Comparison Between Bivariate and Trivariate Flood Frequency Analysis Using the Archimedean Copula Functions, A Case Study of the Karun River in Iran

Mohamad Haytham Klaho
Isfahan University of Technology
Hamid R. Safavi (hasafavi@iut.ac.ir)
Isfahan University of Technology https://orcid.org/0000-0002-9233-9209
Mohamad H. Golmohammadi
Isfahan University of Technology
Maamoun Alkntar
Isfahan University of Technology

Research Article

Keywords: Flood frequency, Copula functions, Archimedean, Hierarchical, Return period

Posted Date: May 28th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-524239/v1

Version of Record: A version of this preprint was published at Natural Hazards on February 12th, 2022.
See the published version at https://doi.org/10.1007/s11069-022-05240-y.

Comparison Between Bivariate and Trivariate Flood Frequency Analysis Using the Archimedean Copula Functions, A Case Study of the Karun River in Iran

Mohamad Haytham Klaho, MSc. Graduated, Dept. of Civil Engineering, Isfahan University of Technology, 8415683111, Isfahan, Iran. E-mail: eng.h.klaho@cv.iut.ac.ir

Hamid R. Safavi (corresponding author), Prof., Dept. of Civil Engineering, Isfahan University of Technology, 8415683111, Isfahan, Iran. Tel: ++983133913826, E-mail: hasafavi@iut.ac.ir

Mohammad H. Golmohammadi, Assist. Prof., Dept. of Civil Engineering, Isfahan University of Technology, 8415683111, Isfahan, Iran. Tel: ++983133913853, E-mail: m.golmohammadi@iut.ac.ir

Maamoun Alkntar, PhD Student, Dept. of Civil Engineering, Isfahan University of Technology, 8415683111, Isfahan, Iran. E-mail: m.alkntar@cv.iut.ac.ir

Abstract

Historically, severe floods have caused great human and financial losses. Therefore, the flood frequency analysis based on the flood multiple variables including flood peak, volume and duration poses more motivation for hydrologists to study. In this paper, the bivariate and trivariate flood frequency analysis and modeling using Archimedean copula functions is focused. For this purpose, the annual flood data over a 55-year historical period recorded at the Dez Dam hydrometric station were used. The results showed that based on goodness of fit criteria, the Frank function built upon the couple of the flood peak-volume and the couple of the flood peak-duration as well as the Clayton function built upon the flood volume-duration were identified to be the best copula families to be adopted. The trivariate analysis was conducted and the Clayton family was chosen as the best copula function. Thereafter, the common and conditional cumulative probability distribution functions were built and analyzed to determine the periodic "and", "or" and "conditional" bivariate and trivariate flood return periods. The results suggest that the bivariate conditional return period obtained for short-term periods is more reliable than the trivariate conditional return period. Additionally, the trivariate conditional return period calculated for long-term periods is more reliable than the bivariate conditional return period.

Keywords: Flood frequency, Copula functions, Archimedean, Hierarchical, Return period.

Introduction

Flood is one of the most important natural disasters annually causing many financial and human losses in different parts of the world, thus, its analysis and forecasting are required to control its possibly acute damages. In the literature, Civil and Ashkar (1980), Silverman (1986), Correia (1987), Gringorten (1963), Sackl and Bergmann (1987), all provided flood peak analysis with limited assessment of flood events. While the study of many hydrological events requires thorough knowledge of flood event (flood peak, flood volume, flood duration and hydrograph shape, etc.), only a handful of researchers have attempted to address this issue. To predict floods in a certain area, it is necessary to understand the three important characteristics of flood namely flood peak, flood volume and flood duration, to perform a joint analysis of the flood data occurring in the past in the area. One of the traditional methods for multivariate flood frequency analysis is the use of classical multivariate distribution functions, such as normal, log normal, Gamma, etc..; however the main problem hindering employing these methods is that these functions face limitations that either reduce the accuracy of the analysis or make it essentially impossible to be conducted. One of the most important limitations to use these functions is the need to specify the parameters of the distribution functions of marginal variables and their uniformity. As a result, when encountering the above limitations, a method for multivariate analysis should replace the classical methods. One of the most appropriate methods available is the use of a special class of multivariate probability functions called Copula Functions In recent years, scientists have used copula functions to analyze flood frequencies. Copula functions were first introduced by Sklar (1959). Different copula functions used in different sciences Joe (1997), Nelsen (2006). Yue et al. (1999) showed that a model for direct use of Gamble's bivariate limit distribution is suitable for analyzing the joint distribution of two random variables. The Application of modeling copula functions in hydrology and environmental modeling was first triggered by De Michele and Salvadori (2003). Favre et al. (2004) concluded that the copula function approach allows one to model dependent structures independently of marginal distributions, while this modeling is impossible when using classical standard distributions. The results show that the use of copula functions is more appropriate because it allows one to consider a wide range of correlations possibly occurring in hydrology. De Michele et al. (2005) used the Archimedean bivariate copula functions to simulate flood peak and flood volume to be able of constructing the artificial flood hydrographs. Modeling joint distributions using copula functions mitigates the intrinsic limitations in the flood frequency analysis by selecting different marginal distributions of the flood characteristics. In general, the copula functions can better fit the joint probability distribution of the certain data to occur to the experimental data really occurring in the nature, as shown by Zhang and Singh (2007). Genest et al. (2007) presented the steps required to construct a copula function model for hydrological purposes and examined the performance of copula function models for modeling the flood peak dependent on the flood volume. Klein et al. (2010) used a method for multivariate probability analysis of flood variables using copula functions to cope with the overestimation of hydrological risk usually caused by performing the univariate probability analysis. Salvadori and De Michele (2010) expressed that the multivariate value-based models are essential tools for evaluation of the potentially dangerous accidents as the target of this research outline how exploiting recent theoretical developments in the theory of copula can easily construct the new multivariate extreme value distributions. Due to the many storms that occurred in Taiwan, (Shiau et al. 2010), concluded that the single univariate analysis could not show a significant relationship between correlated variables. Therefore, this study uses copula functions to construct a common distribution of depth and precipitation duration for storm data. Using Copula to construct a multivariate distribution means that the effects of marginal variables can be separated from dependent variables. They derived the depth-duration-frequency (DDF) formula based on the use of copulas to show the common distribution of depth and duration of precipitation. Placket was chosen to construct the DDF curves. DDF allows rain depth to be estimated for a specific duration of rainfall and return period. This DDF formula improves the understanding of complex hydrological processes and increases the design safety standard of hydraulic structures. Based on their research, an interesting feature of Copula models is that any distribution can be used to display a marginal distribution. Copula functions have been used in various problems in water management and hydrology such as drought and flood frequency analysis. In their study, concluded that between families of Copula, Gumbel functions, there is more correlation for large values and less correlation for small values, whereas the opposite is true for Clayton Copula. Copula Frank has a lot of connection in the middle and little in the bottom.

Volpi and Fiori (2012) stated that in hydrological design and flood management, joint distribution of flood peak and flood volume, on the one hand and that of the flood volume and flood duration, on the other hand, are of particular importance. Therefore, many studies have been conducted to perform multivariate flood frequency analysis considering the relationship between flood variables including flood peak, flood volume and flood duration with restrictive assumptions. In the study of Nashwan et al. (2018), bivariate frequency analysis of the flood in different stations of Kelantan river basin was performed using copula functions to assess the geographical distribution of flood risk. The joint dependent structures of flood variables were modeled using the Gamble copula function. The results showed that different variables are corresponding to different distributions. Also, the correlation analysis between the variables showed a strong relationship between them. The joint distribution functions of flood peak and volume, flood peak and duration, and flood volume and duration showed that the joint return period was much longer than the univariate return period. According to the research Li et al. (2020), copula functions are very useful in flood frequency analysis and can be used for the make the measures necessary to achieve optimal water resources planning and management. Archimedean bivariate functions may not be generalized to the multivariate functions unless additional conditions are imposed on them to construct the Archimedean multivariate types of functions. Accordingly, flood analysis in three variables is more extensively used and the results are more accurate. As a result, the analysis of flood frequency should be carried out upon three variables. The main goal of this study is to investigate the application of the strong Archimedean copula functions in trivariate flood frequency analysis. The relationship between flood variables and frequency analysis of two and three flood variables including flood peak, flood volume and flood duration, were established. The marginal values of these variables and the best family of copula functions were selected to be used in estimating the conditional cumulative distribution as well as the combined return periods.

