Bivariate design flood quantile selection using copulas

Tianyuan Li, Shenglian Guo, Zhangjun Liu, Lihua Xiong and Jiabo Yin

ABSTRACT

Flood event consists of peak discharge and flood volume that are mutually correlated and can be described by a copula function. For a given bivariate joint distribution, a choice of design return period will lead to infinite combinations of peak discharge and flood volume. A boundary identification method is developed to define the feasible ranges of flood peak and volume suitable for combination, and two combination methods, i.e., equivalent frequency combination (EFC) method and conditional expectation combination method for estimating unique bivariate flood quantiles are also proposed. Monte Carlo simulation method is used to evaluate the performance of these combination methods. The Geheyan reservoir in China was selected as case study. It is shown that the joint design values estimated by the two proposed combination methods are both within the feasible range, which means that the methods could be selected for designing unique flood quantiles. The proposed bivariate combination methods are also compared with univariate method, and the reservoir water level estimated by EFC method is higher than the other methods, which means the EFC method is safer for reservoir design. The developed approach provides an applicable way for the identification of feasible range and flood quantile estimation.

Key words | bivariate distribution, boundary identification, copula function, frequency analysis, quantile estimation

Tianyuan Li
Shenglian Guo (corresponding author)
Zhangjun Liu
Lihua Xiong
Jiabo Yin
State Key Laboratory of Water Resources and
    Hydropower Engineering Science,
Wuhan University,
Wuhan 430072,
China
E-mail: slguo@whu.edu.cn

Tianyuan Li
Changjiang Institute of Survey Planning Design and
    Research,
Wuhan 430010,
China

INTRODUCTION

Conventional hydrologic frequency analysis has mainly focused on a single characteristic value such as peak discharge or flood volume by using univariate distribution. However, many hydrological events are characterized by the joint behavior of several correlated random variables; for example, flood hydrograph can be described in terms of flood peak, volume, and duration. When these correlated random variables are relevant to risk analyses or design purposes, the joint study of their probabilistic characteristics leads to a better understanding of the phenomena under investigation and avoids the over- or underestimation of the risks that would be produced by univariate analyses (Salvadori & De Michele 2004; Xiao et al. 2009; Gräler et al. 2013).

Many multivariate analysis methods and case studies have been proposed and discussed in the past decades.

These methods have been applied mainly to the analysis of rainfall storms (De Michele & Salvadori 2003; Zhang et al. 2012; Sun et al. 2013; Abdul & Zeephongsekul 2014), flood frequency (Chowdhary et al. 2011; Chen et al. 2012; Sraj et al. 2015), drought events (Chen et al. 2013a, 2013b; Ganguli 2014; Salvadori & De Michele 2010, 2015), regional frequency analysis (Chebana & Ouarda 2009; Ben et al. 2015; Zhang et al. 2015a), and synthetic flood hydrograph (Gräler et al. 2013; Salvadori et al. 2013).

These multivariate hydrological frequency analysis studies have mainly concentrated on the following aspects: (1) showing the importance and explaining the usefulness of the multivariate framework; (2) fitting the appropriate multivariate distribution (copula and marginal distributions) and estimating the corresponding parameters; (3) estimating multivariate return periods and quantile values. Of these

three aspects, the third one, i.e., the question of estimating multivariate return periods and quantile values, still has not yet been properly addressed in the multivariate frequency analysis framework (Chebana & Ouarda 2011), although the estimation of return periods and quantile values has been extensively studied in univariate hydrological frequency analysis.

One of the main difficulties in the multivariate quantile estimation is how to choose the proper combinations of design values of the concerned random variables for a given multivariate return period of hydrologic structure design. Take the bivariate case (peak discharge Q and flood volume W) as an example. The combinations can differ greatly in terms of their values: moving along the multivariate quantile curve to an asymptote, one of the two variables will approach its marginal value, while the other tends to indefinitely increase (for unbounded random variables). Chebana & Ouarda (2011) proposed the decomposition of the level curve into a naive part (tail) and the proper part (central); they assumed that the naive part was composed of two segments starting at the end of each extremity of the proper part. Salvadori *et al.* (2011) introduced two basic design realizations, i.e., component-wise excess design realization and most-likely design realization. Volpi & Fiori (2012) identified a subset of the critical combinations set that includes a fixed and arbitrarily chosen percentage in probability of the events, on the basis of their probability of occurrence.

Following the ideas of Chebana & Ouarda (2011) and Volpi & Fiori (2012), a developed approach to identify the feasible ranges of flood peak and flood volume combination is suggested. Two combination methods, i.e., the equivalent frequency combination (EFC) method and the conditional expectation combination (CEC) method, are developed to derive the quantiles of flood peak and flood volume for given multivariate return periods. The Geheyan reservoir in China was selected as a case study. The design flood quantiles and design flood hydrographs (DFHs) were derived by bivariate combination methods and the univariate distribution method, respectively. The remainder of this paper is organized as follows. The 'Methods' section gives the methodology used to derive the feasible range and bivariate combination methods. In the 'Statistical procedure' section, we present the statistical procedures used for assessing the

proposed combination methods. The 'Application' section addresses a case study of China's Geheyan Reservoir. Finally, the 'Conclusions' section summarizes the main results and provides necessary explanations.

METHODS

In this study, we propose a developed approach for estimating bivariate feasible ranges of flood peak and flood volume suitable for combination in the critical level curve. To derive the feasible range, a boundary identification method is suggested, which was inspired by the ideas of Chebana & Ouarda (2011) and Volpi & Fiori (2012). Two combination methods for estimating unique bivariate flood quantiles, i.e., the EFC method and the CEC method, are proposed based on the assumption of the relationship between u and v (or q and w).

Copula functions

Multivariate distribution construction using copulas was developed by Nelsen (2006), and Salvadori *et al.* (2007). Every joint distribution can be written in terms of a copula and its univariate marginal distributions. Copula is a function that connects multivariate probability distribution to its one-dimensional marginal distributions (Nelsen 2006). Let $F_{X_i}(x_i)$ $(i = 1, 2, \ldots, n)$ be the marginal cumulative distribution functions (CDFs) of X_i. The objective is to determine the multivariate distribution, denoted as $H_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n)$ or simply H. Thus, the multivariate probability distribution H is expressed in terms of its marginal distributions and the associated dependence function, known as Sklar's theorem (Sklar 1959):

$$C(F_{X_1}(x_1), F_{X_2}(x_2), \ldots, F_{X_n}(x_n))$$
$$= H_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n) \qquad (1)$$

where C, called copula, is uniquely determined whenever $F_{X_i}(x_i)$ are continuous, and captures the essential features of the dependence among the random variables. For details on the copula approach and its hydrological applications, refer to the literature (e.g., Salvadori & De Michele 2015; Zhang *et al.* 2015b).

Different families of copula functions have been proposed and described by Nelsen (2006) and Salvadori et al. (2007). Of all the copula families, the Archimedean family is more desirable for hydrological analyses because it can be easily constructed and can be applied to whether the correlation among the hydrological variables is positive or negative (Nelsen 2006; Chen et al. 2014a). A large variety of copulas belong to this family. Four one-parameter Archimedean copulas, including the Gumbel–Hougaard (G–H), Ali–Mikhai–Haq, Frank and Cook–Johnson copulas, have been applied in frequency analysis by many authors (Zhang & Singh 2006; Salvadori et al. 2007; Chen et al. 2014b; Liu et al. 2015).

Bivariate return period

In the conventional univariate analysis, flood events of interest are often defined by return periods. In the bivariate domain, however, it is still discussed by the community as to which method is most suitable to transform the joint exceedance probability to a bivariate joint return period (JRP). Different JPRs estimated by copula function have been developed for the case of a bivariate flood frequency analysis. Eight types of possible joint events were presented by Salvadori & De Michele (2004) using 'OR' and 'AND' operators, of which two cases are of the greatest interest in hydrological applications (Shiau 2003; Salvadori & De Michele 2004):

(1) (OR case) either $Q > q$, or $W > w$, i.e.,

$$E_{or} = \{Q > q \text{ or } W > w\} \tag{2}$$

(2) (AND case) both $Q > q$ and $W > w$, i.e.,

$$E_{and} = \{Q > q \text{ and } W > w\} \tag{3}$$

In simple words, for E_{or} to happen it is sufficient that either peak discharge Q or flood volume W (or both) exceed given thresholds; instead, for E_{and} to happen it is necessary that both Q and W are larger than prescribed values. Thus, two different JRPs can be defined accordingly

(De Michele et al. 2005):

$$T_{or} = \frac{\mu}{P[Q > q \text{ or } W > w]} = \frac{\mu}{1 - F(q, w)} \tag{4}$$

$$
\begin{aligned}
T_{and} &= \frac{\mu}{P[Q > q \text{ and } W > w]} \\
&= \frac{\mu}{1 - F_Q(q) - F_W(w) + F(q, w)}
\end{aligned}
\tag{5}
$$

where μ is the mean inter-arrival time between two consecutive events (in the case of annual maxima $\mu = 1$ year), and $F(q, w) = P(Q \leq q, W \leq w)$.

