University of Birjand

Water Harvesting Research (2026) Vol. 9, No. 1: 57-70

10.22077/jwhr.2026.11256.1205

ORIGINAL ARTICLE

Application of Nested Copulas for Flood Frequency Analysis (Case Study: Dez Basin, Iran)

Alireza Amirrezaeieh[a], Rasoul Mirabbasi[b&c*], Ali Saremi[a]

[a]Department of Agricultural Systems Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran.
[b]Department of Water Engineering, Shahrekord University, Shahrekord, Iran.
[C]Water Sciences and Hydroinformatics Research Center, Khazar University, Mahsati str. 41, AZ 1096, Baku, Azerbaijan

*Corresponding Author E-mail Address: mirabbasi@sku.ac.ir
Received: 10 April 2026, **Revised**: 21 May 2026, **Accepted**: 02 June 2026

Abstract
One of the hydrological phenomena with a very complex nature is the flood phenomenon, which causes a lot of damage if it occurs. In this study, the joint frequency analysis of flood characteristics at the Sepid Dasht-Zaz station in the Dez basin, Iran, over a 30-year period was investigated using nested copula functions. For this purpose, the flood characteristics such as peak flow, flood volume and base time of the flood were used. By fitting 11 different distribution functions to the studied series, finally, based on statistical tests, the generalized extreme value distribution (GEV) was chosen as a suitable marginal distribution for the studied variables. After selecting the marginal distribution, Archimedean family copulas (including Frank, Ali-Mikhail-Haq, and Clayton) were used for joint frequency analysis of the paired flood characteristics (peak flow and flood volume, peak flow and time base and flood volume and time base). The results showed that Frank and Clayton for the mentioned pair-variables are most consistent with the empirical copulas. The joint return period was used to investigate the trivariate return period of events. The results of joint analysis of flood characteristics in the study area resulted in the typical return period curves for Sepid Dasht-Zaz, which provides the return period of each flood characteristic with different probabilities and given by other flood characteristics. Since the characteristics of floods are different in each flood event, univariate flood analysis does not take into account all characteristics, and hence the use of nested copulas by increasing dimension of analysis can provide more realistic results.
Keywords: Archimedean, Copula, Joint analysis, Marginal distribution, Power law.

1. Introduction

Floods are among the most pervasive natural hazards, occurring with varying likelihood across different regions. Consequently, effective flood management has become a key priority for policymakers worldwide in order to mitigate both human and economic losses. Flood control encompasses a range of measures—such as storage, flow restriction, and flow diversion—aimed at reducing the destructive impacts of flooding, and these measures are selected based on engineering and economic considerations. The design and deployment of various hydraulic structures remain one of the most widely adopted strategies for flood management. In this context, engineers often rely on univariate frequency analyses, typically based on annual maximum discharge, to design such structures. However, floods are inherently multivariate in nature, characterized by interrelated variables such as peak discharge, runoff volume, and time base. Thus, adopting a multivariate perspective is essential to avoid errors arising from the omission of key flood characteristics.

Early attempts to model floods using classical multivariate distribution functions were pioneered by researchers such as Snyder

(1962) and Wong (1963). Nevertheless, these classical methods impose significant limitations, including the requirement of variable independence and the assumption that variables follow a normal distribution with identical marginal distributions. In reality, most hydrological variables exhibit non-normal behavior and often follow different marginal distributions. As a result, conventional multivariate distributions are often ill-suited for analyzing complex hydrological phenomena such as floods.

The copula function, originally developed by Sklar (1959), overcomes these limitations by allowing the construction of multivariate distributions from arbitrary univariate marginals, without imposing constraints on the dependence structure. In recent decades, copulas have attracted growing interest among researchers and have been widely used to accurately describe the dependency structure among correlated hydrological quantities (Salvadori and De Michele 2007). The flexibility and effectiveness of copulas in constructing joint distributions have been confirmed in numerous studies (Ahmadi et al. 2017).

Zhang and Singh (2006) investigated the dependence structure among flood peak, volume, and duration using data from the Amite River (USA) and the Ashuapmushuan River (Canada). They employed various copulas, including Clayton, Ali-Mikhail-Haq, Gumbel-Hougaard, and Frank, to build joint distributions. Their results indicated that the Gumbel-Hougaard copula performed best for modeling the duration–volume and peak–volume pairs, and they subsequently extracted conditional return periods. Xiao et al. (2008) applied copulas to the multivariate frequency analysis of flood peak and volume for the Jhuoshuei River (Taiwan). After determining the marginal distributions—Gumbel for peak flow and log-normal for volume—they tested the Ali-Mikhail-Haq, Gumbel-Hougaard, Plackett, Galambos, and Frank copulas, concluding that the Frank copula best represented the dependence structure.

Reddy and Ganguli (2012) performed a copula-based frequency analysis of flood characteristics (annual peak, volume, and duration) for the Godavari River (India). Among the tested families (Ali-Mikhail-Haq, Clayton, and Frank), the Frank copula was selected based on goodness-of-fit tests for computing conditional return periods. Numerous subsequent studies have successfully applied copulas to flood frequency analysis (Sraj et al. 2015; Saad et al. 2015; Ming et al. 2015; Balistrocchi and Bacchi 2017).

Nazeri Tahroudi et al. (2021a) used a copula-based method to model outflow hydrographs for the Wilson River (USA), Karun River (Iran), and River Wye (UK), demonstrating that the approach accurately simulated both the rising and falling limbs as well as peak flows. Dastourani and Nazeri Tahroudi (2022) applied copulas for the joint frequency analysis of groundwater drawdown and pumping time in a constant-discharge aquifer in northwestern Iran.

Recent advancements in the field have further expanded the scope of copula applications. For instance, Soltaninia and Eskandaripour (2025) proposed a trivariate framework that integrates copula theory with parametric and non-parametric marginals to model flood peak, volume, and duration under non-stationary climate conditions. Similarly, Mukherjee et al. (2025) employed multivariate copula models to assess compound flooding in six Indian river basins, analyzing joint dependencies among precipitation, runoff, and soil moisture and deriving conditional return periods.

Sun et al. (2026) constructed a D-vine copula model for small and medium rivers in arid regions, showing that trivariate joint return periods provide a more realistic representation of extreme flood risk. Additionally, a non-stationary framework integrating GAMLSS and copulas has been developed to address the non-stationarity of flood characteristics and to quantify associated uncertainties. These recent studies underscore the increasing recognition that multivariate approaches are essential for capturing the full complexity of flood behavior.

