Research Article

Deriving Design Flood Hydrograph Based on Conditional Distribution: A Case Study of Danjiangkou Reservoir in Hanjiang Basin

Changjiang Xu,^{1,2} Jiabo Yin,¹ Shenglian Guo,¹ Zhangjun Liu,¹ and Xingjun Hong¹

¹State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan 430072, China

²Bureau of Hydrology, Changjiang Water Resources Commission, Wuhan 430010, China

Correspondence should be addressed to Shenglian Guo; slguo@whu.edu.cn

Received 18 October 2015; Revised 19 January 2016; Accepted 14 February 2016

Academic Editor: Renata Archetti

Copyright © 2016 Changjiang Xu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Design flood hydrograph (DFH) for a dam is the flood of suitable probability and magnitude adopted to ensure safety of the dam in accordance with appropriate design standards. Estimated quantiles of peak discharge and flood volumes are necessary for deriving the DFH, which are mutually correlated and need to be described by multivariate analysis methods. The joint probability distributions of peak discharge and flood volumes were established using copula functions. Then the general formulae of conditional most likely composition (CMLC) and conditional expectation composition (CEC) methods that consider the inherent relationship between flood peak and volumes were derived for estimating DFH. The Danjiangkou reservoir in Hanjiang basin was selected as a case study. The design values of flood volumes and 90% confidence intervals with different peak discharges were estimated by the proposed methods. The performance of CMLC and CEC methods was also compared with conventional flood frequency analysis, and the results show that CMLC method performs best for both bivariate and trivariate distributions which has the smallest relative error and root mean square error. The proposed CMLC method has strong statistical basis with unique design flood composition scheme and provides an alternative way for deriving DFH.

1. Introduction

A reservoir dam can be designed with the design flood hydrograph (DFH), which is a hydrograph adopted according to design standards to ensure the safety of a structure. DFH generally refers to the computed flood hydrograph at the dam site that may be estimated from precipitation or flow records, depending on engineering judgment after consideration of all the pertinent data, including the extent and reliability of the flow data and rainfall data [1]. The DFH is characterized by the joint behavior of several correlated random variables such as flood peak, volume, and duration. Knowledge of the flood peak or volumes alone is not sufficient to design the dam spillway; the entire flood hydrograph must be utilized. The quantiles of the flood peak and volumes corresponding to a particular design return period are necessary for deriving DFH. However, the conventional flood frequency analysis

methods for deriving DFH recommended by many countries are based on the univariate distribution, mainly concentrated on the analysis of annual peak discharge or flood volume series without analyzing the inherent relationship between flood peak and volumes [2]. For example, the Chinese univariate flood frequency analysis method for deriving DFH implicitly assumed that flood peak and flood volumes are independent and amplified with the same frequency, which has shortcomings in practice [3, 4]. Over- or underestimation of risk would be produced by the univariate frequency analysis method because of the unclear relationship between the correlated hydrological variables [5, 6].

To solve this problem, many authors have used statistical methods to derive the DFH by examining the dependence of flood peak, flood volume, and its shape. Xiao et al. [2] proposed a multicharacteristic synthesis index method based

on the MSI quantile and derived the DFH. Pramanik et al. [7] used the probability density functions to fit the shape of hydrographs and subsequently to develop design flood hydrographs for various return periods. Mediero et al. [8] estimated flood volume for a given peak discharge by fitting a regional log-log regression equation over the observed pairs, and a Monte Carlo experiment was conducted to generate an ensemble of DHFs that maintain the statistical properties of marginal distributions of the peaks, volumes, and durations. Serinaldi and Grimaldi [9] studied the advantages and shortcomings of using simple distribution functions with finite support (namely, beta and generalized standard two-sided power distributions) to represent and synthesize direct runoff hydrographs. In their study, the dependence among peak discharge, volume, and duration was explored on a few flood events selected by a recursive digital filter algorithm and an over threshold approach. Domínguez and Arganis [10] introduced and used the IINGEN method to estimate the design floods for Malpaso Dam, and the method makes it possible for considering flood peak, volume, and shape simultaneously. Óscar et al. [11] proposed a method to select the time base of the DFH using spectral analysis method and defined the number of days necessary for taking into account the complete dynamics of the hydrological floods.

Recently, the copula functions have been increasingly applied in exploring the inherent correlation between flood peak and volumes for deriving DFH. Copula functions allow for more flexibility in the marginal distributions and the dependence between peaks and volumes. Chowdhary et al. [12] discussed the identification of suitable copulas for bivariate frequency analysis of flood peak and flood volume data, and Gumbel-Hougaard copula of Archimedean family was found to be the most suitable dependence model for flood peak and volume. Bačová and Halmová [13] explored the analysis and statistical evaluation of the joint probability of the occurrence of peak discharge and volume; they also calculated joint return periods and conditional return periods for the hydrological pair. There have been numerous studies on the degree of the dependence between peaks and volumes that is needed to estimate the multivariate quantiles (e.g., [5, 6]) and the choice of the copula function (e.g., [12, 14]). Gaál et al. [15] studied the peak-volume relationships for the region of Austria, which featured a diverse spectrum of hydrological flood processes and discussed whether flood peak-volume relationships can be typified by comparative hydrology. Sraj et al. [16] compared different bivariate copulas from three families and analyzed the bivariate flood frequency analysis in the Litija station in the Sava River. The previous researches [12–16] derived the multivariate quantile curve under one joint return period, and infinite combinations on the isoline or isosurface could be selected to evaluate the effects of different hydrological loads on the structure [17]. However, one has to select a quantiles combination out of events which all share the same return period for deriving DFH [14]. How to solve this problem has attracted the attention of many researchers (e.g., [6, 14]). Li et al. [6] proposed a copula identical frequency method and used it to estimate bivariate joint flood quantiles in Three Gorges reservoir. Gräler et al.

[14] gave a critical and practical review focusing on synthetic design hydrograph estimation. Their approaches were based on regression analysis, bivariate conditional distributions, bivariate joint distributions, and Kendall distribution function, highlighting theoretic and practical issues of multivariate frequency analysis. Chebana and Ouarda [18] proposed the decomposition of the level curve into a naïve part (tail) and the proper part (central); they assumed that the naïve part was composed of two segments starting at the end of each extremity of the proper part. Salvadori et al. [19] introduced two basic design realizations, that is, componentwise excess design realization and most likely design realization. Volpi and Fiori [20] 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.