The Kendall JRP was introduced by Salvadori & De Michele (2004) in order to identify a univariate critical threshold in a multivariate context, which is given by:

$$\theta_t = \frac{\mu_T}{1 - K_C(t)} \tag{6}$$

where K_C is the Kendall's distribution function associated with the joint CDF of the copula's level curves: $K_C(t) = P[C(u, v) \leq t]$. It allows for the calculation of the probability that a random point (u, v) in the unit square has a smaller (or larger) copula value than a given critical probability level t. In other words, it is related to the probability of occurrence of an event in the area over the copula level curve of value t.

Different definitions of the multivariate return period are available in the literature, based on regression analysis, bivariate conditional distributions, survival Kendall distribution function, and structure performance function. For instance, some studies have focused on a structure-based return period for the design and/or risk assessment of hydrological structures in a bivariate environment (Requena et al. 2013; Volpi & Fiori 2014). A comprehensive review of the JRP estimation methods was given by De Michele et al. (2013), Serinaldi (2015), and Volpi et al. (2015), and references therein.

The OR return period given in Equation (4) has been extensively applied in multivariate hydrological frequency analysis (e.g., Shiau 2003; Salvadori & De Michele 2004; Chebana & Ouarda 2011; Volpi & Fiori 2012; Li et al. 2013; Liu et al. 2015; Sraj et al. 2015). In this study, we focus on the OR case for quantile estimation in a bivariate context.

Feasible range identification for bivariate quantile curve

The critical level curve, as shown in Figure 1, was defined as a bivariate quantile curve by Chebana & Ouarda (2011). As previously stated, for the case of OR return period, the function that describes the level curve for any given return period T or critical probability level p has two asymptotes, $q = q_p$ and $w = w_p$, where $q_p = F_Q^{-1}(p)$ and $w_p = F_W^{-1}(p)$ are the quantiles of the marginal distribution for the given probability level p. According to Equation (4) in the bivariate case, the choice of an appropriate return period T or a critical probability level p for hydraulic structure design will lead to the infinite combinations of flood peak and volume. However, all the bivariate flood events with the same value of T or p along the level curve differ greatly not only in terms of their quantile values, but also in terms of their probability of occurrence, which is measured by the joint probability density function (PDF), i.e., $f(q, w)$, evaluated along the critical level curve (Volpi & Fiori 2012). Meanwhile, different combinations of Q and W are generally not equivalent from a practical point of view, although they all satisfy the flood prevention standards. The boundaries (see points B and C in Figure 1) for selection of design flood peak and volume are necessary in the case that the flood combinations are outside the boundaries with unrealistically low occurrence probabilities.
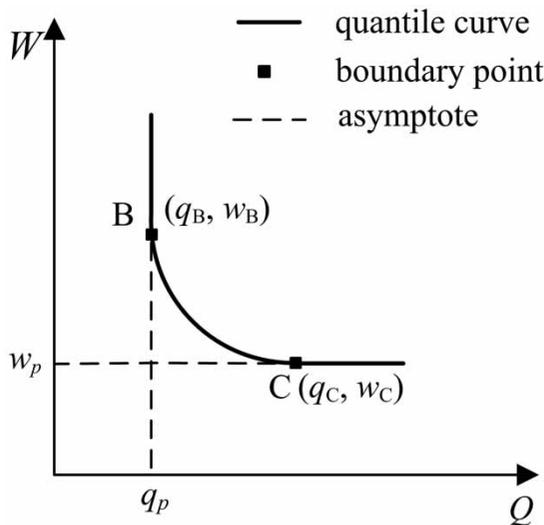

Figure 1 | Bivariate quantile curve with a critical probability level p.

Chebana & Ouarda (2011) proposed a method to decompose the quantile curve in Figure 1 into a naive part (i.e., the subset BC) and a proper part (outside subset BC). They assumed that the naive part is composed of two segments starting at the end of each extremity of the proper part. They also suggested selecting these boundary points according to the empirical version or as close as to the asymptotes (the naive part). Volpi & Fiori (2012) defined the distance of each point along the quantile curve in Figure 1 from its vertex as random variable (s) and derived its PDF. The boundary points of the quantile curve were identified with a chosen percentage in probability of the events. They also proposed a way of decomposition of the quantile curve into naive part and proper part. However, the procedure presented by Volpi & Fiori (2012) is difficult to apply in the curvilinear coordinate system $[s(x,y), n(x,y)]$ or to derive the expression of random variable (s). To overcome these limitations, an approach to identify the boundary points (i.e., B and C) of the quantile curve is developed. A new density function $\varphi(q)$ is used to measure the relative likelihood of flood events, which is a non-curvilinear variable in the procedure.

To derive the new density function with a chosen probability level to decompose the quantile curve, a joint distribution of annual maximum (AM), flood peak (Q) and flood volume (W) should be built by copula functions. The joint distribution function $F(q,w)$ can be expressed in terms of its marginal functions $F_Q(q)$ and $F_W(w)$ by using an associated dependence function C, $F(q,w) = C[F_Q(q), F_W(w)]$.

It is found that flood peak and volumes are usually upper-tailed dependent variables and the G–H copula could reproduce best the observed tail dependence coefficient (e.g., Poulin *et al.* 2007; Ben *et al.* 2015). Therefore, the G–H copula is taken as an example to illustrate the developed boundary identification method because of its easy expression and wide applications (Li *et al.* 2013; Chen *et al.* 2015; Sraj *et al.* 2015).

For the G–H copula function, the relationship of joint distribution $C_\theta(u,v)$ and bivariate return period T can be expressed as ($\mu = 1$ for annual maxima flood series):

$$C_\theta(u, v) = \exp\left\{-[(-\ln u)^\theta + (-\ln v)^\theta]^{1/\theta}\right\} = 1 - \frac{1}{T} \qquad (7)$$

where θ is the dependence parameter of the G–H copula, $u = F_Q(q)$, $v = F_W(w)$.

Thus, the relationship between u and v with the given bivariate return period T can be derived as:

$$v = \exp\left\{ -\left[(-\ln u)^\theta - \left(-\ln\left(1 - \frac{1}{T} \right) \right)^\theta \right]^{1/\theta} \right\} \quad (8)$$

Replacing $u = F_Q(q)$ and $v = F_W(w)$ into the above equation yields:

$$F_W(w) =$$
$$\exp\left\{ -\left[(-\ln F_Q(q))^\theta - \left(-\ln\left(1 - \frac{1}{T} \right) \right)^\theta \right]^{1/\theta} \right\} = \eta(F_Q(q)) \quad (9)$$

in which, $\eta(x) = \exp\left\{ -\left[(-\ln x)^\theta - \left(-\ln\left(1 - \frac{1}{T} \right) \right)^\theta \right]^{1/\theta} \right\}$

Thus, the relationship between Q and W with the fixed bivariate return period T can be derived as:

$$w = F_W^{-1}(v) = F_W^{-1}(\eta(F_Q(q))) = \zeta(q) \quad (10)$$

where $F_W^{-1}(v)$ is the inverse CDF of flood volume W. The above equation reveals that W can be derived by Q if the bivariate return period T is fixed.

It should be noted that other copulas with more complicated formulas sometimes may be needed. For the Frank copula, Clayton copula and several two-parameter copulas, the implicit expression for describing the relationship between Q and W in Equations (7)–(10) could be derived. For copulas with more complicated expressions, the numerical method should be applied. For example, the unique value of w could be obtained with given q by a trial and error method (Liu *et al.* 2015).

After obtaining the corresponding relationship of the values of w and q for the flood events along the critical level curve, the bivariate joint PDF of w and q can be expressed according to Sklar's theory (Nelsen 2006) as:

$$f(q, w) = c_\theta(F_Q(q), F_W(w)) \cdot f_Q(q) \cdot f_W(w) \quad (11)$$

where $f_Q(q)$ and $f_W(w)$ are univariate PDFs of flood peak and volume, respectively, and $c_\theta(u, v)$ is the density of $C_\theta(u, v)$ and defined as:

$$c_\theta(u, v) = \frac{\partial^2 C_\theta(u, v)}{\partial u \partial v} \quad (12)$$

Referring to Equations (9) and (10), the bivariate joint PDF (Equation (11)) of flood peak and volume can be finally described as the function of the single random variable of flood peak Q for the fixed bivariate return period T, i.e.,

$$f(q, w) = c_\theta(F_Q(q), \eta(F_Q(q))) \cdot f_Q(q) \cdot f_W(\zeta(q)) \quad (13)$$

According to Equation (13), there is a curve that can describe the relationship between joint PDF $f(q,w)$ and flood peak Q for a given bivariate return period T or a critical probability level p. Assume that the area between the curve of $f(q,w)$ and the horizontal axis of flood peak Q is A, i.e.,

$$A = \int_{q_p}^{+\infty} f(q, w)dq = \int_{q_p}^{+\infty} c(F_Q(q), \eta(F_Q(q))) \cdot f_Q(q) \cdot f_W(\zeta(q))dq \quad (14)$$

where q_p represents univariate design value of flood peak, i.e., $q_p = F_Q^{-1}(p)$, which is chosen as the lower bound of flood peak in estimation of the bivariate design flood values.