Given the above, and recognizing the need for more realistic statistical methods, the present study focuses on the multivariate analysis of flood characteristics using copulas. By preserving the entire dependence structure among variables, copulas enable the use of bivariate or higher-order joint and conditional distributions. In contrast to many previous studies that, due to methodological difficulties, restricted flood frequency analysis to univariate annual peak flows—thereby ignoring other relevant characteristics—this study incorporates three key random variables: peak flow, time base, and flood volume. This approach aims to achieve maximum reliability in flood frequency analysis by accounting for the full multivariate nature of flood events.

2. Materials and Methods

Figure 1 shows the location of Sepid Dasht-Zaz basin in Lorestan Province, Iran. This is geographically limited between of eastern 48 degrees and 40 minutes to 49 degrees 30 minutes and northern 32 degrees and 56 minutes to 33 degrees 17 minutes. Its total area is 687 km² and the average height is about 2322 m. The hydrometric station of this sub-basin is located in Sepid Dasht, at a distance of about 300 m above the confluence of Zaz and Cesar Rivers, in geographical coordinates of eastern 48 degrees and 52 minutes and northern 33 degrees and 12 minutes in height of 973 m from mean sea level.

This river is one of the important branches of Dez Basin and after crossing Cesar River and narrow and deep valleys, it enters Dez Dam Lake. In this study, in order to analyze the flood characteristics of Zaz Basin, the data of Sepid Dasht-Zaz hydrometric station was used. Out of 1000 floods recorded during 1967-2017, 30 floods were selected with adequate and appropriate data for analysis.

2.1. Extraction of flood characteristics

The most important characteristics of a flood include peak flow, flood volume and duration of floods. In order to determine the continuity of the flood, the start and end dates of the runoff should be specified. In general, flood time boundaries are shown by increasing the flow relative to the base flow (start of runoff) and reducing the flow and returning to base flow (end of runoff). Therefore, base flow is considered as a criterion for determining the flood hydrograph.

In the study of basins, the onset of runoff is usually indicated by the sudden increase in the ascending branch of the hydrograph and its end by the flattening of the descending branch of the hydrograph. In the descending branch of the hydrograph, while turning surface runoff to base flow, a significant change in the slope of the hydrograph occurs. The characteristics of peak flow, flood volume and duration of floods for the year *i* are determined by the following equations and shown in Fig. 2 (Karmakar and Simonovic 2009).

$$D_i = ED_i - SD_i \tag{1}$$

$$V_i = V_i^{Totoal} - V_i^{Base} = \sum_{j=SD_i}^{ED_i} Q_{ij} - \frac{1}{2}(Q_{is} - Q_{ie})(1 + D_i) \tag{2}$$

$$P_i = \max\left[\left(Q_{ij} - Q_{ij}^{Base}\right)\right] \quad j = ED_i, 1 + ED_i, 2 + ED_i \ldots. SD_i \tag{3}$$

where Q_{ij} is the observed river flow on day i of the year i, Q_{is} and Q_{ie} are the daily observational river flows on the start and end dates of the flood of the year i, respectively and Q_{ij}^{Base} is the amount of the basic flow on day j of the year i. Annual peak flow, flood volume and duration of floods are expressed as $P = \{P_i\}$, $V = \{V_i\}$ and $D = \{D_i\}$, respectively.

2.2. Selection of marginal distributions

In this study, for univariate frequency analysis and selection of the appropriate margin distribution for flood characteristics, log-normal (LN), normal (NOR), generalized extreme value (GEV), exponential (EXP), gamma (GAM), logistics (LOG), generalized logistics (GLOG), Weibull (WEI) and generalized Pareto (GPA) distributions were used. After selecting the best statistical distribution, flood characteristics were estimated for the given return period.

In the conventional approach of estimating the flood design by frequency analysis, estimating the distribution parameters is an important step.

When sample size is small or there is a possibility of outlier in the sample data, then the linear momentum method performs better than the prior methods, including the conventional momentum method and the maximum likelihood method (Hosking and Wallis 1988).

After fitting the univariate distributions to the flood characteristics, the goodness of fit test for each distribution has performed by Kolmogorov-Smirnov (KS) test. If at the significance level of 5%, the fit of the models is confirmed by the Kolmogorov-Smirnov test, the given probability distribution is accepted.

After statistical control of the goodness of fit and determination of acceptable distributions, Normalized Root Mean Square Error (NRMSE) (Eq. 4) and Nash-Sutcliffe efficiency coefficient (NSE) (Eq. 5) are calculated to choose the best fitted distribution.

