annual precipitation is 1100 mm and ranges from 670 to 1700 mm with considerable spatial and temporal variability (Xu et al. 2010). The main rainfall season is May–September, with a flooding season from July to August. The annual average temperature in this region is 15.6 °C and ranges from 12 to 20 °C. The Xingshan Hydrologic Station (110°45'0" E, 31°13'0" N) is located on the main stem of

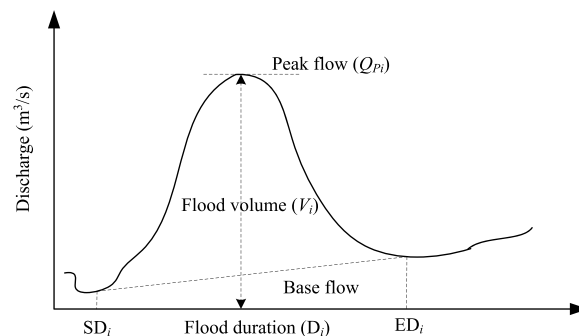


**Fig. 1** The location of the studied watershed

Xiangxi River, with a drainage area of 1900 km<sup>2</sup>. In this study, total fifty years' daily discharge data (1961–2010) from Xingshan Hydrologic Station would be used for probabilistic assessment of flood risks in Xiangxi River.

#### Historical flooding characteristics in Xiangxi River

Based on the daily stream flow data, the annual maximum peak discharge corresponding hydrograph volume and duration values can be obtained. Hence, although the peak discharges are definitely annual maxima, the hydrograph volumes and durations are not necessarily also annual maximums (Sraj et al. 2015). The flood peak applied in this study is defined as the maximum daily flow during the flood event, with flood duration being defined as the total number of days for the flood event, and flood volume being considered as the cumulative flow volume during the flood period. Such flood characteristics are obtained based on the annual scale, meaning in each year one flood would occur. This single-peaked flood hydrograph is shown in Fig. 2. Flood duration (D) can be determined by identifying the time of rise (point “s” in Fig. 2) and fall (point “e” in Fig. 2) of the flood hydrograph. The start of the flood is marked by the sharp rise of the hydrograph and end of the flood runoff is identified by the inflection point on the receding limb of the hydrograph. Between these two points, the total flood volume is estimated. If the rise time of the flood hydrograph is denoted



**Fig. 2** Typical flood hydrograph showing flood flow characteristics (adapted from Ganguli and Reddy 2013)

by SD (day) and fall by ED (day), the flood volume (V) of each flood event is determined using following expression (Yue 2000, 2001; Fan et al. 2018):

$$V_i = (V_i^{total} - V_i^{baseflow}) = \sum_{j=SD_i}^{ED_i} Q_{ij} - \frac{1}{2}(Q_{is} + Q_{ie})(1 + D_i) \quad (1)$$

where  $Q_{ij}$  is the  $j$ th day observed stream flow value for  $i$ th year,  $Q_{is}$  and  $Q_{ie}$  is the observed daily stream flow value on start and end day of the flood hydrograph for  $i$ th year, respectively.  $SD_i$  and  $ED_i$  is the start and end day of a flood event in the  $i$ th year, respectively.  $D_i$  is the flood

duration in the  $i^{\text{th}}$  year. The annual flood peak is obtained by (Fan et al. 2018):

$$Q_i = \max \{Q_{ij}, j = SD_i, SD_{i+1}, \dots, ED_i\}, i = 1, 2, \dots, n. \quad (2)$$

The flood duration can be given by:

$$D_i = ED_i - SD_i, i = 1, 2, \dots, n. \quad (3)$$

Once the flood characteristics are obtained from daily stream flow data, then flood frequency analysis can be analyzed. Table 1 shows some descriptive statistics values of the considered variables (peak discharge,  $Q$ ; hydrograph volume,  $V$ ; and hydrograph duration,  $D$ ). The positive values of kurtosis and skewness suggest that the flood variables can be modeled by sharp and right tailed distributions.

## Methods

### Concept of copula

Copula functions connect univariate marginal distribution functions with the multivariate probability distribution:

$$F(x_1, x_2, \dots, x_n) = C(F_{X_1}(x_1), F_{X_2}(x_2), \dots, F_{X_n}(x_n)) \quad (4)$$

where  $F_{X_1}(x_1), F_{X_2}(x_2), \dots, F_{X_n}(x_n)$  are marginal distributions of random vector  $(X_1, X_2, \dots, X_n)$ . If these marginal distributions are continuous, then single copula function  $C$  exists, which can be written as (Sraj et al. 2015):

$$C(u_1, u_2, \dots, u_n) = F\left(F_{X_1}^{-1}(u_1), F_{X_2}^{-1}(u_2), \dots, F_{X_n}^{-1}(u_n)\right) \quad (5)$$

More details on theoretical background and properties of various copula families can be found in Nelsen (2006). In the following section, brief details of copulas used in the present study are presented.

### One parameter archimedean copulas

A number of copula functions are widely used in practice, mainly including the Archimedean, elliptical, extreme value copulas. Among them, the Archimedean copulas are quite attractive in hydrologic frequency analysis, because they can be easily generated, and are capable of capturing wide range of dependence structure with several desirable properties, such as, symmetry and associativity (Ganguli and Reddy 2013). In the present study, Ali-Mikhail-Haq, Cook-Johnson and Gumbel-Hougaard and Frank copulas are considered for probabilistic assessment of flood risk, which belong to the class of Archimedean copula. In general, a bivariate Archimedean copula can be defined as (Nelsen 2006):

$$C_\theta(u_1, u_2) = \phi^{-1}(\phi(u_1) + \phi(u_2)) \quad (6)$$

where  $u_1$  and  $u_2$  is a specific value of  $U_1$  and  $U_2$ , respectively;  $U_1 = F_{X_1}(x_1)$  and  $U_2 = F_{X_2}(x_2)$ ;  $F_{X_1}$  and  $F_{X_2}$  is the cumulative distribution function (CDF) of random variable  $X_1$  and  $X_2$ , respectively;  $\phi$  is the copula generator that is a convex decreasing function with  $\phi(1)=0$  and  $\phi^{-1}(\cdot)=0$  when  $u_2 \geq \phi(0)$ ; the subscript  $\theta$  of copula  $C$  is the parameter hidden in the generating function. For one parameter copula, the unknown parameter (i.e.,  $\theta$ ) can be estimated using the method of moments with the use of Kendall correlation coefficient (Nelsen 2006). For the copulas with two or more unknown parameters, the maximum likelihood method or maximum pseudo-likelihood method can be selected (Zhang and Singh 2007; Sraj et al. 2015). Table 2 presents some basic characteristics of the applied single-parameter bivariate Archimedean copulas.

**Table 1** Statistical characteristics of flood variables

No.		Flood characteristics		
		Peak (m <sup>3</sup> /s)	Volume (m <sup>3</sup> /s day)	Duration (day)
1	Percentile			
	Minimum	91	72	3
	25%	324	530.8	5
	50%	451.5	713.3	6
	75%	684	1189.5	8
	Maximum	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

### Evaluation of copulas

Since there is a family of copulas, the question is: which copula should be used to obtain joint distributions of flood variables (Zhang and Singh 2006). Various approaches have been proposed for identification of appropriate copulas (Genest and Rivest 1993; Ganguli and Reddy 2013; Sraj et al. 2015). In this study, the procedure for identification of copulas described by Genest and Rivest (1993) would be applied, the detailed steps of such a procedure can be found in Zhang and Singh (2006).

Moreover, the goodness-of-fit statistic tests would to be performed for both marginal and joint distributions through root mean square error (RMSE) and Akaike Information Criterion (AIC). RMSE can be expressed as:



**Table 2** Basic properties of applied copulas

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

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

$$RMSE = \sqrt{E(x_c - x_o)^2} = \sqrt{\frac{1}{N - k} \sum_{i=1}^N [x_c(i) - x_o(i)]^2} \quad (7)$$

where  $E(\cdot)$  is the expectation of the random variable;  $x_c(i)$  denotes the  $i$ th computed value;  $x_o(i)$  denotes the  $i$ th observed value;  $k$  is the number of parameters used in obtaining the computed value;  $N$  is the number of observations. The AIC, developed by Akaike (1974), is also employed to identify the appropriate probability distribution. AIC can be obtained either by calculating the maximum likelihood or by calculating the mean square error of the model, which can be formulated as (Zhang and Singh 2006):

$$\begin{aligned} AIC &= -2 \times \log(\text{maximum likelihood for model}) \\ &\quad + 2 \times (\text{no. of fitted parameters}), \end{aligned} \quad (8)$$

or,

$$AIC = N \times \log(MSE) + 2 \times (\text{no. of fitted parameters}), \quad (9)$$

where,

$$MSE = E(x_c - x_o)^2 = \frac{1}{N - k} \sum_{i=1}^N [x_c(i) - x_o(i)]^2. \quad (10)$$

In the process of the identification of marginal distributions of flood variables, the values of  $x_o$  are presented as the empirical nonexceedance probabilities of the flood variables, and the values of  $x_c$  are presented as the calculated probabilities obtained from the generated marginal distributions. The empirical nonexceedance probabilities for observed values of the flood variables are estimated through the following equation (Gringorten 1963; Cunnane 1978; Adamowski 1985; Zhang and Singh 2006):

$$P_m = \frac{m - 0.44}{N + 0.12} \quad (11)$$

where  $m$  is the index of the  $m$ th smallest observation in the data set arranged in ascending order;  $P_m$  is the probability of the  $m$ th value;  $N$  is the number of the observations.

Meanwhile, in the process of appropriate copula identification, the values of  $x_o$  are presented as the empirical joint frequency (nonexceedance joint probabilities) of the flood variables, and the values of  $x_c$  are presented as the calculated joint probabilities obtained from the generated copula distributions. The joint cumulative frequency can be obtained through the following equation (Zhang and Singh 2006):

$$\begin{aligned} F(x_i, y_i) &= P(X \leq x_i, Y \leq y_i) \\ &= \frac{\text{No. of } (x_j \leq x_i \text{ and } y_j \leq y_i) - 0.44}{N + 0.12}, \end{aligned} \quad (12)$$

$N$  is the number of the observations, and  $i, j = 1, 2, \dots, N$ .

### Conditional distributions

If an appropriate copula function is selected, the conditional joint distribution can then be obtained. Following Zhang and Singh (2006), the conditional distribution function of  $U_1$  given  $U_2 = u_2$  can be expressed as:

$$C_{U_1|U_2=u_2}(u_1) = C(U_1 \leq u_1 | U_2 = u_2) = \frac{\partial}{\partial u_2} C(u_1, u_2) | U_2 = u_2. \quad (13)$$

Similar conditional cumulative distribution for  $U_2$  given  $U_1 = u_1$  can be obtained. Moreover, the conditional cumulative distribution function of  $U_1$  given  $U_2 \leq u_2$  can be expressed as:

$$C_{U_1|U_2 \leq u_2}(u_1) = C(U_1 \leq u_1 | U_2 \leq u_2) = \frac{C(u_1, u_2)}{u_2}. \quad (14)$$

Likewise, an equivalent formula for the conditional distribution function for  $U_2$  given  $U_1 \leq u_1$  can be obtained.