As $f(q,w)$ is a joint density function of q and w, area A does not equal to 1 if only q is taken as integral variable (i.e., $A \neq 1$). A new density function $\varphi(q)$ over the area A which has proper density characters is constructed and expressed as follows:

$$\varphi(q) = \frac{f(q, w)}{A} = \frac{f(q, w)}{\int_{q_p}^{+\infty} f(q, w)dq} \quad (15)$$

Obviously, there is a one-to-one correspondence between the density function $\varphi(q)$ and bivariate PDF $f(q,w)$. The density function $\varphi(q)$ varies with the horizontal axis and $\int_{q_T}^{+\infty} \varphi(q)dq = 1$.

As previously stated, the bivariate design flood combinations near the upper and lower bounds of the quantile curve have lower occurrence probability than that near the middle of the quantile curve. As a consequence, the bivariate

PDF $f(q,w)$ of bivariate design flood combination near the upper and lower bounds of quantile curve is smaller than that near the middle of the quantile curve. The density function $\varphi(q)$ has the same property as the bivariate PDF $f(q,w)$. As the design flood peak (or flood volume) varies from the lower bound, i.e., (q_p) to infinitely great, the density function $\varphi(q)$ increases to the maximum value and then decreases gradually, as shown in Figure 2. The vertex of the density function $\varphi(q)$ describing the full dependence (Chebana & Ouarda 2011; Volpi & Fiori 2012) between peak and volume has the highest density. In other words, this is the most likely bivariate design flood event.

Once the density function $\varphi(q)$ along Q is defined by Equation (13), we can evaluate the lower and upper bounds that contain $\varphi(q)$ with probability of $1-\varepsilon$, for a given probability level ε. The quantiles of lower and upper bounds (q_B and q_C) are specified respectively by (Volpi & Fiori 2012):

$$\int_{q_p}^{q_B} \varphi(q)dq = \alpha_1 \tag{16}$$

$$\int_{q_p}^{q_C} \varphi(q)dq = 1 - \alpha_2 \tag{17}$$

where $\alpha_1 + \alpha_2 = \varepsilon$. The lower and upper bounds q_B and q_C identify a feasible range on the quantile curve, bounded by the points of coordinates $(q_B, \zeta(q_B))$ and $(q_C, \zeta(q_C))$, that excludes the ε percentage in probability of the critical events. The probability levels α_1 and α_2 can be arbitrarily chosen, taking account of the specific problem under investigation (Volpi & Fiori 2012).
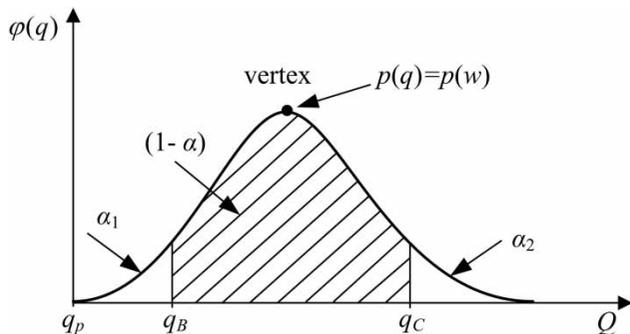

Figure 2 │ Relationship between density function $\varphi(q)$ and flood peak Q.

Bivariate flood quantile selection

For a given bivariate return period T, there are countless combinations of u and v that satisfy Equation (7). To derive the design values of flood peak q and flood volume w, the unique combination of u and v (or q and w) should be determined. Hence besides Equation (7), one more equation that can establish the relationship between u and v (or q and w) is necessary. Two combination methods were proposed to derive the quantiles of flood peak and flood volume for given multivariate return periods, and they are now outlined.

EFC method

With a given bivariate return period T, we assume that the flood peak and flood volume have the same probability of occurrence, i.e., $u = v$ (or $F_Q(q) = F_W(w)$). This assumption is usually taken as a uniform procedure for the derivations of design flood values and DFH in China (MWR 2006; Xiao *et al.* 2008, 2009; Chen *et al.* 2010). Then, the design frequency of bivariate EFC can be obtained by jointly solve the equation $u = v$ and Equation (7).

Taking the G–H copula for example, the relationship between u and v with the given bivariate return period T is described in Equation (7). Based on the assumption that $u = v$, the probabilities of occurrence of flood peak and volume (i.e., u and v) can be estimated by the solution of Equation (18), which is expressed as follows:

$$u = v = \left(1 - \frac{1}{T}\right)^{\zeta} \tag{18}$$

where $\zeta = 2^{-1/\theta}$, and θ is the dependence parameter of the G–H copula.

Consequently, the design value of bivariate EFC can be derived by the inverse function of marginal distributions:

$$q = F_Q^{(-1)}(u) \tag{19a}$$

$$w = F_W^{(-1)}(v) \tag{19b}$$

CEC method

Since the flood peak Q and flood volume W are dependent variables, one may wish to predict the value of W based on an observed value of Q. Let $g(Q)$ be a predictor, i.e., $g \in N =$ {all Borel functions g with $E[g(Q)]^2 < \infty$}. Each predictor is assessed by the 'mean squared prediction error' $E[W - g(Q)]^2$. The conditional expectation $E(W|Q)$ is the best predictor of W in the sense that (Shao 2003):

$$E[W - E(W|Q)]^2 = \min_{g \in \aleph} E[W - g(Q)]^2 \qquad (20)$$

Herein, during a flood event, when the flood peak $Q = q$ takes place, the conditional expectation $E(w|q)$ is used to estimate the value of flood volume, which can be derived as:

$$E(w|q) = \int_{-\infty}^{+\infty} w f_{W|Q}(w) dw \qquad (21)$$

where $f_{W|Q}(w)$ is the density function of the conditional CDF $F_{W|Q}(w)$ and defined as (Zhang & Singh 2006):

$$f_{W|Q}(w) = \frac{f(q, w)}{f_Q(q)} = \frac{c_\theta(u, v) f_Q(q) f_W(w)}{f_Q(q)}$$
$$= c_\theta(u, v) f_W(w) \qquad (22)$$

Hence, Equation (21) could be expressed as:

$$E(w|q) = \int_{-\infty}^{+\infty} w f_{W|Q}(w) dw = \int_{-\infty}^{+\infty} w c_\theta(u, v) f_W(w) dw$$
$$= \int_0^1 F_W^{-1}(v) c_\theta(u, v) dv \qquad (23)$$

where $F_W^{-1}(\cdot)$ is the inverse CDF of W.

Then, the flood peak q and $E(w|q)$ will be the CEC if the following equations are satisfied:

$$\begin{cases} u = F_Q(q) \\ v = F_W[E(w|q)] \\ \dfrac{1}{1 - C_\theta(u, v)} = T \end{cases} \qquad (24)$$

The above equation can be solved by trial and error method with different values of q.

STATISTICAL PROCEDURE

Monte Carlo simulation method

Since the flood volume W is conditioned on the flood peak Q, the conditional distribution is employed to generate the flood data series. The samples are generated from a bivariate distribution composed by Pearson type three (P3) marginal distributions (Li *et al.* 2013) and copula function. The Monte Carlo simulation method is based on the algorithm developed by Chebana & Ouarda (2009), Lee & Salas (2011), and Chen *et al.* (2015). The main generating procedures with the conditional copula are summarized as follows:

1. Fit the marginal distribution $F_Q(q)$ and $F_W(w)$ under recorded data using the P3 marginal distributions, $u = F_Q(q)$ and $v = F_W(w)$.
2. Establish the joint distribution $F(q,w)$ using copulas, $F(q,w) = C(u,v)$, and estimate their parameters based on the Kendall tau method.
3. The conditional distribution function of flood volume and peak are established based on following bivariate joint distribution:

$$F_{W|Q}(w) = P(W \le w | Q = q) = \frac{\partial F(q, w)}{\partial q} \qquad (25)$$

4. One independent random value r_1 is generated from [0, 1] uniform distribution. Let r_1 represent the value of cumulative probability $F_Q(q)$, i.e., $u = F_Q(q) = r_1$, then the peak discharge q can be calculated using the fitted P3 distribution, i.e., $q = F_Q^{(-1)}(u)$.
5. Another independent random value r_2 is also generated from [0, 1] uniform distribution. Let r_2 represent the value of the conditional distribution function $F_{W|Q=q}(w)$, i.e., $F_{W|Q=q}(w) = r_2$, then the flood volume w can be calculated by $w = F_{W|Q=q}^{-1}(r_2)$. Since the solution of Equation (25) is very complex, it could be simplified using copula function as follows:

$$F_{V|U}(v) = P(V \le v | U = u) = \frac{\partial C(u, v)}{\partial u} \qquad (26)$$

After solving $v = F_{V|U=u}^{-1}(r_2)$, the flood volume w can be obtained by $w = F_W^{(-1)}(v)$.

6. Repeat steps (4)–(5) N times, then the data series of peak discharge Q and flood value W with size N are obtained. The parameters of P3 marginal distributions are estimated by L-moment method, respectively. The parameter of copula functions is estimated and the joint distribution is established.