Study Area and Data

Study area and geographical location

Dez Dam Lake is located in the geographical location N323800 E482746 in Khuzestan province (Fig. 1). The lake is located in the northwest of Dezful, 23 km north-east of Andimashk and behind the two mountains Shaydab and Tenguan, and is included in the six provinces of Isfahan, Khuzestan, Lorestan, Markazi, Hamedan and Chaharmahal Bakhtiari. The basic information required for the study area includes the daily inflow of the dam over the last 55 years (from 24/9/1963 to 28/8/2018) which is obtained from the hydrometric station located at the inlet of Dez dam.

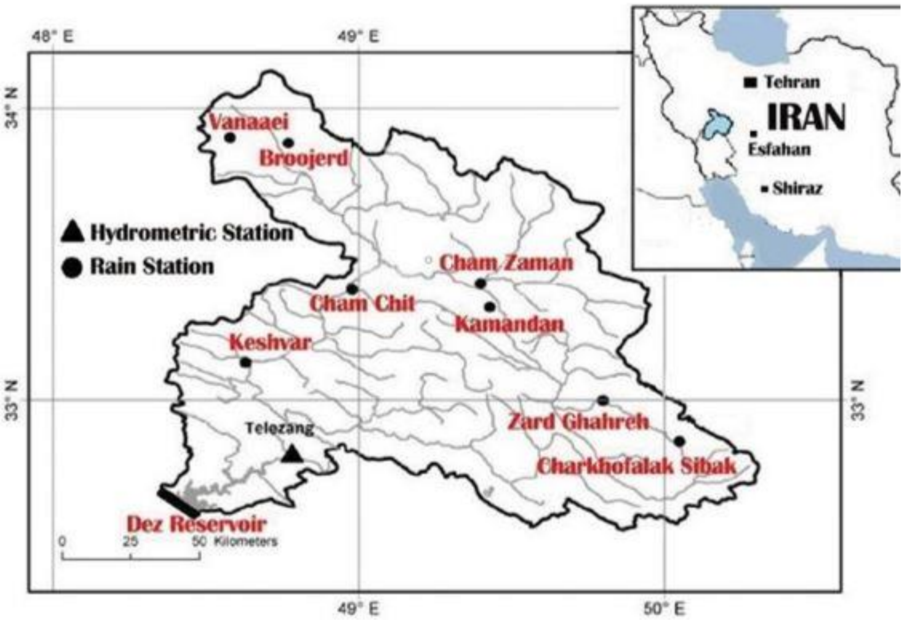


Fig.1 Location of Dez dam catchment and study area

The importance of Dez Dam

The dam irrigates 125,000 hectares of downstream land and has played an important role in controlling floods created upstream, especially in the last five years. The dam has a capacity of 520 megawatts and a final capacity of 3.3 billion cubic meters of water (Felfelani, Movahed, & Zarghami, 2013).

Materials and Methods

Archimedean Copula Functions

Archimedean functions have very important characteristics and play an effective role in hydrological works Salvadori and De Michele (2007).These functions are easily constructed and have suitable properties making these functions be widely used in hydrological analysis Nelsen

(2006). The main reasons why the Archimedean functions are so applicable in the field of hydrological sciences may be listed as follows.

1- Simplicity of making the members of this class.

2- Many families of copula functions belong to this category.

3- This category has many desirable properties.

The multivariate Archimedean copula functions are defined as follows (Nelsen, 2006):

$$C_{\theta(u,v)} = \Phi^{-1}[\Phi(u) + \Phi(v)] \quad (1)$$

In this equation, $\boldsymbol{\theta}$ is a tunable parameter used to form $\boldsymbol{\Phi}$; $\boldsymbol{\Phi}$ is the generating function having a continuous, convex and strictly uniform shape; *u* and *v* are the functions of the marginal cumulative distribution of the studied variables whose probability density function is uniform, expressed as U (0,1).

In this study, five functions of Clayton, Frank, Gamble, Ali Michael-Haq and Joe belonging to the Archimedean family were used for bivariate and trivariate flood analyses. Table 1 describes the Archimedean functions and the relevant relationships. In Table 1, *d* is the number of variables, for each of which a cumulative probability distribution function is already defined.

Table.1. Multivariate Archimedean family

Name	Copula function	Generating function $\Phi(t)$	Parameter θ
Clayton	$(\sum_{i=1}^{d} u_i^{-\theta} - d + 1)^{-1/\theta}$	$\frac{1}{\theta}(t^{-\theta} - 1)$	$(0, \infty)$
Ali–Mikhail–Haq	$\frac{\prod_{i=1}^{d} ui}{1 - \theta \prod_{i=1}^{d}(1 - ui)}$	$\ln(\frac{1 - \theta(1 - t)}{t})$	$(-1,1)$
Gumbel–Hougaard	$\exp(-(\sum_{i=1}^{d}(-lnui)^{\theta})^{1/\theta})$	$(-lnt)^{\theta}$	$(1, \infty)$
Frank	$-1/\theta \ln(1 + \frac{\prod_{i=1}^{d}(e^{-\theta ui} - 1)}{(e^{-\theta} - 1)^{d-1}})$	$-ln\frac{e^{-\theta t} - 1}{e^{-\theta} - 1}$	$(0, \infty)$
Joe	$1 - [1 - \prod_{i=1}^{d}(1 - (1 - ui)^{\theta})]^{1/\theta}$	$-\ln[1 - (1 - t)^{\theta}]$	$(1, +\infty)$

Extract parameters (peak, volume, and duration of a flood event)

A portion of the river discharge is coming from the previous runoff, and to obtain the newly generated flood hydrograph, firstly, it is necessary to subtract the previous river discharge, called the base discharge, from the total existing runoff. This process is called the separation of the hydrograph. The simplest way to separate the base flow hydrograph is drawing a line in the flood hydrograph from the point where the flood begins (A) to the point where the flood ends (B). The part of the hydrograph placed above the line AB denotes the direct runoff hydrograph. Fig. 2 illustrates the duration of a flood occurrence by identifying the start and end time of the flood runoff. The flood volume can be calculated using the Eq. (2) (Yue et al., 1999):

$$V_i = \sum_{j=SDi}^{EDi} q_{ij} - \frac{1}{2}(q_{is} - q_{ie}) \quad (2)$$

where q_{ij} is the jth day observed daily streamflow value for the ith year; q_{is} and q_{ie} are observed daily streamflow values on the start date and end date of flood runoff for the ith year, respectively.

