# A meta-heuristic approach for multivariate design flood quantile estimation incorporating historical information

Jiabo Yin, Shenglian Guo, Xushu Wu, Guang Yang, Feng Xiong and Yanlai Zhou

## ABSTRACT

Design flood quantiles are crucial for hydraulic structures design, water resources planning and management, whereas previous multivariate hydrological quantile estimation methods usually do not consider historical flood information. To overcome such limitations, a meta-heuristic inference function for margins (MHIFM) approach, coupling meta-heuristic algorithm with a modified inference function for margins (IFM) method, is developed for modeling the joint distributions of flood peak and volumes with incorporation of historical flood information. Then, the most likely realization (MLR) and equivalent frequency combination (EFC) methods are employed for selecting multivariate design floods on a quantile iso-surface. The Danjiangkou reservoir located in Hanjiang River basin, the first pilot basin of most regulated water resources management policy in China, is selected as a case study. Application results indicate that the MHIFM approach shows good performance for estimating the parameters of marginal and joint distributions; moreover, the MLR method yields safer design flood quantiles than the EFC method in terms of highest routed reservoir water levels. The proposed MHIFM approach associated with the MLR method is safer and more rational for reservoir design, which would provide rich information as the reference for flood risk assessment, reservoir operation and management.

**Key words** | copula function, historical flood, meta-heuristic algorithm, multivariate frequency analysis, quantile estimation

**Jiabo Yin**
**Shenglian Guo** (corresponding author)
**Xushu Wu**
**Guang Yang**
**Feng Xiong**
**Yanlai Zhou**
State Key Laboratory of Water Resources and
 Hydropower Engineering Science,
Wuhan University,
Wuhan 430072,
China
E-mail: slguo@whu.edu.cn

## INTRODUCTION

Flood frequency analysis is performed to quantify the risk of flooding at different spatial locations and to provide guidelines for determining the design capacity of flood control structures. The design capacities of spillways, levees and several other flood control structures, and reservoir operation or management depend on the estimated magnitude of design flood quantiles (Xiao et al. 2009). Conventional hydrological frequency analysis mainly focuses on unique hydrological characteristics such as peak discharge or flood volume utilizing univariate distribution, which is unable to consider the interdependence between correlated hydrological variables and fails to provide a complete description of hydrological events (Favre et al. 2004). Therefore, a multivariate frequency analysis is needed for characterizing the random hydrological variables' interdependent features, particularly in designing hydraulic structures such as dams (Li et al. 2017).

The copula function, enabling the integration of heterogeneous dependence structures and marginal distributions, has been widely applied in constructing flexible joint distributions of interrelated hydrological variables since the introduction of copulas in hydrology and geosciences by De Michele & Salvadori (2003). This method has been applied mainly to the analysis of rainfall storms (e.g. Goel

*et al.* 2000; De Michele & Salvadori 2003; Wei & Song 2017), drought frequency analysis (e.g. Shiau *et al.* 2006; Chen *et al.* 2016), flood frequency analysis (e.g. Favre *et al.* 2004; Salvadori & De Michele 2004; Brunner *et al.* 2018), synthetic flood hydrograph (e.g. Gräler *et al.* 2013; Xu *et al.* 2016), regional frequency analysis (e.g. Chebana & Ouarda 2009; Zhang *et al.* 2015a), and several other multivariate domains (e.g. De Michele *et al.* 2005; Soltani *et al.* 2018).

While the copula-based methodology for multivariate frequency analysis is well established and explored, a common problem exists in both univariate and multivariate flood frequency analysis, i.e. the gauged record is rarely long enough to yield a reliable estimation of design flood quantiles, especially for high return periods. Furthermore, the existence of a cyclic variation over periods longer than the duration of the records would introduce further bias (Stedinger & Cohn 1986; Guo & Cunnane 1991). To resolve these problems, hydrologists have considerable interest in augmenting flow records, increasing sample size and estimating uncertainty. For example, the peak over threshold sampling method was used to extract flood series and thus significantly increased sample size compared with the annual maximum sampling method (Bhunya *et al.* 2013). Several pioneering approaches such as regionalization procedures (Reed 2002), Bayesian method (Zhang *et al.* 2015b) and bootstrapping approaches (Yin *et al.* 2018a, 2018b) have been suggested to reduce the design flood estimation uncertainty. The widely acknowledged and reliable technique is to incorporate historical or paleological information into flood frequency analysis, which has been recommended as an official procedure (e.g. NERC 1975; MWR 2006). Several methods, such as historically weighted moments, maximum likelihood, probability weighted moments, and L-moments (Guo & Cunnane 1991; Hosking 1995) have been proposed for the incorporation of historical or paleological information into univariate flood frequency analysis. To the authors' knowledge, only a very few studies address the bivariate characteristics of flood events considering historical information (e.g. Li *et al.* 2013), and incorporating flood information within a multivariate framework is still not available.

Besides the problem of incorporating historical floods, the question of estimating joint flood quantiles still has not yet been properly addressed in the multivariate frequency analysis framework (Chebana & Ouarda 2011; Yin *et al.* 2017). Under the multivariate framework, the choice of an appropriate return period for structure design leads to infinite combinations of correlated hydrological variables. These combinations are generally not equivalent, from a practical point of view, i.e. different joint design floods would lead to differences of flood hydrograph and highest water level which occurs in reservoir routing, which likely cause non-negligible discrepancy in spillway capacity design, reservoir operation and management (Volpi & Fiori 2012; Yin *et al.* 2018a, 2018b). To overcome such limitations and derive unique and rational design flood quantiles, many researchers (e.g. Liu *et al.* 2015; Xu *et al.* 2016; Li *et al.* 2017) adopted the equivalent frequency combination (EFC) method, which assumed that different hydrological variables have the same occurrence likelihood under a given Joint Return Period (JRP). Despite the EFC method having ease of implementation, the statistical and theoretical basis of the equal frequency assumption is questionable (Guo *et al.* 2018). To develop a more rational design flood estimation approach, several pioneering works have been conducted. For example, Salvadori *et al.* (2011) proposed a novel concept of 'most-likely design realization' to choose the event with the highest likelihood along the quantile curve. Li *et al.* (2017) developed a conditional expectation combination (CEC) method to derive the quantiles of flood peak and 7-day volume in Geheyan reservoir in China. Yin *et al.* (2017) developed the conditional most likely combination (CMLC) method to depict the interdependence between seasonal floods by using the conditional density function to measure the occurrence likelihood of seasonal flood events. While the CEC and CMLC methods are applicable to estimate the bivariate flood quantiles, they may have less physical realism in the multivariate framework.

Although several studies investigating the interdependence between correlated hydrological variables have been conducted, there is little previous work on deriving multivariate joint design floods with incorporation of historical information, which is extremely crucial in making decisions about flood control and water resources management, particularly for reservoirs with large regulation capacity. This study develops a meta-heuristic inference function for margins (MHIFM) approach to model the joint distributions of flood peak and volumes with incorporation of

historical information. Moreover, we derive the general for-
mulae of most likely realization (MLR) and EFC methods
for selecting multivariate design flood quantiles under a
given JRP.

The remainder of this paper is organized as follows. The
'Study area and data' section introduces the systematic
record and historical floods in Hanjiang River basin. In
the 'Methodology' section, we present the methodology
used to estimate the parameters of marginal and multi-
variate distributions and flood quantile selection methods.
The 'Results and discussion' section addresses a case
study in Danjiangkou reservoir. Finally, the 'Conclusions'
section summarizes the main results and provides necessary
explanations.

## STUDY AREA AND DATA

### Danjiangkou reservoir

The Hanjiang River basin, the first pilot basin of most
regulated water resources management policy in China, is
the largest tributary of the Yangtze River. The Hanjiang

River spans 1,570 km with a basin area of 159,000 km$^2$,
which is dominated by a sub-tropical monsoon climate.
The annual precipitation in Hanjiang River basin varies
from 700 to 1,100 mm, associated with large spatial and sea-
sonal variability. The rainfall from June to October accounts
for 75% of the annual amount, and the convective rain-
storms in the early summer and long-lasting frontal rains
in the autumn often lead to intense flash flood hazards.
The Danjiangkou reservoir marked in Figure 1 is a key
control project located in the middle reach of Hanjiang
River basin, which is the source for the middle route of
the South–North Water Diversion Project in China. The
Danjiangkou reservoir is a multi-purpose reservoir, which
has serious conflicts between flood control, water supply
and hydropower generation. The inflow of Danjiangkou
reservoir is mainly concentrated from late June to early
October, and intense runoff during July and September
plays a dominant role in water resources utilization and
flood risk prevention. The Danjiangkou reservoir has
multi-years regulation ability and the flood control storage
during flood season is over 11.1 billion m$^3$. Since the
Danjiangkou reservoir is the key structural component in
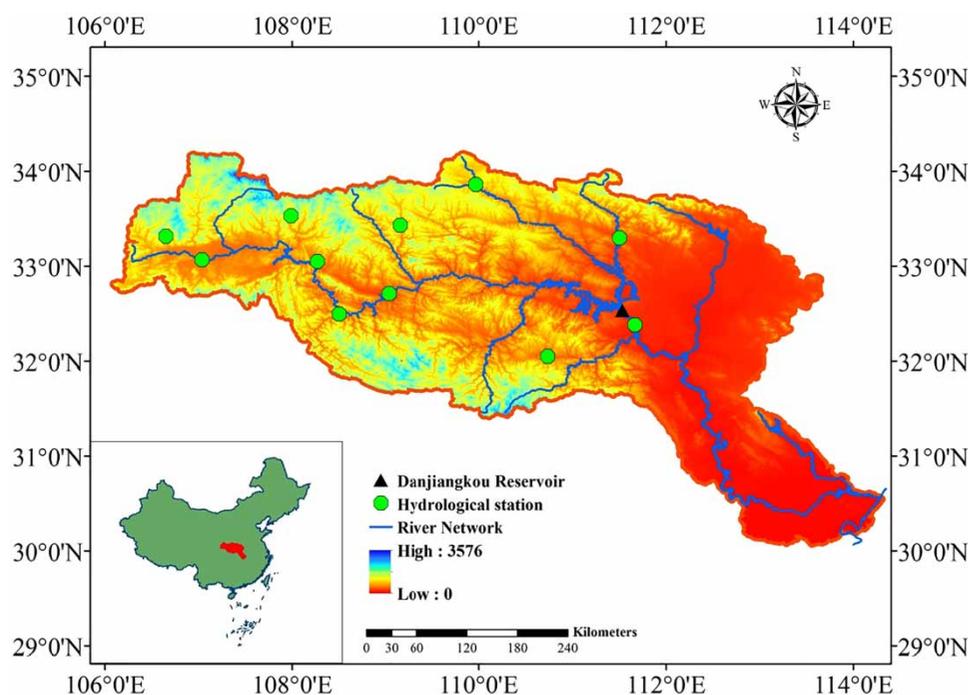the Hanjiang River flood control system and plays a critical



**Figure 1** │ Sketch map of the Danjiangkou reservoir and the Hanjiang River basin.

role in flood control and water supply in China, it is of great significance to investigate its design floods for guiding reservoir operation and management.