7. The flood volume w is estimated under one given peak discharge q using the EFC method and the flood volume w is estimated by Equation (23) for the CEC method.

8. Repeat steps (4)–(7) m_s times.

Performance evaluation

The following evaluation criteria were used to evaluate and compare different methods:

Measure of unbiasedness, denoted by the mean relative estimate error, *MRE*:

$$MRE = \frac{1}{m_s(N-1)} \sum_{i=1}^{m_s} \sum_{j=1}^{N} \left(\frac{w_{i,j} - \tilde{w}_{i,j}}{w_{i,j}} \right) \quad (27)$$

Measure of efficiency, denoted by the relative root mean square error, *RRMSE*:

$$RRMSE = \sqrt{\frac{1}{m_s(N-1)} \sum_{i=1}^{m_s} \sum_{j=1}^{N} \left(\frac{w_{i,j} - \tilde{w}_{i,j}}{w_{i,j}} \right)^2} \quad (28)$$

where N is the sample size; m_s is the number of Monte Carlo experiment iterations; $w_{i,j}$ is the jth simulated flood volume in the ith experiment; $\tilde{w}_{i,j}$ is the jth estimated flood volume derived by the jth simulated peak discharge $q_{i,j}$ in the ith experiment.

APPLICATION

As an illustrative example, the Geheyan reservoir was selected as a case study. The Geheyan reservoir (see Figure 3) is a key control and multi-purpose water resources engineering project in the Qingjiang basin which is one of the main tributaries of the Yangtze River in China. The basin encompasses an area of 17,000 km^2 with an annual average rainfall of 1,500 mm. The annual average discharge and volume at the dam site are 393 m^3/s and 124×10^8 m^3, respectively. The flood season lasts for 5 months, from May to September. Since it is a multi-purpose reservoir, there is serious conflict between flood prevention and hydro-power generation. The reservoir design flood with 1,000-year return period is used as the flood prevention standard. The flood control limiting water level during the flood season and normal water level are 192.2 m and 200 m (a.m.s.l.), respectively. The AM flood peak discharge (Q_p) and 7-day flood volume (W_7) are available with a systematic record of 54 years (1951–2004).
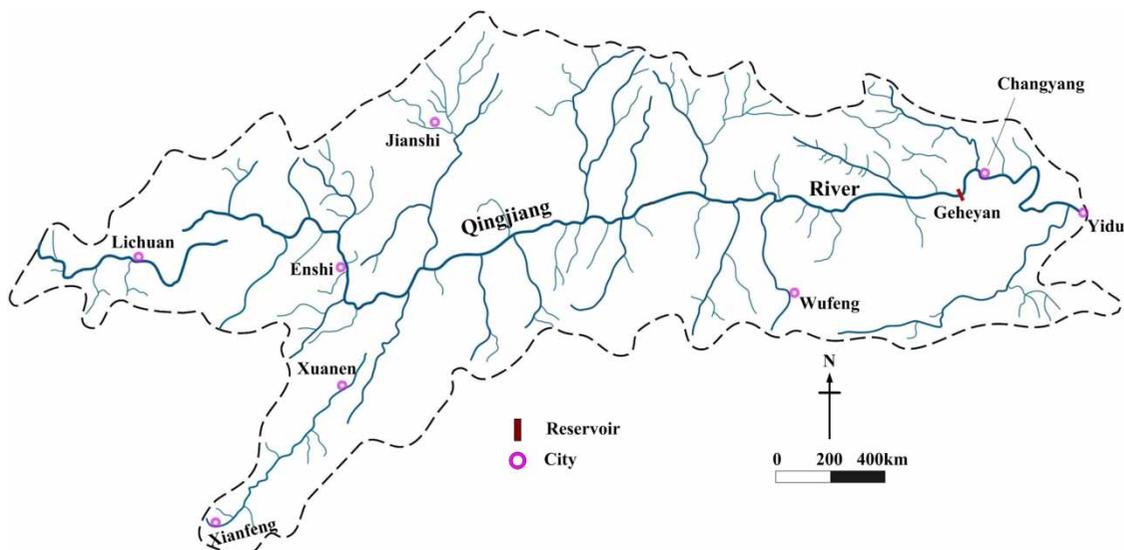

Figure 3 | Location of Geheyan reservoir in the Qingjiang basin.

Parameter estimation and hypothesis test

For AM flood series, the Pearson type three (P3) distributions have been recommended by MWR (2006) as a uniform procedure for flood frequency analysis in China. The P3 distributions are suggested to be applied for peak discharge and flood volumes in China (Chen *et al.* 2014a, 2014b). The PDF of the P3 distribution is defined as:

$$f(x) = \frac{\beta^{\alpha}}{\Gamma(\alpha)}(x - \delta)^{\alpha-1}e^{-\beta(x-\delta)}, \tag{29}$$
$$\alpha > 0, \beta > 0, \delta \leq x < \infty$$

where α, β, and δ are shape, scale, and location parameters of the P3 distribution, respectively.

The parameters of the P3 marginal distributions were estimated by L-moment method and are listed in Table 1. A chi-square goodness-of-fit and the Kolmogorov–Smirnov (K–S) test are used to test the assumption. The results in Table 2 show that the hypothesis could not be rejected at the 5% significance level because $\chi^2 < \chi^2_{0.05}$, and the P3 distributions can also pass the K–S test because $D_n < D_{n,0.95}$. The marginal distribution frequency curves of flood peak and 7-day flood volumes are shown in Figure 4, in which the line represents the theoretical distribution and the crosses represent empirical probabilities. Figure 4 indicates that these theoretical distributions can fit the observed data reasonably well.

Table 1 │ Sample statistic values and estimated parameters of P3 distribution in Geheyan reservoir

	Sample statistic values			P3 parameters		
Variables	Mean	L-C_v	L-C_s	α	B	δ
Q_p (m³/s)	7,820	0.4	1.2	2.78	0.0005	2606.7
W_7 (10⁸ m³)	17	0.5	1.5	1.78	0.1569	5.7

Table 2 │ Hypothesis test results of P3 marginal distribution for flood peak and volume

	Chi-square test		K–S test	
Variables	$\chi^2_{0.05}$	Chi-square statistics, χ^2	$D_{n,0.95}$	D_n
Q_p (m³/s)	7.815	5.313	0.185	0.096
W_7 (10⁸ m³)	9.488	3.396	0.185	0.078

Joint distribution of flood peak and volume based on copulas

The G–H, Clayton, and Frank copulas of the Archimedean family (Table 3) were applied to construct the joint distribution of flood peak and volume. The parameters of bivariate copulas were estimated using the Kendall correlation coefficient (τ), and the expressions which describe the relationship between the estimated parameters and Kendall correlation coefficient are also shown in Table 3. The result of the Kendall tau is 0.66 in Geheyan reservoir. The results of the Cramer–von Mises test S_n^I and K–S test D_n for the pair are presented in Table 4, and P-values were calculated based on the parametric bootstrap or multipliers procedure with 10,000 runs (Kojadinovic *et al.* 2011; Sraj *et al.* 2015). The root mean square error (RMSE) was also calculated and listed in Table 4. The graphical goodness-of-fit test between simulated (sample size 10,000) and observed values is plotted in Figure 5. On the basis of the statistical tests and graphical fitting (Figure 5), it is found that the G–H copula is much more suitable for constructing the joint distribution of peak discharge and flood volume. This finding is in accordance with other authors (e.g., Genest & Mackay 1986; Poulin *et al.* 2007; Sraj *et al.* 2015; Zhang *et al.* 2015b).

Therefore, the G–H copula is used to model the dependence between the extreme maximum annual flood peak and 7-day flood volume in this study. The probability plot of joint distribution is shown in Figure 6, in which the G–H copula could fit the empirical bivariate distribution very well.

Bivariate quantile curve and feasible range identification

The return period of design flood of Geheyan reservoir, i.e., $T = 1,000$-year, was selected as the bivariate return period and $T = 200$-year was also chosen for comparison. The bivariate quantile curves of the two return periods are shown in Figure 7. The relationship between density function $\varphi(q)$ and q is drawn in Figure 8. Even if the G–H copula model is symmetric, the PDF $\varphi(q)$ is not symmetrical due to the difference in the marginal distributions.