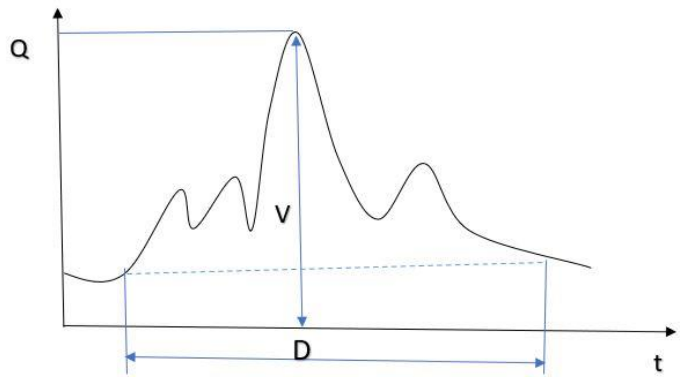


Fig.2. Determining flood characteristics

Correlation Coefficients

Correlation coefficients are the mathematical indicators showing the direction and value between two variables as the observed and computed output of the multivariate distributions. To measure the correlation, various coefficients are commonly used, including Pearson (r), Kendall's tau (τ) and Spearman (ρ) correlation coefficients. In this study, all three mentioned correlation coefficients have been used to examine the correlation between the variables and can be estimated as follows (She & Xia, 2018):

$$r = \frac{\sum_{i=1}^{n}(Xi-\bar{X})(Y-\bar{Y})}{(n-1)\sqrt{Sx^2Sy^2}} \quad (3)$$

$$\tau = \frac{2}{n*(n-1)}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} sgn\big((xi-xj)(yi-yj)\big) \quad (4)$$

$$\rho = 1 - \frac{6*\sum d^2}{n*(n^2-1)} \quad (5)$$

Where $\overline{X}$ and $\overline{Y}$ are the mean of the data (X, Y), Sx^2 and Sy^2 are the variance of the data. n is the number of data and $sgn(x)$ is the signal function. To calculate the Spearman coefficient, all X and Y data are first ranked in terms of their values. Then, the difference between the elements of the same pairs is denoted by d and the number of the elements included in the two data sets is denoted by n.

Empirical Trivariate Probability

The Gringorton experimental equation is one of the equations presenting the Empirical cumulative distribution functions that is commonly used and derived from position mapping relations. The probability of non-experimental encapsulation of the cumulative probability of Gringorton for the three variables is obtained using the Eq. (6) (Zhang & Singh, 2006):

$$P(X \leq x, Y \leq y, Z \leq z) = \frac{\sum_{m=1}^{i}\sum_{l=1}^{i}\sum_{c=1}^{i} n_m - 0.44}{N + 0.12} \tag{6}$$

where n_m is the number of three variables (Xi, Yi, Zi) provided that $Xj < Xi$, $Yj < Yi$ and $Zj < Zi$ and N is sample size.

Estimation of the Copula Parameters

Estimation of the parameters of the copula functions can be done utilizing various methods. To select the best copula function, the best form of correlation relation between the parameters must be obtained. The steps governing the selection of a copula function are generally presented by either of these two methods: (1) Kendall correlation method, and (2) Likelihood method. In this research, the Likelihood method has been used to estimate the model parameters. The basis of this method is to look for the best value of a probability distribution parameter, which should be the value maximizing the likelihood or probability of the observed sample to occur. For simplicity, the logarithm of the likelihood function is used instead of the likelihood itself. This summation of the natural logarithms illustrated in Eq. (7) should be maximized to cause the multiplication of the likelihoods to go to the value 1 as the maximum correlation there may be between the marginal cumulative probability distribution function values.

$$Lnl = \sum_{i=1}^{n} Ln[C(u, v)i] \tag{7}$$

where the maximum value of Lnl appears when at (∂ (Lnl) / $\partial\theta$ = 0.

Goodness of Fit

The purpose of evaluating the goodness of fit of a couple of data sets is to select the most appropriate and best copula function that shows the structure of the dependence between the variables and the behavior of the copula functions well. There are graphical tools and numerical tests to achieve this goal, including Akaike, Root-Mean-Square Error, Nash-Sutcliffe, Max Likelihood and Q_Q plot graph. The mentioned indicators are obtained from the Eqs. (8-11):

$$AIC = N \log(MSE) + 2(P) \tag{8}$$

$$RMSE = \sqrt{\frac{1}{n}} \sum_{i=1}^{n} (Ai - Pi)^2 \tag{9}$$

$$NSE = 1 - \frac{\sum_{i=1}^{n} (Q_m^i - Q_0^i)^2}{\sum_{i=1}^{n} (Q_0^i - \overline{Q_0})^2} \tag{10}$$

$$l(\theta) = \sum_{i=1}^{n} \log[c_\theta(F(Xi), F(Yi))] \tag{11}$$

where N is the number of observations; MSE is the mean squared error and P is the number of fitted parameters; Ai represents observed values; Pi represents the computed values; Q_m^i represents the computed values; Q_0^i expresses the observed values; and $\overline{Q_0}$ expresses the average of the observed values. It is worth mentioning that the lower the values of the AIC and RMSE, the better the copula functions are fitted to the data and the higher the values of the NASH and Likelihood, the better the copula functions are fitted to the data.

Joint and Conditional Cumulative Probability Distribution Functions and return period

After selecting the best copula function for the two- and three-variable modes, the two- and three-variable conditional cumulative distribution functions and return periods for the different modes are obtained. Joint probability distribution is the probability of both the peak flow of the flood and the volume of the flood would be above a certain threshold. Analysis of flood parameters is an important factor in the management of this phenomenon, helping the managers and planners achieve a better understanding of this phenomenon and improve their decision-making process and planning thanks to this understanding. Also, the conditional cumulative distribution function of the three variables $X \leq x$ and $Y \leq y$ for different values of $Z = z$ can be written as follows.

$$F(x, y|\mathrm{Z} \leq \mathrm{z}) = \frac{F_{x,y,z}(x,y,z)}{F_z(z)} \tag{12}$$

In general, the calculation and estimation of the return period is based on the statistical measurement of the historical data and is beneficial to obtain the average repetition time of a phenomenon in a time period and is used to analyze the risk of a phenomenon such as a flood event. The return period of an event upon emergence of either of the three conditions, which here we express it as "or" relation between these conditions, is the return period of the intersection $T_{x,y,z}^{\cup}$ of the three variables representing the events where $(X \geq x \; or \; Y \geq y \; or \; Z \geq z)$ and is defined as follows.

$$T_{x,y,z}^{\cup} = T_{X,y,z}^{or} = \frac{1}{F'(X \geq x \cup Y \geq y \cup Z \geq z)} = \frac{1}{1 - C(x,y,z)} \quad (13)$$

Results and Discussion

Determining the Correlation Coefficients of the Parameters

The correlation coefficients between the variables were calculated and according to the results shown in Table 2, there is a significant and direct correlation between the flood peak and volume and between the flood volume and duration while there is an inverse correlation between flood peak and duration. Furthermore, the correlation between the flood volume and duration is stronger than the correlation between the flood peak and volume as well as the correlation between flood peak and duration.

Table 2 Values of correlation coefficients between parameters

Correlation coefficients	Flood peak and volume	Flood peak and duration	Volume and duration
Kendall's tau (τ)	0.204	-0.2459	0.5106
Spearman (ρ)	0.3047	-0.3476	0.688
Pearson(r)	0.2749	-0.3177	0.6575

Univariate fittings

The selection of marginal distribution functions is done based on Anderson Darling test and Kolmogorov-Smirnov test. The flood peak, volume and duration data were fitted and the results showed that the best probability distribution function on the flood peak data is the log normal distribution function. The gamma distribution function was generalized to the volume data and the Generalized Extreme Value function was generalized to the duration data. Based on this, the parameters of log normal, gamma and Generalized Extreme Value distribution functions are given in Table 3.