## Systematic record and historical floods

According to the regulation and operation guidelines of the Danjiangkou reservoir, peak discharge, 7- and 15-day flood volumes are considered for deriving the design flood quantiles and the corresponding design flood hydrograph (DFH) (Xu et al. 2016). The annual maximum peak discharge, 7-day flood volume, and 15-day flood volume are available with a systematic record of 86 years (1929–2014) from the Bureau of Hanjiang River Water Resources Management. These data series after the dam construction have been restored to natural flow and have been checked for possible long-term trends and abrupt changes, with the test results demonstrating that the dataset is stationary and represented (Xu et al. 2016). The dataset has been widely employed in hydraulic structure capacity design, reservoir operation and water resources management in the Danjiangkou reservoir (e.g. Yang et al. 2017; Zhou et al. 2017; Shen et al. 2018). Besides the systematic observations, the historical flood events have been investigated by Changjiang Water Resources Commission (CWRC) in China. The gathered information is from gauging authority records, historical documents, archives, flood marks and stone inscriptions showing the concrete positions of high water stages recorded. As a result, the six largest historical floods (see Table 1) since 1,583 were quantitatively evaluated by the CWRC.

**Table 1** | Investigated historical floods of the Danjiangkou reservoir in Hanjiang River basin

| Year | Flood peak (m³/s) | 7-day flood volume (10⁸ m³) | 15-day flood volume (10⁸ m³) |
|------|-------------------|------------------------------|-------------------------------|
| 1583 | 61,000 | 171 | 238 |
| 1867 | 45,500 | 131 | 182 |
| 1852 | 45,000 | 130 | 180 |
| 1832 | 44,700 | 129 | 179 |
| 1693 | 42,500 | 124 | 171 |
| 1921 | 38,000 | 112 | 155 |

## METHODOLOGY

### Copula functions and JRP

This study focuses on flood peak discharge ($Q$) and flood volumes ($W_i$($i = 1, 2, \ldots, m$)) to investigate the joint probability of flood occurrence by copulas and to derive the DFHs, in which $m$ represents the number of flood volumes needed for estimating DFHs. The multivariate joint distribution $H$ can be expressed in terms of univariate marginal distributions and the associated dependence function, known as Sklar's theorem (Sklar 1959):

$$H(x_1, x_2, \cdots, x_n) = C_\theta(F_{X_1}(x_1), F_{X_2}(x_2), \cdots, F_{X_n}(x_n)) \qquad (1)$$

where $F_{X_i}(x_i)$ ($i = 1, 2, \cdots, n$) denotes the marginal cumulative density functions (CDFs) of $X_i$, $C_\theta$ is the copula function with corresponding parameter vector $\theta$.

In the conventional univariate framework, flood events of interest are often defined by return periods. In the multivariate context, different definitions of the JRPs have been proposed, such as the OR, AND, Kendall, dynamic, structure-based return periods, and so on (Shiau 2003; Requena et al. 2013; De Michele et al. 2013; Volpi & Fiori 2014; Salvadori et al. 2016). The Danjiangkou reservoir is at risk when either flood peak or flood volume exceeds the design standard, and therefore the multivariate phenomenon for this particular situation is defined based on OR JRP, which can be defined by copulas accordingly (De Michele et al. 2005):

$$T_{OR} = \frac{\mu}{P[Q > q \ or \ W_1 > w_1, .., W_m > w_m]}$$
$$= \frac{\mu}{1 - F(q, w_1, \ldots, w_m)} = \frac{\mu}{1 - C_\theta(u, v_1, \ldots, v_m)} \qquad (2)$$

where $\mu$ is the mean inter-arrival time between two consecutive events (in the case of annual maxima $\mu = 1$ year); $u = F_Q(q)$, $v_i = F_{W_i}(w_i)$ are the marginal distributions of peak discharge and flood volumes, respectively.

### Univariate maximum likelihood estimation with historical information

In univariate hydrological frequency analysis, several methods for incorporating historical information have

been proposed, among which the maximum likelihood estimation method is widely acknowledged and employed (Hald 1949; Stedinger & Cohn 1986; Chen & Singh 2018). In the annual maximum flood series shown in Figure 2, there is a total of $g$ known floods. Of these, $l$ is known to be the largest flood during the maximum historical textual period of $n$ years, while the $c$ large flood occurred during the systematic record (recently gauged data) of $s$ year ($c \leq l$, $c < s$, $s \leq n$ and $g = s + l - c$). A fixed threshold $X_0$ is assumed to be exceeded by the $l$ largest floods and is not exceeded by any of the remaining $n - l$ floods. It is also noted that the $r$ ($r = l - c$) floods in the pre-gauging period $h$ ($h = n - s$) are known as they are included in the $l$ floods which exceed $X_0$, and it is assumed that no other floods exceed the threshold during the pre-gauging period.

Let $f_X$ and $F_X$ denote the probability density function (PDF) and the CDF of variable $X$, respectively. The likelihood function for the mixed sample of $s + r$ known and $h - r$ unknown data is given as (Condie 1986; Stedinger & Cohn 1986; Guo & Cunnane 1991):

$$l(\boldsymbol{\alpha}) = \prod_{i=1}^{s+r} f_X(x_i) \left[ \int_{-\infty}^{X_0} f_X(x)dx \right]^{h-r} \tag{3}$$

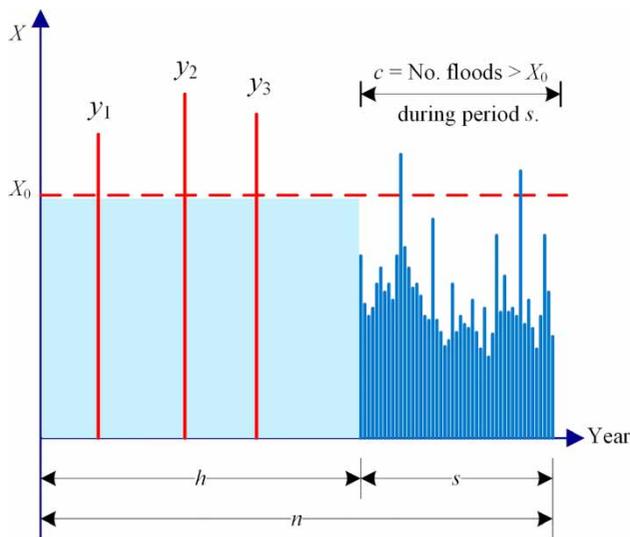where $\boldsymbol{\alpha}$ is the parameter vector of $f_X$ and $F_X$.



**Figure 2** | Sketch of the annual maximum flood series when historical floods are available.

Since $c$ flood events exceeding the perception threshold $X_0$ occur during the systematic period $s$ (analogously to the sketch in Figure 2), the $c$ events are removed from the period $s$ and are treated as historical data (Bayliss & Reed 2001). Then, Equation (3) can be expressed as:

$$l(\boldsymbol{\alpha}) = \prod_{i=1}^{s-c} f_X(x_i) \prod_{j=1}^{l} f_X(y_j) \left[ \int_{-\infty}^{X_0} f_X(x)dx \right]^{h-r} \tag{4}$$

where $x_i(i = 1, 2 \ldots s - c)$ denotes the systematic data less than the threshold $X_0$, $y_j$ ($j = 1, 2 \ldots l$) denotes the $l$ largest floods exceeding the threshold $X_0$.

The log-likelihood function for the univariate distribution can be expressed as (Li *et al.* 2013):

$$L(\boldsymbol{\alpha}) = \sum_{i=1}^{s-c} \log f_X(x_i) + \sum_{j=1}^{l} \log f_X(y_j) + (h - r) \log F_X(X_0) \tag{5}$$

The maximum likelihood estimates are those values of $\boldsymbol{\alpha}$ that maximize Equation (5).

## MHIFM approach

The inference function for margins (IFM) method is the preferred fully parametric method for estimating the parameters of multivariate distributions due to its robustness and ease of implementation (Joe & Xu 1996; Joe 1997; Baharith 2017). The conventional IFM method is only applicable for systematic data series without historical floods, which could not satisfy the requirements of engineering practice in spillway capacity design. Li *et al.* (2013) developed a modified IFM method to estimate the parameters of marginal and bivariate distribution with consideration of historical information. The modified IFM method focused on solving the bivariate problems, failing to incorporate flood information within a multivariate framework, thus limiting the applicability in reservoirs with large capacity. To address such limitations, an MHIFM approach is developed as discussed below.

Let $q^j$ and $w_i^j$ ($j = 1, \ldots, s - c$) denote the systematic data of flood peak ($Q$) and volumes ($W_i$), respectively; $g^k$ and $z_i^k$ ($k = 1, \ldots, l$) denote the flood peak and volumes of

the $l$ largest floods with the same length, respectively; $Q_0$ (or $W_{0i}$) is the fixed threshold exceeded by the $l$ largest flood peaks (or volumes) and not exceeded by any of the remaining $n - l$ flood peaks (or volumes). Furthermore, let $f_Q(q)$, $f_{W_1}(w_i)$ denote the univariate marginal PDFs, and $F_Q(q)$, $F_{W_i}(w_i)$ denote the univariate marginal CDFs of variables $Q$ and $W_i$, respectively; $f_{QW_1,\ldots,W_m}$ denotes the joint PDF.

