Water Resources Management (2022) 36:729–743
https://doi.org/10.1007/s11269-021-03055-3

Multivariate Flood Frequency Analysis Using Bivariate Copula Functions

Homa Razmkhah[1] · **Alireza Fararouie**[1] · **Amin Rostami Ravari**[1]

Received: 25 June 2021 / Accepted: 19 December 2021 / Published online: 19 January 2022

Abstract
Multivariate analysis of flood frequency was used extensively in water resources research. Often the only flood peak or volume is analyzed with statistical distributions, but for a perfect and exact result, the four main characteristics of a flood event, as well as peak, volume, duration, and time-to-peak, are needed. For this reason, multivariate statistical approaches like copula functions developed. This research aims to define and use the bivariate copula (2-copula) probability distribution functions (PDF) for flood characteristics multivariate analysis. When the joint distribution of characteristics such as volume and peak is known, it is possible to define the probability of simultaneous occurrence of design volume and peak flow values.

Keywords Flood characteristic · Bivariate frequency analysis · Copula

1 Introduction

In most hydrological studies several random variables are generally considered independent. For instance, different combinations of rainfall intensity and duration (5), flood peak and volume (Di Michele et al. 2005), drought duration, intensity, magnitude, and so on. Thus, it is often important to link the marginal probability distributions of different variables to describe the main aspect of hydrological events. Conventionally, multivariate probability distributions are used in hydrology (Singh 1987). The most used joint cumulative distribution function (CDF) was the Gaussian, but it has the obvious limitations that the marginal distributions should be normal. The Box-Cox's formula helped to reach condition by data transformation (Box and Cox 1964), however, do not always ensure that the recorded data follow a Gaussian PDF. Thereafter, further bivariate PDFs with non-normal margin distribution, such as the Gumbel bivariate exponential model (Gumbel 1960), with exponential margins, and bivariate Gamma distribution function used in hydrological studies (Grimaldi and Serinaldi 2006). Yue (2001) suggests bivariate Gamma distribution in pairs peak-volume and volume-duration flood frequency analysis. Knowing that extreme hydrological events can be represented by one of the three univariate extreme value

✉ Homa Razmkhah
Homarazmkhah@gmail.com

[1] Department of Water Engineering, Marvdasht Branch, Islamic Azad University, Marvdasht, Iran

distributions (Sklar 1959), bivariate Logistic and mixed models with Gumbel margins proposed (Gumbel and Mustafi 1967).

However, all mentioned models show some limits such as a similar family of marginal distributions, complicated mathematical formulation of an increased number of variables, and distinguish of marginal and joint behavior of variables. The copula PDF functions could overthrow these restraints. Many studies present some recent advances of copula using in hydrological modelings such as calculation of conditional probabilities, level curves of joint distributions, and return periods of multivariate events. Palynchuk and Guo (2011), applied copulas to joint distributions of rainstorm variables (depth, duration, and peak intensity). Serinaldi and Grimaldi (2007) used Fully Nested 3-copula to model the wave height, peak period, and maximum height as three characteristics of sea wave behavior. Renard and Lang (2007) used the Gaussian copula for multivariate extreme value analysis to prepare discharge-duration-frequency (QDF) models. Wang et al. (2009) used copula for flood frequency analysis of the confluence of river systems. Ghizzoni et al. (2010) used copula functions to provide an estimate of the joint probability of flood events in a multisite analysis. Grimaldi and Serinaldi (2006), Dupuis (2007), Genest and Favre (2007), Chen et al. (2010), Reddy and Ganguli (2012), Ganguli and Reddy (2013) used copula in multivariate flood frequency analysis (flood volume, peak, and duration). Razmkhah et al. (2016a, b) used copulas to analyze the spatial dependence of rainfall events and its effect on Rainfall-Runoff (RR) uncertainty modeling. Razmkhah et al. (2017) evaluated the effect of parameter correlation on the RR uncertainty, using 2-copula. Lorenz et al. (2018) used copula for downscale daily precipitation considering spatial correlation. Ayantobo et al. (2019) used a four-variate Archimedean copula in drought frequency analysis. Zhong et al. (2021) analyzed flash flood risk under the compound effect of soil moisture and rainfall using a copula function. Zening et al. (2020) analyzed reservoir inflow for four reservoirs on mainstream and its tributaries using a multivariate copula.

This paper aims to model the bivariate joint distribution of flood peak, volume, base flow duration, and time to peak using a wide range of copula families in the Karoon III basin. The innovation of this work is to analyze flood characteristics such as time to peak (tp), volume (V), peak discharge (Qp), and hydrograph base time (tb), using two variable copula and considering the dependence of the parameters. The accuracy of predicted flood properties such as volume and peak discharge and time to peak is important for flood warning purposes, maintenance, and the operating rules of the dam.

2 Methodology

2.1 Multivariate Flood Frequency Analysis

A probability distribution function (PDF) is a function used to define a particular probability distribution. It is the mathematical function that gives the probabilities of occurrence of different possible outcomes (Everitt 2006; Ash 2008, in Wikipedia) of a random phenomenon in terms of its sample space and the probabilities of events (subsets of the sample space) (Evans and Rosenthal 2010, in Wikipedia).

Flood frequency analysis (FFA) is the means by which flood discharge magnitude (Q) is related to the probability of its being equaled or exceeded in any year or to its frequency of recurrence or return period (T). The return period is used to indicate the average interval between floods of a given magnitude (Archer 1998). Most of the studies, use peak flood frequency analysis to find the occurrence probability of a flood event.

But to design most hydraulic structures, it is more accurate to have information not only about peak discharge (QP, m^3/s), but also the duration of the event (tb, days), in terms of the time interval that discharge exceeds the fixed threshold, volume (V, m3) (Yue et al. 2002), and their joint probabilistic behavior together with flood peak because these flood variables are correlated (Singh and Singh 1991). Hence, many researchers have performed multivariate FFA. For example, Yue (2001) used a bivariate gamma PDF to describe the joint probability distributions, conditional distributions, and joint return periods of two correlated variables. A brief review of univariate and multivariate approaches was explained in Vittal et al. (2015).