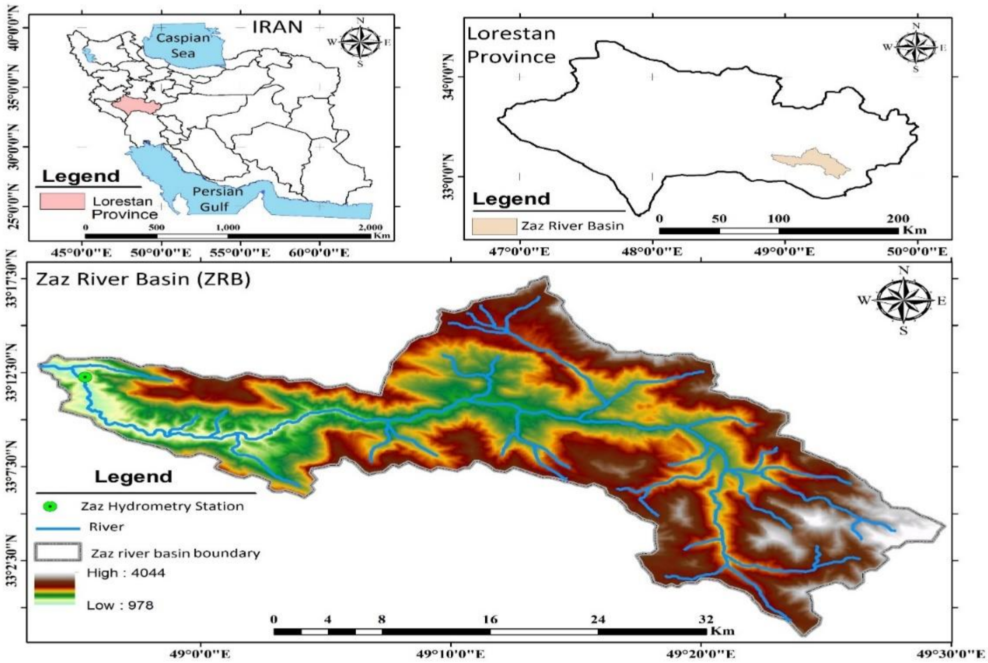

Fig. 1. Location of Zaz Basin in Lorestan Province, Iran

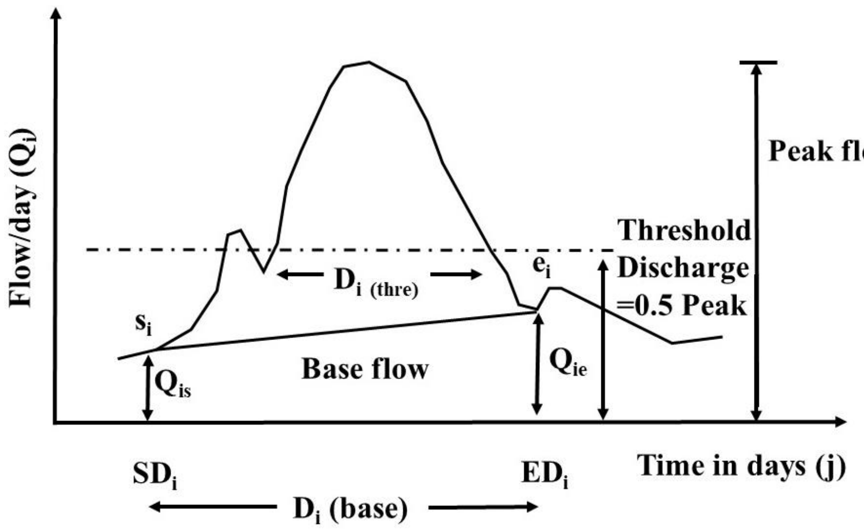

Fig. 2. Characteristics of the ith flood (Karmakar and Simonovic 2009)

Each distribution which had the highest NSE and the lowest NRMSE is selected as the most appropriate fitting distribution.

$$NRMSE = 100 \times \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(F(x_i) - G(x_i))^2}}{F_{\max} - F_{\min}} \quad (4)$$

$$NS = 1 - \frac{\sum_{i=1}^{n}(F(x_i) - G(x_i))^2}{\sum_{i=1}^{n}(F(x_i) - \overline{F})^2} \quad (5)$$

2.3. Copulas and Sklar theory

Copula is a flexible statistical tool to create multivariate distributions with different types of marginal distribution functions. A copula introduction is attributed to Sklar (1959) who described how multivariate distributions can be created by combining the univariate distribution functions. He explained that for d-dimensional continuous random variables $\{X_1,....,X_d\}$ with marginal CDFs $u_j = F_{X_j}(x_j)$, j=1,…,d there exists a d-dimensional copula $C_{U_1,\ldots,U_d}$ as bellow:

$$C_{U_1,\ldots,U_D}(U_1,\ldots,U_D) = H_{X_1,\ldots,X_D}(X_1,\ldots,X_D) \quad (6)$$

where, u_j is the j^{th} margin and $H_{X_1,\ldots,X_d}$ is the joint CDF of $\{X_1,\ldots,X_d\}$. Since the random variables are continuous, the CDF function of margins are non-decreasing from 0 to 1, $C_{U_1,\ldots,U_d}$ can be considered as a conversion $H_{X_1,\ldots,X_d}$ from $[-\infty,\infty]^d$ to $[0,1]^d$. The result of this conversion is that the marginal distributions are separated from $H_{X_1,\ldots,X_d}$ and, therefore, $C_{U_1,\ldots,U_d}$ relate only to the relationship between the variables and provide a complete description of the total dependency structure (Nelsen 2006; Saeidinia et al. 2026).

Copulas are divided into several families. Archimedean copulas are the most widely used functions for multivariate analysis of hydrological events and have explicit equations in their cumulative distribution functions, which has the advantage of using these functions over some functions, for example, elliptic copulas do not have an explicit cumulative distribution. In this study, three different copulas including Ali-Mikhail-Haq, Clayton and Frank were examined for multivariate flood analysis of Dez Basin. Table 1 shows the equations of the used copulas.

The first step of fitting and selecting the copula is to determine the dependency of the two studied variables. In this study, the dependency between flood characteristics (peak flow and flood volume and time base of flood) of Zaz Basin was calculated using Kendall's tau (Eq. 7).

$$\tau = \binom{N}{2}^{-1} \sum_{i<j} sign\left[\left(x_i - x_j\right)\left(y_i - y_j\right)\right] \quad (7)$$

where, N is the sample size, sign(.) is a function of the sign, and x and y are pairs of observational data for the flood characteristics (Nazeri Tahroudi et al. 2021b; Khashei-Siuki et al. 2021, Nazeri Tahroudi 2025; Rehamnia et al. 2026).

The second and very important step of the application of copulas is to estimate the parameter (θ). In order to estimate the parameter of copula dependency, several methods such as momentum, maximum likelihood, canonical maximum likelihood, meta-exploration (such as genetic algorithm), etc. have been developed by researchers, each of which has advantages and disadvantages. Among them, inference functions for margins (IFM) is the most common method for estimating parameters of copulas, which was presented by Joe (1997). IFM is computationally much more efficient than other methods. For the two variables, the assumption is that two random variables X and Y were correlated and follow the functions $f_X(x;\alpha_1,\alpha_2,\ldots,\alpha_p)$ and $f_Y(y;\lambda_1,\lambda_2,\ldots,\lambda_r)$ that $\alpha_1,\alpha_2,\ldots,\alpha_p$ are $f_X(x)$ parameters, and $\lambda_1,\lambda_2,\ldots,\lambda_r$ are $f_Y(y)$ parameters.