In the multivariate domain for deriving DFH, the return period must be defined in terms of the acceptable risk for the reservoir. Different return periods estimated by copula function have been developed for the case of a multivariate flood frequency analysis. Eight types of possible joint events were presented by Yue and Rasmussen [21] and Shiau [22], in which the OR, AND, conditional cases are of greatest interest in hydrological applications. Recently, another definition of the multivariate return period was given by Salvadori and De Michele [23], called Kendall return period. This return period was introduced to identify in a multivariate context a univariate critical threshold. Until now, only a very limited number of studies actually applied this kind of return period (e.g., [24]). Volpi and Fiori [25] proposed a general, structure-based return period for the design and risk assessment of hydrological structures in a bivariate environment. Their work draws attention of practitioners to the importance of considering the structure in hydraulic design and/or risk assessment problems in a multivariate environment, thus advising against the uncritical use of design event-based approaches which neglect the interplay between the structure and the hydrological loads acting on it. More information about the return periods could be found in Shiau [22] and Volpi and Fiori [25]. Until now, a clear assessment of the use of return periods is still a matter of debates [26, 27].

In this paper, we study the peak-volume relationships by copula functions for Danjiangkou reservoir in the Hanjiang basin extensively. Specifically, we investigate the inherent relationship on the basis of their probability of occurrence. Different copula functions from Archimedean family were applied and compared. Then the conditional most likely composition (CMLC) and conditional expectation composition (CEC) methods were proposed, inspired by the idea of Gräler et al. [14] and Salvadori et al. [19]. The main aim of this study is to provide a simple tool for the identification of the relationship between flood peak and volumes and derive DFH.

To achieve these objectives, the paper is organized as follows: introduction of the Chinese method of deriving DFH is illustrated in Section 2. The methodologies used in this study are presented in Section 3. In Section 4, results and discussions of case study are shown, followed by the Conclusions.

2. Chinese Method of Deriving DFH

Pearson type III (P3) distribution is recommended by Ministry of Water Resources of China for deriving DFH [2, 3]. For a given sample series $\{x : \gamma < x < +\infty\}$, the probability density function (PDF) of P3 distribution is expressed as

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} (x - \gamma)^{\alpha-1} e^{-\beta(x-\gamma)}, \quad (1)$$

where α , β , and γ are the shape, scale, and location parameters, respectively, and $\Gamma()$ is the gamma function.

The L-moment (LM) method is preferred for parameter estimation because of its robust properties in the presence of usually small or large values (outliers) [28, 29]. Univariate flood frequency analysis model (P3/LM), that is, P3 distribution and L-moment method, is used to estimate flood quantile for given return periods.

The peak and volume amplitude (PVA) method implicitly assumed that flood peak and flood volumes are amplified with the same frequency (return period). For example, the values of peak discharge and flood volumes are individually estimated by each marginal distribution under given return periods firstly. Then the DFH is constructed by multiplying each discharge ordinate of the typical flood hydrograph (TFH) by an amplifier.

The amplifier is either the ratio of the design peak with a given return period to the TFH peak or the ratio of the design volume with a given return period to the TFH volume. Suppose that the characteristics of annual maximum flood hydrograph consist of flood peak $Q_{\max}$, 1-day maximum flood volume W_{1d} , 3-day maximum flood volume W_{3d} , 7-day maximum flood volume W_{7d} , and 15-day maximum flood volume W_{15d} , the quantiles with design return period T are estimated by P3/LM and denoted by $Q_{\max T}$, W_{1T} , W_{3T} , W_{7T} , and W_{15T} , respectively. The corresponding characteristics of selected TFH are denoted by Q_{TFH} , $W_{1\text{TFH}}$, $W_{3\text{TFH}}$, $W_{7\text{TFH}}$, and $W_{15\text{TFH}}$, respectively. These amplifiers are calculated as follows.

Amplifier K_P for flood peak discharge is as follows:

$$K_P = \frac{Q_{\max T}}{Q_{\text{TFH}}}. \quad (2)$$

Amplifier K_1 for the 1-day maximum flood volume (excluding flood peak) is as follows:

$$K_1 = \frac{W_{1T}}{W_{1\text{TFH}}}. \quad (3)$$

Amplifier K_{3-1} for the 3-day maximum flood volume except for the 1-day maximum flood volume is as follows:

$$K_{3-1} = \frac{W_{3T} - W_{1T}}{W_{3\text{TFH}} - W_{1\text{TFH}}}. \quad (4)$$

Amplifier K_{7-3} for the 7-day maximum flood volume except for the 3-day maximum flood volume is as follows:

$$K_{7-3} = \frac{W_{7T} - W_{3T}}{W_{7\text{TFH}} - W_{3\text{TFH}}}. \quad (5)$$

Amplifier K_{15-7} for the 15-day maximum flood volume except for the 7-day maximum flood volume is as follows:

$$K_{15-7} = \frac{W_{15T} - W_{7T}}{W_{15\text{TFH}} - W_{7\text{TFH}}}. \quad (6)$$

The dividing portion between two contiguous segments of the constructed DFH is not continuous and should be modified by hand smooth treatment in engineering practice [2, 3].

3. Derivation of DFH Based on Copula Function

3.1. Joint Distribution Based on Copula Function. Copula function is a function that describes and models the dependence structure between random variables, independent of the marginal distributions involved (e.g., [26, 28, 30]). Let $F_{X_i}(x_i)$ ($i = 1, 2, \dots, n$) be the cumulative distribution function (CDF) of X_i . The objective is to determine the multivariate distribution, denoted as $H_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n)$ or simply H . Thus, the multivariate probability distribution H is expressed in terms of its marginal and the associated dependence function as Sklar's theorem:

$$C(F_1(x_1), F_2(x_2), \dots, F_n(x_n)) = H_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n), \quad (7)$$

where C , called the copula function, is uniquely determined whenever $F_{X_i}(x_i)$ are continuous and captures the essential features of the dependence among the random variables.