Multivariate frequency analysis makes it possible to jointly evaluate Q_p, $t_{b,}$ and V. Knowing the CDF of (P, V, D, …) and given design P, the pair (V, D) with a specific joint return period can be estimated (Grimaldi and Serinaldi 2006).

2.2 Derivation of Multivariate Distributions

The joint cumulative distribution function (H) of two random variables X and Y can be defined:

$$H(X, Y) = P(X \leq x, Y \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{y} h(U, V) dUdV \tag{1}$$

where x and y are the values of X and Y, and P is the non-exceedance probability. Or the right-hand side represents the probability that the random variable x takes on a value less than or equal to x, and that y takes on a value less than or equal to y (Park 2018, in Wikipedia).

The joint return period of an event (X, Y), $T_{X,Y}(X, Y)$, can be expressed as:

$$T_{X,Y}(X, Y) = \frac{1}{1 - H(X, Y)} \tag{2}$$

2.3 Copula

Let observation $\left(X_{11}, X_{21}, \ldots \ldots .X_{n1}\right), \ldots \ldots \ldots \ldots \ldots \ldots, \left(X_{1n}, X_{2n}, \ldots \ldots .X_{nN}\right)$ be drawn from a multivariate population of $\left(X_1, X_2, \ldots ..X_n\right)$, where N is the number of observations and n is the number of variables. Let$F_{X_i}\left(X_i\right)$, $\mathrm{i} = 1, 2, \ldots$ n, be the marginal CDFs of X_i, $\mathrm{i} = 1, 2, \ldots$ n.

The objective is to determine the multivariate distribution denoted as $H_{X_1, X_{2, \ldots\ldots\ldots} X_n}\left(X_1, X_2, \ldots ..X_n\right)$ or simply H. Copulas are functions that join one-dimensional marginal PDFs to a multivariate PDF (Nelsen 1999). Therefore the multivariate PDF (H) expressed in terms of its marginal distributions and the associated dependence function, C, as:

$$C\left(F_{X_1}\left(X_1\right), \ldots \ldots ..F_{X_n}\left(X_n\right)\right) = H_{X_1, X_{2, \ldots\ldots\ldots} X_n}\left(X_1, X_2, \ldots ..X_n\right) \tag{3}$$

where C, named copula, takes the quality of random variable dependence. The copula functions were developed by Sklar (1959), and different copula families were proposed by Nelsen, (1999). Some copula functions like Archimedean 2-copula e.g. Frank copula (Joe 1997) can model all ranges of variables dependence such as positive, null (independence), or negative (Grimaldi and Serinaldi 2006). Normal and t copulas are a subdivision of the Elliptical, which means that a copula is constructed based on the normal and t PDFs. Both normal and t are symmetrical, and the normal copula is a limiting case of the t when m becomes infinity. The advantage of the t copula is that it can capture lower and upper tail dependence of data

(Goda 2010). Another widely used family of 2-copula is the Archimedean. A detailed list of the Archimedean copulas is presented in Nelsen (2006). Popular Archimedean includes the Gumbel, Frank, and Clayton copulas. A difference between them is the tail dependence: the Gumbel and Clayton capture upper and lower tail dependence, respectively, but Frank does not. One disadvantage of the mentioned 2-copulas is the symmetrical feature concerning diagonal lines of a unit square. To deal with, a class of asymmetrical Archimedean, which are a transformed shape of mentioned family, can be used. Goda (2010) compared simulated samples of the normal and t copula, Gumbel, Frank, and Clayton copulas for different correlation coefficients and parameters and tail dependence characteristics.

The dependence of the variables is usually measured via the canonical Pearson's coefficient of a linear correlation when this parameter only models a linear dependence, and in some cases, it may not exist (De Michele et al. 2005). In copula modeling, the dependence of random data is measured by Kendall's and Spearman's coefficients. The Kendall's stated as the difference between the probability of concordance and discordance, while the Spearman's expressed as the linear correlation coefficient of the probability-transformed variables. The measures are rank-dependent and invariant under strictly monotonic transformations and can be expressed in terms of a copula function (Goda 2010). A copula function can be expressed as:

$$\mathrm{C}\left(\mathrm{u}_1, \ldots, \mathrm{u}_\mathrm{d}\right) = \mathrm{Pr}\left(\mathrm{U}_1 \le \mathrm{u}_1, \ldots, \mathrm{U}_\mathrm{d} \le \mathrm{u}_\mathrm{d}\right) \tag{4}$$

where $\mathrm{U_i} \sim \mathrm{U(O, 1)}$ for $\mathrm{i} = 1, \ldots, \mathrm{d}$. Copulas allow characterizing the dependence structure of random variables from the marginal PDFs. For any multivariate (joint) distribution F of $\mathrm{X}_1, \ldots \ldots .\mathrm{X}_\mathrm{d}$ as random variables, there exists a unique copula C as:

$$\mathrm{C}\left(\mathrm{u}_1, \ldots \ldots, \mathrm{u}_\mathrm{d}\right) = \mathrm{F}\left[\mathrm{F}_1^{-1}\left(\mathrm{u}_1, \ldots \ldots, \mathrm{F}_\mathrm{d}^{-1}\left(\mathrm{u}_\mathrm{d}\right)\right)\right] \tag{5}$$

where F_1^{-1} as the quantile function defined by $\mathrm{F}_1^{-1}(\mathrm{u}) = \inf\left[\mathrm{x} : \mathrm{F_i}(\mathrm{x}) \ge \mathrm{u}\right]$ (Sklar 1959).