For *n* independent pair observations, the log likelihood function was maximized for X and Y i.e. $\ln L_X(x;\alpha_1,\alpha_2,\ldots,\alpha_p)$ and $\ln L_Y(y;\lambda_1,\lambda_2,\ldots,\lambda_r)$ are separately maximized for estimating the parameters. $\hat{\alpha}_1,\hat{\alpha}_2,\ldots,\hat{\alpha}_p$ and $\hat{\lambda}_1,\hat{\lambda}_2,\ldots,\hat{\lambda}_r$ are approximate parameters. The log likelihood function of the joint probability distribution function was considered as follows:

$$\begin{aligned} &Ln\,L(x,y;\hat{\alpha}_1,\hat{\alpha}_2,\ldots,\hat{\alpha}_p,\hat{\gamma}_1,\hat{\gamma}_2,\ldots,\hat{\gamma}_r,\theta) \\ &= Ln\,L_C(x,y;F_x(x),F_y(y),\theta) + \\ &Ln\,L_X(x;\hat{\alpha}_1,\hat{\alpha}_2,\ldots,\hat{\alpha}_p) + \\ &Ln\,L_Y(y;\hat{\gamma}_1,\hat{\gamma}_2,\ldots,\hat{\gamma}_r) \end{aligned} \quad (8)$$

where, $\ln L_C$ is log likelihood function of distribution function of copulas. By replacing the estimated values for $\hat{\gamma}_1,\hat{\gamma}_2,\ldots,\hat{\gamma}_r$ and $\hat{\alpha}_1,\hat{\alpha}_2,\ldots,\hat{\alpha}_p$ in Eq. 8, log likelihood function ln L is maximized to estimate parameter $\hat{\theta}$.

In order to select the most suitable copula, after selecting and fitting the appropriate marginal distribution on each of the studied variables and estimating the distribution parameters by the maximum likelihood method, three different copulas were considered to link these two marginal distributions and the IFM was applied to estimate the dependency parameter of copula (Joe 1997). Then, the results of each copula were compared with the corresponding values of the empirical copula to specify the best fitted copula. NSE, AIC and Cramer-von mises statistics were used to select the best copula.

$$NS = 1 - \frac{\sum_{i=1}^{n} \left(C_{pi} - C_{ei} \right)^2}{\sum_{i=1}^{n} \left(C_{ei} - \bar{C}_e \right)^2} \quad (9)$$

$$AIC = -2LnML + 2k \quad (10)$$

$$S_n = \sum_{i=1}^{n} \left(C_{pi} - C_{ei} \right)^2 \quad (11)$$

where, n is data length, C_p is the values of the theoretical copula, C_e denotes the observational values of the empirical copula, $\bar{C}_e$ is the average of the observational values of the empirical copula, $\ln ML$ is the maximum value of the log likelihood function and k is the number of fitted parameters. The copula was considered more appropriate where AIC and Sn values were the lowest and NSE value was closer to one (Nazeri Tahroudi et al. 2022; Raji et al. 2022).

In this study, nested copulas have been used for trivariate frequency analysis. The use of nested copulas makes it possible to investigate the effect of more than two variables on the phenomenon. Saad et al. (2015) used different combinations of Archimedean copulas to provide nested copulas and provided a range of suitable coefficients for this type of copulas. Figure 3 shows a demonstration of how the nested copula of the Archimedean family were made. As shown in Fig. 3, it is observed that first the primary copulas should be between the two variables and then the nested copula should be provided by combining the CDF of the primary copula and the third variable.

Table 1. The Archimedean Copulas used in this study (Nelsen 2006)

Copula name	Cumulative distribution function	Range of θ
Ali-Mikhail-Haq	$C(u,v) = \frac{uv}{1-\theta(1-u)(1-v)}$	$-1 \le \theta \le 1$
Clayton	$C(u,v) = \max[(u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}, 0]$	$\theta \ge 0$
Frank	$C(u,v) = \frac{-1}{\theta} Ln[1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1}$	$\theta \ne 0$

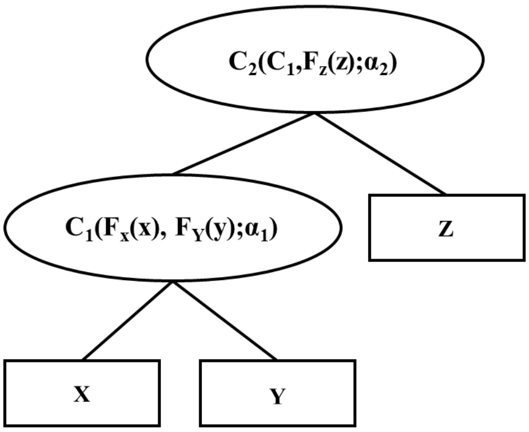

Fig. 3. A scheme of how the nested copula of the Archimedean family were made

2.4. Bivariate return period

Bivariate frequency analysis for two correlated random variables is defined by the joint return period. The joint return periods are defined in two ways as follows (Yue et al. 2001):

The first case is a return period in which the observed value exceeds x or y (i.e. $X>x$ or $Y>y$) indicated by T_{XY}. The second case is the joint return period in which the observed pair has exceeded both x and y (i.e. $X > x$ and $Y > y$) indicated by T'_{XY}. These two types of return periods were presented as copulas and calculated as follows:

$$T_{XY} = \frac{1}{P(X > x \, or Y > y)} = \frac{1}{1 - C\left(F_x(x), F_y(y)\right)} \quad (12)$$

$$T'_{XY} = \frac{1}{P(X > x\ and\ Y > y)} = \frac{1}{1 - F_x(x) - F_y(y) + C(F_x(x), F_y(y))} \quad (13)$$

2.5. Bivariate conditional return period

In the multivariate case, the return period can be expressed as $T_{X<x} = \frac{1}{P(X < x)}$ and x is the risk threshold for an X event. Therefore, the above-mentioned equation can be considered as follows:

$$T_{X<x} = \frac{1}{1 - P(X < x)} \quad (14)$$

Now, to determine the conditional return period by year, for events whose conditional thresholds are conditionally expressed, $P(X \le x)$ can be replaced by conditional joint distribution functions.

For example, to determine the return period of a bivariate event, provided that, using the copulas governing the variables, the following equation can be used.

$$T_{(X|Y)}(x|y) = \frac{1}{1 - C(u|v)} \quad (15)$$

$$C(u|v) = \frac{C(u|v)}{v} \quad (16)$$

where, $C(U|V)$ expresses events in which $U \le u$ provided that $V \le v$ (u and v are marginal distribution functions). Considering the interval of different conditional thresholds and using conditional joint distribution functions, it is possible to achieve a set of points with the same return period and select one of these events as the design event based on the criteria (Yue and Rasmussen 2002).