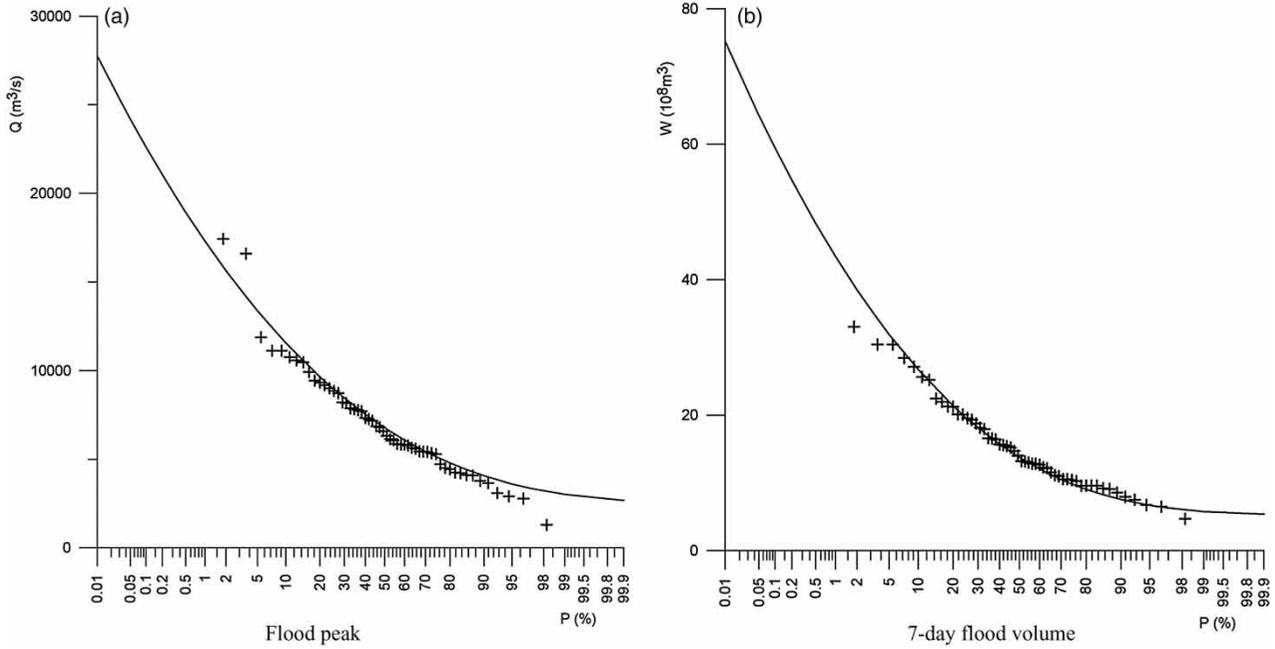The upper and lower bounds on the level curve were estimated numerically by solving Equations (16) and (17),

Figure 4 | Probability curves of flood peak and 7-day flood volume.

**Table 3** | Copula functions and Kendall correlation coefficient

Copula	$C_\theta(u, v)$	$\theta \in$	τ
G–H	$\exp\left\{-[(-\ln u)^\theta + (-\ln v)^\theta]^{1/\theta}\right\}$	$[1, +\infty)$	$1 - \theta^{-1}$
Clayton	$(u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}$	$(0, +\infty)$	$\theta/(\theta + 2)$
Frank	$-\dfrac{1}{\theta}\ln\left[1 + \dfrac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1}\right]$	$R\backslash\{0\}$	$\tau = 1 + \dfrac{4}{\theta}\left[\dfrac{1}{\theta}\displaystyle\int_0^\theta \dfrac{t}{\exp(t) - 1}\,dt - 1\right]$

**Table 4** | Estimated parameters of copulas and statistical tests' results

Copula	Parameter θ	Statistical test		
		RMSE	$S_n^1(p - valve)$	$D_n (D_{n,0.95})$
G–H	2.98	0.068	0.011 (0.44)	0.035 (0.215)
Clayton	3.95	0.074	0.021 (0.17)	0.048 (0.215)
Frank	9.93	0.071	0.015 (0.28)	0.043 (0.215)

and assuming for simplicity (although other assumptions are possible) $\alpha_1 = \alpha_2 = \varepsilon/2$, with $\varepsilon = 0.05$. The upper and lower bounds are denoted as B_1 and C_1, respectively, in Figure 7. It is found that the bounds are close to the horizontal asymptote (i.e., $w_7 = 61.49 \times 10^8 \, \text{m}^3$ for $T = 1,000$ and $w_7 = 50.23 \times 10^8 \, \text{m}^3$ for $T = 200$) and vertical asymptote (i.e., $q_p = 22,800 \, \text{m}^3/s$ for $T = 1,000$ and $q_p =$

19,300 m^3/s for $T = 200$) due to the small value assumed for the probability level ε. The upper and lower bounds were also calculated by the boundary identification method proposed by Volpi & Fiori (2012). The results are also presented in Table 5 and the derived bounds are denoted as B_2 and C_2, as shown in Figure 7. It is shown that the bounds estimated by the proposed
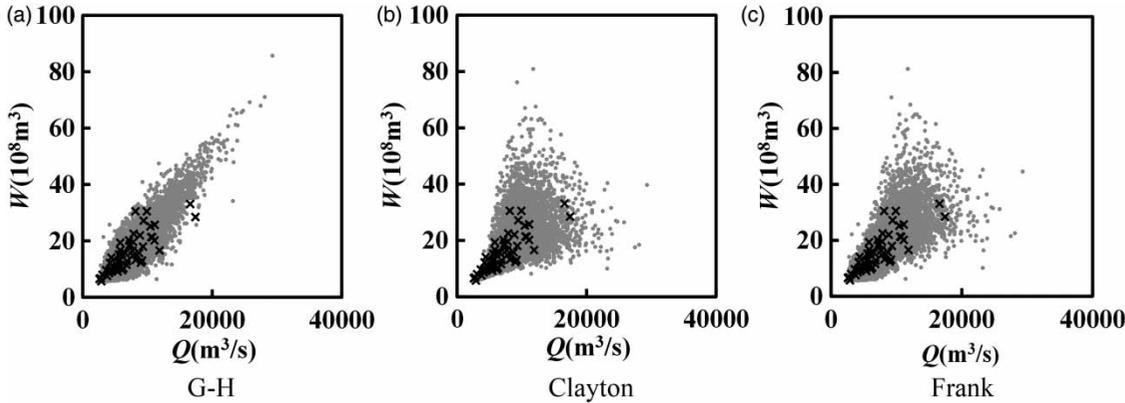
Figure 5 | Comparison between the observed and simulated values for different copula functions.
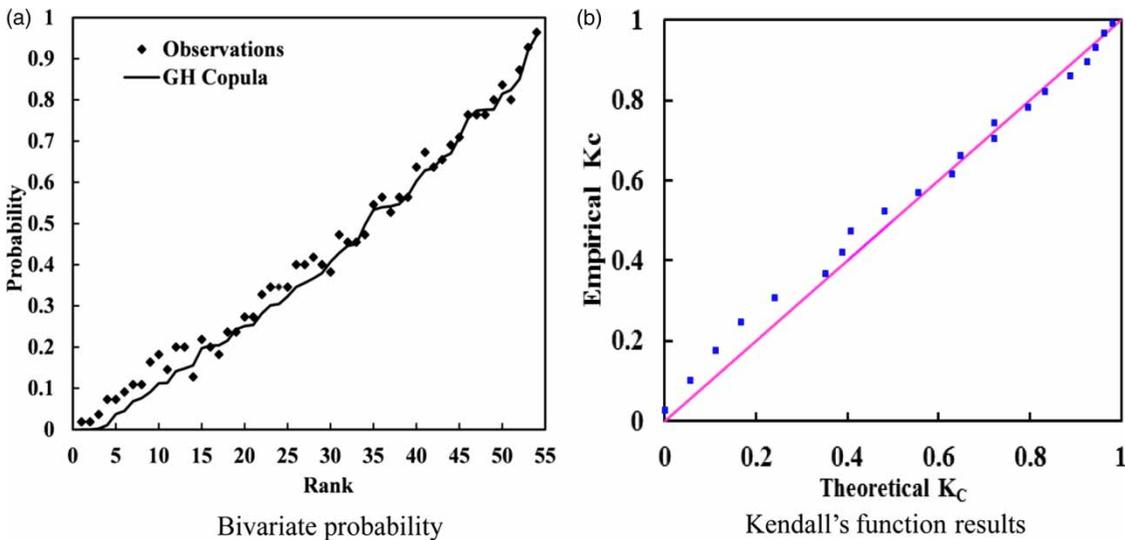
Figure 6 | Comparisons of observed and theoretical bivariate probability distribution.

method and that proposed by Volpi & Fiori (2012) are very similar.