For the bivariate case ($m$ equals 1), referring to Equation (5), the likelihood function with historical floods for bivariate joint distribution can be described as (Li *et al.* 2013):

$$
\begin{aligned}
l(\theta, \alpha_1, \alpha_2) &= F^{h-r} \prod_{j=1}^{s-c} f_{QW_1}(q^j, w_1^j) \prod_{k=1}^{l} f_{QW_1}(g^k, z_1^k) \\
&= C_\theta^{h-r} \prod_{j=1}^{s-c} f_{QW_1}(q^j, w_1^j) \prod_{k=1}^{l} f_{QW_1}(g^k, z_1^k)
\end{aligned}
\tag{6}
$$

where $F = \int_{-\infty}^{Q_0} \int_{-\infty}^{W_{01}} f_{QW_1}(q, w_1) dq dw_1$, $C_\theta = C_\theta[F_Q(Q_0), F_{W_1}(W_{01})]$, $\alpha_1$ and $\alpha_2$ denote the parameter vector of $F_Q(q)$ and $F_{W_1}(w_1)$ respectively.

Then, the log-likelihood function for bivariate joint distribution can be expressed as:

$$
\begin{aligned}
L(\theta, \alpha_1, \alpha_2) &= \sum_{j=1}^{s-c} \log c_\theta[F_Q(q_j), F_{W_1}(w_1^j)] \\
&+ \sum_{k=1}^{l} \log c_\theta[F_Q(g^k), F_{W_1}(z_1^k)] + (h-r) \log C_\theta \\
&+ \sum_{j=1}^{s-c} \log f_Q(q_j) + \sum_{k=1}^{l} \log f_Q(g_k) \\
&+ \sum_{j=1}^{s-c} \log f_{W_1}(w_1^j) + \sum_{k=1}^{l} \log f_{W_1}(z_1^k)
\end{aligned}
\tag{7}
$$

Similarly, for the multivariate case, the likelihood function with historical floods for joint distribution can be derived as:

$$
\begin{aligned}
l(\theta, \alpha_1,, \ldots, \alpha_{m+1}) &= \prod_{j=1}^{s-c} f_{QW_1,\ldots,W_m}(q^j, \ldots, w_m^j) \prod_{k=1}^{l} f_{QW_1,\ldots,W_m}(g^k, \ldots, z_m^k) \\
&\times \left[ \int_{-\infty}^{Q_0} \cdots \int_{-\infty}^{W_{0m}} f_{QW_1,\ldots,W_m}(q^j, \ldots, w_m^j) dq \cdots dw_m \right]^{h-r} \\
&= C_0^{h-r} \prod_{j=1}^{s-c} f_{QW_1,\ldots,W_m}(q^j, \ldots, w_m^j) \prod_{k=1}^{l} f_{QW_1,\ldots,W_m}(g^k, \ldots, z_m^k)
\end{aligned}
\tag{8}
$$

where $C_0 = C_\theta[F_Q(Q_0), \ldots, F_{W_m}(W_{0m})]$, $\alpha_{m+1}$ denotes the parameter vector of $F(w_m)$ respectively.

Under the assumption that the marginal distributions are continuous with the probability density functions, the joint PDF can be expressed by copulas as:

$$
f_{QW_1,\ldots,W_m}(q, w_1, \ldots, w_m) = c_\theta f_Q(q) \cdot f_{W_1}(w_1), \ldots, \cdot f_{W_m}(w_m)
\tag{9}
$$

where $c_\theta$ is the copula density function of $Q$, $W_1, \ldots, W_m$.

Then, the log-likelihood function for multivariate joint distribution can be expressed as:

$$
\begin{aligned}
L(\theta, \alpha_1, \ldots, \alpha_{m+1}) &= \sum_{j=1}^{s-c} \log c_\theta[F_Q(q_j), \ldots, F_{W_m}(w_m^j)] \\
&+ \sum_{k=1}^{l} \log c_\theta[F_Q(g^k), F_{W_1}(z_1^k) \ldots, F_{W_m}(z_m^k)] \\
&+ (h-r) \log C_\theta[F_Q(Q_0), \ldots, F_{W_m}(W_{0m})] \\
&+ \sum_{j=1}^{s-c} \log f_Q(q_j) + \sum_{k=1}^{l} \log f_Q(g_k) \\
&+ \sum_{j=1}^{s-c} \log f_{W_1}(w_1^j) + \sum_{k=1}^{l} \log f_{W_1}(z_1^k) \\
&+ \cdots \\
&+ \sum_{j=1}^{s-c} \log f_{W_m}(w_m^j) + \sum_{k=1}^{l} \log f_{W_m}(z_m^k)
\end{aligned}
\tag{10}
$$

In which, the $m + 1$ log-likelihood functions for the univariate marginal distribution are:

$$
\begin{aligned}
L_1(\alpha_1) &= \sum_{j=1}^{s-c} \log f_Q(q^j) + \sum_{k=1}^{l} \log f_Q(g^k) \\
L_2(\alpha_2) &= \sum_{j=1}^{s-c} \log f_{W_1}(w_1^j) + \sum_{k=1}^{l} \log f_{W_1}(z_1^k) \\
&\cdots \\
L_{m+1}(\alpha_{m+1}) &= \sum_{j=1}^{s-c} \log f_{W_m}(w_m^j) + \sum_{k=1}^{l} \log f_{W_m}(z_m^k)
\end{aligned}
\tag{11}
$$

The parameter estimation procedure consists of $m + 1$ separate optimizations. The optimization of multivariate

likelihood is a function of the dependence parameter vector. More specifically:

(a) the log-likelihood $L_1(\boldsymbol{\alpha_1}) \dots L_{m+1}(\boldsymbol{\alpha_{m+1}})$ of the $m + 1$ univariate marginal distributions are separately maximized by Equation (11) to obtain estimates $\hat{\boldsymbol{\alpha}}_1, \hat{\boldsymbol{\alpha}}_2, \dots, \hat{\boldsymbol{\alpha}}_{m+1}$;

(b) the function $L(\boldsymbol{\theta}, \hat{\boldsymbol{\alpha}}_1, \hat{\boldsymbol{\alpha}}_2, \dots, \hat{\boldsymbol{\alpha}}_{m+1})$ is maximized over $\boldsymbol{\theta}$ to obtain $\hat{\boldsymbol{\theta}}$ in Equation (10). That is, under regularity conditions, $(\hat{\boldsymbol{\alpha}}_1, \hat{\boldsymbol{\alpha}}_2, \dots, \hat{\boldsymbol{\alpha}}_{m+1})$ is the solution of:

$$(\partial L_1/\partial \boldsymbol{\alpha_1}, \partial L_2/\partial \boldsymbol{\alpha_2}, \dots, \partial L_{m+1}/\partial \boldsymbol{\alpha_{m+1}}, \partial L/\partial \boldsymbol{\theta}) = 0 \tag{12}$$

This procedure is computationally simpler than that of estimating all parameters $\boldsymbol{\alpha_1}, \boldsymbol{\alpha_2}, \dots, \boldsymbol{\alpha_m}, \boldsymbol{\theta}$ simultaneously in Equation (12). To maximize the log-likelihood functions, Equations (10) and (11), the conventional Newton–Raphson method may not work due to the complicated high-dimension copula functions and local optimum problem. To overcome these deficiencies, a meta-heuristic algorithm could be incorporated with the modified IFM method considering historical floods to maximize the log-likelihood function. Many meta-heuristic algorithms that combine rules and randomness imitating natural phenomena have been devised to solve difficult and complex engineering optimization problems, which include simulated annealing, tabu search, and evolutionary computation methods (Kang & Zong 2005). The simulated annealing algorithm (SAA) is a stochastic neighborhood search method that is developed for the combinatorial optimization problems based on the analogy of the process of annealing of solids (Kirkpatrick *et al.* 1983). It has the capability of jumping out of the local optima for global optimization (Baykasoğlu & Gindy 2001), so it is used to estimate the parameters of marginal and joint distributions in the MHIFM method. The SAA emulates the gradual cooling process (i.e. annealing) of material, during which the temperature decreases stepwise. At each temperature, an initial configuration is generated and employed to evaluate the corresponding objective function $E_a$ (energy value). Then, a small displacement is imposed on the current configuration to obtain a new objective function $E_b$. If $\delta E = E_b - E_a < 0$, this new configuration is accepted as the best solution. Otherwise, it is accepted

with the probability:

$$p(\delta E) = exp(-\tfrac{\delta E}{KT}) \tag{13}$$

where $T$ is the current temperature and $K$ is the Boltzmann's constant. If this probability is higher than a random value, this configuration is accepted as the best, although it is not better than the previous solution. Once the system has reached equilibrium at the current temperature $T$, the algorithm continues and reduces temperature gradually until the system reaches its minimum temperature (Kirkpatrick *et al.* 1983; Bagherlou & Ghaffari 2018).

In this study, the optimization problem in SAA is a constrained single-objective optimization minimizing the opposite of the log-likelihood function $L_1, L_2, \dots, L_{m+1}$ or $L$ separately in Equations (10) and (11), i.e. $-L_1, -L_2, \dots, -L_{m+1}$ or $-L$. The main procedures of the SAA for the parameter estimation are summarized as follows:

(1) Initialize the temperature $T_0$.
(2) Set a variation interval for the parameters of distribution and generate initial solution configuration randomly.
(3) Calculate the value of objective function $E_a$ (e.g. $-L_a$).
(4) Generate a new solution configuration by imposing a small displacement on current configuration and evaluate the value of objective function $E_b$ (e.g. $-L_b$).
(5) If $\delta E = E_b - E_a = L_a - L_b < 0$, the new solution is accepted. Otherwise, accept the new solution with a probability $p(\delta E)$.
(6) Repeat steps (4) and (5) at current temperature.
(7) Reduce the temperature stepwise and repeat steps (2)–(6) until the stop criterion is satisfied.