Dupuis (2007) defined F as a multivariate PDF with marginal PDFs $F_1, \ldots ., F_d$ as:

$$\mathrm{F}\left(\mathrm{x}_1, \ldots, \mathrm{x}_\mathrm{d}\right) = C\left[F_1\left(x_1\right), \ldots, F_d\left(x_d\right)\right] \tag{6}$$

where C is copula function. Some of the bivariate copula families were presented in Razmkhah et al. (2016b). In this study, several bivariate copula families like Elliptical (Gaussian and Student t), Archimedean (Clayton, Gumble, Frank, Joe), BB_1 (Clayton-Gumbel), BB_6 (Joe-Gumbel), BB_7 (Joe-Clayton), and BB_8 (Frank-Joe)) have been fitted to data to select the most suitable one. For more details about copula families, readers could refer to Genest and MacKay (1986), Joe (1997), Favre et al. (2004), Nelsen (2006), and references therein. For the Archimedean families rotated versions cover negative dependence. The parameters of the 2-copula are estimated using maximum likelihood estimation.

2.4 Copula Model Selection

Copulas are selected according to the Akaike and Bayesian information criteria (AIC and BIC, respectively). First, all available copulas are fitted, then the criteria are computed and the copula with the minimum value chosen.

- AIC

The AIC is a criterion to estimator the relative quality of statistical models (McElreath 2016; Taddy 2019, in Wikipedia). For a statistical model with k as the number of estimated parameters, and $\hat{L}$ as the maximum value of the likelihood function of the model, the AIC value is calculated as (Burnham and Anderson 2002; Akaike 1974, in Wikipedia):

$$\mathrm{AIC} = 2\mathrm{K} - 2\mathrm{Ln}(\hat{\mathrm{L}}) \tag{7}$$

Given a set of candidate models, the preferred model is the one with the minimum AIC value.

- BIC

  The BIC or Schwarz information criterion is another criterion for model selection; the model with the lowest BIC is preferred. It is based, in part, on the likelihood function. It was developed by Schwarz (1978), considering a Bayesian argument.

  The BIC is defined as (Wit et al. 2012):

$$\mathrm{BIC} = \mathrm{KLn(n)} - 2\mathrm{Ln}(\hat{\mathrm{L}}) \tag{8}$$

where

- $(\hat{L})$ = the maximized value of the likelihood function of the model M, i.e. $\left(\hat{L}\right) = = \mathrm{p}\left(x/(\hat{\theta}, \mathrm{M})\right)$, where $\hat{\theta}$ are the parameter values that maximize the likelihood function;
- x = the observed data;
- n = the number of data points in x, the number of observations, or the sample size;
- k = the number of parameters estimated by the model

2.5 Conditional Probabilities

The conditional probabilities can be calculated, differentiating the corresponding copula (C) of random variables U and V (Nelsen 1999):

$$p\{V \leq v|U = u\} = \frac{\partial C(U, V)}{\partial U} \tag{9}$$

$$p\{U \leq u|V = v\} = \frac{\partial C(U, V)}{\partial V} \tag{10}$$

2.6 Copulas and Return Periods

Return period, as the average time between two successive realizations of the event, is a criterion for sizing the hydraulic structures and risk analysis (Salvadori and De Michele 2007). The univariate frequency analysis may lead to an under/overestimation of the risk because many hydrological events are characterized by the joint behavior of several random variables. Using copula can cover the mentioned disadvantage.

2.7 Bivariate Return Periods

Bivariate return period is discussed for x and y variables with increasing marginals Fx and Fy and C copula. Having U and V as Eq. (5) we have:

$$\begin{cases} U = F_x(x) \\ V = F_y(y) \end{cases} <=> \begin{cases} x = F_x^{-1}(U) \\ U = F_y^{-1}(V) \end{cases} \tag{11}$$

In practice, an event could be dangerous if either U or V exceed given thresholds, Eq. (12), or both U and V are longer than a prescribed value, Eq. (13):

$$E_{U,V}^{\cup} = \{U > u\} \cup \{V > v\}(or\ case) \tag{12}$$

$$E_{U,V}^{\cap} = \{U > u\} \cap \{V > v\}(and\ case) \tag{13}$$

where $\cup$ *and* $\cap$ are or and operators and $E_{U,V}^{\cup}$ and $E_{U,V}^{\cap}$ are the events with the greatest interest. Let the $P_{U,V}^{\cup}$ and $P_{U,V}^{\cap}$ be the probability of the events $E_{U,V}^{\cup}$ and $E_{U,V}^{\cap}$ respectively. Using copula we have (Nelsen 1999):

$$P_{U,V}^{\cup} = P\{U > u \cup V > v\} = 1 - C(U, V) \tag{14}$$

$$P_{U,V}^{\cap} = P\{U > u \cap V > v\} = \bar{C}(1 - U, 1 - V) \tag{15}$$

where

$$\bar{C}(U, V) = U + V - 1 + C(1 - U, 1 - V) \tag{16}$$

Then the return periods $T_{U,V}^{\cup}$ and $T_{U,V}^{\cap}$ of the events $E_{U,V}^{\cup}$ and $E_{U,V}^{\cap}$ can be defined as:

$$T_{U,V}^{\cup} = \frac{\mu}{P_{U,V}^{\cup}} \tag{17}$$

$$T_{U,V}^{\cap} = \frac{\mu}{P_{U,V}^{\cap}} \tag{18}$$

where μ = average interval time between two successive events E_i …. $T_{U,V}^{\cup}$ and $T_{U,V}^{\cap}$ are decreasing functions so we have (Salvadori and De Michele 2007):

$$P_{U,V}^{\cup} \geq P_{U,V}^{\cap}, \tag{19}$$

$$T_{U,V}^{\cup} \leq T_{U,V}^{\cap}. \tag{20}$$

Using the above equations, we can calculate the return periods of the events $E_{U,V}^{\cup}$ and $E_{U,V}^{\cap}$, in terms of the pair (x, y).