Estimation of bivariate flood quantiles

The proposed bivariate EFC and CEC methods were used to estimate flood peak and 7-day flood volume quantiles with return periods of $T = 1,000$ and $T = 200$ years, respectively. For the purpose of comparison, the univariate flood quantiles (called marginal quantiles by Chebana & Ouarda (2011)) were estimated by marginal distributions, assuming that the univariate return periods (T_Q and T_W) were equal to the bivariate return period (i.e., $T_Q = T_W = T$). The univariate

flood quantiles can be obtained from the equations $q = F_Q^{-1}(p) = F_Q^{-1}(1 - 1/T)$ and $w = F_W^{-1}(p) = F_W^{-1}(1 - 1/T)$. The results of the component-wise excess realization and the most likely realization proposed by Salvadori *et al.* (2011) were also estimated. The estimation results of bivariate and univariate quantiles were listed in Table 6. It is shown that the design values of bivariate quantiles are larger than those of univariate quantiles. The quantiles estimated by the four bivariate event selection methods are also shown in Figure 7, and the estimation points of the EFC method are denoted as point E, while the quantiles estimated by the CEC method are denoted as point F. For the results of selection approaches proposed by Salvadori *et al.* (2011),
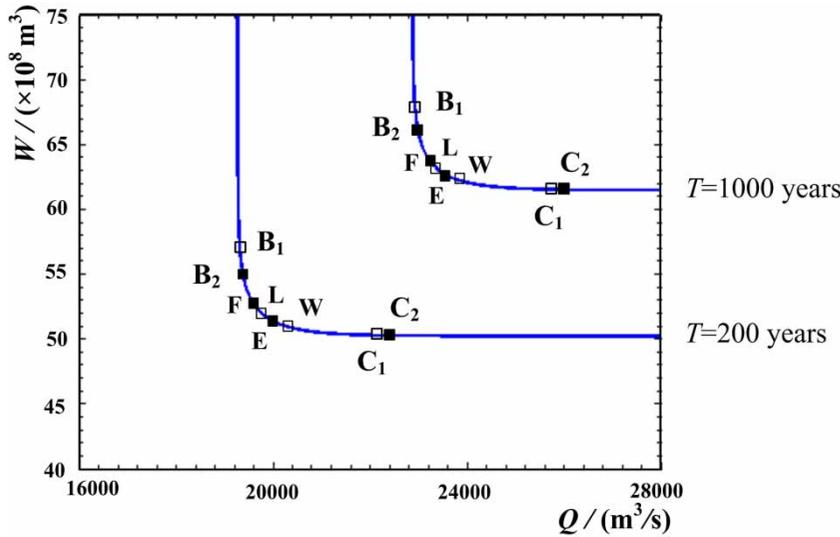
Figure 7 │ Bivariate quantile curve of joint distribution of flood peak and 7-day flood volume.
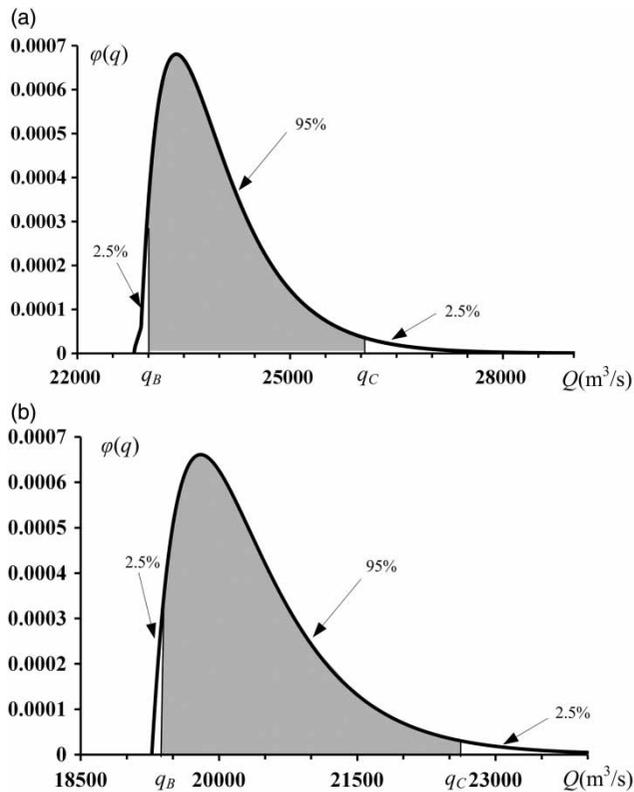

Figure 8 │ Relationship between density function $\varphi(q)$ and flood peak Q under different bivariate return periods.

Table 5 │ Comparison of the lower and upper bounds of the quantile curve

Boundary identification method	Return period	Lower bound		Upper bound	
		Q_p (m³/s)	W_7 (10⁸ m³)	Q_p (m³/s)	W_7 (10⁸ m³)
Volpi & Fiori (2012)	1,000	22,930	65.84	26,080	61.54
	200	19,350	50.27	22,460	55.86
Developed method	1,000	23,000	65.76	26,100	61.52
	200	19,400	54.49	22,500	50.26

Table 6 │ Design flood values and corresponding highest water levels estimated by bivariate quantile combinations and univariate distribution

T	Method	Q_p (m³/s)	W_7 (×10⁸ m³)	Z_{max} (m)
1,000	EFC	23,390	63.09	202.97
	CEC	23,420	62.98	202.92
	Component-wise excess realization	23,510	62.78	202.90
	Most-likely realization	23,400	63.05	202.95
	Univariate distribution	22,800	61.49	202.58
200	EFC	19,800	51.87	198.10
	CEC	20,130	51.11	197.79
	Component-wise excess realization	20,200	51.03	197.59
	Most-likely realization	19,940	51.50	197.82
	Univariate distribution	19,300	50.23	197.30

the events of component-wise excess realization are denoted as point W, and the events of most likely realization are denoted as point L. From Figure 7, we find that the joint

design values estimated by the four event selection methods are within the feasible regions. Consequently, the two proposed methods and selection approaches proposed by

Salvadori *et al.* (2011) could be selected as an option of deriving unique flood quantiles, and they could satisfy the inherent law of hydrologic events and have statistical basis to some degree. It can be seen from Table 6 and Figure 7 that the estimated events of the EFC method and that of the most likely realization are similar. The bivariate EFC results have larger flood volume and smaller flood peak than bivariate CEC results. As well, the results estimated by the component-wise excess realization have larger flood peak and smaller flood volume than the other three methods.

Results of statistical test

A Monte Carlo experiment was carried out to assess the performance of the proposed combination (EFC and CEC) methods. In order to describe the correlation between Q and W, the parameter of G–H copula is taken to be $\theta = 2.98$, while the Kendall correlation coefficient is 0.66. In order to test the validation of the experiment method, 10,000 samples of flood pair $(Q\text{-}W)$ were generated and the statistic values calculated based on the simulated data and given in Table 7. The Kendall correlation coefficient of the simulated samples was also calculated as 0.63. It can be seen that there is no significant difference between the generated and observed values. Thus, the generated samples could be used for statistical testing hereafter.

The flood samples of flood pair $(Q\text{-}W)$ with different sample sizes varying from $N = 30$ to $N = 100$ were generated. All experiment iterations were conducted $m_s = 500$ times for these simulations. The evaluation criteria of *RRMSE* and *MRE* are calculated and plotted in Figure 9. It is shown that with increasing data size N, there is a slight decrease in absolute value of *RRMSE* and *MRE*. When the sample size increases from 30 to 100, the *RRMSE* value decreases from 0.25% to 0.20% by 20% and
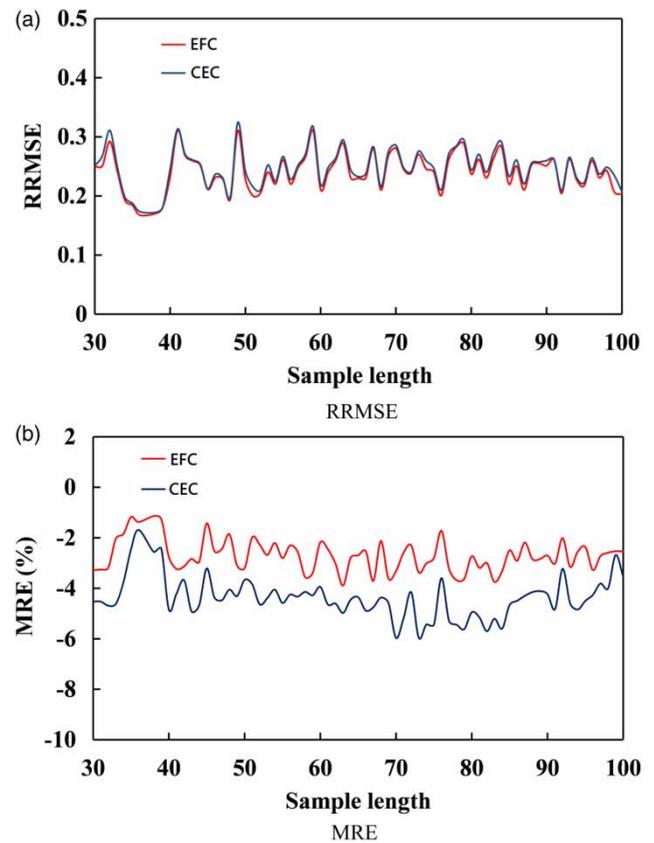

Figure 9 | Results of statistical experiment for EFC method and CEC method.

the absolute value of *MRE* decreases from 3.3% to 2.5% by 24% for the EFC method. In the case of the CEC method, the *RRMSE* value decreases by 16% and the absolute value of *MRE* decreases by 22%. It is shown in Figure 9 that smaller RRMSE and MRE absolute values are estimated by the EFC method. When the sample size is greater than 60 years, the values of *RRMSE* are almost less than 0.28 and the values of *MRE* are less than 4.0%; while for the CEC method, the *RRMSE* is less than 0.30 and the *MRE* is less than 6.0%. Monte Carlo experimental results show that the EFC method performs a little better than the conditional expectation method.