## EFC method

After modelling the joint distribution with incorporation of historical information by the developed MHIFM approach, we need to estimate the design flood quantiles, which is necessary for flood control system design, water resources management and reservoir operation. However, for a given JRP $T_{OR}$, there are innumerable combinations of flood peak and volumes that satisfy Equation (2). Take the bivariate case as an example: the critical level curve as shown in Figure 3 is defined as a bivariate quantile curve
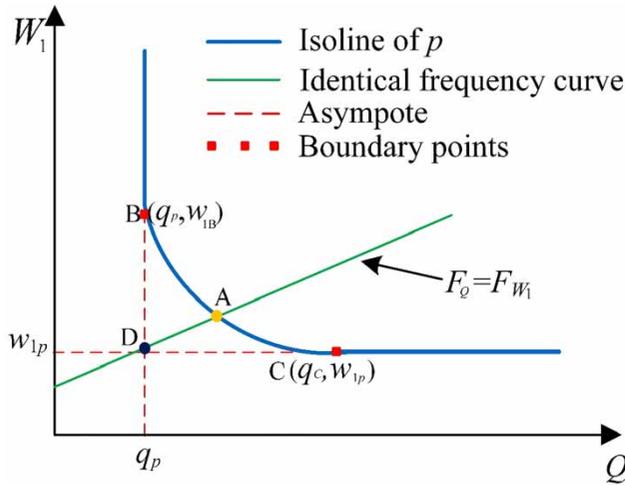
**Figure 3** | Bivariate quantile curve with a critical probability level *p*.

under the OR JRP, with countless quantile combinations satisfying the flood prevention standard. Chebana & Ouarda (2011) segmented the *p*-level curve into a central part (i.e. the subset *BC* in Figure 3) and a tail part by employing its geometric properties. All the bivariate events along the quantile curve have the same JRP; however, these events may not be equivalent from a practical point of view (Volpi & Fiori 2012; Yin *et al.* 2017).

To meet the needs in engineering, a unique combination of flood peak and volumes should be determined. Hence, besides Equation (2), one or more constraints characterizing the interdependence between hydrological variables are necessary. The EFC method recommended by the Chinese design flood guidelines (MWR 2006) is widely employed due to its ease of implementation, which assumes that the flood peak and volumes have the same probability of occurrence, i.e. $u = v_1 = , \ldots, v_m$ (or $F_Q(q) = F_{W_1}(w_1) = , \ldots, F_{W_m}(w_m)$). As shown in Figure 3, the tangent curve on which $F_Q(q) = F_{W_1}(w_1)$ is depicted as a chosen constraint on the critical level curve and the intersection *A* point demonstrates the unique estimator of the EFC method. Moreover, point *D* depicted in Figure 3 indicates the estimated design flood quantiles by the conventional univariate method, in which only the marginal distribution is used under the assumption that the design frequency of different variables is equal to the joint design frequency.

Taking the Asymmetric Gumbel–Hougaard (A-GH) copula within trivariate framework as an example, the relationship of joint distribution $C_\theta(u, v_1, v_2)$ and trivariate JRP $T_{OR}$ can be expressed as:

$$C_\theta(u, v_1, v_2) = \exp\{ - ([(-\ln u)^{\theta_2} + (-\ln v_1)^{\theta_2}]^{\theta_1/\theta_2} + (-\ln v_2)^{\theta_1})^{1/\theta_1}\} = 1 - \frac{1}{T_{OR}} \tag{14}$$

where $\theta_1$, $\theta_2$ is the parameter of A-GH copula function.

Based on the assumption that $u = v_1 = v_2$, the probabilities of occurrence of design flood quantiles (i.e. $u$, $v_1$ and $v_2$) can be estimated by the solution of Equation (14), which is expressed as follows:

$$u = v_1 = v_2 = (1 - 1/T_{OR})^{[2^{\theta_1/\theta_2}+1]^{-1/\theta_1}} \tag{15}$$

It should be noted that other copulas with more complicated formulas or higher-dimension copulas are sometimes needed and the values of $u, v_1, \ldots, v_m$ could be estimated by the numerical method such as a trial and error method (Liu *et al.* 2015).

After obtaining the values of $u, v_1, \ldots, v_m$ for the flood events along the critical level curve, the design value of the EFC method can be derived by the inverse function of marginal distributions with corresponding parameter vector:

$$\begin{cases} q = F_Q^{-1}(u; \boldsymbol{\alpha}_1) \\ w_1 = F_{W_1}^{-1}(v_1; \boldsymbol{\alpha}_2) \\ \cdots \\ w_m = F_{W_m}^{-1}(v_m; \boldsymbol{\alpha}_{m+1}) \end{cases} \tag{16}$$

### MLR of flood quantile

As previously stated, all possible flood events under a given JRP $T_{OR}$ differ in terms of their probability of occurrence. Salvadori *et al.* (2011) measured the likelihood of flood events with the joint PDF and proposed the novel concept of 'most likely design realization' to choose the point with the highest likelihood along the *p*-level isolines. In this paper, the general formulae of MLR are derived as discussed below.

For a given JRP, the most likely flood event $(q*, w_1^*, \ldots, w_m^*)$ at the quantile iso-surface is derived by selecting the event with the highest joint probability density:

$$\begin{cases} \mathrm{argmax} f_{QW_1 \cdots W_m}(q, w_1, \ldots, w_m) = c_\theta \cdot f_Q(q) \cdot f_{W_1}(w_1), \cdots f_{W_m}(w_m) \\ C_\theta[F_Q(q), F_{W_1}(w_1), \ldots, F_{W_m}(w_m)] = 1 - 1/T_{\mathrm{OR}} \end{cases}$$
(17)

The Lagrange multiplier method (Chen & Singh 2017a, 2017b; Yin *et al.* 2018a, 2018b) is applied for solving Equation (17) to obtain the MLR. For a given JRP $T_{\mathrm{OR}}$, the Lagrange function is constructed as:

$$\varphi(q, w_1, \ldots, w_m) = c_\theta f_Q(q) f_{W_1}(w_1), \ldots, f_{W_m}(w_m) \\ + \lambda \{C_\theta[F_Q(q), F_{W_1}(w_1), \ldots, F_{W_m}(w_m)] - (1 - 1/T_{\mathrm{OR}})\}$$
(18)

where $\lambda$ is the Lagrange multiplier.

The first order derivative equal to zero will reach the maximum value. For the bivariate case, the following equation should be satisfied:

$$\begin{cases} \varphi_q = & f_{W_1}(w_1)[c_u f_Q^2(q) + c(u, v_1) f'_Q(q)] \\ & + \lambda \dfrac{\partial c[F_Q(q), F_{W_1}(w_1)]}{\partial u} f_Q(q) = 0 \\ \varphi_{w_1} = & f_Q(q)[c_{v_1} f_{W_1}^2(w_1) + c(u, v_1) f'_{W_1}(w_1)] \\ & + \lambda \dfrac{\partial C[F_Q(q), \ F_{W_1}(w_1)]}{\partial v_1} f_{W_1}(w_1) = 0 \\ \varphi_\lambda = & C[F_Q(q), F_{W_1}(w_1)] - (1 - 1/T_{OR}) = 0 \end{cases}$$
(19)

where $c_u = \dfrac{\partial c(u, v_1)}{\partial u}$, $c_{v1} = \dfrac{\partial c(u, v_1)}{\partial u}$ and $f'_Q(q)$ and $f'_{W_1}(w_1)$ is the derivative function of $f_Q(q)$ and $f_{W_1}(w_1)$, respectively.

For $(m + 1)$ dimension case, the following equation should be satisfied:

$$\begin{cases} \varphi_q = (c_u f_Q^2(q) + cf'_Q(q)) \prod\limits_{i=1}^m f_{W_i}(w_i) + \lambda C_u = 0 \\ \varphi_{w_1} = f_Q(q) \prod\limits_{i=2}^m f_{W_i}(w_i) \cdot (c_{v_1} f_{W_1}^2(w_1) + cf'_{W_1}(w_1)) + \lambda C_{v_1} = 0 \\ \cdots\cdots \\ \varphi_{w_n} = f_Q(q) \prod\limits_{i=1}^{m-1} f_{W_i}(w_i) \cdot (c_{v_m} f_{W_m}^2(w_m) + cf'_{W_m}(w_m)) + \lambda C_{v_m} = 0 \\ \varphi_\lambda = C[F_Q(q), F_{W_1}(w_1), \cdots, F_{W_m}(w_m)] - (1 - 1/T_{\mathrm{OR}}) = 0 \end{cases}$$
(20)

where $c_{v_m} = \partial c(u, v_1, \cdots, v_m)/\partial v_m$, $f'_{W_m}(w_m)$ is the derivative function of $f_{W_m}(w_m)$,

$$C_u = \dfrac{\partial C[F_Q(q), F_{W_1}(w_1), \cdots, F_{W_m}(w_m)]}{\partial u} f_Q(q),$$

$$C_{v_m} = \dfrac{\partial C[F_Q(q), F_{W_1}(w_1), \cdots, F_{W_m}(w_m)]}{\partial v_m} f_{W_m}(w_m).$$

The nonlinear Equation (20) can be solved by numerical methods, such as the incomplete Jacobian Newton method proposed by Liu & Ni (2008).

## RESULTS AND DISCUSSION

### Parameter estimation for marginal distributions

For annual maximum flood series, the Pearson type three (P-III) distributions have been recommended by MWR (2006) as a uniform procedure for flood frequency analysis in China. The PDF of the P-III distribution is defined as:

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} (x - \delta)^{\alpha-1} e^{-\beta(x-\delta)}, \alpha > 0, \beta > 0, \delta \le x < \infty \qquad (21)$$

where $\alpha$, $\beta$ and $\delta$ are shape, scale and location parameters of the P-III distribution, respectively.

The series of flood peak discharge ($Q$), 7-day flood volume ($W_1$) and 15-day flood volume ($W_2$) were considered in the Danjiangkou reservoir following a previous study (Xu *et al.* 2016). The length of the systematic observations is unequivocally given: $s = 86$ years. Since one flood event occurring in 1935 ranked second among the historical and systematic records, this flood is usually taken as an extraordinary flood during the systematic period, and thus $c = 1$ and $r = l - 1$. While six historical floods during the pregauge period $h = 347$ (i.e. from 1583 to 1929) were investigated, the total largest floods $l$ equals 7 and $r$ equals 6; the perception thresholds of peak discharge, 7-day flood volume and 15-day flood volume are $Q_0 = 3,700 \ \mathrm{m^3/s}$, $W_{01} = 110 \ 10^8 \ \mathrm{m^3}$ and $W_{02} = 150 \ 10^8 \ \mathrm{m^3}$, respectively.