3.2. Conditional Most Likely Composition (CMLC) Method. The P3 distribution is selected as marginal distributions of peak discharge (Q) and flood volumes (W_i ($i = 1, 2, \dots, n$)), which are denoted as $F_Q(q)$, $F_{W_i}(w_i)$. Then the corresponding density functions are denoted as $f_Q(q)$ and $f_{W_i}(w_i)$, in which n represented the number of flood volumes needed for estimating DFH. For a given design standard (return period), it is necessary to find an appropriate combination of flood quantiles since different combinations would result in different DFH [18, 20]. How to select appropriate combinations (i.e., flood quantiles composition method) is extremely important in practice. The conventional univariate hydrological frequency analysis method for deriving DFH implicitly assumed that flood peak and volumes are independent and amplified with the same frequency, which could not explore the inherent correlation of hydrological variables [2]. Under given peak discharge, the combinations differ in terms of their probability of occurrence.

Based on copula function, the joint distribution function $F(q, w_1, w_2, \dots, w_n)$ can be expressed as [30]

$$F(q, w_1, w_2, \dots, w_n) = C(u, v_1, v_2, \dots, v_n), \quad (8)$$

where $C(u, v_1, v_2, \dots, v_n)$ is copula function; $u = F_Q(q)$, $v_i = F_{W_i}(w_i)$, $i = 1, 2, \dots, n$, are marginal distribution of peak discharge and flood volumes, respectively.

The corresponding conditional probability distribution is

$$\begin{aligned}
 F_{W_n|Q,W_1,W_2,\dots,W_{n-1}}(w_n) &= P(W_n \leq w_n | Q = q, W_1 \\
 &= w_1, W_2 = w_2, \dots, W_{n-1} = w_{n-1}) \\
 &= \frac{\partial^n F(q, w_1, w_2, \dots, w_n) / \partial q \partial w_1 \partial w_2 \cdots \partial w_{n-1}}{\partial^n F(q, w_1, w_2, \dots, w_{n-1}) / \partial q \partial w_1 \partial w_2 \cdots \partial w_{n-1}} \quad (9) \\
 &= \frac{\partial^n C(u, v_1, v_2, \dots, v_{n-1}, v_n) / \partial u \partial v_1 \partial v_2, \dots, \partial v_{n-1}}{c(u, v_1, v_2, \dots, v_{n-1})}.
 \end{aligned}$$

The density function of conditional probability distribution is

$$\begin{aligned}
 f_{W_n|Q,W_1,W_2,\dots,W_{n-1}}(w_n) &= \frac{dF_{W_n|Q,W_1,W_2,\dots,W_{n-1}}(w_n)}{dw_n} \\
 &= \frac{c(u, v_1, v_2, \dots, v_n)}{c(u, v_1, v_2, \dots, v_{n-1})} \cdot f_{W_n}(w_n), \quad (10)
 \end{aligned}$$

where $c(u, v_1, v_2, \dots, v_{n-1}, v_n) = \partial^{n+1} C(u, v_1, v_2, \dots, v_{n-1}, v_n) / \partial u \partial v_1 \partial v_2, \dots, \partial v_{n-1} \partial v_n$ is density function of copula.

Maximize $f_{W_n|Q,W_1,W_2,\dots,W_{n-1}}(w_n)$; the corresponding values of $(q, w_1, w_2, \dots, w_{n-1}, w_M)$ are conditional most likely composition (CMLC).

$f_{W_n|Q,W_1,W_2,\dots,W_{n-1}}(w_n)$ is a single function of w_n ; its derivation to w_n is given by

$$\begin{aligned}
 \frac{df_{W_n|Q,W_1,W_2,\dots,W_{n-1}}(w_n)}{dw_n} \\
 = \frac{1}{c(n-1)} [c_1 \cdot f_{W_n}^2(w_n) + c(n) \cdot f_{W_n}'(w_n)] \quad (11)
 \end{aligned}$$

in which $c_1 = \partial c / \partial v_n$, $c(n-1) = c(u, v_1, v_2, \dots, v_{n-1})$, $c(n) = c(u, v_1, v_2, \dots, v_n)$, and $f_{W_n}'(w_n)$ is the derivative function of $f_{W_n}(w_n)$.

To maximize $f_{W_n|Q,W_1,W_2,\dots,W_{n-1}}(w_n)$, the following equation should be satisfied:

$$\frac{1}{c(n-1)} [c_1 \cdot f_{W_n}^2(w_n) + c(n) \cdot f_{W_n}'(w_n)] = 0. \quad (12)$$

If W_n is following P3 distribution as (1), then the following equation should be satisfied:

$$f_{W_n}'(w_n) = f_{W_n}(w_n) \left[-\beta + \frac{\alpha - 1}{w_n - \gamma} \right], \quad (13)$$

where α , β , and γ are shape, scale, and location parameters of marginal distribution function $F_{W_n}(w_n)$, respectively.

Substitute (13) into (12) and simplify it as follows:

$$c_1 \cdot f_{W_n}(w_n) + c(n) \cdot \left(-\beta + \frac{\alpha - 1}{w_n - \gamma} \right) = 0. \quad (14)$$

Equation (14) is the general formula of the CMLC method. If the peak discharge and flood volumes are independent, then the value of c_1 would be 0, and the value of $c(n)$

would be 1. Substituting the values of c_1 and $c(n)$ into (14), it could be simplified as follows: $w_n = (\alpha - 1) / \beta + \gamma$. However, the independent assumption is only a special case in practice.

The CMLC method is an approach to describe the compositions of flood peak and volumes by using the conditional density function to measure the occurrence likelihood of flood events. It is the mode of the conditional distribution $F_{W_n|Q,W_1,W_2,\dots,W_{n-1}}(w_n)$. Based on the principle of maximizing conditional joint probability density function, (14) can be solved by numerical computation methods, such as harmonic mean Newton's method [31].