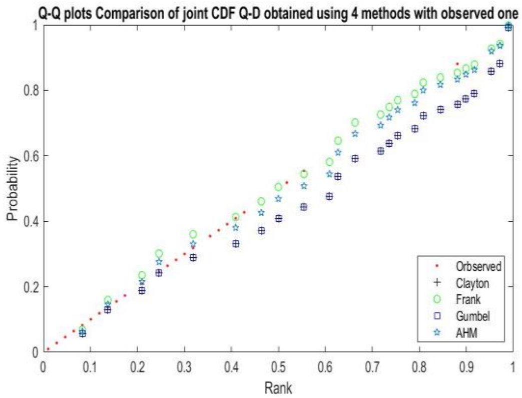
Table 3 Parameters of selected probability distribution functions

Variable	Selected distribution function	Parameters of selected probability distribution functions				
		μ	σ	α	β	k
Peak flood	Log normal	7.478.2	0.576188	----	----	----
Volume	Gamma	----	----	2.00274	12843	----
Duration	Generalized Extreme Value	63.6923	35.449	----	----	0.0107922

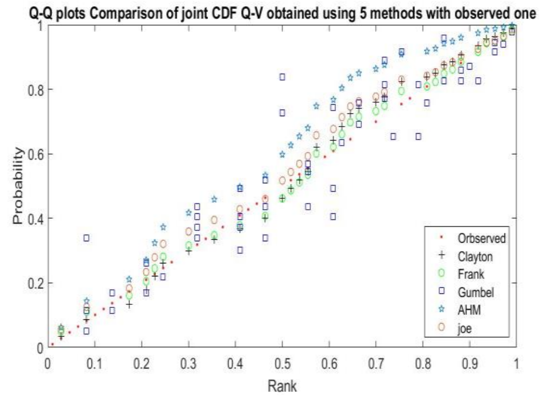
Estimation Parameters and Goodness of Fit

Since the main purpose of this research is to model two and three flood variables, the parameters of five family of copula functions including (Clayton, Frank, Gamble, Ali Mikhail -Haq and Joe) were calculated. How to estimate the parameters of these functions are fully described in Section 3.5. The values estimated for the parameters for the bivariate model through the Maximum Likelihood Method with an approximate 95% confidence interval can be seen in Tables 4, 5 and 6. Also, the Q_Q graphs are shown in Fig. 3 (a, b, and c). Also, the assumed range of the parameter

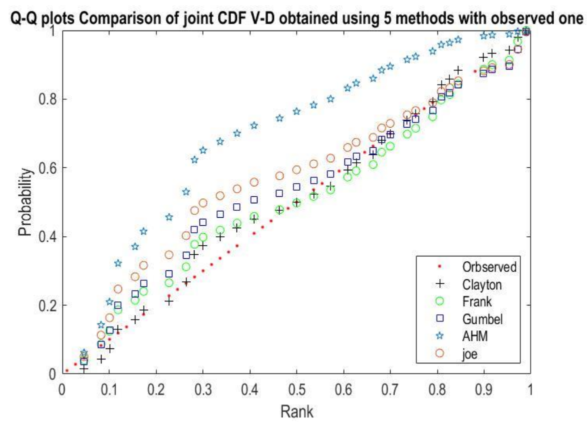
$\boldsymbol{\theta}$ is given in these tables. Since the parameter $\boldsymbol{\theta}$ of the Joe function coupling the probability of the flood peak and flood duration was outside of previously adopted range of the parameter, the modeling of the Joe function for this condition was avoided. The final selection of the copula function was done upon evaluating the performance of every copula function generated with respect to the measures described in section 3.6. As implicitly mentioned before, the lower the AIC and RMSE; the higher the estimation accuracy; and the higher the NSE criterion; the more accurate the model. Regarding the Maximum Likelihood criterion, the higher the value of this measure, the more acceptable it is. Having compared the performance of the different copula functions when applied to fitting a certain couple of the marginal distributions, the best bivariate copula between flood peak and volume and between flood peak and duration was revealed to be the Frank family, while the Clayton family was identified as the best bivariate copula between the flood volume and duration.



(a)



(b)



(c)

Fig. 3 Curve Q-Q plot. **(a)** between flood peak and flood volume. **(b)** between flood peak and flood duration. **(c)** between the volume and duration of the flood

Table 4 Goodness of fit test results of fitting the bivariate copula functions between flood peak and volume

Copula function	parameter	Max log likelihood	AIC	RMSE	NSE
Clayton	0.35	200.02	-40	0.195	0.988
Frank	**1.72**	**208.2**	**-41.64**	**0.168**	**0.991**
Gamble	1.15	200.84	-40.17	0.192	0.989
Ali Mikhail Haq	-0.77	76.8	-15.36	1.835	-0.021
Joe	1.15	191.17	-38.23	0.229	0.984

Table 5 Goodness of fit test results of fitting the two-variable copula functions between flood peak and duration

Copula function	parameter	Max log likelihood	AIC	RMSE	NSE
Clayton	0	165.89	-33.18	0.363	0.92
Frank	**-2.14**	**182.37**	**-36.47**	**0.269**	**0.956**
Gamble	1	165.89	-33.18	0.363	0.92
Ali Mikhail Haq	-0.96	67.65	-13.53	2.168	-1.875
Joe	-----	----	-----	----	----

Table 6. Goodness of fit test results of fitting the two-variable copula functions between flood volume and duration

Copula function	parameter	Max log likelihood	AIC	RMSE	NSE
Clayton	**1.76**	**184.63**	**-36.93**	**0.258**	**0.981**
Frank	5.58	179.15	-35.83	0.285	0.976
Gamble	1.76	170.67	-34.13	0.333	0.968
Ali Mikhail Haq	-0.71	74.52	-14.9	1.913	-0.061
Joe	1.84	157.15	-31.43	0.278	0.947

In addition to modeling the bivariate copulas, the purpose of this study is also to determine the best trivariate copula function with respect to the Maximum Likelihood values calculated for each of the copulas shown in Table 7. It is worth mentioning that which data (flood peak or volume or duration) is sorted in an ascending order, the results of the copula parameter are the same. Because the Gamble copula function parameter is not in the range of its variations, it cannot be modeled in the trivariate form. According to the goodness-of-fit results presented in Table 7 and the corresponding plots depicted in Fig 4, the Clayton function is the best trivariate function for fitting the joint probability to all the flood variables involved in this study.

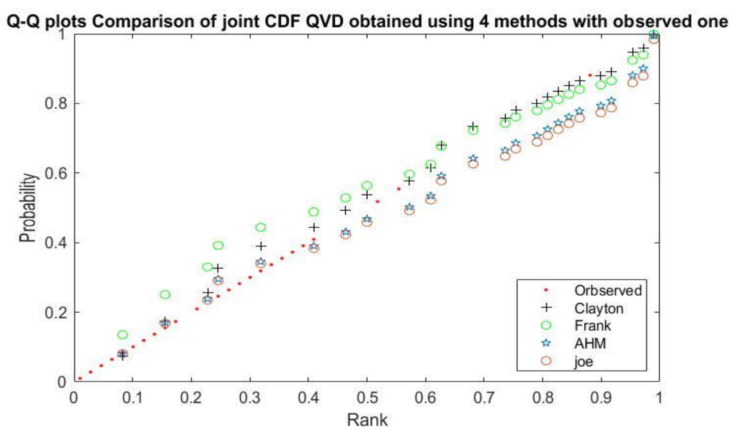


Fig.4 Curve Q_Q plot for three variables

Table.7. Goodness test results of fitting copula functions of three variables between flood peak, volume and flood duration

Copula function	parameter	Max log likelihood	AIC	RMSE	NSE
Clayton	**0.31**	**165.58**	**-33.12**	**0.365**	**0.916**
Frank	1.27	163.19	-32.64	0.382	0.908
Gamble	----	----	----	----	----
Ali Mikhail Haq	0.08	150.59	-30.12	0.48	0.855
Joe	1.13	155.57	-31.11	0.438	0.879

Joint Probability Distribution of Two and Three Variables

Some important information for flood management can be obtained from the joint probability distribution resulting from copula functions. The probability of both the flood peak and the flood volume to be above a certain threshold is an important condition triggering a flood warning system and would be the start point for emergency flood planning. After selecting the best copula function with reference to the goodness of fit criteria, the joint probability distribution curves/surfaces of the pairwise variables involved in the flood were plotted. In detail, joint probability distribution in addition to contour lines of the joint probability distribution, between flood peak and volume, between flood peak and duration, and between volume and duration of flood as well as joint probability of three variables for the fixed 60-day flood duration, are shown in Figs. 5, 6, 7, and 8.