The probability density function (pdf) of a copula function can be expressed as:

$$c(u_1, u_2) = \frac{\partial^2 C(u_1, u_2)}{\partial u_1 \partial u_2}, \quad (15)$$

and the joint pdf of the two random variables can be obtained as:

$$\begin{aligned} f(x_1, x_2) &= \frac{\partial^2 C(u_1, u_2)}{\partial x_1 \partial x_2} = \frac{\partial^2 C(u_1, u_2)}{\partial u_1 \partial u_2} \frac{\partial u_1}{\partial x_1} \frac{\partial u_2}{\partial x_2} \\ &= f_{X_1}(x_1) f_{X_2}(x_2) c(u_1, u_2). \end{aligned} \quad (16)$$

Consequently, the conditional pdf of  $X_1$ , given the value of  $X_2$ , can be formulated as:

$$f(x_1|x_2) = \frac{f(x_1, x_2)}{f_{X_2}(x_2)} = f_{X_1}(x_1) c(u_1, u_2). \quad (17)$$

And the conditional pdf of  $X_2$ , given the value of  $X_1$ , can be expressed as:

$$f(x_2|x_1) = \frac{f(x_1, x_2)}{f_{X_1}(x_1)} = f_{X_2}(x_2) c(u_1, u_2). \quad (18)$$

### Primary and secondary return period

If appropriate copula functions are specified to reflect the joint probabilistic characteristics among peak, duration and volume of the flood, some conditional, primary and secondary return periods can be obtained. Specifically, Joint (primary) return periods called OR and AND can be formulated as (Salvadoriet al. 2007; Graler et al. 2013; Sraj et al. 2015):

$$T_{u_1, u_2}^{OR} = \frac{\mu}{1 - C_{U_1 U_2}(u_1, u_2)}, \quad (19)$$

$$T_{u_1, u_2}^{AND} = \frac{\mu}{1 - u_1 - u_2 + C_{U_1 U_2}(u_1, u_2)}, \quad (20)$$

where  $\mu$  is the mean inter arrival time of the two consecutive flooding events.

The secondary return period, called Kendall's return period, is defined as follows (Salvadoriet al. 2011; Sraj et al. 2015):

$$\bar{T}_{u_1, u_2} = \frac{\mu}{1 - K_C(t)}, \quad (21)$$

where  $K_C$  is the Kendall's distribution, associated with theoretical Copula function  $C_\theta$ . For Archimedean copulas,  $K_C$  can be expressed as (Nelsen 2006):

$$K_C(t) = t - \frac{\phi(t)}{\phi'(t^+)}, \quad (22)$$

where  $\phi'(t^+)$  is the right derivative of the copula generator function  $\phi(t)$ .

### Bivariate hydrologic risk analysis

Risk is the probability of occurrence of an extreme, dangerous, hazardous, or (more generally) undesirable event (Kite 1988). In engineering design of hydrologic infrastructures, risk can be explained as the chance of downstream flooding attributable to uncontrolled water release from upstream flooding facilities (e.g., a reservoir), leading to life and property losses (Gebregiorgis and Hossain 2012). Yen (1970) proposed a formulation for the risk of failure associated with the return period of a flooding event, which can be expressed as:

$$R = 1 - (1 - p)^n = 1 - q^n = 1 - (1 - 1/T)^n, \quad (23)$$

where  $R$  is the risk of failure;  $p$  and  $q$  is the exceedance and nonexceedance probability, respectively;  $T$  is the return period of a flooding event;  $n$  is the design life of the hydraulic structure.

In practical flooding control practice, it is necessary to characterize the flooding event through multiple aspects (e.g., peak and duration) rather than only one flooding variable (e.g., peak). For example, a flood event with high peak flow and long duration may result in serious loss in property, while a short-duration event with high peak may only cause a flash flood. Consequently, bivariate hydrologic risk would be much helpful in taking nonstructural safety measures and developing flood mitigation strategies. In this study, the joint return period in "AND" case is applied to define the bivariate risk analysis as follows:

$$R_{u_1, u_2} = 1 - \left(1 - \frac{1}{T_{u_1, u_2}^{AND}}\right)^n. \quad (24)$$

## Results

### Marginal probability distribution functions of flood variables

First, the univariate flood frequency analyses would be performed based on the historical flooding records. Many parametric distributions have been used to estimate flood frequencies from observed annual flood series, such as the general extreme value distribution in the United Kingdom, Log-Pearson Type-III in the U.S. and Pearson III in China (Adamowski 1989; Kidson and Richards 2005; Wu et al. 2013). In this study, three parametric distribution functions, including Gamma, GEV and Log-normal were used to fit the observed flooding data. The parameters in these three distributions were estimated through maximum likelihood estimation (MLE) method. The expressions for probability functions (PDFs) and the values of their associated unknown parameter estimated through MLE are presented in Table 3.

Figure 3 illustrates the fitted marginal distributions for the three flood variables through Gamma, GEV and Log-normal distribution functions. For empirical CDF of the observations was calculated through Eq. (9). The CDFs for the marginal distributions of flood variables (in Fig. 3) show good agreement between theoretical distributions and the empirical distributions. In detail, all three CDFs (i.e., Gamma, GEV, Lognormal) fit the flood peak and volume better than duration.

The performance of each marginal distribution is evaluated against the empirical nonexceedance probability, which is calculated through Eq. (11), using root mean square error (RMSE) and Akaike Information Criterion (AIC) criteria. The results are presented in Table 4, which provides a comparison of performances for various marginal distributions. From Table 4, the model results indicate that, based on the historical flooding records from 1961 to 2010, the log-normal distribution is the best fit model for peak flow, volume and duration. Although the differences among relative performances of Gamma, GEV and Lognormal are very small on fitting the flood duration (i.e., the RMSE values of Gamma, GEV and Lognormal is 0.0663, 0.0673, and 0.0655, respectively), the Lognormal distribution would be chosen due to its lowest RMSE and AIC values.

#### Dependence of flood variables

The dependence of flood variables was evaluated through the Pearson's linear correlation ( $r$ ), and one non-parametric dependence measure, Kendall's tau. The Pearson's linear correlation, measures the linear dependence between two random variables, but assumes that the underlying distribution is normal, and it is not invariant under monotonic non-linear transformation (Reddy and Ganguli 2012). The Kendall's tau is calculated using ranking of variable values rather than actual values. Therefore, the value of Kendall's tau is invariant under monotonic non-linear transformations and no distributional assumption

is required. Hence, Kendall's tau is more preferred to evaluate the dependence between two random variables with nonlinear relationship in hydrology.

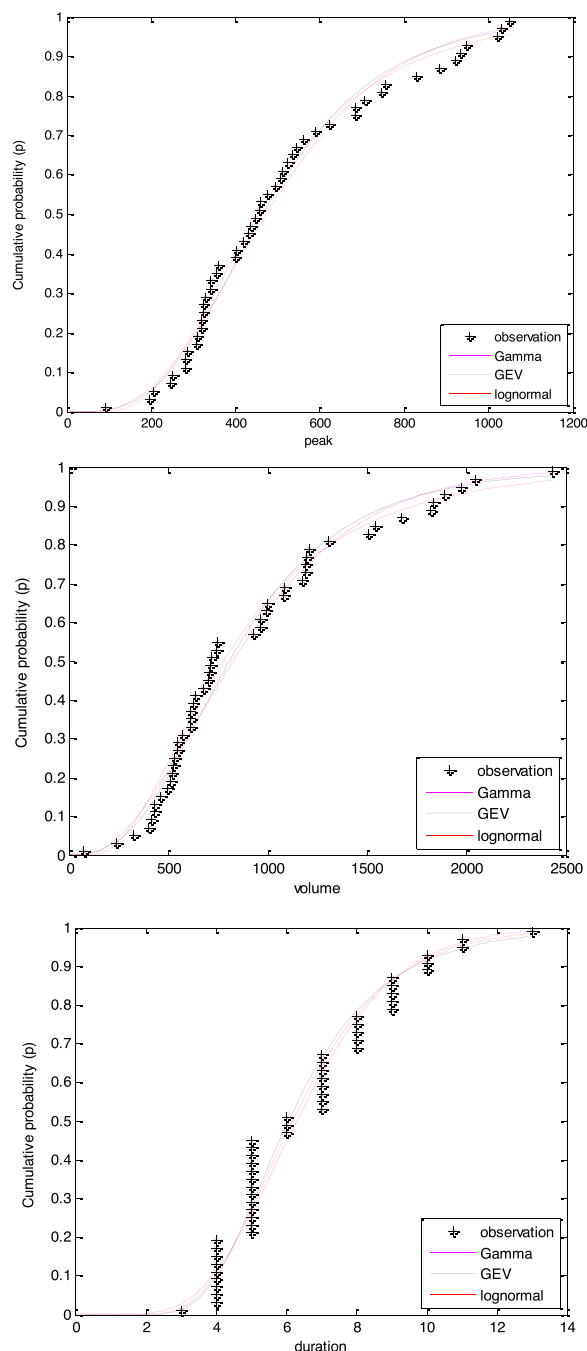
Table 5 presents the values of Pearson's linear correlation coefficient and Kendall's tau among flood variables—flood peak, volume and duration. It can be seen that the values of Pearson's  $r$  and Kendall's tau between peak and volume are highest, followed by those between volume and duration, and then flood peak and duration. In detail, the Pearson, Kendall correlation coefficient values were 0.75 and 0.63 for peak-volume, 0.46 and 0.52 for volume-duration, and  $-0.06$  and  $0.15$  for peak and duration. These results indicate that the correlation between the flooding components of peak and volume would be higher than that for volume and duration. In our case, the correlation coefficient for peak and duration is much smaller than for the other two pairs (i.e., peak-volume and volume-duration), which is consistent with conclusion from previous studies (Grimaldi and Serinaldi 2006; Karmakar and Simonovic 2009; Reddy and Ganguli 2012; Sraj et al. 2015).

#### Joint distributions based on copula method

Four Archimedean families of copulas, including Ali-Mikhail-Haq, Cook-Johnson (Clayton), Gumbel-Hougaard and Frank copulas are applied to model the dependence among flood variables. Since all the three copulas are single-parameter Archimedean copula, the unknown parameter can be estimated by method-of-moments-like (MOM) estimator based on inversion of Kendall's tau. For our current study, the values of Kendall's tau for flood peak-volume and volume-duration are 0.63 and 0.52, respectively. Consequently, the Ali-Mikhail-Haq could not be applied for the pairs of peak-volume and volume-duration since it can only be used with the Kendall's tau value varied between  $-0.18$  and  $0.33$  (Nelsen 2006). Therefore, the Ali-Mikhail-Haq copula may not be applicable for dependence analysis of

**Table 3** Parameters of marginal distribution functions of flood variables

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



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

peak-volume and volume-duration. The joint distribution functions for flood peak and volume, obtained through the four above-mentioned copulas, are shown in Fig. 4; the joint distributions for volume-duration, and peak-duration are shown in Figs. 5 and 6, respectively.