3. Results and Discussion

After extracting the flood characteristics (Table 2) of Zaz Basin, different marginal distributions were used to perform the frequency analysis of flood characteristics. To fit the probabilistic distributions to the characteristics of peak flow, flood volume, and time base of flood, first the linear momentum for the extraction series is calculated and then by equating them with the linear momentum of the probabilistic distributions, the coefficients of the given distributions were calculated. At first, the fitness of the used distributions should be statistically confirmed using the KS test. Then, using the RMSE and NSE criterion, the accuracy and efficiency of statistical distributions in modeling flood characteristics were investigated and the best distribution at each station was selected for frequency analysis.

The results of KS test are given in Table 3. According to Table 3, it can be seen that the fitness of the distributions used for frequency analysis of the flood characteristics of the Zaz Basin are confirmed at the significance level of 5%. After this step, the accuracy and performance of the probabilistic used distributions should be investigated. According to Table 3, it can be seen that at Zaz station, the values of NRMSE and NSE statistics for GEV distribution are the lowest and highest values, respectively, hence, it was selected as the best marginal distribution for frequency analysis of flood characteristics in Zaz station.

3.1. Selection of copulas

After determining the best marginal functions, the Archimedean copulas of Ali-Michail-Haq, Clayton and Frank were used to provide the joint distribution functions. First, the correlation between flood characteristics (peak flow, flood volume, and time base) at Zaz station were calculated using Kendall's tau and the results are presented in Table 4.

According to this Table, it can be seen that the extracted series have a good correlation and are suitable for multivariate analysis. At the next step, IFM was used to estimate the parameter of dependency. AIC, NSE and Sn statistics were used to choose the most suitable copula. Table 5 shows the results of the goodness of fit tests for each of the copulas. According to this Table, it can be seen that for the pair-variables of flood volume and base time of floods, Frank copula and for the pair-variables of flood volume and peak flow, Clayton copula and for the pair-variables of time base and peak flow of floods, Frank copula with the lowest (highest) values of AIC and Sn (NSE) have the best performance in estimating both probabilities and are selected as the appropriate copulas. Xiao et al. (2006) and Reddy and Ganguli (2012) also used Frank and Clayton for joint analysis of the flood characteristics and reported the results very well.

3.2. Provision of nested copulas

In order to study the nested copulas, different structures were considered as shown in Fig. 4. In this case, for the three characteristics of flood at the studied station, three different structures were investigated (Fig. 4). At the first step, the selected copulas in Table 5 were used. At the second step, copulas were investigated for the second structure.

The study results of the selected copulas at the second step are given in Table 6. As shown in Fig. 4, at each step, first two parameters were extracted using copulas and the distribution of the best margins, and the parameter c consisting of the studied pair parameter. At the next step, the third parameter is connected with CDF. In this way, nested copulas are obtained for all three states in Fig. 4.

Table 2. The flood characteristics for Zaz Basin during the statistical period 1967-2017

Row	From	To	Vol (m^3)	Tb (hr)	Qp (m^3/s)
1	1968/11/16	1968/11/17	1631880	32	32.3
2	1969/03/14	1969/03/15	6738480	34	138.8
3	1973/02/22	1973/02/23	2831760	34	58.4
4	1983/12/11	1983/12/12	1285992	34	26.9
5	1984/03/26	1984/03/27	2951280	46	39.7
6	1986/02/06	1986/02/07	1355040	44	27.4
7	1986/05/03	1986/05/03	20714400	80	288
8	1987/03/02	1987/03/05	15935040	90	129
9	1987/12/23	1987/12/24	4206240	38	52.5
10	1988/03/01	1988/03/05	6328440	57	84
11	1989/03/10	1989/03/11	1077120	32	20.3
12	1989/03/28	1989/03/29	1375200	30	29.2
13	1989/12/4	1989/12/5	1293840	30	34.8
14	1992/11/08	1992/11/09	1215360	38	23.2
15	1993/01/07	1993/01/09	7708320	48	128
16	1993/02/21	1993/02/25	1854000	36	39
17	1993/03/05	1993/03/04	3787920	38	58.4
18	1994/04/30	1994/05/01	2183040	42	44
19	1994/11/07	1994/11/10	22790160	92	257
20	1994/11/23	1994/11/27	30168000	80	342
21	1995/01/09	1995/01/10	400320	40	5.8
22	1999/01/28	1999/01/30	12852000	62	161
23	2001/11/18	2001/11/19	3456000	42	50
24	2001/12/20	2001/12/21	4849200	44	68
25	2003/04/02	2003/04/03	5053680	36	79.4
26	2004/01/03	2004/01/04	3908160	40	44.8
27	2004/11/23	2004/11/24	560160	40	8.9
28	2005/03/10	2005/03/15	39909600	78	309
29	2006/01/25	2006/01/29	6405840	110	34.3
30	2006/02/08	2006/02/14	37382400	104	353

Table 3. Values of NSE and NRMSE statistics for univariate distributions fitted to the flood characteristics of Zaz Basin

Probability Distribution	NSE			NRMSE		
	Vol	Tb	Qp	Vol	Tb	Qp
NOR	0.712	0.785	0.765	15.612	13.500	14.103
LN	0.978	0.868	0.970	4.315	10.564	5.002
EXP	0.861	0.526	0.938	10.849	20.036	7.276
GAM	0.918	0.837	0.919	8.333	11.772	8.267
GEV	**0.984**	**0.978**	**0.978**	**3.688**	**4.315**	**4.347**
LOG	0.818	0.848	0.852	12.427	11.338	11.192
GLOG	0.980	0.909	0.971	4.153	8.786	4.920
GPA	0.974	0.768	0.946	4.686	14.031	6.780
WEI	0.944	0.825	0.930	6.873	12.167	7.709

Table 4. Kendall's tau correlation coefficients for dependency between flood characteristics of Zaz Basin

Parameter	Vol	Tb	Qp
Vol	1.00	0.57	0.76
Tb	0.57	1.00	0.51
Qp	0.76	0.51	1.00

Table 5. Results of goodness of fit tests and copula parameters for flood characteristics of Zaz Basin (bold values are best)