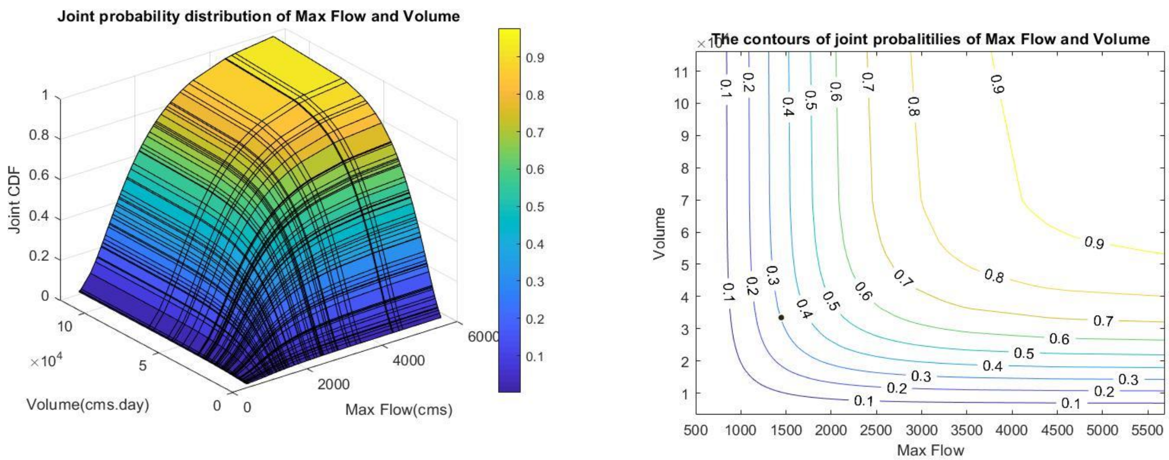


Fig.5 Joint Probability and contour lines of **Frank** distribution between flood peak and flood volume

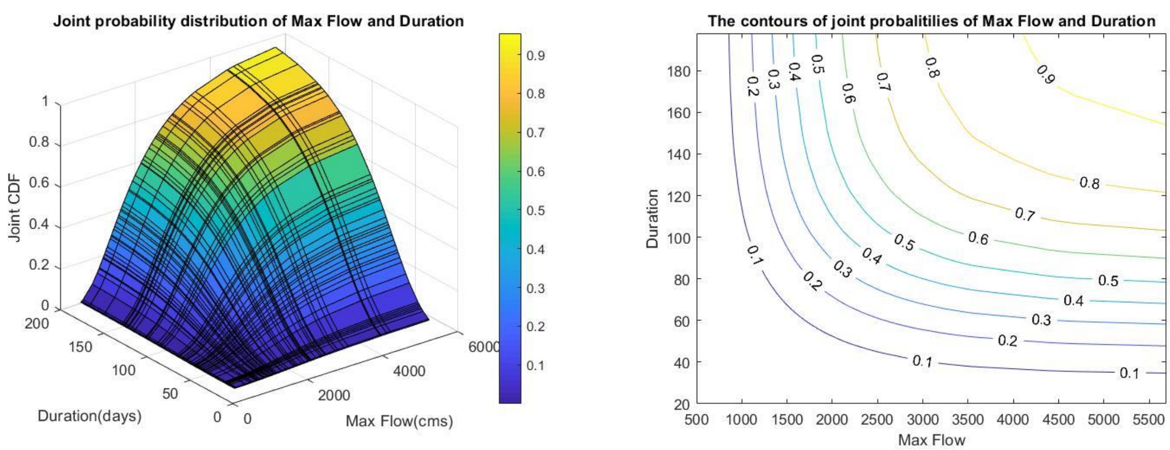


Fig.6 Joint Probability and contour lines of **Frank** distribution between flood peak and flood duration

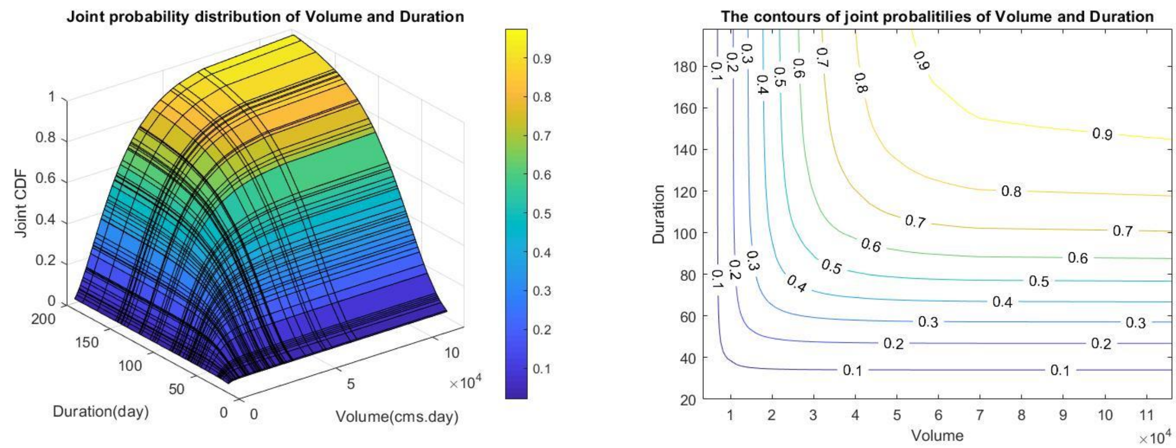


Fig.7 Joint Probability and contour lines of **Clayton** distribution between flood volume and flood duration

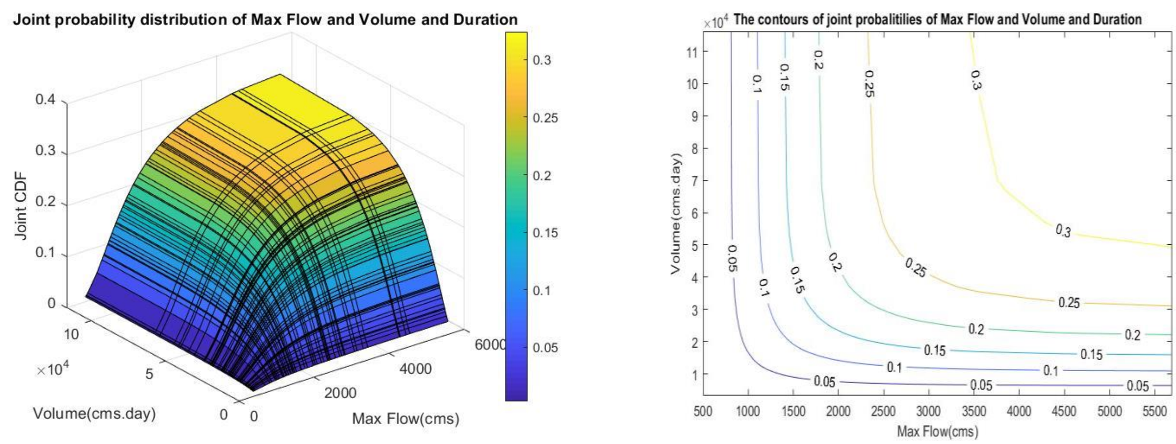


Fig.8 Joint Probability and contour lines of **Clayton** distribution between flood peak, volume, and duration

Return periods

In multivariate flood frequency analysis, flood variables are sometimes assumed to be independent variables without considering the variance/covariance structure of flood variables. Using the copula function of the selected two variables, the return period of type "or" corresponding to both series of flood parameters as well as the contour lines can be calculated and plotted, as shown in Figs. 9, 10, and 11. Comparing the return periods of the pairwise variables for different cases illustrated in Figs. 9, 10, and 11, the maximum return period is obtained to be 44 years based on the copula function built upon the case of the flood peak,-flood volume. This return period is taken into account as the highest and thus, the most critical return period of the flood events in a historical period and can be used to predict the flood risks at the future. In order to detect the difference in the return period obtained based on different hypotheses, the conditional return periods were calculated upon the bivariate and trivariate copula functions and compared. As an example, the standard conditional return periods of the flood peak limited by a certain volume and a certain duration of the flood, are depicted in Table 8. The results show that the flood peak extracted from a bivariate copula is greater than that extracted from a trivariate copula for the return periods less than or equal to 20 years. Thus, the bivariate conditional return period s for a short-term period is more reliable than the trivariate conditional return period. While, for the return periods longer than 20 years and up to 100 years, the flood peak derived from a trivariate copula is greater than that derived from a bivariate copula, meaning that the trivariate conditional return period for a long-term period is more reliable than the bivariate conditional return period. Also, the flood peak extracted from the Standard "or" bivariate and trivariate conditional return periods when the volume and duration of the flood are limited are shown in detail in Table 9. Note that in this table, the return period is "or", after one year. The flood peak derived from the trivariate conditional copula of type "or" is greater than that derived from the bivariate conditional copula of type "or", meaning that if decided to achieve more reliable flood peaks and thus, to make the more reasonable decisions, the conditional return periods should be calculated based on the conditional copulas of type "or". In addition, the conditional return periods calculated by the trivariate Clayton function for a fixed 60-day flood duration versus the flood peak and volume are shown in both forms of the surface and contours in Fig. 12.