Since there is a class of copulas, investigating the differences among the four chosen copulas and identifying the

most appropriate copulas for further analysis are necessary. In this study, the method for copula identification is based on the process provided by Zhang and Singh (2006). Figures 7, 8, 9 show the comparison of joint cumulative probabilities obtained through empirical equation and copula for flood peak-volume, volume-duration and peak-duration, in which the empirical probabilities were obtained through Eq. (12). For flood peak-volume, all three copulas (excluding Ali-Mikhail-Haq copula) produced a good graphical fit to the empirical probabilities, as shown in Fig. 7. However, the Gumbel-Hougaard and Frank copulas showed better results for plotting the joint probability of flooding peak and volume. As can be seen from Fig. 8, the Gumbel-Hougaard and Frank copulas produced better fits to the empirical probabilities than that of Cook-Johnson copula (Ali-Mikhail-Haq copula was excluded). For flood peak and duration, all the four copulas can be applied. As shown in Fig. 9, Gumbel-Hougaard, Frank and Ali-Mikhail-Haq copulas produced good results, while the Cook-Johnson resulted in under-estimation for the joint probabilities.

To further identify the best copula, the root mean square error (RMSE) (expressed by Eq. (7)), and Akaike information criterion (AIC) (expressed by Eq. (9)) are used to test the goodness of fit of sample data to the theoretical joint distribution obtained using copula functions. Table 6 shows the comparison of RMSE and AIC values for joint distributions obtained through different copula functions for flood peak-volume, peak-duration and volume-duration. The results indicate that the Gumbel-Hougaard and Frank copulas performed better for modelling the joint distributions for flood peak-volume and volume-duration than Cook-Johnson copula. The differences between Gumbel-Hougaard and Frank copulas in quantifying the joint probabilities of flood peak-volume and volume-duration are rarely small. For example, the RMSE value for the Gumbel-Hougaard and Frank copula of peak-volume is 0.0312 and 0.0304, respectively, while the AIC value is  $-344.8576$  and  $-347.4879$ , respectively. Based on the values of RMSE and AIC, it can be concluded that the Frank copula would be best for quantifying the joint distributions of flood peak-volume and volume-duration, while the Ali-Mikhail-Haq copula performs better in modelling the joint distribution of flood peak-duration than the other three copulas.

#### Conditional cumulative distribution functions and return periods of flood characteristics

Based on the results presented in Table 6, the conditional cumulative distribution functions (CDFs) for peak-volume, and volume-duration can be obtained through the Frank copula, while the conditional CDFs for peak-duration can be quantified by the Ali-Mikhail-Haq copula.



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

PDF	RSME			AIC		
	Peak	Volume	Duration	Peak	Volume	Duration
Gamma	0.0378	0.0445	0.0663	− 323.5512	− 307.1904	− 267.3406
GEV	0.0340	0.0428	0.0673	− 332.2144	− 309.1165	− 263.8541
Lognormal	0.0265	0.0361	0.0655	− 358.8192	− 328.2335	− 268.6465

**Table 5** Values of correlation coefficients for flooding characteristics

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

Figure 10 shows the conditional CDFs of flooding variables for the Xiangxi River at Xingshan Station. It can be seen that, among the flooding pairs peak-volume, peak-duration and volume-duration, the values of conditional CDF for one flooding variable would decrease as the value of other flooding variable increase. This indicates positive correlation structures between peak-volume, peak-duration, and volume-duration. Besides, the decreasing trend of conditional CDFs for peak-duration is less than the other two pairs, indicating less correlation structures between peak and duration. This is consistent with the results presented in Table 5.

The multivariate flooding frequency analysis is helpful in understanding critical hydrologic behavior of flood through analyzing the concurrence probabilities of various combinations of flooding characteristics. The joint return period and second return period are calculated based on Eqs. (19)–(21) to reflect the historical flooding characteristics. Table 7 presents the primary return periods of peak, volume, and duration obtained by univariate marginal distributions, and joint return periods for “AND” and “OR” cases for bivariate distributions. In general, the joint return period in “AND” case is longer than the joint return period in “OR” case when same univariate return period is assumed. For three pairs of flooding variables (i.e., peak-volume, peak-duration, volume-duration), the joint period in “OR” case do not vary significantly with the same primary joint periods. Conversely, the joint period in “AND” case for peak-volume is much longer than the other two flooding pairs. For example, consider the primary return periods of peak, volume, duration are 100 years. The joint return period  $T_{PV}^{AND}$  and  $T_{DV}^{AND}$  is 1255.64 and 1710.98 years, respectively, while the joint return period  $T_{DV}^{AND}$  is 5532.43 years. This is due

to the less correlation between flooding peak and duration, as shown in Table 4. Also, the secondary return periods are also presented in Table 7. The secondary return period can be useful for analyzing risk of supercritical flood events, which is defined as the average time between the occurrence of two supercritical flooding events. As the primary return period increases, the probability of supercritical flooding events decreases, leading to increase of the secondary return period. Furthermore, the secondary return period is always higher than that of the primary return period and the join return periods of  $T^{OR}$  and  $T^{AND}$ .

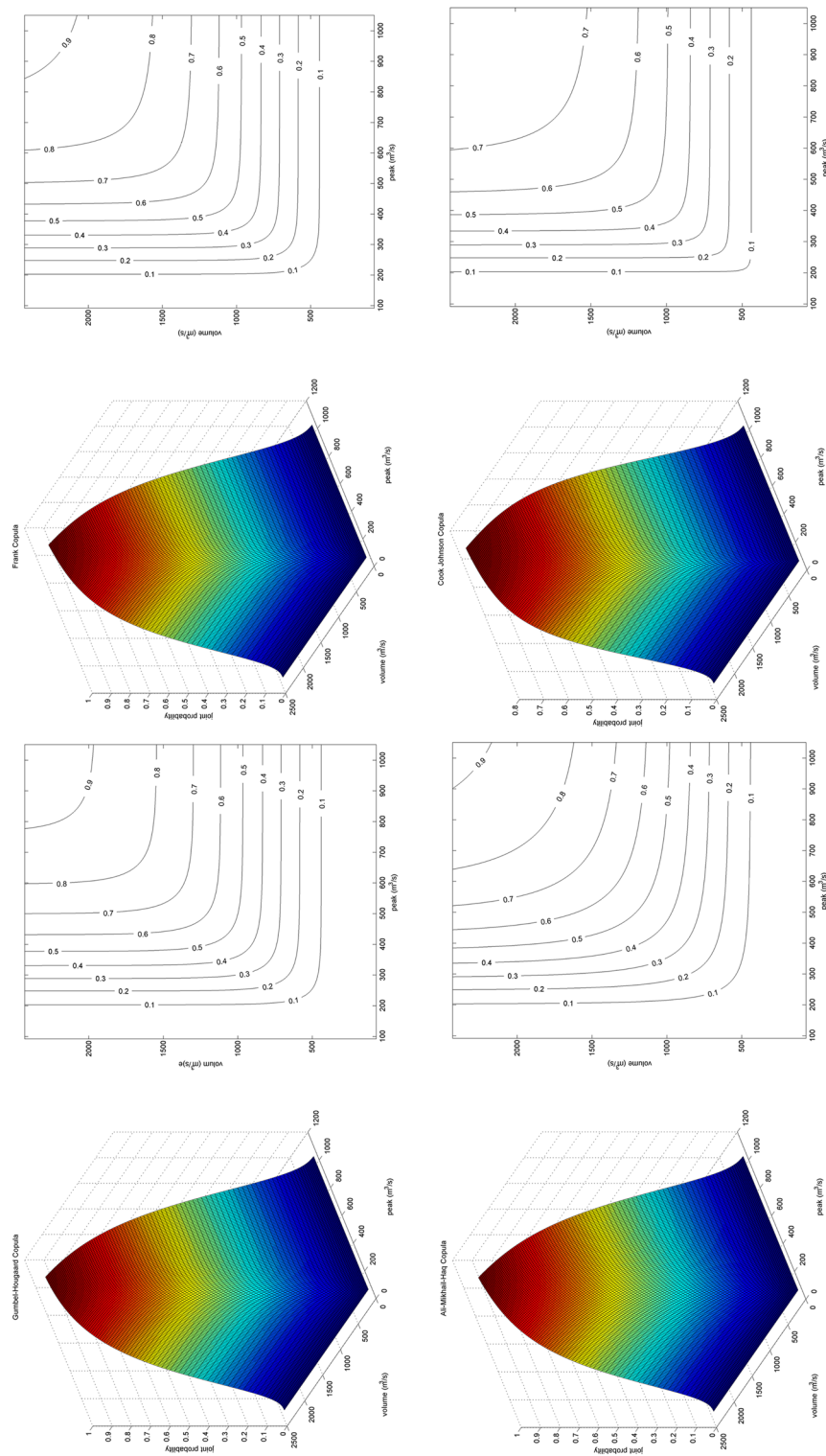
#### Bivariate hydrologic risk analysis for Xiangxi River

For one flooding event, the failure of hydraulic structures is mainly due to high peak flow. Therefore, the flooding peak flow would be the essential factor to be considered for analyzing hydrologic risks. However, other flooding variables (i.e., flooding duration and volume) are also quite critical for actual flooding control and mitigation. In detail, the flooding duration is the vital factor for decision maker in characterizing the flooding control pressure, while the flooding volume is related to the diversion of flooding. Consequently, an integrated risk analysis framework to consider more flooding variables, can be more helpful for actual flood control than the traditional risk analysis in which only the flooding peak is considered. Therefore, in this study, bivariate hydrologic risk analysis method in considering peak-duration and peak-volume would be proposed to identify the inherent flooding characteristics in the Xiangxi river basin.