DFH based on joint distribution

The two combination methods were applied to derive the DFH, and the resulting highest reservoir water level was selected as an index to evaluate the effects of different hydrological loads on structure. The DFH for a dam is the

Table 7 | Results of statistical values for generated flood samples

	Sample statistic values		
Variables	**Mean**	C_v	C_s
Q_p (m^3/s)	7,803	0.39	1.18
W_7 (10^{8}m^3)	16.75	0.48	1.47

flood of suitable probability and magnitudes adopted to ensure safety of the dam in accordance with appropriate design standards. The AM flood hydrograph of 1997, which has high peak and large volume with posterior-peak shape, was selected as a typical flood hydrograph (TFH). The DFH with bivariate combinations was amplified from a TFH by the following method (Xiao *et al.* 2008):

$$DFH(t) = (TFH(t) - Q_{TFH}) \times (w/DT - q)/(W_{TFH}/DT - Q_{TFH}) + q \qquad (30)$$

where $DFH(t)$ and $TFH(t)$ are the flood discharges of the DFH and TFH for time t respectively; Q_{TFH} is flood peak discharge of TFH; W_{TFH} is 7-day flood volume of TFH for flood duration DT; q and w are flood peak and 7-day flood volume of bivariate design flood combination, respectively. Nevertheless, other DFH generation methods based on flood peak and volume are also available and can be applied with the bivariate design value combinations.

The DFHs of 1,000-year and 200-year return periods were constructed, respectively, with the bivariate EFC method and bivariate CEC method as shown in Figure 10. It is found in Figure 10 that only a few differences exist between the DFHs estimated by the EFC and CEC methods. This is due to the fact that the differences between the bivariate design values vary within a small range. Volpi & Fiori (2012) found that the feasible range on a *p*-level curve strongly depends on the correlation coefficient of *Q* and *W*. In the limiting case of full dependence, the level curve reduces to its vertex and the width of the feasible range tends to 0 (Volpi & Fiori 2012). Since the Kendall correlation coefficient between flood peak and 7-day volume in Geheyan reservoir equals to 0.66, the differences of quantiles estimated by EFC and CEC methods are relatively small in this case study.

The DFH rescaled by univariate distribution design values and two realizations proposed by Salvadori *et al.* (2011) was also derived from TFH by Equation (30). These DFHs were routed through the Geheyan reservoir with initial water level (flood control limiting water level, 192.2 m). The corresponding highest reservoir water levels (Z_{max}) were calculated and are listed in Table 6.

It is shown in Table 6 that the design values of flood peak and 7-day flood volume obtained by univariate

distribution method are both smaller than those obtained by four bivariate methods. The resulting Z_{max} of the univariate method is relatively lower than those of bivariate approaches. Since flood events are naturally multivariate phenomena and flood peak and flood volume are mutually correlated, the quantiles estimated by bivariate distribution are more rational than these by univariate distribution (Chebana & Ouarda 2011).

The comparison results listed in Table 6 also show that Z_{max} obtained by bivariate EFC method is larger than that obtained by the other three bivariate methods, while the component-wise excess method reaches the lowest Z_{max}. The results of Z_{max} calculated by most-likely realization are a little lower than the EFC method and the CEC method obtains a slightly higher Z_{max} than that of component-wise excess method. Comparing the results of 200-year and 1,000-year return period, it is found that the differences among the four bivariate methods decrease as the return period increases. The water level reaches 202.97 m by the EFC method and is slightly higher than other methods for the 1,000-year return period. Since the Geheyan reservoir has a large amount of flood control storage with annual regulation ability, the design flood volume is relatively more important than peak discharge for flood prevention safety. As a consequence, the bivariate EFC method with slightly larger 7-day flood volume is safer for reservoir design than other methods.

CONCLUSIONS

The quantile is an important and extensively studied notion in hydrological frequency analysis, but it is not properly addressed in the multivariate frequency analysis frameworks. Inspired by the idea of Chebana & Ouarda (2011) and Volpi & Fiori (2012), a developed approach to derive the feasible range and for estimating bivariate flood quantiles is proposed and applied in the Geheyan reservoir. The following conclusions were drawn from this study:

1. The developed quantile curve boundary identification method using the normalized joint PDF values to measure the relative occurrence likelihood of a
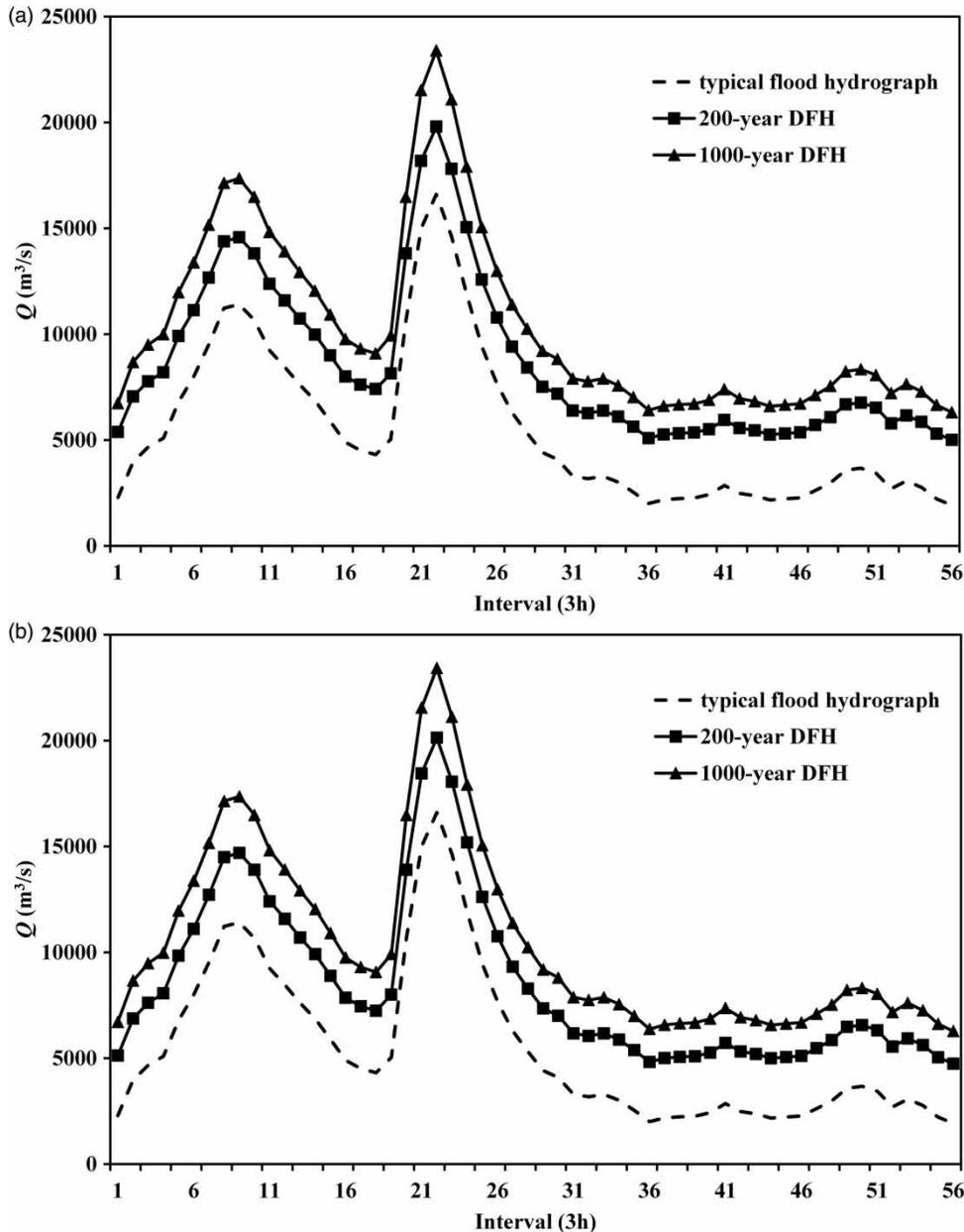
Figure 10 | DFHs derived by EFC method and CEC method.

seasonal flood event provides a developed applicable rule to narrow down the infinite number of possible choices for bivariate hydrological design values. The developed methods can be applied to bivariate OR return period and given the probability level ε.

2. The quantile combination methods provide a simple but effective way for bivariate quantile estimation

with given bivariate return period. The results illustrate that the joint design values estimated by the two proposed combination methods are within the feasible regions, and the EFC method perform satisfactorily.

3. The application results of Geheyan reservoir show that the bivariate EFC results with slightly larger 7-day flood volume can obtain the highest reservoir water level

(Z_{max}) and, therefore, EFC is safer for reservoir design than other methods.

ACKNOWLEDGEMENTS

REFERENCES

Abdul, R. U. & Zeephongsekul, P. 2014 Copula based analysis of rainfall severity and duration: a case study. *Theor. Appl. Climatol.* **115** (1–2), 153–166.

Ben, A., Chebana, F., Ouarda, T. B., Bruneau, P. & Barbet, M. 2015 Bivariate index-flood model: case study in Québec, Canada. *Hydrolog. Sci. J.* **60** (2), 247–268.