Table.8. The return period of the two and three-variable conditional standard and its peak discharge

Conditional return period	Peak flood Q (CMS)	Peak flood Q (CMS)	Peak flood Q (CMS)
	(V=v, D=d) (v=80000 CMS*day, d=60day)	**d=60day**	**v=80000 CMS*day**
1.01	385.877	461.94	460.8
2	1514.2	1768.7	1764.8
5	2564.6	2872.5	2867.4
10	3404.4	3701.3	3695.6
20	4350.9	4563.1	4556.9
50	6007.5	5775.4	5768.6
100	8794..4	6757.7	6750.4

Table.9. The return period of the two and three-variable “or” standard and its peak discharge

“or” return period	Peak flood Q (CMS)	Peak flood Q (CMS)	Peak flood Q (CMS)
	(V=v, D=d) (v=80000 CMS*day, d=60day)	**d=60day**	**v=80000 CMS*day**
1.01	462.88	462.23	464.4
2	1810.2	1783.4	1800.6
5	3063.4	2938.1	2995.6
10	4227.5	3866.4	4009.1
20	6416.4	4973.4	5391.2

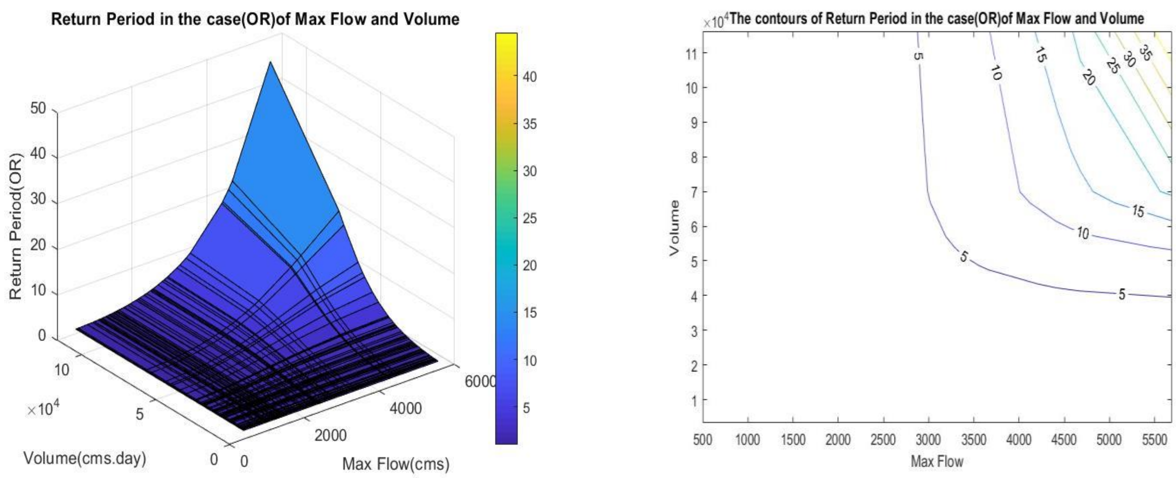


Fig.9 Return period and contour lines of the return period obtained by bivariate Frank function of type "or" between peak flow and flood volume

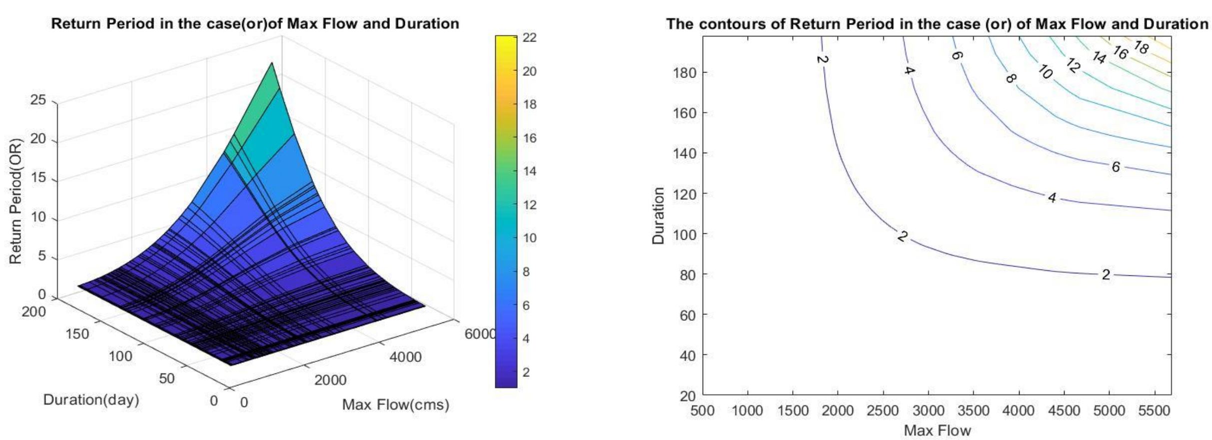


Fig.10 Return period and contour lines of the return period obtained by bivariate Frank function of type "or" between peak flow and duration

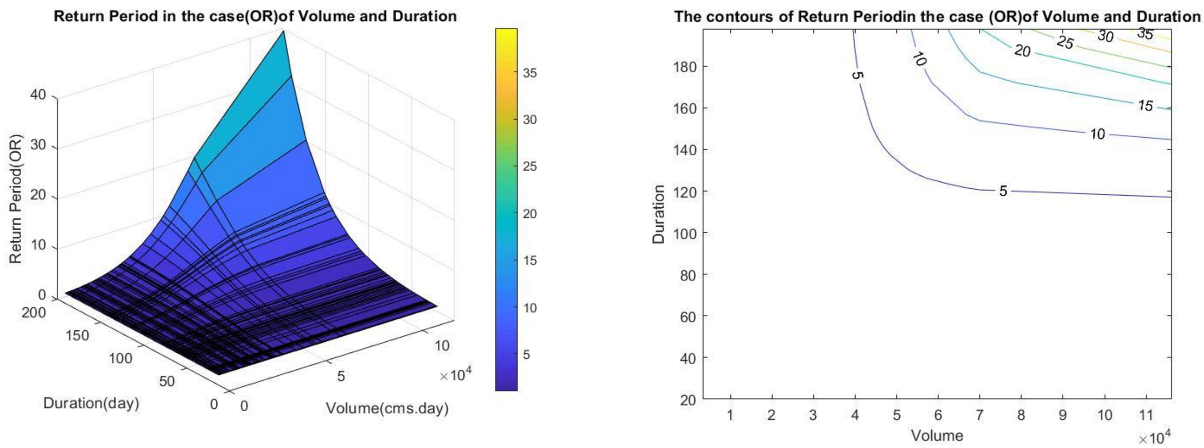


Fig.11 Return period and contour lines of the return period obtained by bivariate Clayton function of type "or" between volume and duration