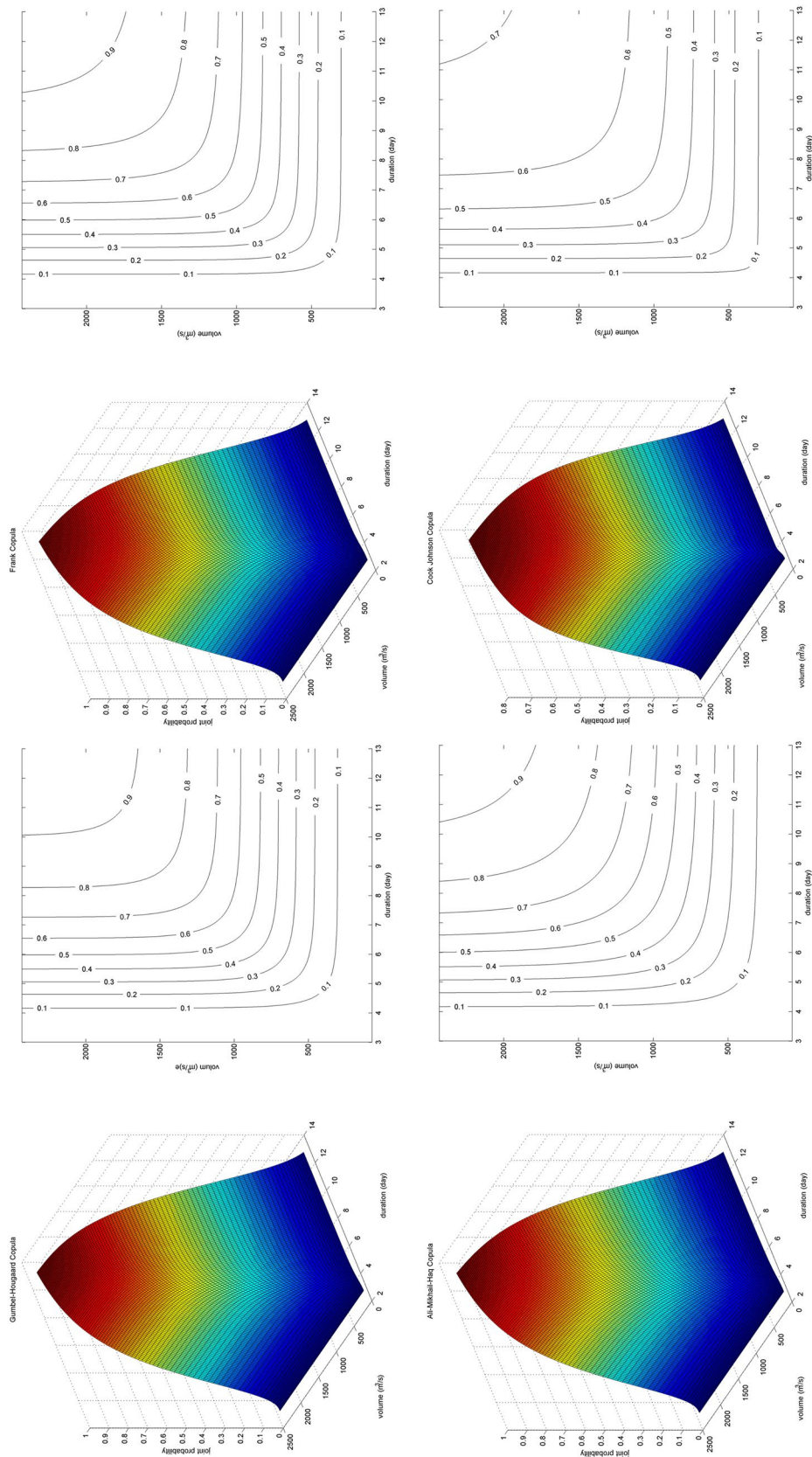
There are no reservoirs in the Xiangxi River near Xingshan station. Consequently, we mainly analyze the failure risk for river levee around Xingshan Station. Three designed flows are considered for the river levee of Xiangxi River near Xingshan station, which are 1000, 1200 and 1500 m<sup>3</sup>/s with the return period being about 20, 50, and 100 years, respectively. Four service time scenarios are also assumed for the river levee, namely 10, 20, 50 and 100 years.

#### Bivariate risk analysis for flooding peak flow and duration

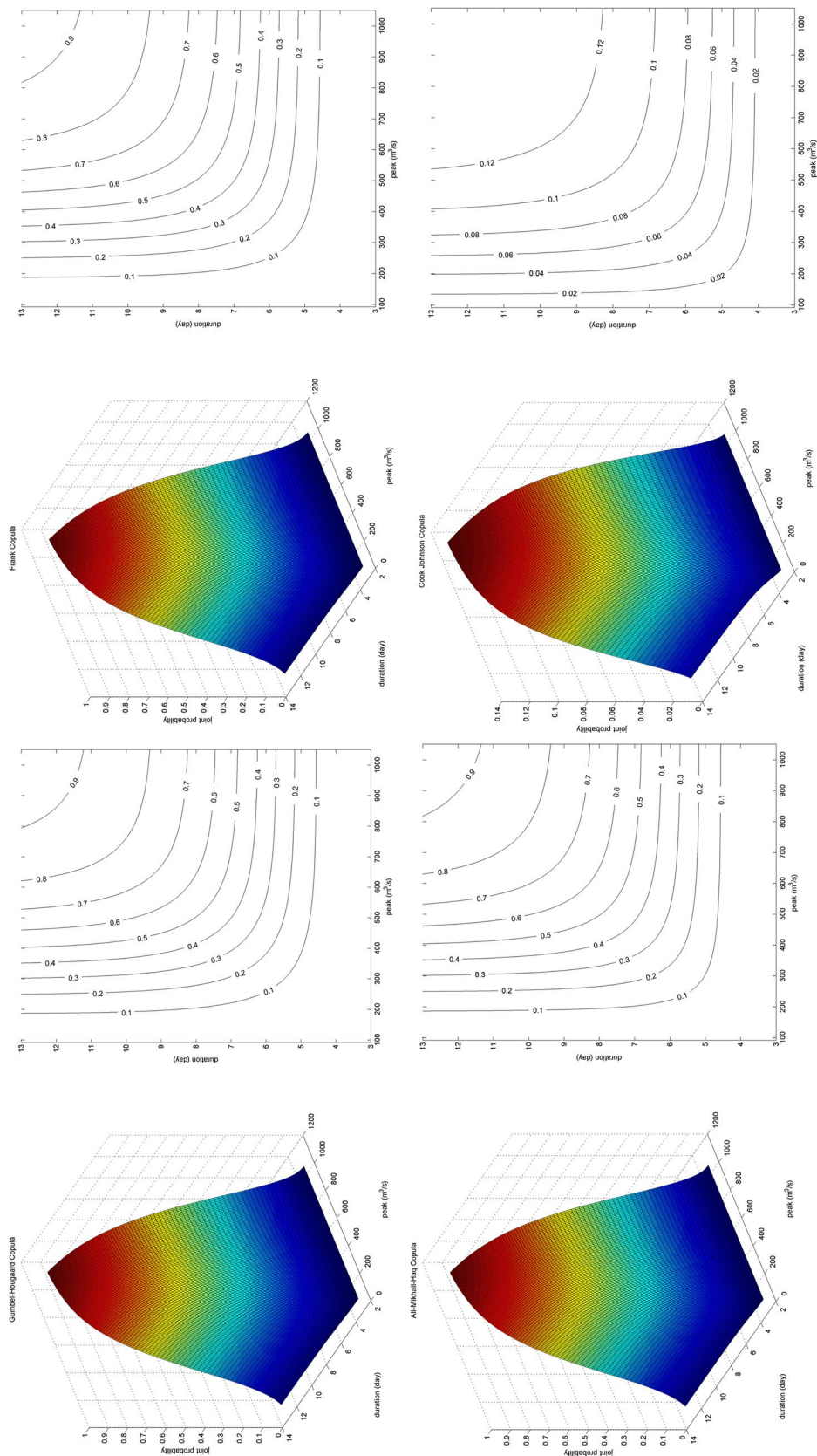
Figure 11 shows the variations in the failure risk of river levee around Xingshan Station under different flooding