The parameters of the P-III marginal distributions were estimated by the first stage of the MHIFM approach in Equation (11) and are listed in Table 2. To compare the estimated design flood quantiles by the developed MHEFM approach with that by the conventional IFM method, we also estimated the parameters of marginal

**Table 2** │ Estimated parameters of P-III marginal distributions and hypothesis test results for flood peak and volumes

| Variables | P-III parameters | | | K-S Test | | Chi-Square Test | | P-III parameters | | | K-S Test | | Chi-Square Test | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\alpha$ | $\beta$ | $\delta$ | $D_{KS}$ | $D_{0.05}$ | $\chi^2$ | $\chi^2_{0.05}$ | $\alpha$ | $\beta$ | $\delta$ | $D_{KS}$ | $D_{0.05}$ | $\chi^2$ | $\chi^2_{0.05}$ |
| $Q$ (m$^3$/s) | 1.91 | 0.0001 | 2,452.02 | 0.103 | 0.142 | 3.251 | 5.016 | 2.78 | 0.0002 | 1,038.87 | 0.115 | 0.147 | 4.154 | 5.224 |
| $W_1$ (10$^8$ m$^3$) | 1.68 | 0.0473 | 10.16 | 0.124 | 0.142 | 3.651 | 5.016 | 2.43 | 0.0606 | 5.03 | 0.125 | 0.147 | 4.243 | 5.224 |
| $W_2$ (10$^8$ m$^3$) | 2.18 | 0.0401 | 17.86 | 0.095 | 0.142 | 3.025 | 5.016 | 3.21 | 0.0479 | 3.61 | 0.108 | 0.147 | 3.985 | 5.224 |

distributions by the conventional IFM method which only enables the consideration of systematic data.

A Chi-Square goodness-of-fit and the Kolmogorov–Smirnov (K-S) test are used to test the assumption $H_0$, that the flood magnitudes follow the P-III distribution. The results in Table 2 show that hypothesis could not be rejected at the 5% significance level because $\chi^2 < \chi^2_{0.05}$, and the P-III distributions can also pass the K-S test because $D_{KS} < D_{0.05}$. The marginal distribution frequency curves of flood peak, 7-day and 15-day flood volumes by employing MHIFM and IFM methods are illustrated in Figure 4, and indicates that these theoretical distributions can fit the observed data reasonably well.

## Joint distribution of flood peak and volumes

Different families of copula functions have been proposed and described by Nelsen (2006), among which the Archimedean family is more desirable for hydrological analyses because it can be easily constructed and applied to whether the correlation among the hydrological variables is positive or negative (Kim *et al.* 2018; Lei *et al.* 2018). The Gumbel–Hougaard (G-H), Frank and Clayton copulas of the Archimedean family were applied to construct the bivariate joint distributions between $Q$ and $W_1$, and the symmetric and asymmetric copulas were selected as candidates for constructing the trivariate dependence of $Q$–$W_1$–$W_2$. The expressions of the copulas can be found in Xu *et al.* (2016). The parameters of bivariate and trivariate copulas were estimated by MHIFM and IFM, respectively. The root mean square errors (RMSE) and AIC values were calculated and listed in Table 3 to assess the performance of the parameter estimation approaches. The statistical results in Tables 2 and 3 both

indicate that the statistical values under MHIFM method are smaller than those of the IFM method. This implies that incorporating historical information in multivariate frequency analysis is possible and the developed MHIFM approach has good performance for estimating the parameters of marginal and joint distributions.

Table 3 also indicates that the bivariate G-H copula has the smallest RMSE and AIC values for constructing the joint distribution of $Q$–$W_1$, and the trivariate asymmetric G-H copula performs best for $Q$–$W_1$–$W_2$. Therefore, the bivariate and trivariate asymmetric G-H copula is used to model the dependence between the extreme maximum annual flood peak and volumes in this study. Figure 5 demonstrates the bivariate joint CDFs of flood peak and volumes by MHIFM and IFM methods respectively, which show the CDFs constructed by G-H copula is continuous.

## Bivariate design flood quantile estimation

After constructing the joint distribution by MHIFM and IFM methods, the EFC and MLR methods are used to estimate the bivariate design flood quantiles of flood peak and 7-day volume under JRP $T_{OR} = 200$, $T_{OR} = 100$, $T_{OR} = 50$, $T_{OR} = 20$ and $T_{OR} = 10$ year, respectively. For the purpose of comparison, the univariate method considering historical floods (Chinese design flood guidelines) was also used to estimate the design flood quantiles by assuming that the univariate return period was to the bivariate return period. Figure 6 demonstrates the bivariate quantile curves and the estimated design flood quantiles under the G-H copula.

To assess the rationality of the estimated design flood quantiles by the EFC and MLR methods, the 95% confidence interval (CI) bounds were derived by the approach
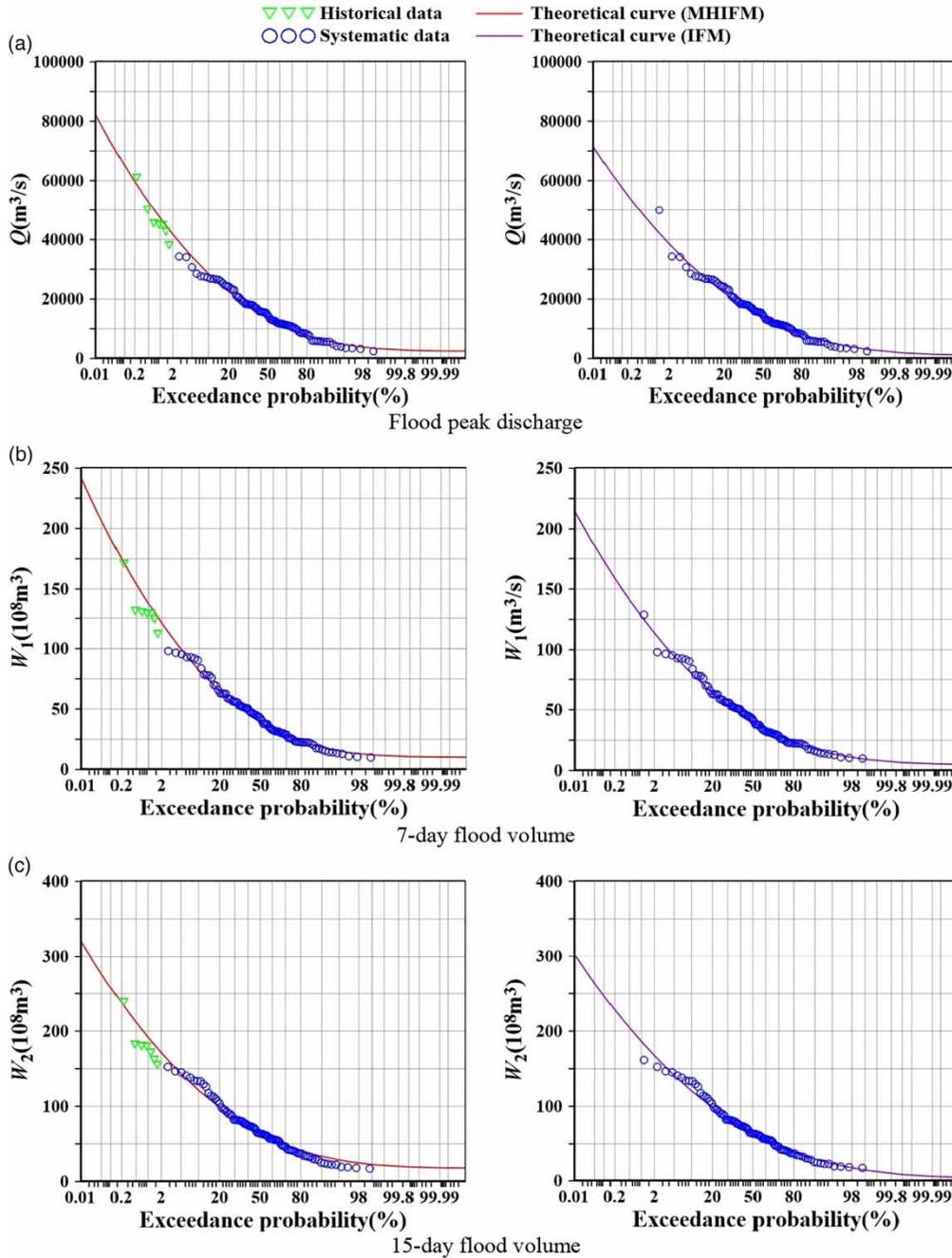
**Figure 4** | Probability curves of flood peak discharge and flood volumes by MHIFM and IFM methods.

proposed by Li *et al.* (2017). The selected flood events and coordinate pairs of the lower and upper bounds of the 95% CI are given in Figure 6, and it is revealed that the estimated design floods by the EFC and MLR methods are both within the 95% CI. This means that the EFC and MLR methods are reasonable for estimating the bivariate flood quantiles in the Danjiangkou reservoir.

The estimated quantiles of univariate and bivariate methods are listed in Table 4. The results in Table 4 indicate that the design flood quantile based on the univariate

**Table 3** | Estimated parameters of bivariate and trivariate copulas and results of statistical test

| Copula functions | MHIFM method | | | IFM method | | |
|---|---|---|---|---|---|---|
| | $\theta/[\theta_1, \theta_2]$ | RMSE | AIC | $\theta/[\theta_1, \theta_2]$ | RMSE | AIC |
| $Q\text{-}W_1$ | | | | | | |
| G-H | 3.64 | 0.040 | −590 | 3.40 | 0.042 | −543 |
| Clayton | 6.62 | 0.048 | −557 | 3.75 | 0.052 | −507 |
| Frank | 15.37 | 0.045 | −569 | 13.56 | 0.048 | −520 |
| $Q\text{-}W_1\text{-}W_2$ | | | | | | |
| Symmetric G-H | 3.42 | 0.055 | −276 | 3.37 | 0.055 | −251 |
| Symmetric Clayton | 5.05 | 0.061 | −267 | 4.95 | 0.063 | −241 |
| Symmetric Frank | 12.35 | 0.059 | −270 | 11.68 | 0.059 | −243 |
| Asymmetric G-H | [3.29,3.68] | 0.051 | −279 | [3.18,3.52] | 0.053 | −254 |
| Asymmetric Clayton | [4.65,5.79] | 0.057 | −270 | [4.68,5.35] | 0.055 | −247 |
| Asymmetric Frank | [11.37,17.25] | 0.053 | −274 | [11.18,16.85] | 0.058 | −244 |

distributions considering historical floods are less than the EFC and MLR methods. This especially illustrates that the bivariate methods associated with the OR JRP is safer than the univariate distribution method for the engineering practice, and this viewpoint has been confirmed by many authors (e.g. Shiau *et al.* 2006; Zhang *et al.* 2015b; Salvadori *et al.* 2016).