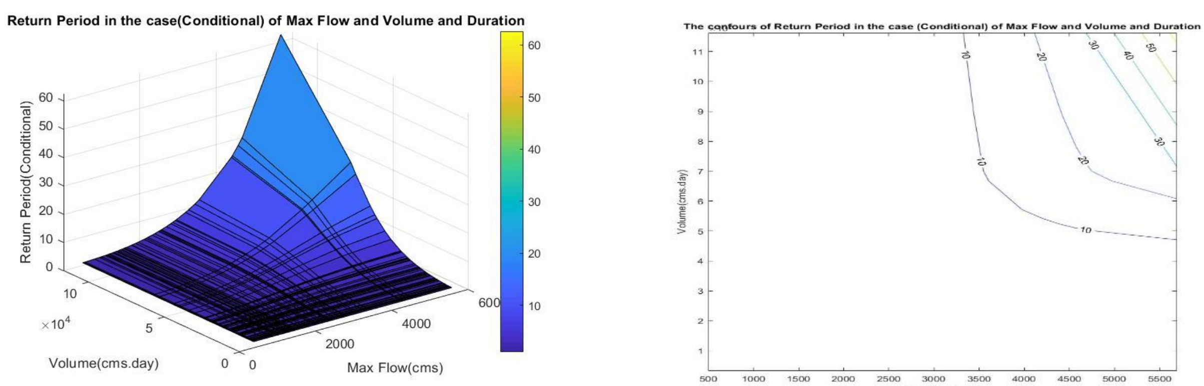


Fig.12 Conditional return period of three variables and contour lines for the 60-day flood duration

Conclusion

In recent decades, the phenomenon of flood has caused a lot of human and financial losses and irreparable damages. Knowing the flood encompasses several characteristics considered as its influential variables such as flood peak, flood volume and flood duration, the univariate analysis of flood frequency may make some errors. Therefore, trivariate flood frequency analysis should be considered as a method to thoroughly characterize the flood events and the probability of them to occur in the future. It is worth mentioning that the abovementioned flood variables (peak, volume, and duration) are random in nature and are correlated in pairs. The copula functions as the alternative functions could significantly raise the accuracy in the flood frequency predictions as a result of lacking all the limitations of the classical functions. Among the copula functions, the Archimedean functions are of three major advantages:(1) the ability to make multivariate analysis; (2) the ability to involve the correlation between the multiple studied variables in the analysis; and (3) the ability to define the marginal distributions of different families for different multiple variables. With respect to these benefits of using the Archimedean copulas functions were used to analyze the flood frequency and determine the return periods in different cases of the variables combinations. Each variable can be better fitted and estimated by a specific probability distribution function. These functions are experimentally determined to be the log normal, gamma, and Generalized Extreme Value functions for the flood peak, flood volume and flood duration, respectively. Thereafter, the parameters of the distribution functions were estimated and tuned using the maximum likelihood method. The Frank copula family was adopted for modeling and coupling the probability functions fitted to the three pairs of the flood variables including the peak-volume, the peak-duration, also the volume-duration Moreover, the Clayton family was selected for the volume- duration couple. Due to the importance of considering the trivariate form for any flood frequency analysis, in this study, this type of the analysis was also conducted. We concluded that the best trivariate copula function for the flood frequency analysis in the case study of this study is the Clayton family. Estimating the joint probability of the flood occurrence affected by two and/or three variables was calculated and plotted to be used for providing the future management plans related to water resources, risk analysis, contingency planning and flood warning. Return period is the average time that an event, such as flood, is expected to occur at the maximum magnitude. To better understand the concept of return period, the return period of two and three variables of the type "or" and "Conditional" was defined and discussed. The results of comparing the return period of the type "or" for two variables for different cases of the variable combination of showed that the return period "or" for the bivariate flood peak-flood volume) is less than the other cases, making the flood occurrence more reliable with these return periods. Therefore, the return period is proposed to be estimated for this combination (flood peak-flood volume), whenever decided to conduct the bivariate analysis. Furthermore, the results of conducting the trivariate analysis suggested that the estimated risk of the flood occurrence considering three variables is higher than that considering resulting from the univariate or bivariate analysis in the long-term, as Nashwan et al. (2018) concluded in their research. We conclude that the return period calculations for the bivariate cases and in the short-term period are more reliable than those in the case of the trivariate analysis, but the trivariate conditional return period calculated for the long-term periods is more reliable to come true than that calculated in the case of bivariate analysis. In addition, the flood peak occurring in the same return periods calculated by the copulas of type "or" is greater than that while occurring in the "conditional" return periods. Thus, the return periods of the type "or" is of higher risk of occurrence as they are lower than the other return periods calculated by another methods and thus, may be more reliable to be used for flood management purposes in the future. As the overall results suggest, it is recommended to perform similar flood frequency analysis in other important basins of the Karun River. These analyses can be useful for risk assessment related to hydrological issues including overflow design and flood management to take more reasonable and reliable control measures.

Conflict of Interest:

There is no conflict of interest.

References

Civil, É. p. D. d. g., & Ashkar, F. 1980. ***Partial duration series models for flood analysis***: École polytechnique de Montréal.

Correia, F. N. 1987. Multivariate partial duration series in flood risk analysis, ***Hydrologic Frequency Modeling***: 541-554: Springer.

De Michele, C., & Salvadori, G. 2003. A generalized Pareto intensity-duration model of storm rainfall exploiting 2-copulas. ***Journal of Geophysical Research: Atmospheres***, 108(D2).

De Michele, C., Salvadori, G., Canossi, M., Petaccia, A., & Rosso, R. 2005. Bivariate statistical approach to check adequacy of dam spillway. ***Journal of Hydrologic Engineering***, 10(1): 50-57.

Favre, A. C., El Adlouni, S., Perreault, L., Thiémonge, N., & Bobée, B. 2004. Multivariate hydrological frequency analysis using copulas. ***Water Resources Research***, 40(1).

Felfelani, F., Movahed, A. J., & Zarghami, M. 2013. Simulating hedging rules for effective reservoir operation by using system dynamics: a case study of Dez Reservoir, Iran. ***Lake and Reservoir Management***, 29(2): 126-140.

Genest, C., Favre, A. C., Béliveau, J., & Jacques, C. 2007. Metaelliptical copulas and their use in frequency analysis of multivariate hydrological data. ***Water Resources Research***, 43(9).

Gringorten, I. I. 1963. A plotting rule for extreme probability paper. ***Journal of Geophysical Research***, 68(3): 813-814.

Joe, H. 1997. ***Multivariate models and multivariate dependence concepts***: CRC Press.

Klein, B., Pahlow, M., Hundecha, Y., & Schumann, A. 2010. Probability analysis of hydrological loads for the design of flood control systems using copulas. ***Journal of Hydrologic Engineering***, 15(5): 360-369.

Li, T., Wang, S., Fu, B., & Feng, X. 2020. Frequency analyses of peak discharge and suspended sediment concentration in the United States. ***Journal of Soils and Sediments***, 20(2): 1157-1168.

Nashwan, M. S., Ismail, T., & Ahmed, K. 2018. Flood susceptibility assessment in Kelantan river basin using copula. ***Int. J. Eng. Technol***, 7(2): 584-590.

Nelsen, R. B. 2006. An Introduction to Copulas. Springer, New York. ***MR2197664***.

Sackl, B., & Bergmann, H. 1987. A bivariate flood model and its application, ***Hydrologic Frequency Modeling***: 571-582: Springer.

Salvadori, G., & De Michele, C. 2007. On the use of copulas in hydrology: theory and practice. ***Journal of Hydrologic Engineering***, 12(4): 369-380.

Salvadori, G., & De Michele, C. 2010. Multivariate multiparameter extreme value models and return periods: A copula approach. ***Water Resources Research***, 46(10).

She, D., & Xia, J. 2018. Copulas-based drought characteristics analysis and risk assessment across the loess plateau of China. ***Water Resources Management***, 32(2): 547-564.

Shiau, J.-T., Wang, H.-Y., & Tsai, C.-T. 2010. Copula-based depth-duration-frequency analysis of typhoons in Taiwan. ***Hydrology Research***, 41(5): 414-423.