**Fig. 4** Joint probability distribution functions for peak flow and volume through different copulas

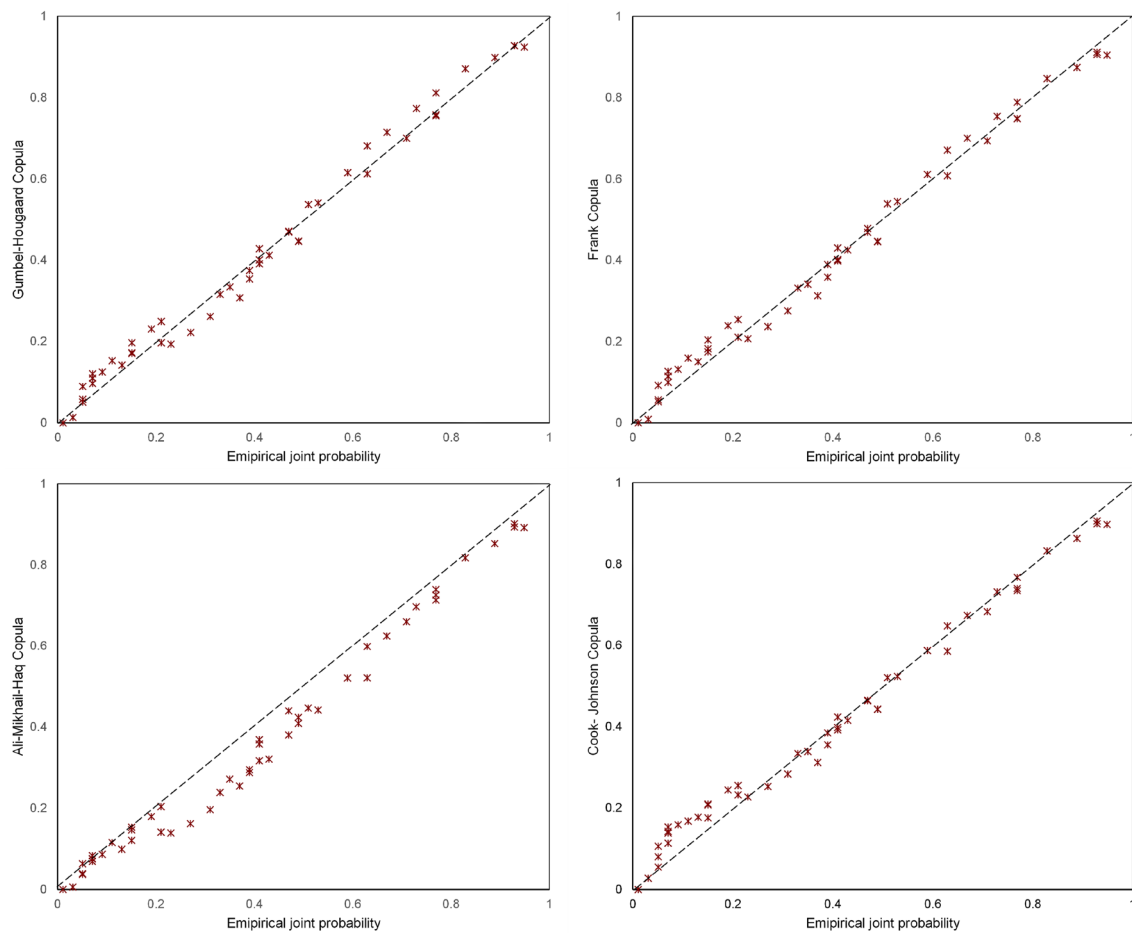


**Fig. 5** Joint probability distribution functions for flood duration and volume through different copulas



**Fig. 6** Joint probability distribution functions for flood peak and duration through different copulas





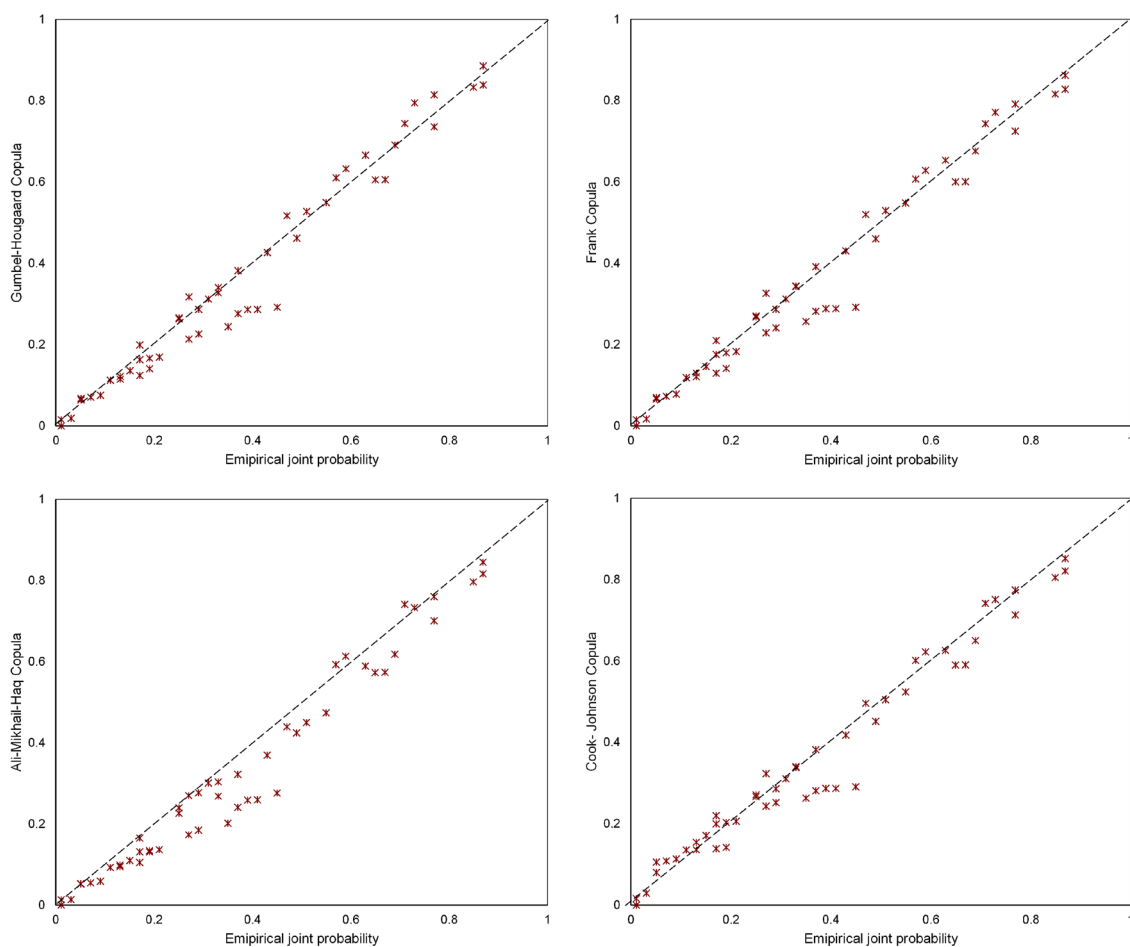
**Fig. 7** Comparison of joint cumulative probabilities obtained through empirical equation and copula for flood peak and volume

peak-duration scenarios. From this figure, it can be seen that for the same service time, the risk would decrease as the increase of the designed flow. Similarly, for the same designed flow, the failure risk of river levee would increase as the increase of service time. For example, as presented in Table 8, if the service time of the river levee is designed to be 10 years, the failure risk of the river levee for the designed flow 1000, 1200 and 1500  $\text{m}^3/\text{s}$  would be 43.6, 22.34 and 10.97%, respectively. Meanwhile, the potential risk of designed flow 1000, 1200, 1500  $\text{m}^3/\text{s}$  under 10-year service time would be 43.6, 22.34, 10.97%, respectively.

For bivariate hydrologic risk analysis, one of the major characteristics is that the variation of the hydrologic risk can be reflected with respect to flooding duration or volume. Figure 11 also shows the changing trends of the flooding risk with respect to flooding duration under different designed flows and service time. In this figure, the initial risk values (points on the y-coordinate) are obtained through Eq. (23) without considering the flooding duration scenarios, while the points on the solid,

dashed and asterisk lines are derived based on Eq. (24). The solid, dashed and asterisk lines in Fig. 11 indicate that, the bivariate risk of flooding peak flow and duration would keep constant for some time and then decrease with the increase of flooding duration for all designed flows and service time. However, the detailed decreasing points for different designed flow under different service time are also different. As presented in Table 8, for a designed flow of 1000  $\text{m}^3/\text{s}$  and a service time of 10 years, the failure risk of the river levee would keep at about 40% and then decrease significantly if the flooding duration is larger than 6 days. Conversely, the risk for the designed flow of 1000  $\text{m}^3/\text{s}$  and service time of 100 years would not change significantly until the duration larger than 8 days.

The bivariate risk of the flooding peak flow and duration is much meaningful for actual hydrologic facility design and potential flooding control. In practical engineering hydrologic facility construction, the return period of peak flow would be the key factor to be considered. Moreover, in addition to the hydrologic facility construction, materials for flood defense such sand, wood,



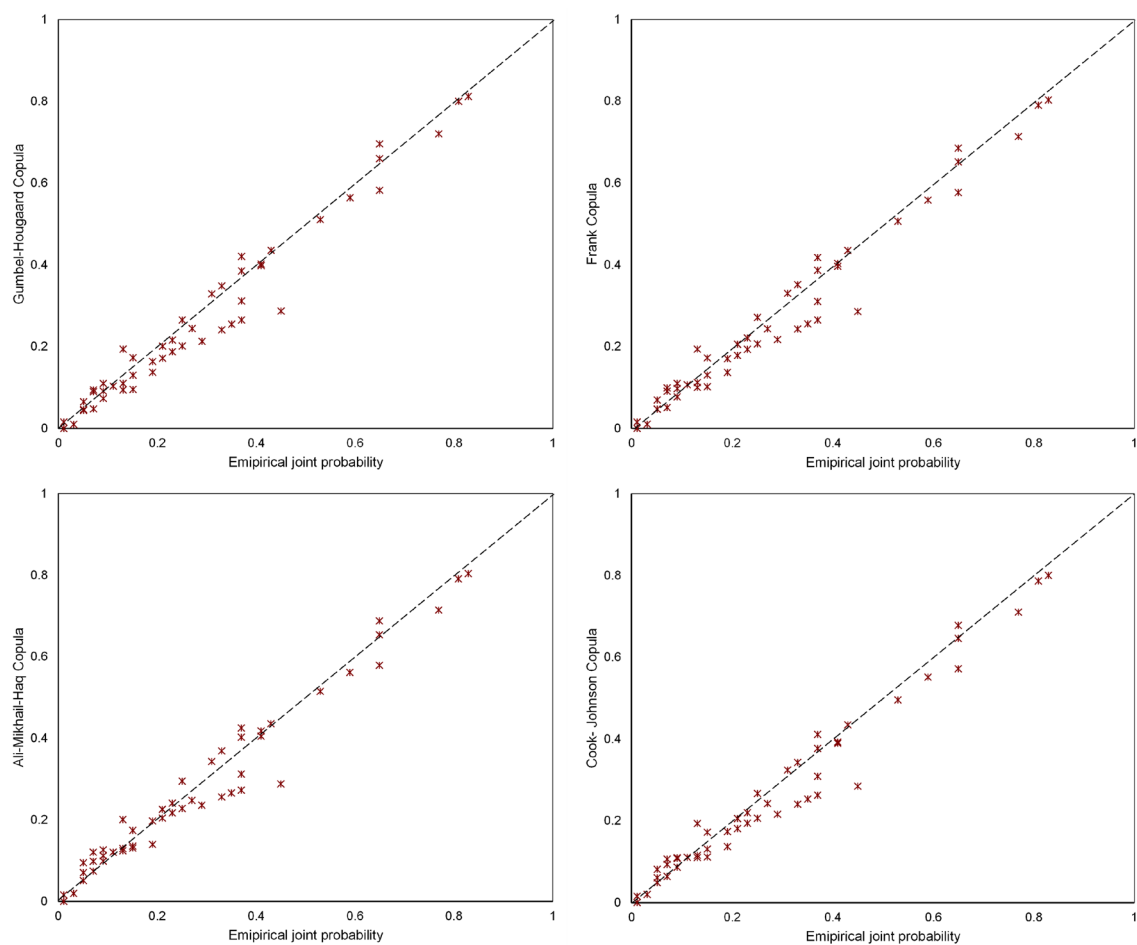
**Fig. 8** Comparison of joint cumulative probabilities obtained through empirical equation and copula for flood duration and volume

bags and so on, usually need to be prepared for flooding control at some important spots (e.g., near cities) of the river levee. In the preparation of those flood defense materials, the flood risk with respect to the flooding duration can be a useful reference. For example, as presented in Fig. 11 and Table 8, if the river levee is constructed with a designed flow of  $1000 \text{ m}^3/\text{s}$  and 10-year service time, the flooding risk with the flooding duration of 2, 4, 6, 8, 10-day would be 43.6, 42.67, 34.08, 18.52, 7.45%, respectively. Such risk values can be considered as references for decision makers to determine how much materials would be prepared for flood defense.

#### **Bivariate risk analysis for flooding peak flow and volume**

The bivariate hydrologic risk for flood peak flow and volume indicates the probability of co-occurrence of flood peak flow and volume values. Similar with the bivariate hydrologic risk for flood peak-duration, the initial risk values are derived based on Eq. (23) for different designed flows and service time, and such risk values

would decrease as the increase of flood volume, as shown in Fig. 12. Table 9 presents the detailed flooding risk values for different designed flows and service time, with respect to different flood volume scenarios. In general, the bivariate risk values for flood peak-volume would not decrease significantly for all designed flows and service time periods at low flooding volumes. This suggests that the occurrence of one flooding peak flow would usually be accompanied with some flood volumes, e.g.  $1500 \text{ m}^3/\text{s}$ , as presented in Table 9. However, as shown in Fig. 12, for one designed flow and service time period, the increase of flood volume would lead to decrease for the bivariate risk for flood peak flow and volume. As can be found in Table 8, for a designed flow of  $1000 \text{ m}^3/\text{s}$  and a service time period of 100 years, the failure risk of the river levee would be more than 95% with the flood volume being less than  $2000 \text{ m}^3/\text{s}$ . Such a risk value would be decrease to 87.81, 66.43, 43.33 and 26.18 with the flood volume increasing to 2500, 3000, 3500 and  $4000 \text{ m}^3/\text{s}$ , respectively.