Comparing the results estimated by the MHIFM and IFM approaches in Table 4 and Figure 6, it is shown that for the IFM method the design flood quantiles selected by the EFC

and MLR methods are both underestimated compared with the MHIFM method. This is due to the fact that the conventional IFM method fails to consider the historical floods and leads to underestimation for flood frequency analysis. Table 4 also demonstrates that the estimated flood quantiles under the IFM method are less than the univariate estimations when the return period is over 20 years. However, for the 10-year return period, both the EFC and MLR designed events under the IFM method are larger than the univariate design floods. This may imply that, even though the historical
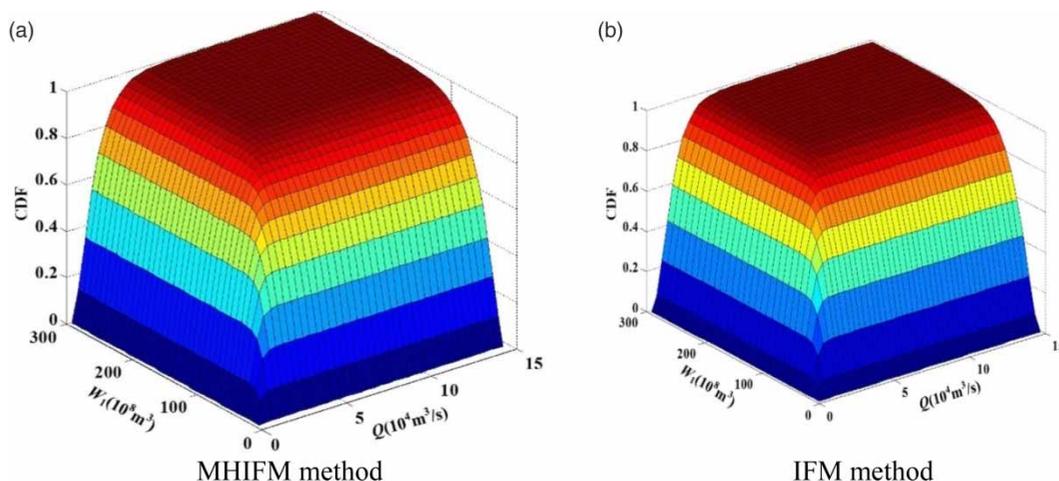


**Figure 5** | Bivariate joint distributions of peak discharge and 7-day flood volume by MHIFM and IFM methods.
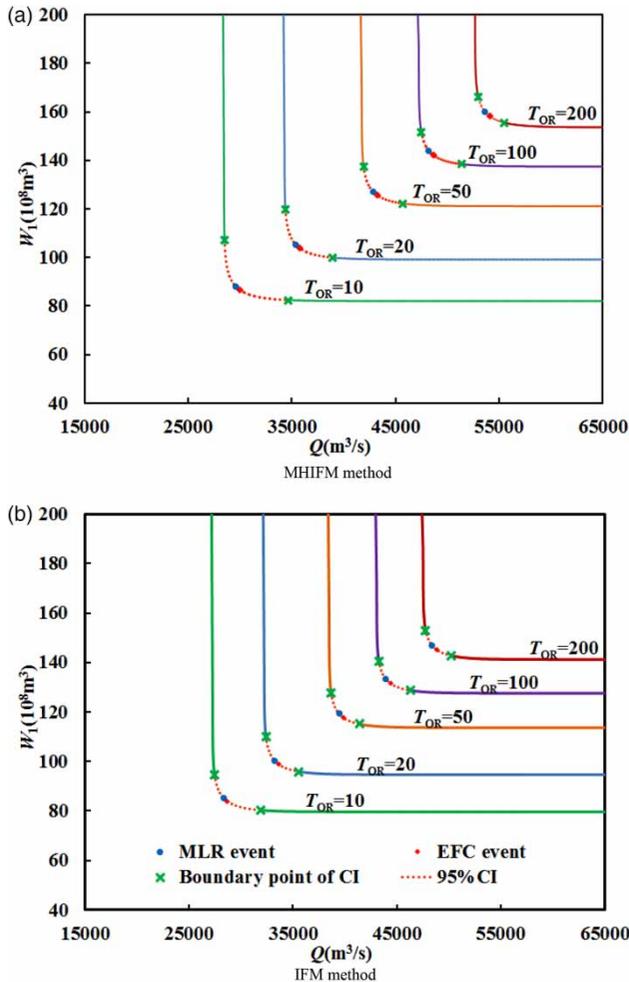
floods may have less impact on design floods under the low return period, considering historical floods is essential for hydrologic structure capacity design since the conventional IFM method could not satisfy the flood control standard, particularly for the high return period.
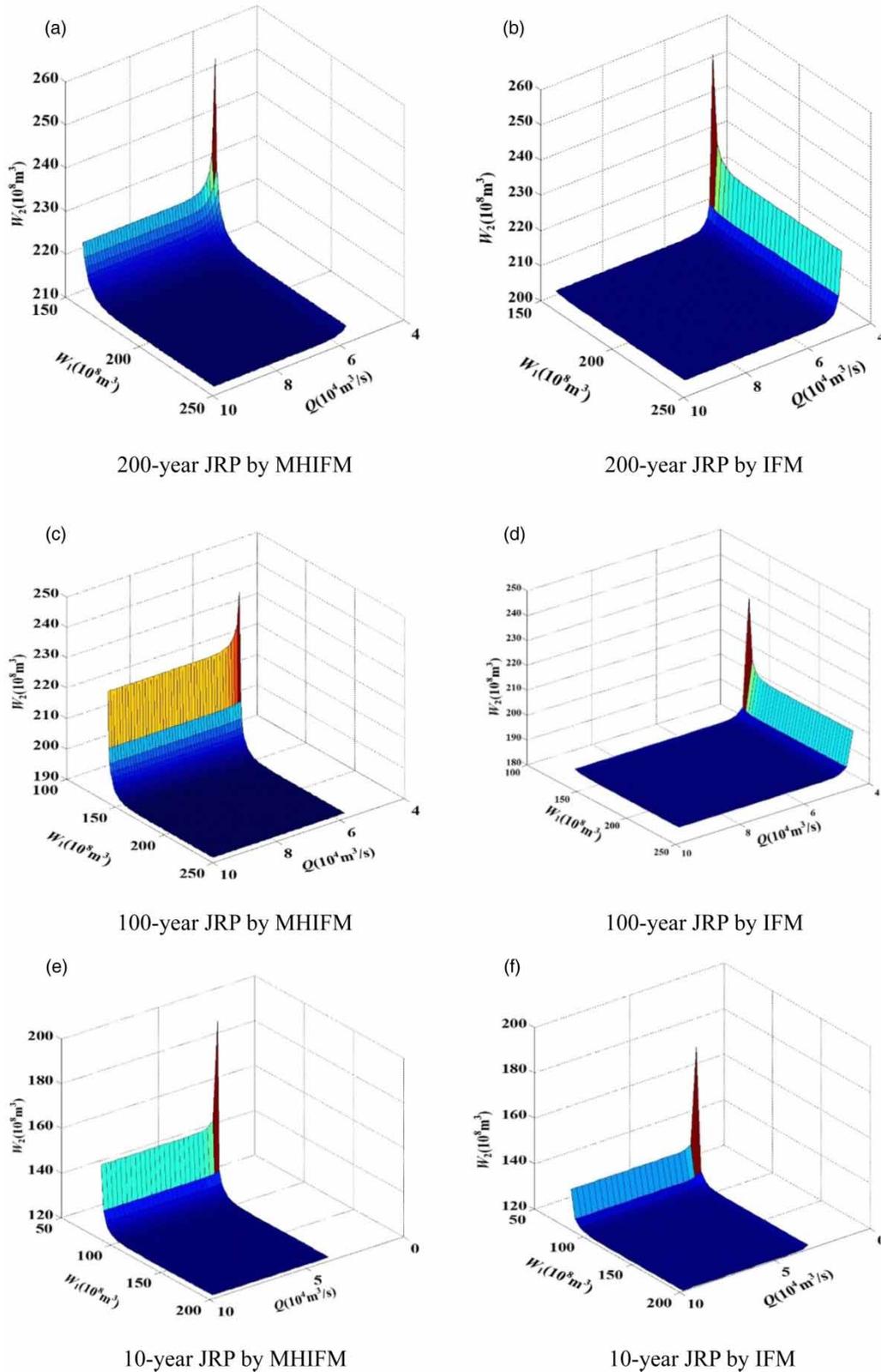
## Multivariate design flood quantile estimation

The Danjiangkou reservoir has large flood control storage during the flood season, and peak discharge, 7-day and 15-day flood volumes are necessary for deriving the design flood quantiles and the corresponding DFH. To meet this need, the trivariate surfaces of $Q$–$W_1$–$W_2$ under the JRP of 200, 100, 50, 20 and 10 years are estimated by MHIFM and IFM methods. Figure 7 demonstrates the trivariate surfaces based on the asymmetric G-H copula function under three JRPs for the sake of saving space. It is demonstrated that numerous schemes of design flood quantiles of flood peak and volumes can satisfy a given flood prevention standard, and the surfaces formed by the IFM method are lower than that by the MHIFM approach. The EFC and MLR methods are employed to select design flood quantiles on the surfaces under both MHIFM and IFM approaches, and the estimated values are listed in Table 5.