Chebana, F. & Ouarda, T. B. M. J. 2009 Index flood-based multivariate regional frequency analysis. *Water Resour. Res.* **45** (10). doi: 10.1029/2008WR007490.

Chebana, F. & Ouarda, T. B. M. J. 2011 Multivariate quantiles in hydrological frequency analysis. *Environmetrics* **22**, 63–78.

Chen, L., Guo, S. L., Yan, B. W., Liu, P. & Fang, B. 2010 A new seasonal design flood method based on bivariate joint distribution of flood magnitude and date of occurrence. *Hydrol. Sci. J.* **55** (8), 1264–1280.

Chen, L., Singh, V. P., Guo, S. L., Hao, Z. C. & Li, T. Y. 2012 Flood coincidence risk analysis using multivariate copula functions. *J. Hydrol. Eng.* **17** (6), 742–755.

Chen, L., Singh, V. P., Guo, S. L., Ashok, K. M. & Guo, J. 2013a Drought analysis using copulas. *J. Hydrol. Eng.* **18** (7), 797–808.

Chen, Y. D., Zhang, Q., Xiao, M. Z. & Singh, V. P. 2013b Evaluation of risk of hydrological droughts by the trivariate Plackett copula in the East River basin (China). *Nat. Hazard.* **68** (2), 529–547.

Chen, L., Ye, L., Singh, V. P., Zhou, J. Z. & Guo, S. L. 2014a Determination of input for artificial neural networks for flood forecasting using the copula entropy method. *J. Hydrol. Eng.* **19** (11), 04014021.

Chen, L., Singh, V. P., Guo, S. L. & Zhou, J. Z. 2014b Copula entropy coupled with artificial neural network for rainfall–runoff simulation. *Stoch. Environ. Res. Risk Assess.* **28** (7), 1755–1767.

Chen, L., Singh, V. P., Guo, S. L., Zhou, J. Z. & Zhang, J. H. 2015 Copula-based method for multisite monthly and daily streamflow simulation. *J. Hydrol.* **528**, 369–384.

Chowdhary, H., Escobar, L. & Singh, V. P. 2011 Identification of suitable copulas for bivariate frequency analysis of flood peak and flood volume data. *Hydrol. Res.* **42** (2–3), 193–216.

De Michele, C. & Salvadori, G. 2003 A generalized Pareto intensity-duration model of storm rainfall exploiting 2-copulas. *J. Geophys. Res.* **108** (2D). doi:10.1029/2002JD 002534.

De Michele, C., Salvadori, G., Canossi, M., Petaccia, A. & Rosso, R. 2005 Bivariate statistical approach to check adequacy of dam spillway. *J. Hydrol. Eng.* **1**, 50–57.

De Michele, C., Salvadori, G., Vezzoli, R. & Pecora, S. 2013 Multivariate assessment of droughts: frequency analysis and dynamic return period. *Water Resour. Res.* **49**, 6985–6994.

Ganguli, P. 2014 Probabilistic analysis of extreme droughts in Southern Maharashtra using bivariate copulas. *J. Hydrol. Eng.* **20** (1), 90–101.

Genest, C. & Mackay, L. 1986 The joy of copulas: bivariate distributions with uniform marginals. *Am. Stat.* **40**, 280–283.

Gräler, B., Van den Berg, M. J., Vandenberghe, S., Petroselli, A., Grimaldi, S., De Baets, B. & Verhoest, N. E. C. 2013 Multivariate return periods in hydrology: a critical and practical review focusing on synthetic design hydrograph estimation. *Hydrol. Earth Syst. Sci.* **17**, 1281–1296.

Kojadinovic, I., Yan, J. & Holmes, M. 2011 Fast large-sample goodness-of-fit tests for copulas. *Statistica Sinica* **21**, 814–817.

Lee, T. & Salas, J. 2011 Copula-based stochastic streamflow generation of hydrological data applied to Nile River flows. *Hydrol. Res.* **42** (4), 318–330.

Li, T. Y., Guo, S. L., Chen, L. & Guo, J. L. 2013 Bivariate flood frequency analysis with historical information based on Copula. *J. Hydrol. Eng.* **18** (8), 1018–1030.

Liu, P., Li, L., Guo, S., Xiong, L., Zhang, W., Zhang, J. & Xu, C.-Y. 2015 Optimal design of seasonal flood limited water levels and its application for the Three Gorges Reservoir. *J. Hydrol.* **527**, 1045–1053.

MWR (Ministry of Water Resources) 2006 Regulation for calculating design flood of water resources and hydropower projects, Water Resources & Hydropower Press, Beijing (in Chinese).

Nelsen, R. B. 2006 *An Introduction to Copulas,* 2nd edn. Springer, New York, USA.

Poulin, A., Huard, D., Favre, A. C. & Pugin, S. 2007 Importance of tail dependence in bivariate frequency analysis. *J. Hydrol. Eng.* **12** (4), 394–403.

Requena, A. I., Mediero, L. & Garrote, L. 2013 A bivariate return period based on copulas for hydrologic dam design: accounting for reservoir routing in risk estimation. *Hydrol. Earth Syst. Sci.* **17**, 3023–3038.

Salvadori, G. & De Michele, C. 2004 Frequency analysis via copulas: theoretical aspects and applications to hydrological events. *Water Resour. Res.* **40**, W12511. doi:10.1029/2004WR003133.

Salvadori, G. & De Michele, C. 2010 Multivariate multiparameter extreme value models and return periods: a copula approach. *Water Resour. Res.* **46**, W10501. doi:10.1029/2009WR009040.

Salvadori, G. & De Michele, C. 2015 Multivariate real-time assessment of droughts via copula-based multi-site hazard trajectories and fans. *J. Hydrol.* **526**, 101–115.

Salvadori, G., De Michele, C., Kottegoda, N. T. & Rosso, R. 2007 *Extremes in Nature: An Approach Using Copulas.* Springer Water Science and Technology Library, Berlin, Germany.

Salvadori, G., De Michele, C. & Durante, F. 2011 Multivariate design via Copulas. *Hydrol. Earth Syst. Sci. Discuss.* **8**, 5523–5558.

Salvadori, G., Durante, F. & Michele, C. 2013 Multivariate return period calculation via survival functions. *Water Resour. Res.* **49** (4), 2308–2311.

Serinaldi, F. 2015 Dismissing return periods! *Stoch. Env. Res. Risk Assess.* **29** (4), 1179–1189.

Shao, J. 2003 *Mathematical Statistics,* 2nd edn. Springer, New York, USA.

Shiau, T. H. 2003 Return period of bivariate distributed extreme hydrological events. *Stoch. Env. Res. Risk A.* **17** (1–2), 42–57.

Sklar, A. 1959 Fonctions de répartition à n dimensions et leurs marges. Publications de l'Institut de Statistique de L'Université, Paris, 8, pp. 229–231.

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.

Sun, S., Khu, S. T. & Djordjevic, S. 2013 Sampling rainfall events: a novel approach to generate large correlated samples. *Hydrol. Res.* **44** (2), 351–361.

Volpi, E. & Fiori, A. 2012 Design event selection in bivariate hydrological frequency analysis. *Hydrolog. Sci. J.* **57** (8), 1506–1515.

Volpi, E. & Fiori, A. 2014 Hydraulic structures subject to bivariate hydrological loads: return period, design, and risk assessment. *Water Resour. Res.* **50**, 885–897.

Volpi, E., Fiori, A., Grimaldi, S., Lombardo, F. & Koutsoyiannis, D. 2015 One hundred years of return period: strengths and limitations. *Water Resour. Res.* **51**, 8570–8585. doi:10.1002/2015WR017820.

Xiao, Y., Guo, S. L., Liu, P. & Fang, B. 2008 A new design flood hydrograph method based on bivariate joint distribution. In: *Hydrological Sciences for Managing Water Resources in the Asian Developing World* (X. H. Chen, Y. D. Chen, J. Xia & H. Zhang, eds). IAHS Publications 319. IAHS Press, Wallingford, UK, pp. 75–82.

Xiao, Y., Guo, S. L., Liu, P., Yan, B. W. & Chen, L. 2009 Design flood hydrograph based on multi-characteristic synthesis index method. *J. Hydrol. Eng.* **14** (12), 1359–1364.

Zhang, L. & Singh, V. P. 2006 Bivariate flood frequency analysis using the copula method. *J. Hydrol. Eng.* **11** (2), 150–164.

Zhang, Q., Singh, V. P., Li, J., Jiang, F. & Bai, Y. 2012 Spatio-temporal variations of precipitation extremes in Xinjiang, China. *J. Hydrol.* **434**, 7–18.

Zhang, Q., Qi, T., Singh, V. P., Chen, Y. D. & Xiao, M. 2015a Regional frequency analysis of droughts in China: a multivariate perspective. *Water Resour. Manage.* **29** (6), 1767–1787.

Zhang, Q., Xiao, M. & Singh, V. P. 2015b Uncertainty evaluation of copula analysis of hydrological droughts in the East River basin, China. *Global Planet Change.* **129**, 1–9.