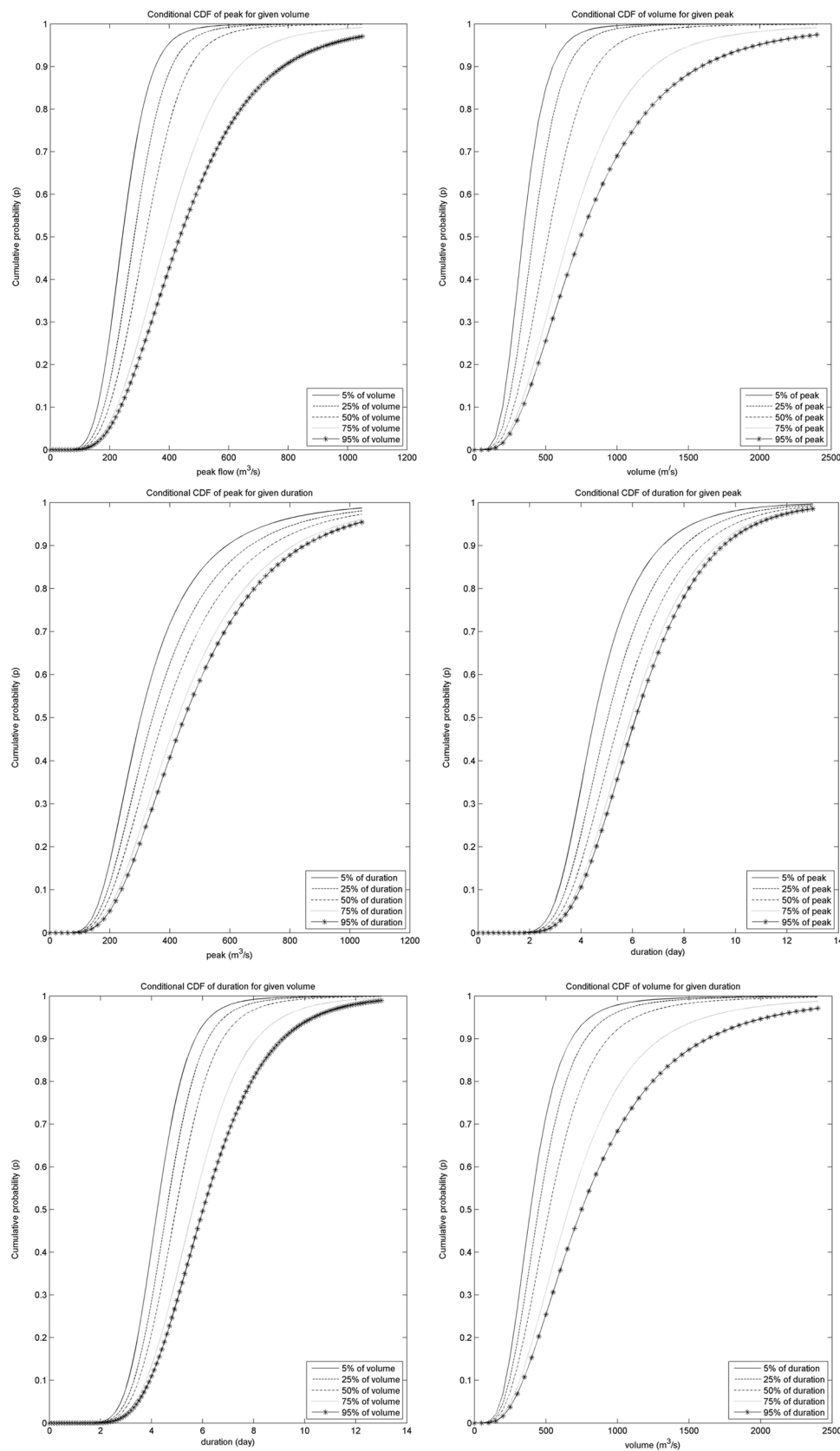
**Fig. 9** Comparison of joint cumulative probabilities obtained through empirical equation and copula for flood peak and duration

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

Copula	RMSE			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

The implication for the bivariate risk of flooding peak flow and volume is to provide decision support for hydrologic facility design and establishment of flooding diversion areas. The actual flooding control practices, the excess water of floods can be redirected temporary holding ponds or other bodies of water with a lower risk or impact to flooding. For example, in China, the flood diversion areas are rural areas that are deliberately flooded in emergencies in order to protect cities. In flood

diversion practice, the bivariate risk for flood peak flow and volume would be an important reference for the design of flooding diversion areas. For example, as shown in Fig. 12 and Table 9, for the river levee with a designed flow of 1000 m<sup>3</sup>/s and 10-year service period, the flooding risk value would be 43.60, 43.55, 41.39, 31.44, 18.98, 10.34% with a flood volume being 500, 1000, 1500, 2000, 2500, and 3500 m<sup>3</sup>/s, respectively. These risk values suggest the flood diversion area be at least designed based

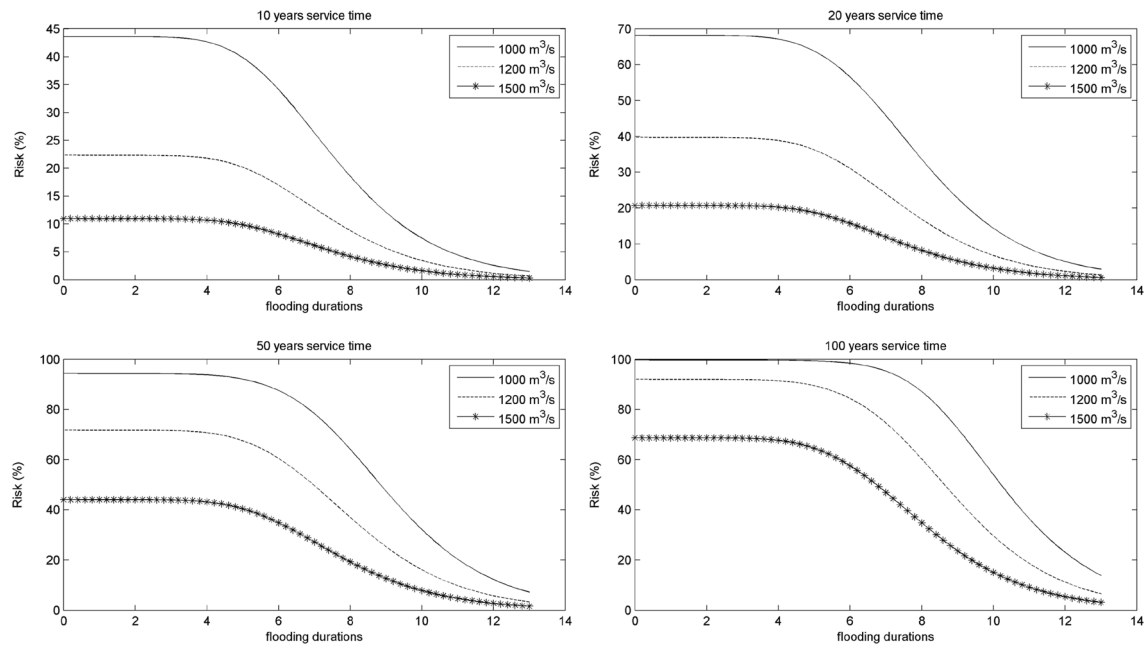


**Fig. 10** Conditional cumulative distribution function of flooding variables



**Table 7** Comparison of univariate, bivariate return periods for flood characteristics (year)

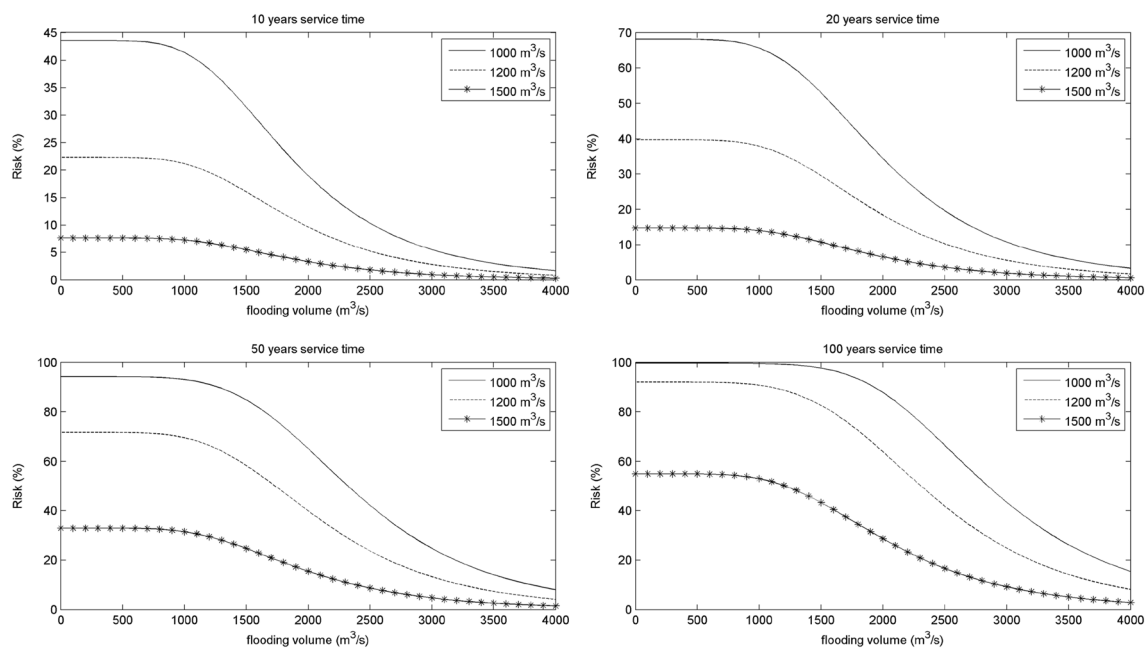
T	Peak (m <sup>3</sup> /s)	Volume (m <sup>3</sup> /s day)	Duration (day)	$T_{PV}^{AND}$	$T_{PD}^{AND}$	$T_{DV}^{AND}$	$T_{PD}^{OR}$	$T_{DV}^{OR}$	$T_{PV}^{OR}$	$\bar{T}_{DV}$	$\bar{T}_{PD}$	$\bar{T}_{PV}$
5	689.54	1313.48	8.25	7.66	16.18	8.80	2.96	3.49	3.71	6.54	9.56	6.06
10	857.38	1730.89	9.60	21.19	59.78	25.78	5.46	6.20	6.54	16.83	32.61	14.70
20	1026.38	2173.89	10.87	65.75	229.22	84.03	10.46	11.35	11.79	49.39	119.82	40.48
50	1256.75	2809.50	12.51	338.46	1395.49	452.35	25.46	26.46	26.99	243.62	710.43	186.97
100	1438.39	3333.42	13.74	1255.64	5532.43	1710.98	50.46	51.51	52.07	889.62	2791.36	662.29



**Fig. 11** Bivariate flooding risk under different flooding peak-duration scenarios

**Table 8** The flooding risk in Xiangxi river under different peak-duration scenarios (%)

Service time	10 years			20 years			50 years			100 years		
Designed flow (m <sup>3</sup> /s)	1000	1200	1500	1000	1200	1500	1000	1200	1500	1000	1200	1500
Duration (day)												
2	43.60	22.34	10.97	68.19	39.69	20.73	94.29	71.76	44.06	99.67	92.02	68.71
4	42.67	21.80	10.69	67.13	38.85	20.24	93.81	70.76	43.18	99.62	91.45	67.71
6	34.08	16.95	8.21	56.55	31.03	15.75	87.55	60.50	34.85	98.45	84.39	57.55
8	18.52	8.81	4.19	33.60	16.85	8.20	64.08	36.96	19.25	87.10	60.26	34.79
10	7.45	3.45	1.62	14.34	6.78	3.21	32.09	16.09	7.83	53.88	29.59	15.05
12	2.58	1.18	0.55	5.09	2.35	1.10	12.25	5.77	2.72	23.00	11.20	5.38
14	0.85	0.39	0.18	1.69	0.77	0.36	4.16	1.92	0.90	8.16	3.80	1.78
16	0.28	0.13	0.06	0.55	0.25	0.12	1.37	0.63	0.29	2.73	1.25	0.58
18	0.09	0.04	0.02	0.18	0.08	0.04	0.46	0.21	0.10	0.91	0.41	0.19
20	0.03	0.01	0.01	0.06	0.03	0.01	0.15	0.07	0.03	0.31	0.14	0.07



**Fig. 12** Bivariate flooding risk under different flooding peak-volume scenarios

**Table 9** The flooding risks in Xiangxi river under different peak-volume scenarios (%)

Service time	10 years			20 years			50 years			100 years		
Designed flow (m <sup>3</sup> /s)	1000	1200	1500	1000	1200	1500	1000	1200	1500	1000	1200	1500
Volume (m <sup>3</sup> /s)												
500	43.60	22.34	7.67	68.19	39.70	14.76	94.29	71.76	32.92	99.67	92.03	55.00
1000	43.55	22.32	7.67	68.13	39.65	14.74	94.27	71.71	32.88	99.67	92.00	54.95
1500	41.39	21.19	7.28	65.64	37.90	14.03	93.08	69.61	31.46	99.52	90.76	53.03
2000	31.44	16.06	5.51	52.99	29.54	10.71	84.85	58.32	24.67	97.70	82.63	43.25
2500	18.98	9.67	3.31	34.36	18.41	6.52	65.09	39.86	15.51	87.81	63.83	28.61
3000	10.34	5.26	1.80	19.61	10.25	3.57	42.06	23.68	8.69	66.43	41.76	16.63
3500	5.52	2.81	0.96	10.74	5.54	1.91	24.72	13.27	4.71	43.33	24.78	9.21
4000	2.99	1.52	0.52	5.89	3.02	1.04	14.08	7.37	2.57	26.18	14.20	5.08

on a volume of 1500 m<sup>3</sup>/s, such the bivariate risk would not decrease significantly for the flood volume less than 1500 m<sup>3</sup>/s.