It is shown in Table 5 that the MHIFM method has larger design flood quantiles of flood peak, 7-day and 15-day volumes compared with the estimated values by the IFM

**Figure 6** │ Bivariate quantile of flood peak and 7-day volume by MHIFM and IFM methods.

**Table 4** │ Results of design flood quantiles estimated by univariate and bivariate methods

| Parameter estimation method | Quantile selection approach | Variable | Return period (year) | | | | |
|---|---|---|---|---|---|---|---|
| | | | 200 | 100 | 50 | 20 | 10 |
| Univariate | | $Q(\text{m}^3/\text{s})$ | 52,683 | 47,256 | 41,748 | 34,295 | 28,467 |
| | | $W_1(10^8\,\text{m}^3)$ | 153.6 | 137.5 | 121.2 | 99.2 | 82.1 |
| Bivariate MHIFM | EFC | $Q(\text{m}^3/\text{s})$ | 54,159 | 48,748 | 43,257 | 35,827 | 30,013 |
| | | $W_1(10^8\,\text{m}^3)$ | 158.0 | 141.9 | 125.6 | 103.7 | 86.6 |
| | MLR | $Q(\text{m}^3/\text{s})$ | 53,650 | 48,228 | 42,822 | 35,381 | 29,623 |
| | | $W_1(10^8\,\text{m}^3)$ | 160.0 | 143.9 | 127.2 | 105.3 | 88.0 |
| Bivariate IFM | EFC | $Q(\text{m}^3/\text{s})$ | 48,822 | 44,383 | 39,844 | 33,635 | 28,701 |
| | | $W_1(10^8\,\text{m}^3)$ | 145.2 | 131.6 | 117.8 | 98.9 | 84.0 |
| | MLR | $Q(\text{m}^3/\text{s})$ | 48,346 | 43,896 | 39,396 | 33,240 | 28,311 |
| | | $W_1(10^8\,\text{m}^3)$ | 147.2 | 133.6 | 119.5 | 100.4 | 85.4 |

**Figure 7** | Trivariate surfaces of flood peak and volumes by MHIFM and IFM methods.

**Table 5** | Results of trivariate design flood quantiles and corresponding highest routed reservoir water levels estimated by MHIFM and IFM methods

| Parameter estimation method | Quantile selection method | Variable | Joint return period (year) | | | | |
|---|---|---|---|---|---|---|---|
| | | | 200 | 100 | 50 | 20 | 10 |
| Trivariate MHIFM | EFC | $Q(\text{m}^3/\text{s})$ | 55,171 | 49,770 | 44,290 | 36,877 | 30,073 |
| | | $W_1(10^8\,\text{m}^3)$ | 161.0 | 144.9 | 128.7 | 106.8 | 89.7 |
| | | $W_2(10^8\,\text{m}^3)$ | 221.1 | 201.0 | 180.7 | 153.1 | 131.3 |
| | | $Z_{max}(\text{m})$ | 170.31 | 167.44 | 167.25 | 166.38 | 165.11 |
| | MLR | $Q(\text{m}^3/\text{s})$ | 54,462 | 49,351 | 43,784 | 36,358 | 29,678 |
| | | $W_1(10^8\,\text{m}^3)$ | 162.2 | 145.6 | 129.5 | 107.5 | 90.3 |
| | | $W_2(10^8\,\text{m}^3)$ | 222.4 | 202.3 | 181.4 | 153.9 | 131.8 |
| | | $Z_{max}(\text{m})$ | 170.35 | 167.54 | 167.32 | 166.47 | 165.18 |
| Trivariate IFM | EFC | $Q(\text{m}^3/\text{s})$ | 49,593 | 45,168 | 40,645 | 34,460 | 28,547 |
| | | $W_1(10^8\,\text{m}^3)$ | 147.6 | 134.0 | 120.2 | 101.4 | 86.5 |
| | | $W_2(10^8\,\text{m}^3)$ | 213.1 | 194.9 | 176.2 | 150.5 | 130.1 |
| | | $Z_{max}(\text{m})$ | 168.29 | 166.95 | 166.68 | 166.19 | 164.84 |
| | MLR | $Q(\text{m}^3/\text{s})$ | 49,075 | 44,768 | 40,056 | 33,983 | 28,067 |
| | | $W_1(10^8\,\text{m}^3)$ | 148.3 | 134.9 | 121.4 | 102.1 | 86.9 |
| | | $W_2(10^8\,\text{m}^3)$ | 214.2 | 195.7 | 177.0 | 151.2 | 130.6 |
| | | $Z_{max}(\text{m})$ | 168.31 | 167.02 | 166.72 | 166.24 | 164.88 |

method. This is true under two quantile selection methods, which means that the parameter estimation approach has far heavier impacts on design flood quantiles than different quantile selection methods. Comparing the results of the EFC and MLR methods, it is found that the MLR method has smaller flood peak and larger volumes compared with the EFC method. Despite the design flood quantiles derived from the EFC and MLR methods, both can meet the flood prevention standards and these different joint design floods may lead to discrepancy of the flood hydrograph and highest water level which occurs in reservoir routing, which supports the necessity for comparing the hydrological loads on the structure of different joint design floods under a given JRP.

## DFH and corresponding reservoir routing

The estimated trivariate design flood quantiles of flood peak, 7-day and 15-day volumes under different methods are applied to derive the DFH, and the resulting highest reservoir water level is selected as an index to evaluate the effects of different hydrological loads on structure. The DFH for a dam is the flood of suitable probability and magnitudes adopted to ensure safety of the dam in accordance with appropriate design standards. The DFH could be derived by amplifying the typical flood hydrographs (TFH),

which has been widely used by practitioners (e.g. Yue & Rasmussen 2002; Tan *et al.* 2017). The flood hydrograph with high flood peak or large flood volume is usually selected as a TFH, and thus the flood hydrograph of 1964 is selected as a TFH following a previous study (Yin *et al.* 2017). The peak and volume amplitude method based on the selected TFH is used to derive the DFH (MWR 2006; Xiao *et al.* 2009), and the results are shown in Figure 8. The DFHs are routed through the reservoir with initial water level (flood control limiting water level, 160.0 m) according to the reservoir operation rule in Danjiangkou reservoir, and the corresponding highest reservoir water levels ($Z_{max}$) are calculated and listed in Table 5.

The comparison results listed in Table 5 show that $Z_{max}$ obtained by the trivariate MHIFM method is higher than that by the conventional IFM method. Comparing the results of 200-year JRP, it is found that the differences under two selection methods among the MHIFM and IFM methods are both over 2.0 m. For 20- and 10-year return periods, the $Z_{max}$ of MHIFM is higher, 0.19–0.30 m, than that of the IFM method, which implies that multivariate flood frequency analysis considering historical floods has a significant impact on flood control safety, especially for the high return period. The conventional IFM method, which fails to incorporate the historical floods, may lead to heavier flood hazards for hydraulic structure.
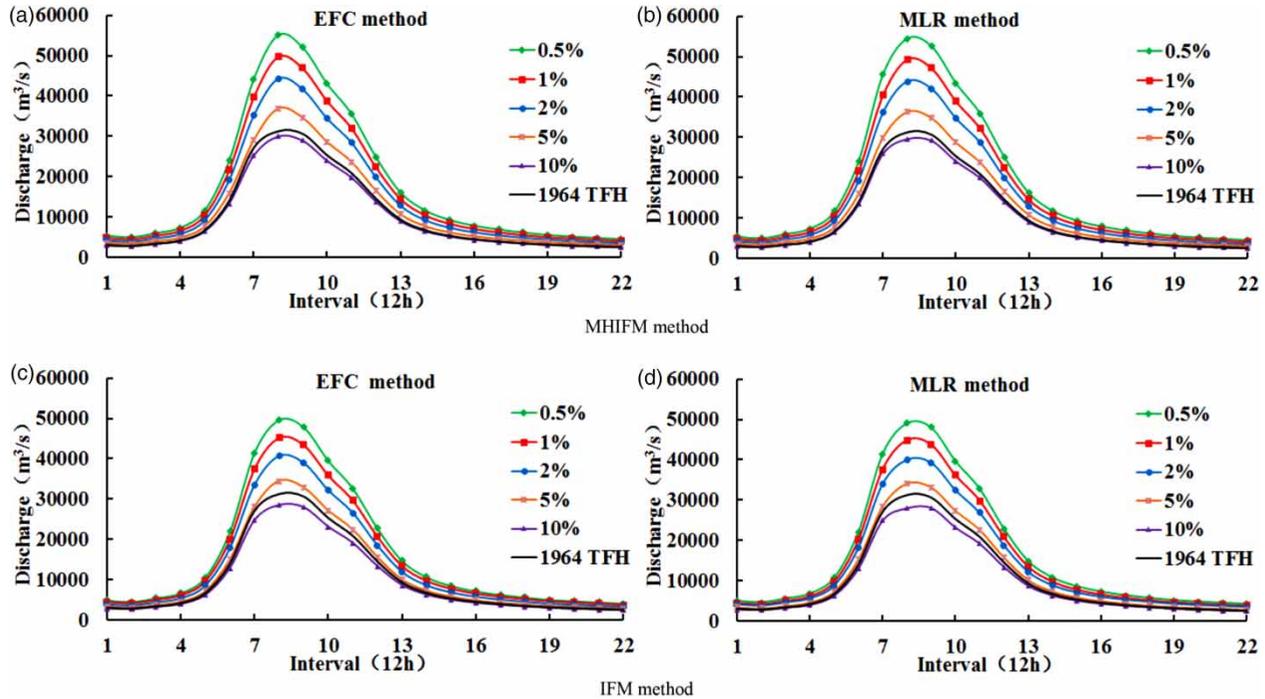
**Figure 8** | Derived DFH under different trivariate estimation methods in the Danjiangkou reservoir.

Table 5 also demonstrates the comparative results of $Z_{max}$ estimated by the EFC and MLR methods, and it indicates that these different quantile selection methods have slight impacts on the corresponding $Z_{max}$. Comparing the $Z_{max}$ by the EFC and MLR methods, it is found that the $Z_{max}$ of the EFC method is lower than that of the MLR method, which is true under the same JRP and parameter estimation approach. For example, for the MHIFM approach, the $Z_{max}$ of the EFC method (170.30 m) is lower than that of the MLR method (170.35 m) under 200-year JRP. This may be due to the fact that the Danjiangkou reservoir has a large amount of flood control storage with annual regulation ability, and therefore the design flood volume is relatively more important than peak discharge for flood prevention safety. As a consequence, the trivariate MLR method with slightly larger 7- and 15-day flood volume is safer for reservoir design compared with the EFC method.

## CONCLUSIONS

Flood quantile estimation is essential and important for deriving the DFH and determining reservoir operation

rules. This study developed an MHIFM approach to model the joint distribution of flood peak and volumes with incorporation of historical information. The MLR and EFC method were employed for selecting unique design floods under a given JRP. The DFHs in Danjiangkou reservoir were derived based on trivariate design floods, and through reservoir routing the corresponding highest reservoir water levels ($Z_{max}$) were obtained. The main conclusions of this study are as follows.