2.8 Univariate Versus Multivariate Analysis

In a multivariate hydrological event, univariate frequency analysis cannot evaluate the occurrence of an event completely. Salvadori and De Michele (2007) showed the desire return period (π) followed this equation:

$$T^{\cup}_{U,V} \le \pi \le T^{\cap}_{U,V} \tag{21}$$

So if the desire return period (π) was not determined using joint PDFs design could be under dimensioned leading to risk increase, or overestimated leading to waste of money. De Michele et al. (2005) using the value of real design showed the return period of the event $E^{\cup}_{U,V}$ was about 20% smaller than π, while the return period of $E^{\cap}_{U,V}$ was about 30% larger.

Another problem is the identification of the events having an assigned return period (T). In general, the solution is not unique, because different choices of U and V may yield the same T. Thus iso-lines of $T^{\cup}_{U,V}$ and $T^{\cap}_{U,V}$ should be identified (Salvadori and De Michele 2007).

2.9 Level Curves

The level curves of a joint distribution can provide practical information in the design process. Given $0 \le t \le 1$, the curves defined by the Eq. (22) are called the level curves of C:

$$L_t = \{C(U,V) = t\} \tag{22}$$

The use of copulas may yield analytical solutions to the problem of calculating the isolines (Salvadori and De Michele 2007).

2.9.1 Isolines of $T^{\cup}_{U,V}$

In the equation $T^{\cup}_{U,V} = \frac{\mu}{P^{\cup}_{U,V}} = \frac{\mu}{1-C_{U,V}}$, $P^{\cup}_{U,V}$ reduces by increasing C, while the $T^{\cup}_{U,V}$ increases. Increasing C is equivalent to choosing more extreme events with a longer return period. Salvadori and De Michele (2007) showed that iso-line plots would be the same for all of the distributions F_{xy} generated by C.

2.9.2 Isolines of $T^{\cap}_{U,V}$

We told that $P^{\cap}_{U,V} = \bar{C}(1-U, 1-V) = 1 - U - V + C(U,V)$ where $0 \le C(U,V) \le 1$. The authors referred to Salvadori and De Michele (2007) for preparing $T^{\cap}_{U,V}$ isolines.

2.10 Tail Dependence

The most common definition of tail dependence between two random variables X and Y is:

$$\lambda_U = \lim_{U\to 1} P\{G_X(X) > U | H_Y(Y) > V\} \tag{23}$$

$$\lambda_L = \lim_{U\to 0^+} P\{G_X(X) \le U | H_Y(Y) \le V\} \tag{24}$$

where λ_U and λ_L are called the upper and lower tail dependence. $G_X(X)$ and $H_Y(Y)$ are the marginal distributions and U and V are threshold levels. The tail dependence corresponds to the probability that one margin to be large/small conditioned on the other margin being large/small. The tail dependence can also be defined by a copula function. The authors referred to Yue et al. (2014) for a detailed description.

2.11 Case Study

The Karoon III is a sub-basin of large Karoon, located in the southwest of Iran. The watershed boundaries limit to 49°, 30' to 52° E and 30° to 32°, 30' N. The area is approximately 24,200 (km^2) with 30 reliable climatology and synoptic gauges. The elevation varies from 700 (m) at the outlet of the basin to 4500 (m) on the Kouhrang and Dena mountains. The digital elevation model (DEM) and major drainage network of Karoon III are shown in Fig. 1. About 50% of the watershed area has a higher elevation than 2500 (m). The watershed receives an average annual precipitation of 767 (mm) that about 55% of precipitation is as snowfall. The average daily discharge flow of the basin is about 384 (m^3/s). The sub-basins physiographic characteristics were presented in Razmkhah et al. (2016a, b) . Pole-Shaloo was the only existed streamflow gauge, at the outlet of the basin with recorded daily data used in this study.

3 Results

Since Q_p, t_b, $t_{p,}$ and V are variables defined by the same physical phenomenon, they should be mutually correlated. Table 1 summarized the statistics of the Pearson's correlation coefficient (r), Kendall's tau (τ), and Spearman's rho (ρ_s) of t_b, t_p, $Q_{p,}$ and V random variables. A two-sided test with 5% and 1% of significance levels confirmed that variables are correlated. Pearson's correlation coefficient represents the linear dependence between two variables. Both τ and ρ_s are based on the ranks which represent the degree of association

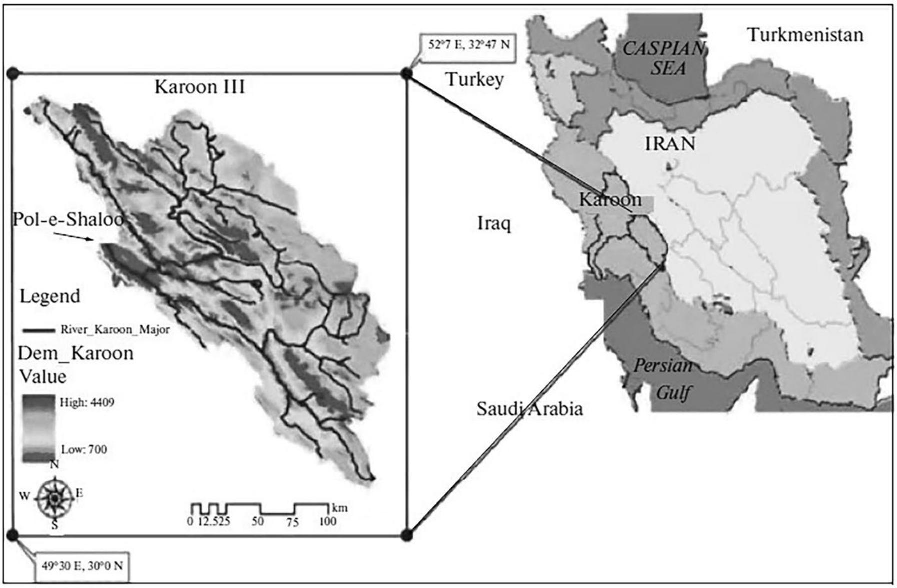


Fig. 1 DEM and major drainage network of Karoon III (Razmkhah 2018)