3.3. Conditional Expectation Composition (CEC) Method.

The CEC method is also proposed to estimate the volumes under given hydrological variables, which is another way to analyze the inherent dependence between correlated flood peak and volumes. If the variables $Q, W_1, W_2, \dots, W_{n-1}$ are known, then the conditional expectation $E(W_n | Q, W_1, W_2, \dots, W_{n-1})$ is used to estimate flood volume; that is,

$$\begin{aligned}
 E(W_n | Q, W_1, W_2, \dots, W_{n-1}) \\
 = \int_{-\infty}^{+\infty} w_n \cdot f_{W_n|Q,W_1,W_2,\dots,W_{n-1}}(w_n) dw_n \\
 = \int_{-\infty}^{+\infty} w_n \cdot \frac{c(u, v_1, v_2, \dots, v_n)}{c(u, v_1, v_2, \dots, v_{n-1})} \cdot f_{W_n}(w_n) dw_n \quad (15) \\
 = \int_0^1 \frac{c(u, v_1, v_2, \dots, v_n)}{c(u, v_1, v_2, \dots, v_{n-1})} \cdot F_{W_n}^{-1}(v_n) dv_n,
 \end{aligned}$$

where $F_{W_n}^{-1}(v_n)$ is $v_n = F_{W_n}(w_n)$ reverse function.

Equation (15) is the general formula of conditional expectation composition (CEC) method, which can be solved by Gauss-Legendre quadrature rule [32].

3.4. Confidence Interval of Conditional Probability Distribution.

If $Q, W_1, \dots, W_{n-1}$ are known, let

$$g(v_n) = \frac{c(u, v_1, v_2, \dots, v_n)}{c(u, v_1, v_2, \dots, v_{n-1})} \cdot F_{W_n}^{-1}(v_n). \quad (16)$$

For a given significance level α , the lower and upper limits W_n^l and W_n^u identifying the confidence interval of flood volume W_n are derived by

$$\begin{aligned}
 W_n^l &= \int_0^{\alpha/2} g(v_n) dv_n, \\
 W_n^u &= \int_0^{1-\alpha/2} g(v_n) dv_n. \quad (17)
 \end{aligned}$$

Specifically, if $\alpha = 0.10$, then the 5% and 95% quantiles compose the 90% confidence intervals of conditional probability distribution.

TABLE 1: Statistics of flood peak and volume series at Danjiangkou reservoir.

	Flood peak ($10^3 \text{ m}^3/\text{s}$)	3-day flood volume (10^8 m^3)	7-day flood volume (10^8 m^3)	15-day flood volume (10^8 m^3)
[Min, max]	[2.3, 34.3]	[4.61, 66.43]	[10.04, 93.12]	[20.96, 134.40]
Median	15.50	18.45	37.67	67.95
Mean	15.77	28.03	46.07	71.82
Variance	8.32	15.27	24.07	35.17
Coefficient of skewness	0.384	0.628	0.551	0.481

FIGURE 1: Location of Danjiangkou reservoir in the Hanjiang basin.

3.5. Evaluation Indexes. To assess and compare different methods, the root mean square error (RMSE) and relative error (Bias) were selected as evaluation indexes:

$$\text{RMSE} = \sqrt{\frac{1}{L} \sum_{i=1}^L \left(\frac{\hat{X}_i - X_i}{X_i} \right)^2}, \quad (18)$$

$$\text{Bias} = \frac{1}{L} \sum_{i=1}^L \left(\frac{\hat{X}_i - X_i}{X_i} \right) \times 100\%,$$

where X_i and $\hat{X}_i$ are observed and estimated values and L is the length of sample series. The smaller these indexes are, the better the fitting or estimating methods are.

4. Case Study

4.1. Danjiangkou Reservoir. The Hanjiang basin is the largest tributary of the Yangtze River; it passes through the provinces of Shanxi and Hubei of China and merges into the Yangtze River at Wuhan city. The river's length is 1570 km and the basin area is 159,000 km² with a subtropical monsoon climate. The regional precipitation and flow are characterized by high seasonal variability, with 75% of rainfall occurring from June to October, where the rare rainstorms in the early summer and long-lasting rains in the autumn often result in great floods.

The Danjiangkou reservoir located in the middle reach of the basin (Figure 1) is the source of water for the middle route

of the South-North Water Diversion Project in China. The available water resources in Hanjiang basin and the impact of water diversion have been discussed by many authors [33]. Since it is a multipurpose reservoir, there is serious conflict between flood prevention, water supply, and hydropower generation. The normal pool level is 170 m and corresponding storage capacity is 29.05 billion m³. The annual maximum peak discharges and flood volumes of the inflow of Danjiangkou reservoir are available with a systematic record of 61 years (1954–2014). The flood data series was provided by the Bureau of Danjiangkou Water Resources Management. This data series has been restored and checked for possible long-term trends and abrupt changes, and the test results demonstrate that the data set is stationary and represented. The sample statistics of the flood peak and volume series were listed in Table 1.

4.2. Flood Quantiles Estimated by P3/LM. Univariate hydrological frequency analysis model (P3/LM) was used to estimate flood quantile for given return periods. The estimated parameters of the P3 distribution for peak discharge and flood volumes series were listed in Table 2. A Chi-Square (χ^2) goodness-of-fit test was performed to test the assumption, H_0 , that the flood magnitudes follow the P3 distribution. Table 2 shows that the assumption could not be rejected at the 5% significance level.

For given different return periods ($T = 5, 10, 20, 50, 100, 200, 500$, and 1000 years), the corresponding quantiles

TABLE 2: Estimated parameters of the P3 distribution and hypothesis test.

Variable	Sample statistic values			P3 parameter			χ^2 statistics	$\chi^2_{0.05}$
	Mean	$L-Cv$	$L-Cs$	α	β	γ		
$Q_{\max}$ (m ³ /s)	15768	0.54	0.58	11.8906	0.0004	-13593	0.28	3.84
W_{1d} (10 ⁸ m ³)	11.94	0.55	0.64	9.7656	0.4759	-8.5819	0.76	3.84
W_{3d} (10 ⁸ m ³)	28.03	0.56	0.82	5.9488	0.1554	-10.2549	0.88	3.84
W_{7d} (10 ⁸ m ³)	46.07	0.54	0.84	5.6689	0.0957	-13.1629	1.03	3.84
W_{15d} (10 ⁸ m ³)	71.82	0.5	0.71	7.9349	0.0784	-29.3349	1.29	3.84

TABLE 3: Comparison of design floods estimated by P3/LM and CMLC methods.