Parameters	Statistics	Clyton	AMH	Frank
U (Vol) and V (Tb)	AIC	-9.016	-8.939	**-9.041**
	NS	0.959	0.775	**0.970**
	Sn	0.101	0.562	**0.074**
	teta	8.349	0.985	**12.523**
U (Vol) and V (Qp)	AIC	**-9.159**	-9.066	-9.158
	NS	**0.968**	0.761	0.962
	Sn	**0.096**	0.596	0.084
	teta	**11.068**	1.000	10.621
U (Tb) and V (Qp)	AIC	-9.066	-7.689	**-9.159**
	NS	0.961	0.596	**0.966**
	Sn	0.103	0.618	**0.086**
	teta	13.721	0.741	**16.712**

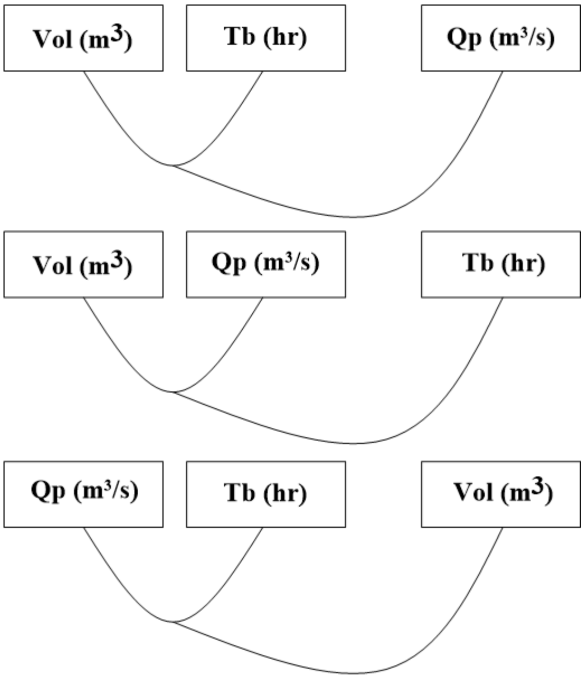

Fig. 4. Different states of investigations the nested copula structure (first step, two boxes on the left and second step: box on the right)

According to Table 6, it can be seen that in the case of pair variable selection (Vol, Tb) and (Qp), Clayton with a coefficient of 19.56 was selected as the best copula to estimate the joint return period. For the pair variable (Vol, Qp) and (Tb) as well as the pair variable (Tb, Qp) and (Vol), Frank and Clayton were selected as best copulas with coefficients of 20 and 4.80, respectively, according to the Bias, NSE and RMSE criteria. Finally, using selected copulas, the multivariate and joint return period of studied variables were presented in Figs. 5, 6, and 7.

Since one of the important objectives of the present study is the application of nested copulas for multivariate frequency analysis of flood characteristics at Zaz station, frequency analysis of flood characteristics at Zaz station was performed using copulas, which led to the presentation of typical curves for the study area.

Table 6. Results of goodness of fit tests and the copula parameters for flood characteristics of Zaz Basin at the second step (Bold values are the best choice)

Parameters	Statistics	Clyton	AMH	Frank
U (Vol, Tb) and V (QP)	Bias	**0.06**	0.18	0.15
	NS	**0.96**	0.70	0.89
	RMSE	**0.06**	0.14	0.06
	Teta	**19.56**	1.00	4.79
U (Vol, QP) and V (Tb)	Bias	0.07	0.14	**0.07**
	NS	0.07	0.70	**0.93**
	RMSE	0.08	0.16	**0.08**
	Teta	8.45	1.00	**20.00**
U (Tb, Qp) and V (Vol)	Bias	**0.07**	0.14	0.09
	NS	**0.93**	0.71	0.90
	RMSE	**0.08**	0.16	0.09
	Teta	**4.80**	1.00	20.00

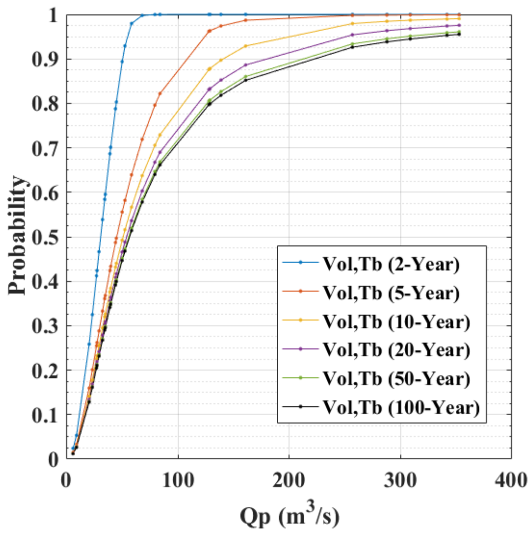

Fig. 5. Joint return period of peak flow of Zaz station according to the time base and flood volume in 2-year to 100-year return period.

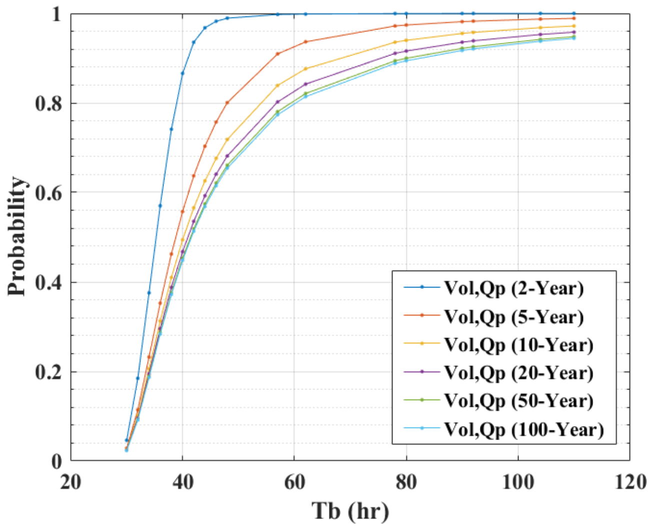

Fig. 6. Joint return period of flood time base given peak flow and flood volume in 2-year to 100-year return period.