Table 1 Dependence between variables as expressed by Pearson's (r), Kendall's tau (τ), and Spearman's rho (ρ_s) correlation coefficients

Variables	r	τ	ρ_s
tb-Qp	0.223**	0.222**	0.298**
	0.009	0.000	0.009
tp-Qp	0.201*	0.163*	0.206*
	0.019	0.012	0.016
tp-tb	0.457**	0.490**	0.541**
	0.000	0.000	0.000
V-Qp	0.947**	0.879**	0.979**
	0.000	0.000	0.000
V-tb	0.350**	0.322**	0.430**
	0.000	0.000	0.000
V-tp	0.278**	0.233**	0.302**
	0.001	0.000	0.000

*Correlation is significant at 0.05 level; **Correlation is significant at 0.01 level

between two random variables (Yue et al. 2014). The copula (P, V) is more correlated than the others. Equation (25) shows the rank of correlation between flood characteristics:

$$(V, Qp) > (t_p, t_b) > (V, t_b) > (V, t_p) > (t_b, Q_p) > (t_p, Q_p) \tag{25}$$

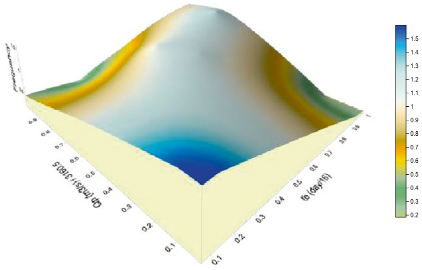
The order of correlation between variables could be explained by the physical governing low of rainfall-runoff processes. Extreme rainfalls resulted in a large amount of runoff volume and consequently peak discharge; long-based time floods have a longer time to peak; larger base time floods belong to the larger volume of floods and so on.

Table 2 shows copula families fitted to the flood characteristics, and related parameters, using R software (Brenchmann and Schepsmeier 2013). The P-value of the independent test showed that all of the variables are dependent. Quantities (V, Qp) and (t_p, t_b), seem to follow Clayton's family of copula while the others follow BB_8 copula. The (V, Qp) and (t_p, t_b) were the most correlated ones and tail dependence analysis showed that they are lower tail dependent, (Table 2). The BB_8 copula family fitted to the other variables, did not show any lower or upper tail dependence.

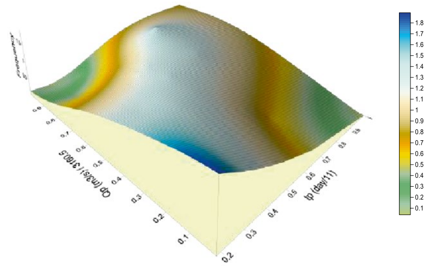
Figure 2a–f represent 3D plots of copula functions fitted to the flood characteristics. The variation of correlated parameters could be seen in the figures. It could be seen that (V, Qp) and (t_p, t_b) are the most correlated variables with lower tail dependence (Fig. 2c, d).

Table 2 Copula families fitted to the flood characteristics

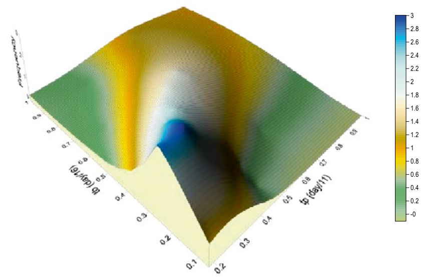
Flood Characteristic	P-value of Independent Test	Copula Family	Parameter 1	Parameter 2	Lower tail Dependence	Upper tail dependence
t_b-Q_p	0.000134	BB_8	1.908471	0.9323931	0	0
t_p-Q_p	0.005030	BB_8	3.203957	0.8607931	0	0
t_p-t_b	0	Clayton	2.980911	0	0.7925271	0
V-Q_p	0	Clayton	4.705008	0	0.8630168	0
V-t_b	3.11e-08	BB_8	2.005408	0.924485	0	0
V-t_p	6.26e-05	BB_8	3.329927	0.8346275	0	0



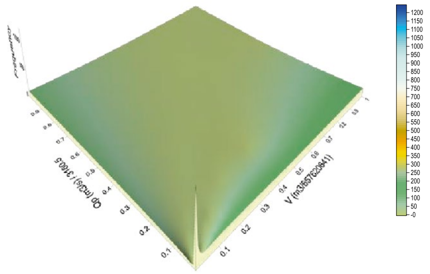
a. (t_b, Q_p)



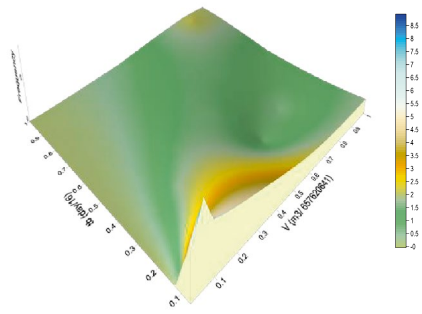
b. (t_p, Q_p)



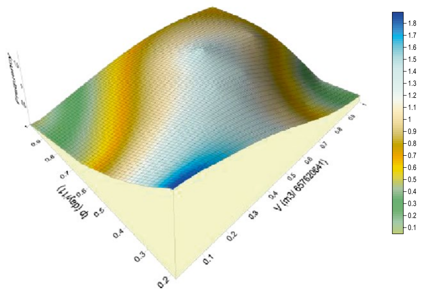
c. (t_p, t_b)



d. (V, Q_p)