Return period (year)	$Q_{\max}$ (m ³ /s)		W_{7d} (10 ⁸ m ³)		W_{15d} (10 ⁸ m ³)		
	P3/LM	P3/LM	CMLC	Bias (%)	P3/LM	CMLC	Bias (%)
5	22625	65.4	64.7	-1.07	100.2	98.1	-0.02
10	27135	79.3	78.6	-0.88	119.7	116.6	-0.03
20	31122	92.0	91.4	-0.65	137.2	134.2	-0.02
50	35901	107.5	107.0	-0.47	158.4	154.6	-0.02
100	39261	118.7	118.2	-0.42	174.5	170.7	-0.02
200	42461	129.4	128.0	-1.09	189.9	184.3	-0.03
500	46498	143.0	140.9	-1.47	208.2	201.8	-0.03
1000	49437	156.1	148.3	-4.50	224.6	211.7	-0.06

TABLE 4: Bivariate copula functions and the relationship between their parameter and Kendall correlation coefficient.

Copula	$C_{\theta}(u, v)$	$\theta \in$	τ
Gumbel-Hougaard (G-H)	$\exp \left\{ - \left[(-\ln u)^{\theta} + (-\ln v)^{\theta} \right]^{1/\theta} \right\}$	$[1, \infty)$	$1 - \theta^{-1}$
Clayton	$(u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}$	$(0, \infty)$	$\frac{\theta}{(\theta + 2)}$
Ali-Mikhail-Haq (AMH)	$\frac{uv}{[1 - \theta(1-u)(1-v)]}$	$[-1, 1)$	$\tau = \left(1 - \frac{2}{3\theta}\right) - \frac{2}{3} \left(1 - \frac{1}{\theta}\right)^2 \ln(1 - \theta)$
Frank	$-\frac{1}{\theta} \ln \left[1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1} \right]$	$R \setminus \{0\}$	$\tau = 1 + \frac{4}{\theta} \left[\frac{1}{\theta} \int_0^{\theta} \frac{t}{\exp(t) - 1} dt - 1 \right]$

of flood peaks and volumes were estimated based on the univariate P3/LM model, and the results were listed in Table 3.

The empirical and P3 distribution frequency curves of flood peaks and flood volumes were drawn in Figure 2, on which the Weibull plotting position formula [34] was used. It is shown that theoretical values can fit the observed values very well.

4.3. Joint Distribution of Flood Peak and Volumes Based on Copulas. Different families of copula functions have been proposed and described by Nelsen [30]. The Archimedean copula family is more desirable for hydrologic analyses, because most of the family can easily be constructed and some of the family can be applied whether the correlation among the hydrological variables is positive or negative [16, 26]. The copula functions (Table 4) were applied to construct the bivariate joint distributions between flood peak ($Q_{\max}$) and flood volumes (W_{1d} , W_{3d} , W_{7d} , and W_{15d}). The values of Kendall correlation coefficient (τ) are 0.86, 0.80, 0.75, and 0.73 for the four pairs $Q_{\max}$ - W_{1d} , $Q_{\max}$ - W_{3d} , $Q_{\max}$ - W_{7d} , and

$Q_{\max}$ - W_{15d} , respectively. The parameters of bivariate copulas were estimated by the method of the inversion of Kendall's τ . Based on the results of Kendall's τ , the AMH copula could not be used because it is only suitable when Kendall's coefficient is below 0.34 [30].

The G-H, Clayton, and Frank copula functions were compared with different statistical tests. The p values were calculated based on the parametric bootstrap or multipliers [35] procedure with 10 000 runs. The results of the Cramer-von Mises test (S_n^I) and Kolmogorov-Smirnov test (D_n) were listed in Table 5.

The comparison between simulated (sample size 10 000) and observed values for three chosen copula functions was plotted in Figure 3. On the basis of statistical tests and graphical goodness-of-fit tests, the G-H copula is much suitable for constructing the bivariate joint distribution of peak discharge and flood volumes. This finding is in accordance with Li et al. [6] and Sraj et al. [16].

Therefore, the G-H copula was used to model the dependence between the flood peak and flood volumes in this study.

FIGURE 2: The fit of P3 probability curves with observed flood peak and flood volumes.

The linear correlation between empirical and theoretical frequency values of bivariate G-H copulas for flood peak and volumes was plotted in Figure 4, on which the linear correlation coefficients are 0.9994, 0.9991, 0.9991, and 0.9991, respectively. The comparison of empirical plots and theoretic frequency curves of bivariate G-H copulas was also plotted in Figure 5, which shows that the theoretical frequency line

(or curve) fit empirical values very well with high correlation coefficients.

The G-H, Clayton, and Frank copulas from Archimedean family were also used to establish trivariate symmetric and asymmetric joint distributions of $Q_{\max}-W_{1d}-W_{3d}$, $Q_{\max}-W_{3d}-W_{7d}$, and $Q_{\max}-W_{7d}-W_{15d}$, respectively. Table 6 shows selected trivariate copula functions, where parameter of

FIGURE 3: Comparison between the observed and simulated values for three chosen copula functions.

TABLE 5: Estimated parameters of copulas and results of statistical test.

Copula Pair	θ	G-H S_n^l (p value)	D_n	θ	Clayton S_n^l (p value)	D_n	θ	Frank S_n^l (p value)	D_n
$Q_{\max}-W_{1d}$	6.91	0.010 (0.39)	0.023	11.81	0.022 (0.21)	0.037	25.87	0.019 (0.22)	0.035
$Q_{\max}-W_{3d}$	5.08	0.012 (0.38)	0.027	8.17	0.026 (0.17)	0.040	18.53	0.022 (0.19)	0.030
$Q_{\max}-W_{7d}$	4.15	0.011 (0.35)	0.026	6.30	0.024 (0.18)	0.039	14.75	0.020 (0.21)	0.032
$Q_{\max}-W_{15d}$	3.77	0.013 (0.36)	0.028	5.55	0.021 (0.20)	0.035	13.21	0.017 (0.27)	0.031

FIGURE 4: Linear correlation between empirical and theoretical frequency values of bivariate G-H copulas for flood peak and volumes.

symmetric copulas is θ and asymmetric copulas are θ_1 and θ_2 ($\theta_2 > \theta_1$). Parameters of trivariate copula functions were estimated by the maximum likelihood method [26]. To assess and compare six trivariate copula functions, the Kolmogorov-Smirnov test value D_n and Akaike information criterion (AIC) [29] were selected and the results were listed in Table 7. It is shown that AIC and D_n values of asymmetrical G-H copulas are less than those of other copulas. Therefore, the asymmetrical G-H copulas are selected to establish 3-dimensional joint distribution of flood peak and volumes in this study. Similar to bivariate joint distribution, the linear correlation relationship and comparison of the empirical

plots and theoretical frequency line (or curve) of asymmetrical trivariate G-H copulas were plotted in Figures 6 and 7, which show that the theoretical frequency line (or curve) fit empirical values very well with high correlation coefficients.