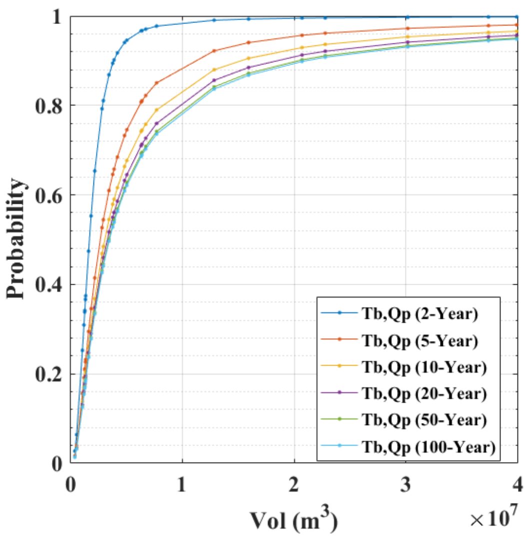

Fig. 7. Joint return period of flood volume in Zaz station according to the time base and flood peak flow in 2-year to 100-year return period

Figures 5 to 7 can be used as typical curves to manage the flood characteristics at Zaz station as well as the flood management. By selecting the appropriate nested copulas, trivariate conditional probabilities can be calculated, as in the bivariate case. In Figs. 5 to 7, two flood characteristics at the studied station were conditioned and the corresponding probability values were calculated for the third characteristic. This means that if the calculation of the return period of the flood volume event at Zaz station is expected, it is sufficient to determine the return period and the corresponding probability of time base and peak flow at the same station.

As shown in Fig. 5, it can be seen that in case of floods with a return period of 50 years at Zaz station, according to the time base and flood volume in this basin, peak flow with a probability of 90% will be equal to 200 cubic meters per second.

In the same way, different states can be investigated and possible climatic scenarios can be analyzed using similar curves and the most appropriate conditions for the designs can be selected. Since these curves are provided based on flood volume and time base, they will have significant accuracy compared to the univariate case. Figure 6 shows the different return periods of the time base of flood given by peak flow and flood volume.

As shown in Fig. 6, it can be seen that in case of a 50-year flood at the studied station, with an 85% probability, it will last for about 80 hours. With different probabilities, the time base of the flood in the study area can be estimated with different return periods. Figure 7 shows the return period and its different probabilities for flood volume values according to the time base and peak flow at Zaz station. Similarly, in a 50-year flood, the flood volume at this station will be 20 million cubic meters with a probability of about 90%. These curves can be used for flood frequency analysis for different structures at different levels.

It should be noted that the present analysis is based on 30 flood events at the study station. Although this number is common for applying copula methods in hydrological studies, the relatively small sample size may lead to uncertainty in the estimation of copula parameters and marginal distributions. Therefore, it is recommended that further analyses be conducted if longer-term data become available. Moreover, the presented results should be interpreted with caution, taking this limitation into account.

4. Conclusion

Floods and the resulting hazards have always led to the utmost caution for the design and operation of hydraulic structures. Due to the complex nature of the flood phenomenon, there is a need for methods that can provide a more comprehensive view by considering various factors. For this purpose, in this study, the nested copulas were used for joint analysis of the flood in Dez Basin. These copulas were implemented for the three parameters of flood volume, peak flow, and time base of floods at Zaz station. For each of the mentioned parameters, the copulas were investigated as pair variables. Using nested copulas in joint frequency analysis makes it possible to include more than 2 dimensions in studies.

The study results of the statistical distributions showed that GEV distribution has the best fit for all three variables of flood volume, peak flow, and time base of floods. In this study, three copulas of Ali-Mikhail-Haq, Frank and Clayton were used to provide nested copulas. The results of assessing the provided nested copulas showed that for the studied pair-variables (flood volume and time base of floods, flood volume and peak flow, and time base and peak flow), Frank, Clayton and Frank were selected as appropriate nested copulas to calculate the conditional probabilities, respectively.

By selecting the best copulas, the joint frequency analysis of each of the studied variables were investigated given by the other two variables. The results showed that considering the flood volume and return period of the studied floods, the 2-year return period of peak flow, which is not expected to occur every year, is equal to 50 m^3/s, which is 90% likely to occur. The 2-year return period of the time base of flood given by the peak flow and flood volume of the studied floods are considered, with a 90% probability, is estimated at about 42 hours, which is acceptable considering the average time base during the statistical period. The volume of floods with a return period of 2-year and given by pair-variables of time base and peak flow with a probability of 90%, was estimated about 5 million cubic meters. According to the curves, which are presented as the typical curve for the given station, the peak flow, time base and flood volume can be estimated with different probabilities and durations.

These curves are specific to the study station and provided according to the floods that occurred and are reliable as long as the joint distribution of these three parameters does not change. The results indicated that the nested copula had a good performance in the studied stations and the use of this method in any area or station can provide local curves that help to manage floods in the area.

5. Conflict of Interest

The authors declare no conflict of interest over this paper.

6. References

Ahmadi, F., Radmaneh, F., Parham, G., & Mirabbasi Najafabadi, R. (2017). Application of Archimedean Copula Functions in Flood Frequency Analysis (Case Study: Dez Basin). *Iranian Journal of Soil and Water Research*, *48*(3), 477-489. https://doi.org/10.22059/ijswr.2017.217805.667551

Balistrocchi, M., & Bacchi, B. (2017). Derivation of flood frequency curves through a bivariate rainfall distribution based on copula functions: application to an urban catchment in northern Italy's climate. *Hydrology Research*, *48*(3), 749-762. https://doi.org/10.2166/nh.2017.109

Dastourani, M., & Nazeri Tahroudi, M. (2022). Toward coupling of groundwater drawdown and pumping time in a constant discharge. *Applied Water Science*, *12*(4), 1-13. https://doi.org/10.1007/s13201-022-01606-6

Hosking, J. R. M., & Wallis, J. R. (1988). The effect of intersite dependence on regional flood frequency analysis. *Water Resources Research*, *24*(4), 588-600. https://doi.org/10.1029/WR024i004p00588

Joe, H. (1997). Multivariate models and multivariate dependence concepts. London: Chapman & Hall, 399 pp. ISBN: 0-412-07331-5

Karmakar, S., & Simonovic, S. P. (2009). Bivariate flood frequency analysis. Part 2: a copula-based approach with mixed marginal distributions. *Journal of Flood Risk Management,* *2*(1), 32-44. https://doi.org/10.1111/j.1753-318X.2009.01020.x