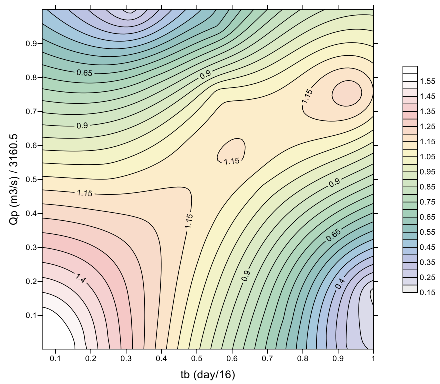
e. (V, t_b)



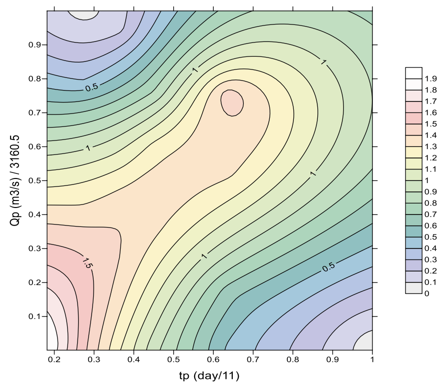
f. (V, t_p)

Fig. 2 3D plots of flood characteristics copula functions

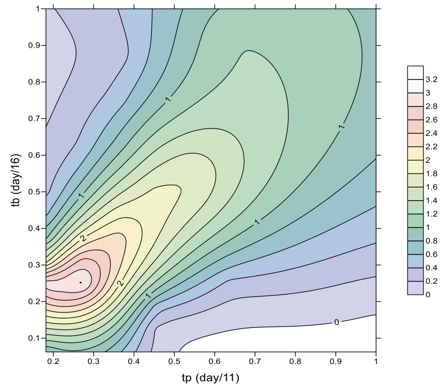
Figure 3a–f show iso-lines (contour plots) of the fitted copula functions. As we told the identification of the events having an assigned return period (T) in the multivariate analysis does not have a unique solution, because different choices of U and V may yield the same T. Thus iso-lines of T prepared. The figures could be used to determine



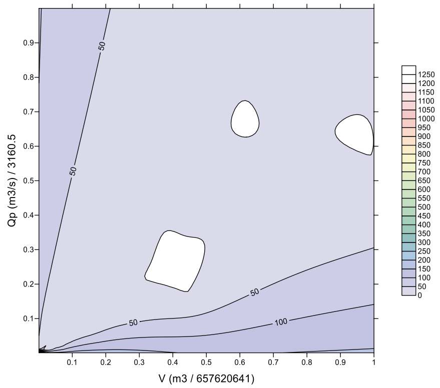
a. (t_b, Q_p)



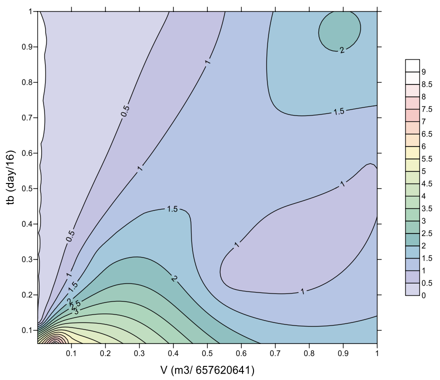
b. (t_p, Q_p)



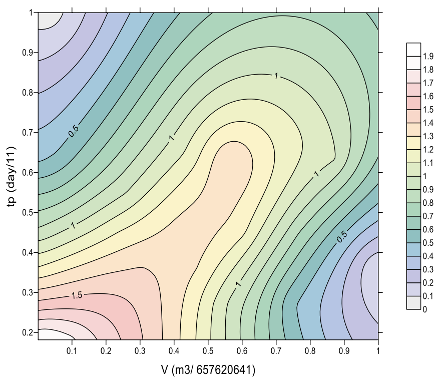
c. (t_p, t_b)



d. (V, Q_p)