4.4. CMLC and CEC Estimation and Assessment. Flood volumes were estimated by the CMLC, CEC methods and compared with P3/LM model. For bivariate distribution, the flood volumes (W_{1d} , W_{3d} , W_{7d} , and W_{15d}) based on peak discharge were estimated by (14) and the results were denoted as $\text{Prob}(W_{3d} | Q_{\max})$, $\text{Prob}(W_{7d} | Q_{\max})$, and $\text{Prob}(W_{15d} | Q_{\max})$, respectively. Similarly, the CEC method was used to

FIGURE 5: Comparison of the empirical plots and theoretical frequency curves of bivariate G-H copula for flood peak and volumes.

estimate $E(W_{3d} | Q_{\max})$, $E(W_{7d} | Q_{\max})$, and $E(W_{15d} | Q_{\max})$ by (15) for given peak discharge. As the P3/LM model is based on the assumption that volumes and peak discharge were in the same frequency, flood volumes (W_{1d} , W_{3d} , W_{7d} , and W_{15d}) were estimated for given return period.

The RMSE and Bias were selected as evaluation indexes to assess the performance of different methods. Table 8 shows that the CMLC method has smallest RMSE and Bias while CEC method has the largest RMSE and Bias values. For a given significance level $\alpha = 0.10$, then 90% confidence intervals of flood volumes were evaluated by (17). Figure 8 shows

7-day and 15-day flood volumes as well as confidence intervals (CI) estimated by CMLC and CEC methods. Figure 8 shows that most of observations are within confidence intervals. It could be concluded that CMLC method is reasonable and performs best in bivariate distribution.

To compare the performance of different methods in trivariate distribution, 15-day volumes were estimated by CMLC and CEC method, respectively, based on peak discharges and 7-day volumes. $\text{Prob}(W_{15d} | Q_{\max}, W_{7d})$ and $E(W_{15d} | Q_{\max}, W_{7d})$ were estimated by CMLC and CEC methods, respectively. Table 9 shows that CMLC method

TABLE 6: Trivariate symmetric and asymmetric Archimedean copula functions.

Copula functions	$C_\theta(u, v, w)$ or $C_{\theta_1}(w, C_{\theta_2}(u, v))$
Symmetric Gumbel-Hougaard (S-G-H)	$\exp \left\{ - \left[(-\ln u)^\theta + (-\ln v)^\theta + (-\ln w)^\theta \right]^{1/\theta} \right\}$
Symmetric Clayton (S-Clayton)	$\left\{ u^{-\theta} + v^{-\theta} + w^{-\theta} - 2 \right\}^{-1/\theta}$
Symmetric Frank (S-Frank)	$-\frac{1}{\theta} \ln \left\{ 1 + \frac{(\exp(-\theta u) - 1)(\exp(-\theta v) - 1)(\exp(-\theta w) - 1)}{(\exp(-\theta) - 1)^2} \right\}$
Asymmetric Gumbel-Hougaard-M6 (A-G-H)	$\exp \left\{ - \left(\left[(-\ln u)^{\theta_2} + (-\ln v)^{\theta_2} \right]^{\theta_1/\theta_2} + (-\ln w)^{\theta_1} \right)^{1/\theta_1} \right\}$
Asymmetric Clayton-M4 (A-Clayton)	$\left\{ \left(u^{-\theta_2} + v^{-\theta_2} - 1 \right)^{\theta_1/\theta_2} + w^{-\theta_1} - 1 \right\}^{-1/\theta_1}$
Asymmetric Frank-M3 (A-Frank)	$-\frac{1}{\theta_1} \ln \left\{ 1 - (1 - \exp(-\theta_1))^{-1} \left(1 - \left[1 - (1 - \exp(-\theta_2))^{-1} \cdot (1 - \exp(u\theta_2))(1 - \exp(v\theta_2)) \right]^{\theta_1/\theta_2} \right) (1 - \exp(-w\theta_2)) \right\}$

FIGURE 6: Linear correlation between the empirical and theoretical frequency values of trivariate asymmetric G-H copulas for flood peaks and volumes.

TABLE 7: Estimated parameters of trivariate copulas and results of statistical test.

Pair	$Q_{\max}-W_{1d}-W_{3d}$			$Q_{\max}-W_{3d}-W_{7d}$			$Q_{\max}-W_{7d}-W_{15d}$		
Copula	$\theta/[\theta_1, \theta_2]$	AIC	D_n	$\theta/[\theta_1, \theta_2]$	AIC	D_n	$\theta/[\theta_1, \theta_2]$	AIC	D_n
S-G-H	4.79	204.3	0.042	3.82	198.4	0.045	3.63	218.4	0.046
S-Clayton	7.25	210.3	0.047	5.84	206.4	0.049	5.24	225.7	0.050
S-Frank	15.34	208.9	0.044	12.54	204.8	0.047	11.24	220.4	0.049
A-G-H	[4.56, 6.45]	200.5	0.035	[3.77, 4.03]	193.7	0.037	[3.51, 3.69]	206.5	0.040
A-Clayton	[7.01, 7.89]	204.6	0.043	[5.72, 6.24]	197.2	0.046	[5.12, 5.65]	213.9	0.047
A-Frank	[15.03, 16.02]	202.5	0.039	[11.89, 12.67]	200.4	0.041	[11.09, 12.34]	209.6	0.044

FIGURE 7: Comparison of the empirical plots and theoretical frequency curves of trivariate asymmetric G-H copulas for flood peaks and volumes.

TABLE 8: Comparison of flood volumes estimated by univariate and bivariate distributions.