Khashei-Siuki, A., Shahidi, A., Ramezani, Y., & Nazeri Tahroudi, M. (2021). Simulation of potential evapotranspiration values based on vine copula. *Meteorological Applications*, *28*(5), e2027. https://doi.org/10.1002/met.2027

Ming, X., Xu, W., Li, Y., Du, J., Liu, B., & Shi, P. (2015). Quantitative multi-hazard risk assessment with vulnerability surface and hazard joint return period. *Stochastic environmental research and risk assessment*, *29*(1), 35-44. https://doi.org/10.1007/s00477-014-0935-y

Mukherjee, A., Poonia, V., & Swarnkar, S. (2025). Assessing compound flood drivers in Peninsular India: Multivariate copula-based approach. *Journal of Environmental Management*, *395*, 127875. https://doi.org/10.1016/j.jenvman.2025.127875

Nazeri Tahroudi, M. (2025). Comprehensive global assessment of precipitation trend and pattern variability considering their distribution dynamics. *Scientific Reports*, *15*(1), 22458. https://doi.org/10.1038/s41598-025-06050-5

Nazeri Tahroudi, M., Ramezani, Y., De Michele, C., & Mirabbasi, R. (2022). Trivariate joint frequency analysis of water resources deficiency signatures using vine copulas. *Applied Water Science*, *12*(4), 67. https://doi.org/10.1007/s13201-022-01589-4

Nazeri Tahroudi, M., Ramezani, Y., De Michele, C., & Mirabbasi, R. (2021a). Flood routing via a copula-based approach. *Hydrology Research*, *52*(6), 1294-1308. https://doi.org/10.2166/nh.2021.008

Nazeri Tahroudi, M., Ramezani, Y., De Michele, C., & Mirabbasi, R. (2021b). Multivariate analysis of rainfall and its deficiency signatures using vine copulas. *International Journal of Climatology*. *42*(4): 2005-2018. https//:doi.org/10.1002/joc.7349

Nelsen. R. B. (2006). An introduction to copulas. Springer, New York, 269p. https://doi.org/10.1007/0-387-28678-0_5

Raji, M., Tahroudi, M. N., Ye, F., & Dutta, J. (2022). Prediction of heterogeneous Fenton process in treatment of melanoidin-containing wastewater using data-based models. *Journal of Environmental Management*, *307*, 114518. https://doi.org/10.1016/j.jenvman.2022.114518

Reddy, M. J., & Ganguli, P. (2012). Bivariate flood frequency analysis of Upper Godavari River flows using Archimedean copulas. *Water Resources Management,* *26*(14), 3995-4018. https://doi.org/10.1007/s11269-012-0124-z

Rehamnia, I., Nazeri Tahroudi, M. & Saeidinia, M. Evaluating the role of dimensionality and complexity structure in time series models for precipitation simulation. Model. Earth Syst.

Environ. 12, 4 (2026). https://doi.org/10.1007/s40808-025-02649-9

Saad, C., El Adlouni, S., St-Hilaire, A., & Gachon, P. (2015). A nested multivariate copula approach to hydrometeorological simulations of spring floods: the case of the Richelieu River (Québec, Canada) record flood. *Stochastic Environmental Research and Risk Assessment*, *29*(1), 275-294. https://doi.org/10.1007/s00477-014-0971-7

Saeidinia, M., Haghiabi, A. H., Nazeri Tahroudi, M., Nasrolahi, A., & De Michele, C. (2025). Deep learning and vine copula-based sequencing: Approaches under investigation for forecasting soil temperature dynamics. *Stochastic Environmental Research and Risk Assessment*, *39*(11), 4949-4983. https://doi.org/10.1007/s00477-025-03061-6

Salvadori, G., & De Michele, C. (2007). On the use of copulas in hydrology: theory and practice. *Journal of Hydrologic Engineering*, *12*(4), 369-380. https://doi.org/10.1061/(ASCE)1084-0699(2007)12:4(369)

Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges. In *Annales de l'ISUP, 8,* 229-231.

Snyder, W. M. (1962). Some possibilities for multivariate analysis in hydrologic studies. *Journal of geophysical research*, *67*(2), 721-729. https://doi.org/10.1029/JZ067i002p00721

Soltaninia, S., & Eskandaripour, M. (2025). Trivariate flood frequency analysis using copula functions: Assessing non-stationary hydrological risks under climate change. *Journal of Water and Climate Change*, *16*(10), 3130–3156. https://doi.org/10.2166/wcc.2025.068

Sraj, M., Bezak, N., & Brilly, M. (2015). Bivariate flood frequency analysis using the copula function: a case study of the Litija station on the Sava River. *Hydrological Processes, 29*(2), 225-238. https://doi.org/10.1002/hyp.10145

Sun, B., Liu, X., Wang, G., Miao, P., Xie, K., & Ma, H. (2026). Multivariate Joint Risk Assessment of Small- and Medium-Sized River Flood in Arid and Semi-Arid Regions Based on Vine Copula. *Water*, *18*(9), 1098. https://doi.org/10.3390/w18091098

Wong, S. T. (1963). A multivariate statistical model for predicting mean annual flood in new england1. *Annals of the Association of American Geographers*, *53*(3), 298-311. https://doi.org/10.1111/j.1467-8306.1963.tb00451.x

Xiao, Y., Guo, S., Liu, P., & Fang, B. (2008). A new design flood hydrograph method based on bivariate joint distribution. *IAHS Publications-Series of Proceedings and Reports*, *319*, 75-82. https://doi.org/10.5555/20093172552

Yue, S., & Rasmussen, P. (2002). Bivariate frequency analysis: discussion of some useful concepts in hydrological application. *Hydrological Processes*, *16*(14), 2881-2898. https://doi.org/10.1002/hyp.1185

Yue, S., Ouarda, T. B., & Bobée, B. (2001). A review of bivariate gamma distributions for hydrological application. *Journal of Hydrology*, *246*(1-4), 1-18. https://doi.org/10.1016/S0022-1694(01)00374-2

Zhang, L. S. V. P., & Singh, V. P. (2006). Bivariate flood frequency analysis using the copula method. *Journal of hydrologic engineering*, *11*(2), 150-164. https://doi.org/10.1061/(ASCE)1084-0699(2006)11:2(150)