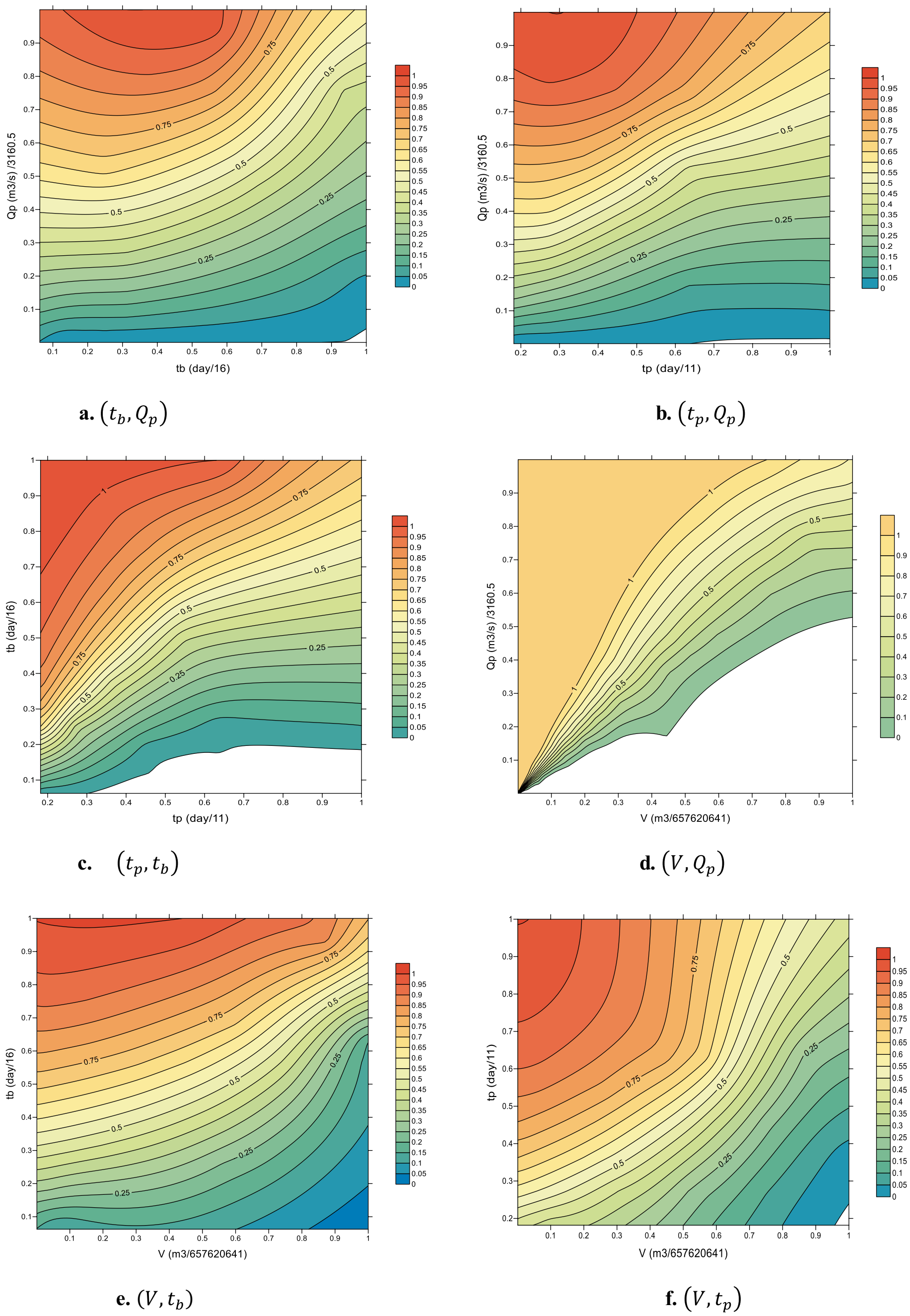


e. (V, t_b)



f. (V, t_p)

Fig. 3 Contour plots (iso-lines) of copula functions



a. (t_b, Q_p) **b.** (t_p, Q_p)

c. (t_p, t_b) **d.** (V, Q_p)

e. (V, t_b) **f.** (V, t_p)

Fig. 4 Conditional probability iso-lines of copula functions (V/U) for (U, V) different $\left(t_b, Q_p\right)$ with a specified T or return period (figure a); $\left(t_p, Q_p\right)$ from figure b; $\left(t_p, t_b\right)$ from figure c; $\left(V, Q_p\right)$ from figure d; $\left(V, t_b\right)$ from figure e, and $\left(V, t_p\right)$ from figure f. Then another flood characteristics could be extracted from the other graphs, and we have many possible variables with different probability to make a decision more precisely.

Figure 4a–f shows conditional probability iso-lines of the bi-variates copula functions of (V/U) for (U, V). The conditional probability iso-lines of (U/V) figure was not presented because of the large volume of the article. The figures could be used to determine a Q_p for a specified t_b and t_p (figure a, b); t_p for a specified t_b (figure c); Q_p for a specified V (figure d); V for a specified t_b, and t_p for a specified V (figure e, f) and so on.

4 Conclusion

In this study, a multivariate flood frequency analysis was performed using 2-copula functions. By this multivariate tool, bivariate density and distribution of peak, volume, duration, and time to peak of flood events are determined without assuming the same type of marginal distributions. A brief analysis of observed data shows that peak, volume, duration, and time to peak variables are dependent in several manners. Properties of the copulas are described numerically and graphically. Advantages in using copulas are highlighted and confirmed in the case study.

The use of copulas yields analytical or graphical solutions to the problem of calculating the iso-lines of a joint distribution. Once the design return period has been decided, it is easy to calculate all the joint events such as $E^{\cup}_{U,V}$ and $E^{\cap}_{U,V}$ in terms of the pair (x, y).

In particular, we consider events, depending upon the joint behavior of two dependent random variables. This approach can be easily generalized to the multivariate case, using copula functions without traditional multivariate joint distributions. These results yield a copula-based procedure for the estimation of the probability of correlated random variables. The results help hydrologists for stochastic joint dependence of the variables and offer the possibilities of correctly sizing the structures.

Author Contributions Homa Razmkhah: Conceptualization, data curation, methodology, analysis, interpretation of the results, software, visualization, and writing. Alireza Fararouie and Amin Rostami Ravari: Supervised the study and reviewed the whole content. All authors read and approved the manuscript.

Declarations

Ethics Approval Not applicable.

Consent to Participate Authors have agreed to submit the article.

Consent for Publication Authors have agreed to submit the article in this Journal.

Conflicts of Interest The authors have no conflict of interest.