(1) The investigated data series consist of systematic record and six historical flood events, and the fitting results indicate that the developed MHIFM approach that incorporates historical information into multivariate distribution has a good performance for estimating the parameters of marginal and joint distributions.

(2) The conventional IFM method that fails to consider historical floods produces underestimated design flood quantiles compared with the MHIFM method, and is also inferior to the univariate method with incorporation of historical information, implying that the conventional IFM method is unable to satisfy the

flood prevention standard. The reservoir routing results show that $Z_{max}$ obtained by the trivariate MHIFM approach is higher than that obtained by the conventional IFM approach and the trivariate MLR method yields safer results for reservoir design than the EFC method.

(3) The developed MHIFM approach associated with the MLR method could provide a more comprehensive and effective way for multivariate quantile estimation to incorporate historical information, particularly when the recorded data series is relatively short, which would provide rich information in engineering practice as the references for flood risk assessment, reservoir operation and management.

## ACKNOWLEDGEMENTS

## REFERENCES

Bagherlou, H. & Ghaffari, A. 2018 A routing protocol for vehicular ad hoc networks using simulated annealing algorithm and neural networks. *Supercomput.* **74** (6), 2528–2552.

Baharith, L. A. 2017 Parametric and semiparametric estimations of bivariate truncated type I generalized logistic models driven from copulas. *Int. J. Stat. Probab.* **7**, 72–85.

Baykasoğlu, A. & Gindy, N. N. Z. 2001 A simulated annealing algorithm for dynamic layout problem. *Comput. Oper. Res.* **28**, 1403–1426.

Bayliss, A. C. & Reed, D. W. 2001 *The Use of Historical Data in Flood Frequency Estimation*. Centre for Ecology and Hydrology, UK, Wallingford, UK.

Bhunya, P. K., Berndtsson, R., Jain, S. K. & Kumar, R. 2013 Flood analysis using negative binomial and generalized Pareto models in partial duration series (PDS). *J. Hydrol.* **497**, 121–132.

Brunner, M. I., Sikorska, A. E. & Seibert, J. 2018 Bivariate analysis of floods in climate impact assessments. *Sci. Total Environ.* **616**, 1392–1403.

Chebana, F. & Ouarda, T. B. M. J. 2009 Index flood-based multivariate regional frequency analysis. *Water Resour. Res.* **45**, W10435.

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

Chen, L. & Singh, V. P. 2017a Entropy based derivation of generalized distributions for hydrometeorological frequency analysis. *J. Hydrol.* **557**, 699–712.

Chen, L. & Singh, V. P. 2017b Generalized beta distribution of the second kind for flood frequency analysis. *Entropy* **19**, 1–17.

Chen, L. & Singh, V. P. 2018 Entropy-based derivation of generalized distributions for hydrometeorological frequency analysis. *J. Hydrol.* **557**, 699–712.

Chen, Y. D., Zhang, Q., Xiao, M., Singh, V. P. & Zhang, S. 2016 Probabilistic forecasting of seasonal droughts in the pearl river basin. *China. Stoch. Env. Res. Risk Assess.* **30**, 2031–2040.

Condie, R. 1986 Flood samples from a three-parameter lognormal population with historical information: the asymtotic standard error of estimate of the three year flood. *J. Hydrol.* **85**, 139–150.

De Michele, C. & Salvadori, G. 2003 A generalized Pareto intensity-duration model of storm rainfall exploiting 2-copulas. *J. Geophys. Res.* **108** (D2), 4067.

De Michele, C., Salvadori, G., Canossi, M., Petaccia, A. & Rosso, R. 2005 Bivariate statistical approach to check adequacy of dam spillway. *J. Hydrol. Eng.* **10**, 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.

Favre, A. C., El Adlouni, S., Perreault, L., Thiémonge, N. & Bobée, B. 2004 Multivariate hydrological frequency analysis using copulas. *Water Resour. Res.* **40**, W01101.

Goel, N. K., Kurothe, R. S., Mathur, B. S. & Vogel, R. M. 2000 A derived flood frequency distribution for correlated rainfall intensity and duration. *J. Hydrol.* **228**, 56–67.

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

Guo, S. L. & Cunnane, C. 1991 Evaluation of the usefulness of historical and paleological floods in quantile estimation. *J. Hydrol.* **129**, 245–262.

Guo, S. L., Muhammad, R., Liu, Z. J., Xiong, F. & Yin, J. B. 2018 Design flood estimation methods for cascade reservoirs based on copulas. *Water* **10**, 560.

Hald, A. 1949 Maximum likelihood estimation of the parameters of a normal distribution which is truncated at a known point. *Skand. Aktuarietidskrift.* **32**, 119–134.

Hosking, J. R. M. 1995 The use of L-moments in the analysis of censored data. In: *Recent Advances in Life-Testing and Reliability* (N. Balakrishnan, ed.). CRC Press, Boca Raton, FL, pp. 545–564.

Joe, H. 1997 *Multivariate Models and Dependence Concepts*. Chapman & Hall, London.

Joe, H. & Xu, J. J. 1996 *The Estimation Method of Inference Functions for Margins for Multivariate Models*. Technical Report no. 166, Department of Statistics, University of British Columbia.

Kang, S. L. & Zong, W. G. 2005 A new meta-heuristic algorithm for continuous engineering optimization: harmony search theory and practice. *Comput. Method. Appl. M.* **194**, 3902–3933.

Kim, J. Y., So, B. J., Kwon, H. H., Kim, T. W. & Lee, J. H. 2018 Estimation of return period and its uncertainty for the recent 2013–2015 drought in the Han River watershed in South Korea. *Hydrol. Res.* **49** (5), 1313–1329.

Kirkpatrick, S., Gelatt, C. D. & Vecchi, M. P. 1983 Optimization by simulated annealing. *Science* **220**, 671–680.

Lei, X. H., Tan, Q. F., Wang, X., Wang, H., Wen, X., Wang, C. & Zhang, J. W. 2018 Stochastic optimal operation of reservoirs based on copula functions. *J. Hydrol.* **557**, 265–275.

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

Li, T., Guo, S., Liu, Z., Xiong, L. & Yin, J. 2017 Bivariate design flood quantile selection using copulas. *Hydrol. Res.* **48**, 997–1013.

Liu, H. & Ni, Q. 2008 Incomplete Jacobian Newton method for nonlinear equations. *Comput. Math. Appl.* **56**, 218–227.

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 *Regulations for Calculating Design Flood of Water Resources and Hydropower Projects*. Water Resources & Hydropower Press, Beijing (in Chinese).

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

NERC (Natural Environment Research Council) 1975 *Flood Studies Report*, vol. 1. NERC, London.

Reed, D. W. 2002 Reinforcing flood–risk estimation. *Phil. Trans. R. Soc. Lond. A* **360** (1796), 1373–1387.

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. Sc.* **17**, 3023–3038.

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

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

Salvadori, G., Durante, F., De Michele, C., Bernardi, M. & Petrella, L. 2016 A multivariate copula-based framework for dealing with hazard scenarios and failure probabilities. *Water Resour. Res.* **52**, 3701–3721.

Shen, M., Chen, J., Zhuan, M., Chen, H., Xu, C.-Y. & Xiong, L. 2018 Estimating uncertainty and its temporal variation related to global climate models in quantifying climate change impacts on hydrology. *J. Hydrol.* **556**, 10–24.

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

Shiau, J. T., Wang, H. & Tsai, C. 2006 Bivariate frequency analysis of floods using copulas. *J. Am. Water Resour. Assoc.* **42**, 1549–1564.

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

Soltani, M., Laux, P., Mauder, M. & Kunstmann, H. 2018 Spatiotemporal variability and empirical copula-based dependence structure of modeled and observed coupled water and energy fluxes. *Hydrol. Res.* **49** (5), nh2018163.

Stedinger, J. R. & Cohn, T. A. 1986 The value of historical and paleoflood information in flood frequency analysis. *Water Resour. Res.* **22**, 785–793.

Tan, Q. F., Wang, X., Liu, P., Lei, X. H., Cai, S. Y., Wang, H. & Ji, Y. 2017 The dynamic control bound of flood limited water level considering capacity compensation regulation and flood spatial pattern uncertainty. *Water Resour. Manage.* **31**, 143–158.

Volpi, E. & Fiori, A. 2012 Design event selection in bivariate hydrological frequency analysis. *Hydrol. Sci. J.* **57**, 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.

Wei, T. & Song, S. 2017 Copula-based composite likelihood approach for frequency analysis of short annual precipitation records. *Hydrol. Res.* **49** (5), 1498–1512.

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

Xu, C. J., Yin, J. B., Guo, S. L., Liu, Z. J. & Hong, X. J. 2016 Deriving design flood hydrograph based on conditional distribution: a case study of Danjiangkou reservoir in Hanjiang Basin. *Math. Probl. Eng.* **2016**, 4319646.

Yang, G., Guo, S., Liu, P., Li, L. & Xu, C. 2017 Multiobjective reservoir operating rules based on cascade reservoir input variable selection method. *Water Resour. Res.* **53**, 3446–3463.

Yin, J. B., Guo, S. L., Liu, Z. J., Chen, K. B., Chang, F.-J. & Xiong, F. 2017 Bivariate seasonal design flood estimation based on copulas. *J. Hydrol. Eng.* **22** (12), 0001594.

Yin, J. B., Guo, S. L., Liu, Z. J., Yang, G., Zhong, Y. X. & Liu, D. D. 2018a Uncertainty analysis of bivariate design flood estimation and its impacts on reservoir routing. *Water Resour. Manage.* **32** (5), 1795–1809.

Yin, J. B., Guo, S. L., He, S. K., Guo, J. L., Hong, X. J. & Liu, Z. J. 2018b A copula-based analysis of projected climate changes to bivariate flood quantiles. *J. Hydrol.* **566**, 23–42.

Yue, S. & Rasmussen, P. 2002 Bivariate frequency analysis: discussion of some useful concepts in hydrological application. *Hydrol. Process.* **16**, 2881–2898.

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. *Glob. Planet Change* **129**, 1–9.

Zhou, Y., Guo, S., Hong, X. & Chang, F. J. 2017 Systematic impact assessment on inter-basin water transfer projects of the Hanjiang River Basin in China. *J. Hydrol.* **553**, 584–595.