#### Conditional probability density functions of flood characteristics

In addition to derive the conditional cumulative distribution functions and conditional return periods based on the best-fitted copula for the historical flooding data, the conditional probability density functions of the flooding variable can also be generated based on Eqs. (15)–(18). In flooding risk analysis, the peak flow would be the critical factor to judge whether a flood would occur.

However, once the flood occurred, the severity of the flood would also influenced by other flooding variables such as flood duration and volumes. In detailed, the flood duration would be related to the flooding control pressure in which flood defense materials should prepared for strengthening the river levee and some inspectors should cruise along the river to confirm the safety of the levee. In another side, the flood volume would highly influence the flood diversion practices, in which excess water would be diverted to temporary holding ponds with lower risk in order to protect cities.

Figure 13 shows the distributions of flood volume conditional on the flood peak flows with different return

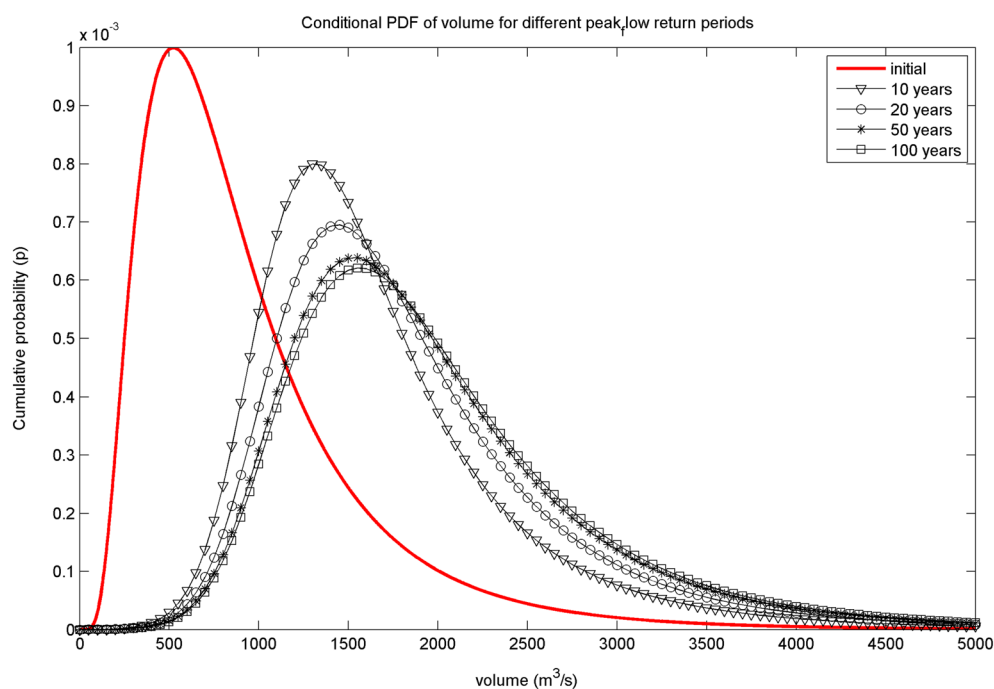
periods. In this study, the flood peak flows with return periods of 10, 20, 50, and 100-year are under consideration. Each curve represents the probability distribution function (PDF) of flood volume associated with a flood peak flow with a particular return period. According to the PDFs of flood volume, as the increase of flood return period, which is equivalent to higher peak flow of the flood, the flood volume would be expected to be higher as well. For example, as shown in Table 10, if a flood with 10-year return period occurred, the flood volume is likely less than those accompanied with a flood with 20-year return period. Moreover, as can be seen from Table 10, the higher peak flows would usually lead to increases in the mean and standard deviation values of the flood volume PDFs. However, the increasing rate would generally decrease. For example, the mean values of the flood volume PDFs accompanied with the flood peaks with 10 and 20-year return period would be 1658.09 and 1814.52  $\text{m}^3/\text{s}$ , respectively, and the standard deviations would be 682.97 and 735.76  $\text{m}^3/\text{s}$ , respectively. In comparison, the mean values associated with the flood peaks with 50 and 100-year return period would be 1913.47 and 1947.16  $\text{m}^3/\text{s}$ , respectively, and the standard deviation values would be 765.90 and 775.58  $\text{m}^3/\text{s}$ , respectively. Such PDFs of flood volume conditional on different flood peak flows can be considered as the references for flooding diversion practices and be involved in the flood

optimization models to determine the capacities of flooding diversion.

Figure 14 shows the distributions of flood duration conditional on the flood peak flows with different return periods. It is indicated that, the increase of flood return period would not significantly change the PDFs of the flood duration. As presented in Table 10, the mean and standard deviation values of the PDFs associated with the flood with 10-year return period would be 7.44 and 2.32 days, while those values accompanied with the flood with 100-year return period would be 7.57 and 2.33 days. Such low impact of flood peak on the flood duration is due to the low correlation between flood peak and duration, as presented in Table 5. The engineering implications of the PDFs of flood duration conditional on flood peak can be useful for flood control. Once a flood occurred, the PDF of flood duration conditional on the flood can be provided as a reference to determine how much flood defense materials should prepared for strengthening the river levee and how long the inspectors should cruise along the river to confirm the safety of the levee.

## Discussion and conclusions

In the present work, bivariate hydrologic risk for the Xiangxi River in the Three Gorges Reservoir area, China was analyzed based on bivariate copula methods. In the bivariate hydrologic risk analysis framework,



**Fig. 13** Probability density functions of volume under different peak flow return periods

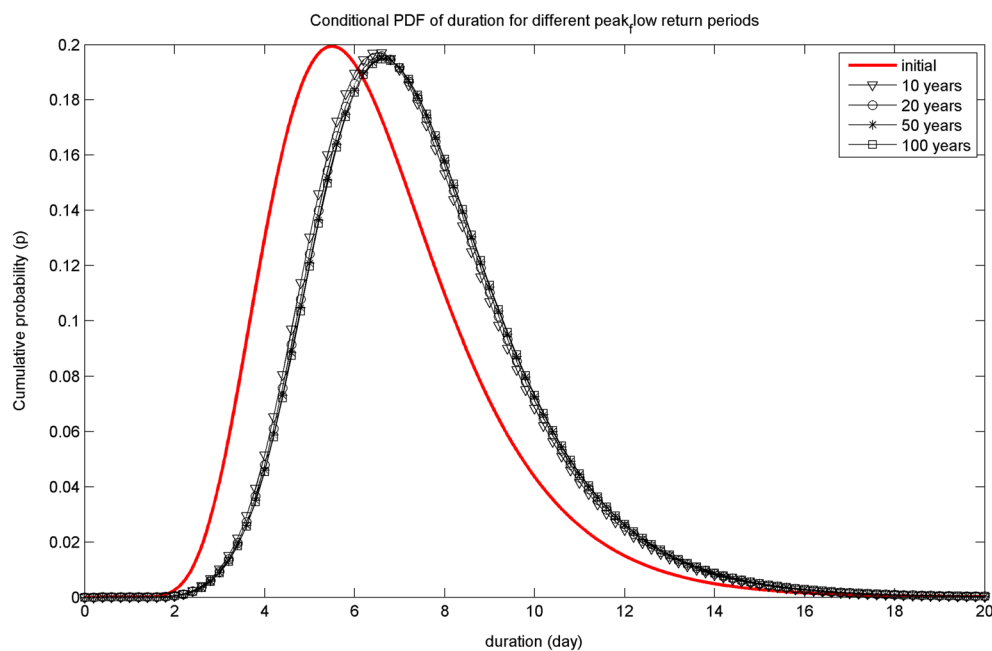
**Table 10** Statistical characteristics of the conditional PDFs of flooding duration and volume under different peak flow return periods

Flooding variables	Index	Initial	Return periods of peak flow (year)			
			10	20	50	100
Duration	Mean	6.55	7.44	7.51	7.55	7.57
	Std	2.30	2.32	2.33	2.33	2.33
	Kurtosis	1.72	1.51	1.47	1.45	1.44
	Skewness	1.04	0.95	0.94	0.93	0.93
Volume	Mean	934.21	1658.09	1814.52	1913.47	1947.16
	Std	622.80	682.97	735.76	765.90	775.58
	Kurtosis	5.05	2.74	1.85	1.41	1.28
	Skewness	1.87	1.43	1.25	1.15	1.12

the bivariate frequency analysis, which considered the flooding variables pairs of flood peak, duration and volume, was firstly conducted through the Ali-Mikhail-Haq (AMH), Cook-Johnson, GumbelHougaard (GH), and Frank copulas. The root mean square error (RMSE) and Akaike Information Criterion (AIC) values were then employed to choose the most appropriate copula in modelling joint distributions of the flooding variable pairs. The primary, conditional and secondary return periods were then derived based on the selected copula. The bivariate hydrologic risk was defined based on the joint return period of flooding variables to reflect the hydrologic risks of flood peak-duration and flood peak-volume pairs. Besides, the conditional probability distribution

functions (PDFs) of flood volume and duration under different flood peak scenarios were also derived to explore the variation in PDFs of flood volume and duration corresponding to different flood peak flows.

The results indicated that the correlation coefficient for flood peak and duration is much smaller than the other two pairs of flood variables (i.e., flood peak-volume and flood volume-duration). For the four Archimedean copulas, the Frank copula would be best for quantifying the joint distributions of flood peak-volume and volume-duration, while the Ali-Mikhail-Haq copula performed better in modelling the joint distribution of flood peak-duration. The joint return period in “AND” case was much longer than the joint return



**Fig. 14** Probability density functions of duration under different peak flow return periods



period in “OR” case when same univariate return period was assumed. Moreover, the secondary return period was always higher than that of the primary return period and the joint return periods of  $T^{OR}$  and  $T^{AND}$ , indicating the low probability of the occurrence of the supercritical flooding event. The bivariate risk for flood peak flow-duration indicated that, as the increase of the flood duration, the probability of the flood peak and duration would keep constant for some time and then decrease for all designed flows and service time. Such bivariate risk values could provide decision support for flood control. Moreover, the bivariate risk for flood peak-volume exhibited similar trend with that for flood peak-duration, which could be helpful for establishment of flooding diversion areas. Finally, the conditional probability density functions of the flood duration and volume for given flood peak flows could be applied to reflect the severity of a flood. The results indicated that the distributions of flood volume would be highly influenced by the flood peak flows, in which the flood volume would be expected to increase as the increase of flood return period. Those distributions could support the flood diversion practices once a flood occurred. The probability of flood duration conditional on flood peak flows stated that the increase of flood return period would not significantly change the statistical characteristics of the flood duration, which is due to low correlation between flood peak and duration.