References

Akaike H (1974) A new look at the statistical model identification. IEEE Trans Autom Control 19(6):716–723
Archer D (1998) Flood frequency analysis. In: Encyclopedia of Hydrology and Lakes. Encyclopedia of Earth Science. Springer, Dordrecht
Ash RB (2008) Basic probability theory (Dover ed.), Mineola, N.Y.: Dover Publications. pp 66–69
Ayantobo O, Li Y, Song S (2019) Multivariate drought frequency analysis using four-variate symmetric and asymmetric Archimedean copula functions. Water Resour Manag 33:103–127
Box GEP, Cox D (1964) An analysis of transformation (with discussion). J Roy Statist Soc Ser B 26:211–246
Brenchmann EC, Schepsmeier U (2013) Modeling dependence with C- and V-Vine copulas: the R package CDVine. J Statist Soft 52(3):1–27
Burnham KP, Anderson DR (2002) Model selection and multimodel inference: a practical information-theoretic approach (2nd ed.), Springer-Verlag
Chen L, Guo S, Yan B, Liu P, Fang B (2010) A new seasonal design flood method based on bivariate joint distribution of flood magnitude and date of occurrence. Hydrologic Sci 55(8):1264–1280
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(1):50–57
Dupuis DJ (2007) Using copulas in hydrology: benefits, cautions, and issues. J Hydrol Eng 12(4):381–393
Evans M, Rosenthal JS (2010) Probability and statistics: the science of uncertainty (2nd ed.), New York: W.H. Freeman and Co
Everitt BE (2006) The Cambridge dictionary of statistics (3rd ed.), Cambridge, UK: Cambridge University Press
Favre A-C, El Adlouni S, Perreault L, Thiémonge N, Bobée B (2004) Multivariate hydrological frequency analysis using copulas. Water Resour Res 40(1):W01101
Ganguli P, Reddy MJ (2013) Probabilistic assessment of flood risks using trivariate copulas. Theoretic Appl Climatol 111:341–360
Genest C, MacKay RJ (1986) Copules archimédiennes et familles de lois bidimensionnelles dont les marges sont données. Canad J Statat 14(2):145–159
Genest C, Favre A-C (2007) Everything you always wanted to know about copula modeling but were afraid to ask. J Hydrol Eng 12(4):347–368
Ghizzoni T, Roth G, Rudari R (2010) Multivariate skew-t approach to the design of accumulation risk scenarios for the flooding hazard. Adv in Water Resour 33:1243–1255
Goda K (2010) Statistical modeling of joint probability distribution using copula: Application to peak and permanent displacement seismic demands. Struc Safet 32:112–123
Grimaldi S, Serinaldi F (2006) Asymmetric copula in multivariate flood frequency analysis. Adv in Water Resour 29(8):1155–1167
Gumbel EJ (1960) Bivariate exponential distributions. J Americ Statist Assoc 55:698–707
Gumbel EJ, Mustafi CK (1967) Some analytical properties of bivariate extreme distributions. J Americ Statist Assoc 62:596–688
Joe H (1997) Multivariate models and dependence concepts. Chapman and Hall, CRC, Boca Raton
Lorenz M, Bliefernicht J, Haese B, Kunstmann H (2018) Copula-based downscaling of daily precipitation fields. Hydrolog Proc 32:3479–3494
McElreath R (2016) Statistical rethinking: a Bayesian course with examples in R and Stan. CRC Press
Nelsen RB (1999) An introduction to copulas. Springer
Nelsen RB (2006) An introduction to copulas. Springer
Palynchuk BA, Guo Y (2011) A probabilistic description of rainstorms incorporating peak intensities. J Hydrology 409:71–80
Park KII (2018) Fundamentals of probability and stochastic processes with applications to communications, Springer
Razmkhah H (2018) Parameter uncertainty propagation in a Rainfall – Runoff model; Case study: Karoon III river basin. Water Resour 45(1):34–49
Razmkhah H, Saghafian B, AkhoundAli AM, Radmanesh F (2016a) Rainfall-Runoff modeling considering soil moisture accounting algorithm, case study: Karoon III river basin. Wat Resource 43(4):699–710
Razmkhah H, AkhoundAli AM, Radmanesh F, Saghafian B (2016b) Evaluation of rainfall spatial correlation effect on rainfall-runoff modeling uncertainty, considering 2-copula. Arab J Geoscience 9:323
Razmkhah H, AkhoundAli AM, Radmanesh F (2017) Correlated parameters uncertainty propagation in a Rainfall-Runoff model, considering 2-copula, case study: Karoon III river basin. Env Model Assess 22:503–521

Reddy MJ, Ganguli P (2012) Bivariate flood frequency analysis of upper Godavari river flow using Archimedean copulas. Water Resour Manag 26(14):3995–4018
Renard B, Lang M (2007) Use of a Gaussian copula for multivariate extreme value analysis: Some case studies in hydrology. Adv Water Resour 30:897–912
Salvadori G, De Michele C (2007) On the use of copulas in hydrology: Theory and practice. J Hydrol Eng 12(4):369–380
Schwarz GE (1978) Estimating the dimension of a model. Ann Stat 6(2):461–464
Serinaldi F, Grimaldi S (2007) Fully nested 3-copula: Procedure and application on hydrological data. J Hydrol Eng 12(4):420–430
Singh VP (1987) Hydrological frequency modeling. Reidal, Boston
Singh K, Singh VP (1991) Derivation of bivariate probability density functions with exponential marginals. Stoch Hydrol Hydraul 5:55–68
Sklar A (1959) Fonctions de repartition an dimensions et leurs marges. Publications De L'institut Statistique De L'universite De Paris 8:229–231
Taddy M (2019) Business data science: Combining machine learning and economics to optimize, automate, and accelerate business decisions. McGraw-Hill, New York
Vittal H, Singh J, Kumar P, Karmakar S (2015) A framework for multivariate data-based at-site flood frequency analysis: Essentiality of the conjugal application of parametric and nonparametric approaches. J Hydrol 525:658–675
Wang C, Chang N-B, Yeh G-T (2009) Copula-based flood frequency (COFF) analysis at the confluences of river systems. Hydrol Process 23(10):1471–1486
Wit E., Heuvel EVD, Romeijin JW (2012) All models are wrong...': an introduction to model uncertainty. Statistica Neerlandica 66(3):217–236
Yue S (2001) A bivariate gamma distribution for use in multivariate flood frequency analysis. Hydrol Process 15(6):1033–1045
Yue S, Quarda T, Bobee B, Legendre P, Bruneau P (2002) Approach for describing statistical properties of flood hydrograph. J Hydrologic Eng 7(2):147–153
Yue K-X, Xiong L, Gottschalk L (2014) Derivation of low flow distribution functions using copulas. J Hydrology 508:273–288
Zening W, Chentao H, Wang H, Zhang Q (2020) Reservoir inflow synchronization analysis for four reservoirs on a multivariate copula model. Water Resour Manag 34:2753–2770
Zhong M, Zeng T, Jiang T, Wu H, Chen X, Hong Y (2021) A copula-based multivariate probability analysis for flash flood risk under the compound effect of soil moisture and rainfall. Water Resour Manag 35(4):1–16