Flood volumes	W_{1d}		W_{3d}		W_{7d}		W_{15d}	
Index	RMSE	Bias (%)	RMSE	Bias (%)	RMSE	Bias (%)	RMSE	Bias (%)
CMLC	0.1580	1.14	0.1833	1.41	0.2070	1.59	0.2299	2.53
CEC	0.1680	2.78	0.2025	4.08	0.2205	4.70	0.2503	5.70
P3/LM	0.1620	2.21	0.1878	2.49	0.2074	2.67	0.2359	4.21

FIGURE 8: Estimated volumes and confidence intervals for given peak discharges.

performs best with the smallest RMSE and Bias values, while CEC method has the largest values of indexes and univariate P3/LM model performs reasonably well.

4.5. Derivation of Design Flood Hydrograph. All of these results demonstrate that the CMLC method performs best for analyzing the inherent relationship between flood peak and volumes. The proposed method was used to derive the DFH in Danjiangkou reservoir. There are so many definitions of return period, such as AND, OR, and Kendall return periods [23, 24]. The conditional return period [15, 27] was selected just as an example to demonstrate the proposed method to derive DFH. In Danjiangkou reservoir, the peak discharge, 7-day and 15-day volumes were always indicated as the control hydrological variables for deriving DFH. For different return periods (5, 10, 20, 50, 100, 200, 500, and 1000 years), the peak discharge was estimated by marginal distribution firstly. The second step is to estimate 7-day volumes under the calculated peak discharge by bivariate CMLC method. The third step is to estimate 15-day volumes under the calculated peak

discharge and 7-day volumes by trivariate CMLC method. The design flood volumes estimated by CMLC and Chinese method were listed in Table 3. It is shown that 7-day and 15-day flood volumes estimated by CMLC method are less than those of P3/LM method for different return periods.

After getting the values of flood peak and volumes under each return period, the DFH could be derived by amplifying the typical flood hydrographs (TFH). The observed annual maximum flood hydrographs of 1975 and 1983 that have high peaks and large volumes with posterior-peak shape at the Danjiangkou reservoir were selected as TFH, respectively. The amplifiers of flood peak and volumes were calculated by (2) to (6), respectively. The DFH is derived by multiplying discharge ordinate of the TFH by corresponding amplification factor. As an illustrative example, only 1000-year DFH derived by the proposed CLMC and conventional Chinese methods were plotted in Figure 9 based on 1975 and 1983 TFH. It is shown that the flood quantiles estimated by the Chinese method are larger than that by the proposed CLMC method.

TABLE 9: Comparison of flood volumes estimated by univariate and trivariate distributions.

Method	W_{3d}		W_{7d}		W_{15d}	
Index	RMSE	Bias (%)	RMSE	Bias (%)	RMSE	Bias (%)
3-dimension CMLC	0.1525	0.87	0.1501	-0.58	0.1518	-0.33
3-dimension CEC	0.1976	4.58	0.2741	4.71	0.2807	6.69
P3/LM	0.1833	1.41	0.2070	1.59	0.2299	2.53

FIGURE 9: 1000-year DFH derived by the CMLC and Chinese methods based on 1975 and 1983 TFH.

5. Conclusions

The derivation of design flood hydrograph is very important in multivariate flood frequency analysis frameworks. Unique design floods under given standard are necessary for engineering practice. The simple regression method may be preferred due to its ease for application. To avoid the arbitrariness of design floods, the proposed conditional methods which have statistical basis might be one of the possible options for deriving flood quantiles and DFH. In this study, the general formulae of CMLC and CEC methods were proposed and developed to consider the relationship between the flood peaks and volumes. The proposed methods were applied in the Danjiangkou reservoir and compared with conventional Chinese univariate flood frequency analysis method. The following conclusions were drawn from this study.

- (1) The Chinese method for deriving DFH implicitly assumed that flood peak and flood volumes are independent and amplified with the same frequency, which might produce over- or underestimation of risk.
- (2) The Gumbel-Hougaard copula is more suitable for describing the correlation between flood peak and volumes. The results of case study in Danjiangkou reservoir demonstrate that the CMLC method performs best while the CEC method performs worst.
- (3) The proposed CMLC method not only has strong statistical basis, but also has rational results with unique composition scheme and will provide a new approach for derivation of design flood hydrograph.

Competing Interests

The authors declare that there are no competing interests regarding the publication of this paper.

Acknowledgments

This study was supported by the Chinese National Nature Research Foundation (Grant nos. 51539009, 51579183).