The accuracy of the copula method in modelling the joint probability of flood variables is influenced by many factors, such as the performance of the marginal distribution of the flood variable, the algorithm used for estimating the unknown parameters in the marginal distributions and copulas. In current study, the lognormal distribution was employed to model the marginal distributions of the flood peak flow, volume and duration. Consequently, further studies are still required to improve the performance of the copula methods through using more accurate methods to model the marginal distributions of the three flood variables.

#### Acknowledgements

The authors are grateful to the editors and the anonymous reviewers for their insightful comments and suggestions.

#### Author contributions

YF: design of the study, data collection, coding the model and paper draft. The author read and approved the final manuscript.

#### Funding

This study is supported by the Brunel University Open Access Publishing Fund.

#### Availability of data and materials

The datasets used or analyzed during the current study are available from the corresponding author upon reasonable request.

## Declarations

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

Y. R. Fan is an editorial board member of *Environmental Systems Research*. The author has no further conflicts of interest to declare.

Received: 4 July 2021 Accepted: 15 September 2022

Published: 29 September 2022

## References

- Adamowski K (1985) Nonparametric kernel estimation of flood frequencies. *Water Resour Res* 21(11):1585–1590
- Adamowski K (1989) A Monte Carlo comparison of parametric and nonparametric estimation of flood frequencies. *J Hydrol* 108:295–308
- Akaike H (1974) A new look at the statistical model identification. *IEEE Trans Autom Control* 19(6):716–723. <https://doi.org/10.1109/TAC.1974.1100705>
- Chebana F, Dabo-Niang S, Ouarda TBMJ (2012) Exploratory functional flood frequency analysis and outlier detection. *Water Resour Res* 48:W04514
- Cunnane C (1978) Unbiased plotting positions—a review. *J Hydrol* 37(3):205–222
- Fan YR, Huang GH, Veawab A (2012) A generalized fuzzy linear programming approach for environmental management under uncertainty. *J Air Waste Manag Assoc* 62(1):72–86
- Fan Y, Huang G, Zhang Y, Li Y (2018) Uncertainty quantification for multivariate eco-hydrological risk in the Xiangxi River within the Three Gorges Reservoir Area in China. *Engineering* 4(5):617–626
- Fan Y, Huang K, Huang G, Li Y (2020a) A factorial Bayesian copula framework for partitioning uncertainties in multivariate risk inference. *Environ Res* 183:109215. <https://doi.org/10.1016/j.envres.2020.109215>
- Fan Y, Huang K, Huang G, Li Y, Wang F (2020b) An uncertainty partition approach for inferring interactive hydrologic risks. *Hydrol Earth Syst Sci* 24(9):4601–4624. <https://doi.org/10.5194/hess-24-4601-2020>
- Fan Y, Yu L, Shi X (2021a) Uncertainty quantification and partition for multivariate risk inferences through a factorial multimodel Bayesian copula (FMBC) system. *J Hydrol* 598:126406. <https://doi.org/10.1016/j.jhydrol.2021.126406>
- Fan YR, Yu L, Shi X, Duan QY (2021b) Tracing uncertainty contributors in the multi-hazard risk analysis for compound extremes. *Earth's Future* 9:e2021EF002280
- Ganguli P, Reddy MJ (2013) Probabilistic assessment of flood risks using trivariate copulas. *Theoret Appl Climatol* 111(1–2):341–360
- Gebregiorgis AS, Hossain F (2012) Hydrological risk assessment of old dams: case study on Wilson dam of Tennessee River basin. *J Hydrol Eng* 17:201–212
- Genest C, Favre AC (2007) Everything you always wanted to know about copula modeling but were afraid to ask. *J Hydrol Eng* 12(4):347–368
- Genest C, Rivest LP (1993) Statistical inference procedure for bivariate Archimedean copulas. *J Am Stat Assoc* 88:1034–1043
- Genest C, Rémillard B, Beaudoin D (2009) Goodness-of-fit tests for copulas: A review and a power study. *Insur Math Econ* 44(2):199–213. <https://doi.org/10.1016/j.insmatheco.2007.10.005>
- Graier B, van den Berg MJ, Vandenbergh S, Petroselli A, Grimaldi S, De Baets B, Verhoest NEC (2013) Multivariate return periods in hydrology: a critical and practical review focusing on synthetic design hydrograph estimation. *Hydrol Earth Syst Sci* 17:1281–1296
- Grimaldi S, Serinaldi F (2006) Asymmetric copula in multivariate flood frequency analysis. *Adv Water Resour* 29:1281–1296
- Gringorten II (1963) A plotting rule of extreme probability paper. *J Geophys Res* 68(3):813–814

- Han JC, Huang GH, Zhang H, Li Z, Li YP (2014a) Bayesian uncertainty analysis in hydrological modeling associated with watershed subdivision level: a case study of SLURP model applied to the Xiangxi River watershed, China. *Stoch Env Res Risk Assess* 28(4):973–989
- Han JC, Huang GH, Zhang H, Li Z (2014b) Heterogeneous precipitation and streamflow trends in the Xiangxi River watershed, 1961–2010. *J Hydrol Eng* 19(6):1247–1258
- Hu YM, Liang ZM, Solomatine DP, Wang HM, Liu T (2021) Assessing the impact of precipitation change on design annual runoff in the Headwater Region of Yellow River, China. *J Environ Inform* 37(2):122–129
- Huang K, Fan YR (2021) Parameter uncertainty and sensitivity evaluation of copula-based multivariate hydroclimatic risk assessment. *J Environ Inf* 38(2):131–144
- Huang K, Dai LM, Yao M, Fan YR, Kong XM (2017) Modelling dependence between traffic noise and traffic flow through an entropy-copula method. *J Environ Inf* 29(2):134–151
- Kao SC, Govindaraju RS (2010) A copula-based joint deficit index for droughts. *J Hydrol* 380:121–134
- Karmakar S, Simonovic SP (2009) Bivariate flood frequency analysis. Part 2: a copula-based approach with mixed marginal distributions. *J Flood Risk Manag* 2:32–44
- Kidson R, Richards KS (2005) Flood frequency analysis: assumption and alternatives. *Prog Phys Geogr* 29(3):392–410
- Kite GW (1988) Frequency and risk analysis in water resources. Water Resources Publications, Littleton
- Kong XM, Huang GH, Fan YR, Li YP (2015) Maximum entropy-Gumbel-Hougaard copula method for simulation of monthly streamflow in Xiangxi river, China. *Stoch Env Res Risk Assess* 29:833–846
- Lee T, Salas JD (2011) Copula-based stochastic simulation of hydrological data applied to Nile River flows. *Hydrol Res* 42(4):318–330
- Li YP, Huang GH, Nie SL (2008) Inexact multistage stochastic integer programming for water resources management under uncertainty. *J Environ Manage* 88(1):93–107
- Li Z, Huang GH, Fan YR, Xu JL (2015) Hydrologic risk analysis for nonstationary streamflow records under uncertainty. *J Environ Inf* 26(1):41–51
- Lyu X, Fan Y (2021) Characterizing impact factors on the performance of data assimilation for hydroclimatic predictions through multilevel factorial analysis. *J Environ Inf* 38(1):68–82
- Ma M, Song S, Ren L, Jiang S, Song J (2013) Multivariate drought characteristics using trivariate Gaussian and Student copula. *Hydrol Process* 27:1175–1190
- Mei H, Li YP, Lv J, Chen XJ, Lu C, Suo C, Ma Y (2021) Development of an integrated method (MGCMs-SCA-FER) for assessing the impacts of climate change: a case study of Jing-Jin-Ji Region. *J Environ Inf* 38(2):145–161
- Moftakhari HR, Salvadori G, AghaKouchak A, Sanders BF, Matthew RA (2017) Compounding effects of sea level rise and fluvial flooding. *Proc Natl Acad Sci* 114(37):9785–9790. <https://doi.org/10.1073/pnas.1620325114>
- Nelsen RB (2006) An introduction to copulas. Springer, New York
- Reddy JM, Ganguli P (2012) Bivariate flood frequency analysis of upper Godavari river flows using Archimedean copulas. *Water Resour Manage* 26(14):3995–4018
- Sadegh M, Moftakhari H, Gupta HV, Ragno E, Mazdiyasni O, Sanders B et al (2018) Multi-hazard scenarios for analysis of compound extreme events. *Geophys Res Lett* 45:5470–5480. <https://doi.org/10.1029/2018GL07731>
- Salvadori G, De Michele C, Kottegoda NT, Rosso R (2007) Extremes in nature: an approach using copula. Springer, Dordrecht, p 292
- Salvadori G, De Michele C, Durante F (2011) On the return period and design in a multivariate framework. *Hydrol Earth Syst Sci* 15:3293–3305
- Song S, Singh VP (2010) Meta-elliptical copulas for drought frequency analysis of periodic hydrologic data. *Stoch Env Res Risk Assess* 24(3):425–444
- Sraj M, Bezak N, Brilly M (2015) Bivariate flood frequency analysis using the copula function: a case study of the Litija station on the Sava River. *Hydrol Process* 29(2):225–238
- Vandenberghe S, Verhoest NEC, De Baets B (2010) Fitting bivariate copulas to the dependence structure between storm characteristics: a detailed analysis based on 105 year 10 min rainfall. *Water Resour Res*. <https://doi.org/10.1029/2009wr007857>
- Water Conservancy Bureau of Xingshan (2004), Water conservancy annuals of Xingshan, 1949–2003. Xingshan, Hubei, China
- Wu ZY, Lu GH, Liu ZY, Wang JX, Xiao H (2013) Trends of extreme flood events in the Pearl river basin during 1951–2010. *Adv Clim Chang Res* 4(2):110–116
- Xu H, Taylor RG, Kinston DG, Jiang T, Thompson JR, Todd MC (2010) Hydrological modeling of River Xiangxi using SWAT2005: a comparison of model parameterizations using station and gridded meteorological observations. *Quatern Int* 226:54–59
- Ye L, Cai Q, Liu R, Cao M (2009) The influence of topography and land use on water quality of Xiangxi River in Three Gorges Reservoir region. *Environ Geol* 58(5):937–942
- Yen BC (1970) Risk analysis in design of engineering projects. *J Hydrol Eng* 96(4):959–966
- Yue S (2000) The bivariate lognormal distribution to model a multivariate flood episode. *Hydrol Process* 14(14):2575–2588
- Yue S (2001) A bivariate gamma distribution for use in multivariate flood frequency analysis. *Hydrol Process* 15(6):1033–1045
- Zhang L, Singh VP (2006) Bivariate flood frequency analysis using the copula method. *J Hydrol Eng* 11:150–164
- Zhang L, Singh VP (2007) Bivariate rainfall frequency distributions using Archimedean copulas. *J Hydrol* 332(1–2):93–109
- Zhang Q, Xiao MZ, Singh VP, Chen XH (2013) Copula-based risk evaluation of hydrological droughts in the East River basin, China. *Stoch Env Res Risk Assess* 27:1397–1406

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)