References

- [1] ICOLD (International Congress on Large Dams), "Selection of design flood," Bulletin Report 82, ICOLD, Paris, France, 1992.
- [2] Y. Xiao, S. L. Guo, P. Liu, B. Yan, and L. Chen, "Design flood hydrograph based on multicharacteristic synthesis index method," *Journal of Hydrologic Engineering*, vol. 14, no. 12, pp. 1359–1364, 2009.
- [3] MWR (Ministry of Water Resources), *Regulation for Calculating Design Flood of Water Resources and Hydropower Projects*, Water Resources & Hydropower Press, Beijing, China, 2006 (Chinese).
- [4] P. Liu, L. Li, S. Guo et al., "Optimal design of seasonal flood limited water levels and its application for the Three Gorges Reservoir," *Journal of Hydrology*, vol. 527, pp. 1045–1053, 2015.
- [5] L. Chen, V. P. Singh, G. Shenglian, Z. Hao, and T. Li, "Flood coincidence risk analysis using multivariate copula functions," *Journal of Hydrologic Engineering*, vol. 17, no. 6, pp. 742–755, 2012.
- [6] T. Y. Li, S. L. Guo, L. Chen, and J. Guo, "Bivariate flood frequency analysis with historical information based on copula," *Journal of Hydrologic Engineering*, vol. 18, no. 8, pp. 1018–1030, 2013.
- [7] N. Pramanik, R. K. Panda, and D. Sen, "Development of design flood hydrographs using probability density functions," *Hydrological Processes*, vol. 24, no. 4, pp. 415–428, 2010.
- [8] L. Mediero, A. Jiménez-Álvarez, and L. Garrote, "Design flood hydrographs from the relationship between flood peak and volume," *Hydrology and Earth System Sciences*, vol. 14, no. 12, pp. 2495–2505, 2010.
- [9] F. Serinaldi and S. Grimaldi, "Synthetic design hydrographs based on distribution functions with finite support," *Journal of Hydrologic Engineering*, vol. 16, no. 5, pp. 434–446, 2011.
- [10] M. R. Domínguez and J. M. L. Arganis, "Validation of methods to estimate design discharge flow rates for dam spillways with large regulating capacity," *Hydrological Sciences Journal*, vol. 57, no. 3, pp. 460–478, 2012.
- [11] A. Óscar, D. M. Ramon, A. J. Maritza et al., "Estimate of design hydrographs for the Angostura Dam, Sonora, using statistical and spectral methods," *Water Resources Management*, vol. 29, no. 11, pp. 4021–4043, 2015.
- [12] H. Chowdhary, L. A. Escobar, and V. P. Singh, "Identification of suitable copulas for bivariate frequency analysis of flood peak and flood volume data," *Hydrology Research*, vol. 42, no. 2-3, pp. 193–216, 2011.
- [13] M. V. Bačová and D. Halmová, "Joint modeling of flood peak discharges, volume and duration: a case study of the Danube River in Bratislava," *Journal of Hydrology and Hydromechanics*, vol. 62, no. 3, pp. 186–196, 2014.
- [14] B. Gräler, M. J. VandenBerg, S. Vandenbergh et al., "Multivariate return periods in hydrology: a critical and practical review focusing on synthetic design hydrograph estimation," *Hydrology and Earth System Sciences*, vol. 17, no. 4, pp. 1281–1296, 2013.
- [15] L. Gaál, J. Szolgay, S. Kohnová et al., "Dependence between flood peaks and volumes: a case study on climate and hydrological controls," *Hydrological Sciences Journal*, vol. 60, no. 6, pp. 968–984, 2015.
- [16] M. Sraj, N. Bezak, and M. Brilly, "Bivariate flood frequency analysis using the copula function: a case study of the Litija station on the Sava River," *Hydrological Processes*, vol. 29, no. 2, pp. 225–238, 2015.
- [17] D. Nijssen, A. Schumann, M. Pahlow, and B. Klein, "Planning of technical flood retention measures in large river basins under consideration of imprecise probabilities of multivariate hydrological loads," *Natural Hazards and Earth System Science*, vol. 9, no. 4, pp. 1349–1363, 2009.
- [18] F. Chebana and T. B. Ouarda, "Multivariate quantiles in hydrological frequency analysis," *Environmetrics*, vol. 22, no. 1, pp. 63–78, 2011.
- [19] G. Salvadori, C. De Michele, and F. Durante, "Multivariate design via Copulas," *Hydrology and Earth System Sciences Discussions*, vol. 8, no. 3, pp. 5523–5558, 2011.
- [20] E. Volpi and A. Fiori, "Design event selection in bivariate hydrological frequency analysis," *Hydrological Sciences Journal*, vol. 57, no. 8, pp. 1506–1515, 2012.
- [21] S. Yue and P. Rasmussen, "Bivariate frequency analysis: discussion of some useful concepts in hydrological application," *Hydrological Processes*, vol. 16, no. 14, pp. 2881–2898, 2002.
- [22] J. T. Shiau, "Return period of bivariate distributed extreme hydrological events," *Stochastic Environmental Research and Risk Assessment*, vol. 17, no. 1-2, pp. 42–57, 2003.
- [23] G. Salvadori and C. De Michele, "Frequency analysis via copulas: theoretical aspects and applications to hydrological events," *Water Resources Research*, vol. 40, no. 12, Article ID W12511, pp. 1–17, 2004.
- [24] S. Vandenbergh, N. E. C. Verhoest, C. Onof, and B. De Baets, "A comparative copula-based bivariate frequency analysis of observed and simulated storm events: a case study on Bartlett-Lewis modeled rainfall," *Water Resources Research*, vol. 47, no. 7, Article ID W07529, 2011.
- [25] E. Volpi and A. Fiori, "Hydraulic structures subject to bivariate hydrological loads: return period, design, and risk assessment," *Water Resources Research*, vol. 50, no. 2, pp. 885–897, 2014.
- [26] G. Salvadori and C. De Michele, "Multivariate real-time assessment of droughts via copula-based multi-site Hazard Trajectories and Fans," *Journal of Hydrology*, vol. 526, pp. 101–115, 2015.
- [27] S. Yue, T. B. M. J. Ouarda, B. Bobée, P. Legendre, and P. Bruneau, "The Gumbel mixed model for flood frequency analysis," *Journal of Hydrology*, vol. 226, no. 1-2, pp. 88–100, 1999.
- [28] Q. Zhang, T. Qi, V. P. Singh, Y. D. Chen, and M. Xiao, "Regional frequency analysis of droughts in China: a multivariate perspective," *Water Resources Management*, vol. 29, no. 6, pp. 1767–1787, 2015.
- [29] L. Chen, S. L. Guo, B. W. Yan et al., "A new seasonal design flood method based on bivariate joint distribution of flood magnitude and date of occurrence," *Hydrological Sciences Journal*, vol. 7, no. 11, pp. 1264–1280, 2015.
- [30] R. B. Nelsen, *An Introduction to Copulas*, Springer Series in Statistics, Springer, New York, NY, USA, 2nd edition, 2006.
- [31] A. Y. Özban, "Some new variants of Newton's method," *Applied Mathematics Letters*, vol. 17, no. 6, pp. 677–682, 2004.

- [32] E. Babolian, M. J. Masjed, and M. R. Eslahchi, "On numerical improvement of Gauss-Legendre quadrature rules," *Applied Mathematics and Computation*, vol. 160, no. 3, pp. 779–789, 2005.
- [33] S. L. Guo, J. X. Wang, L. H. Xiong, A. Ying, and D. Li, "A macro-scale and semi-distributed monthly water balance model to predict climate change impacts in China," *Journal of Hydrology*, vol. 268, no. 1-4, pp. 1–15, 2002.
- [34] S. L. Guo and C. Cunnane, "Evaluation of the usefulness of historical and palaeological floods in quantile estimation," *Journal of Hydrology*, vol. 129, no. 1-4, pp. 245–262, 1991.
- [35] I. Kojadinovic, J. Yan, and M. Holmes, "Fast large-sample goodness-of-fit tests for copulas," *Statistica Sinica*, vol. 21, no. 2, pp. 841–871, 2011.

Hindawi

Submit your manuscripts at
<http://www.hindawi.com>