Silverman, B. W. 1986. ***Density estimation for statistics and data analysis***: CRC press.

sklar. 1959. Fonctions de repartition an dimensions et leurs marges. . ***Publ. Inst. Stat. Univ. Paris***, 229-231.

Volpi, E., & Fiori, A. 2012. Design event selection in bivariate hydrological frequency analysis. ***Hydrological Sciences Journal***, 57(8): 1506-1515.

Yue, S. 2000. The bivariate lognormal distribution to model a multivariate flood episode. ***Hydrological Processes***, 14(14): 2575-2588.

Yue, S., Ouarda, T., Bobée, B., Legendre, P., & Bruneau, P. 1999. The Gumbel mixed model for flood frequency analysis. ***Journal of Hydrology***, 226(1-2): 88-100.

Zhang, L., & Singh, V. 2006. Bivariate flood frequency analysis using the copula method. ***Journal of Hydrologic Engineering***, 11(2): 150-164.

Zhang, L., & Singh, V. P. 2007. Trivariate flood frequency analysis using the Gumbel–Hougaard copula. ***Journal of Hydrologic Engineering***, 12(4): 431-439.

Figures

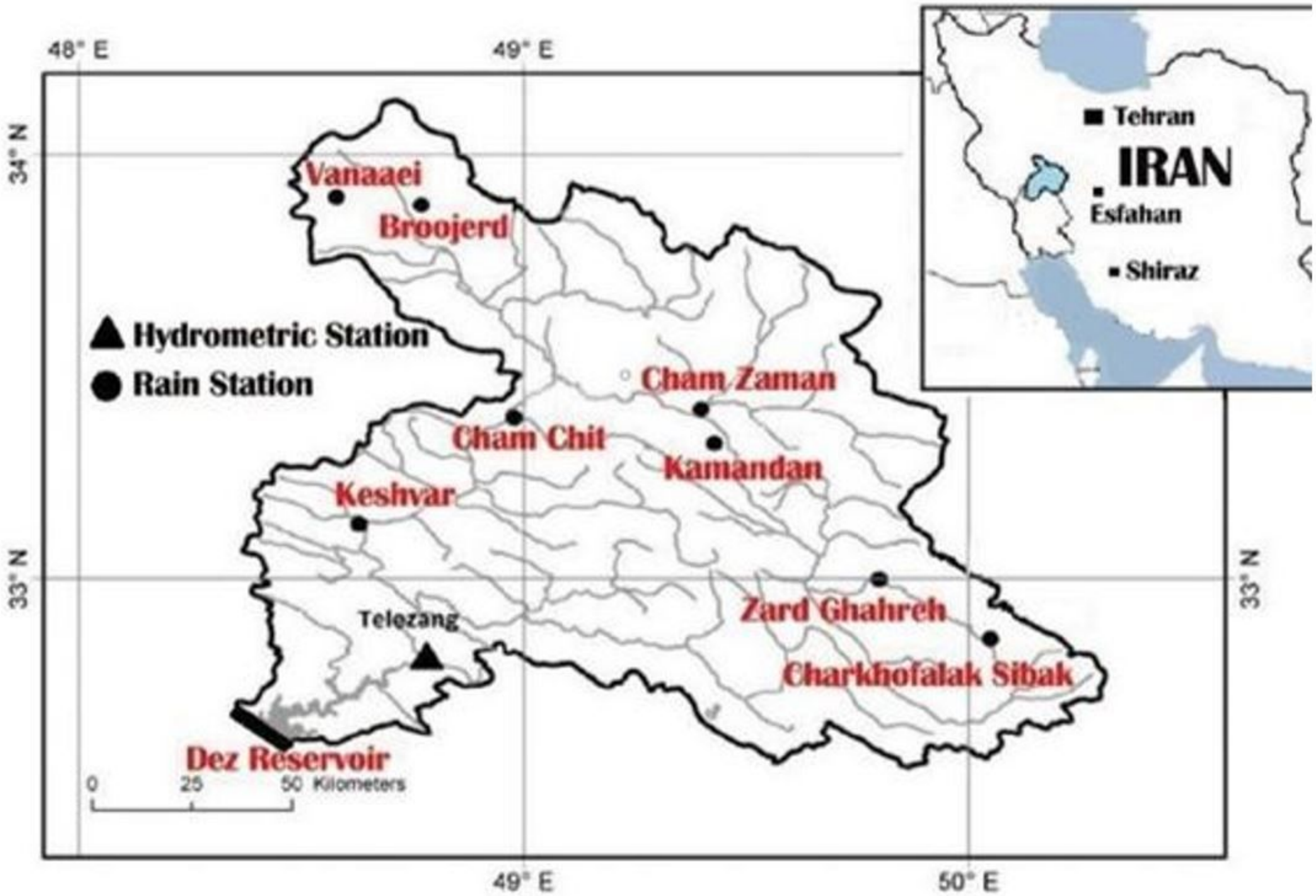


Figure 1

Location of Dez dam catchment and study area

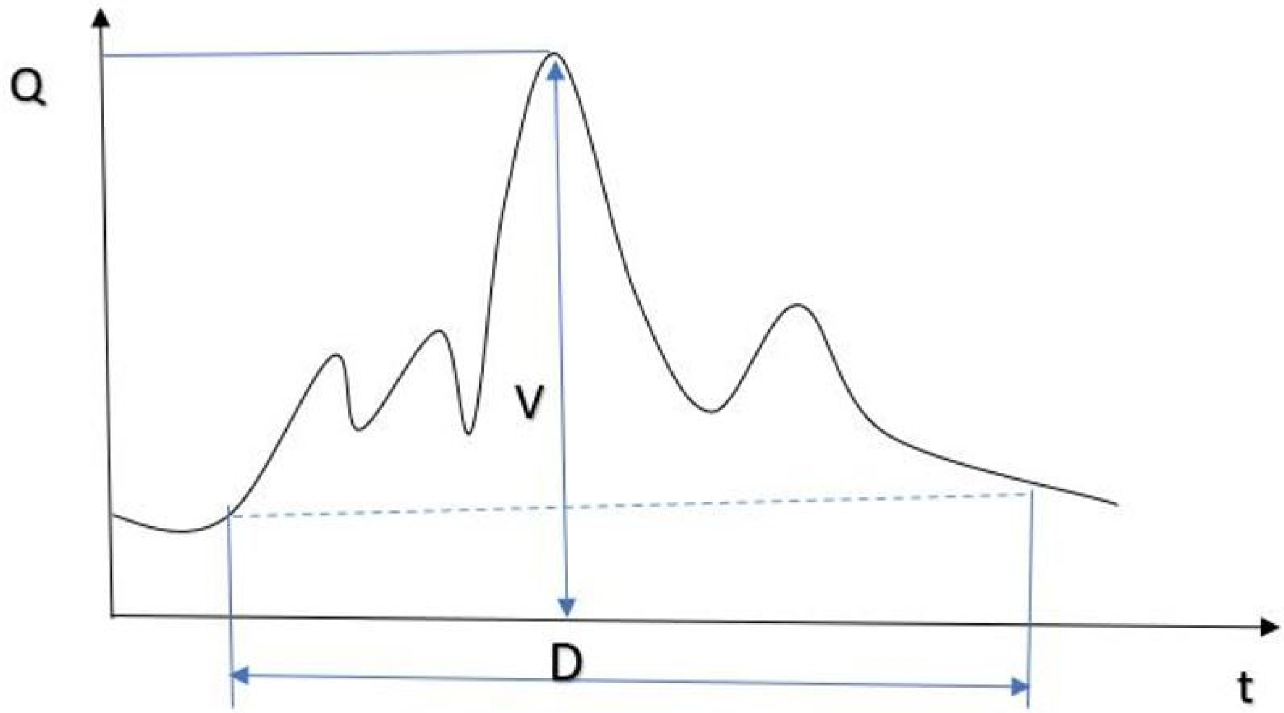


Figure 2

Determining flood characteristics

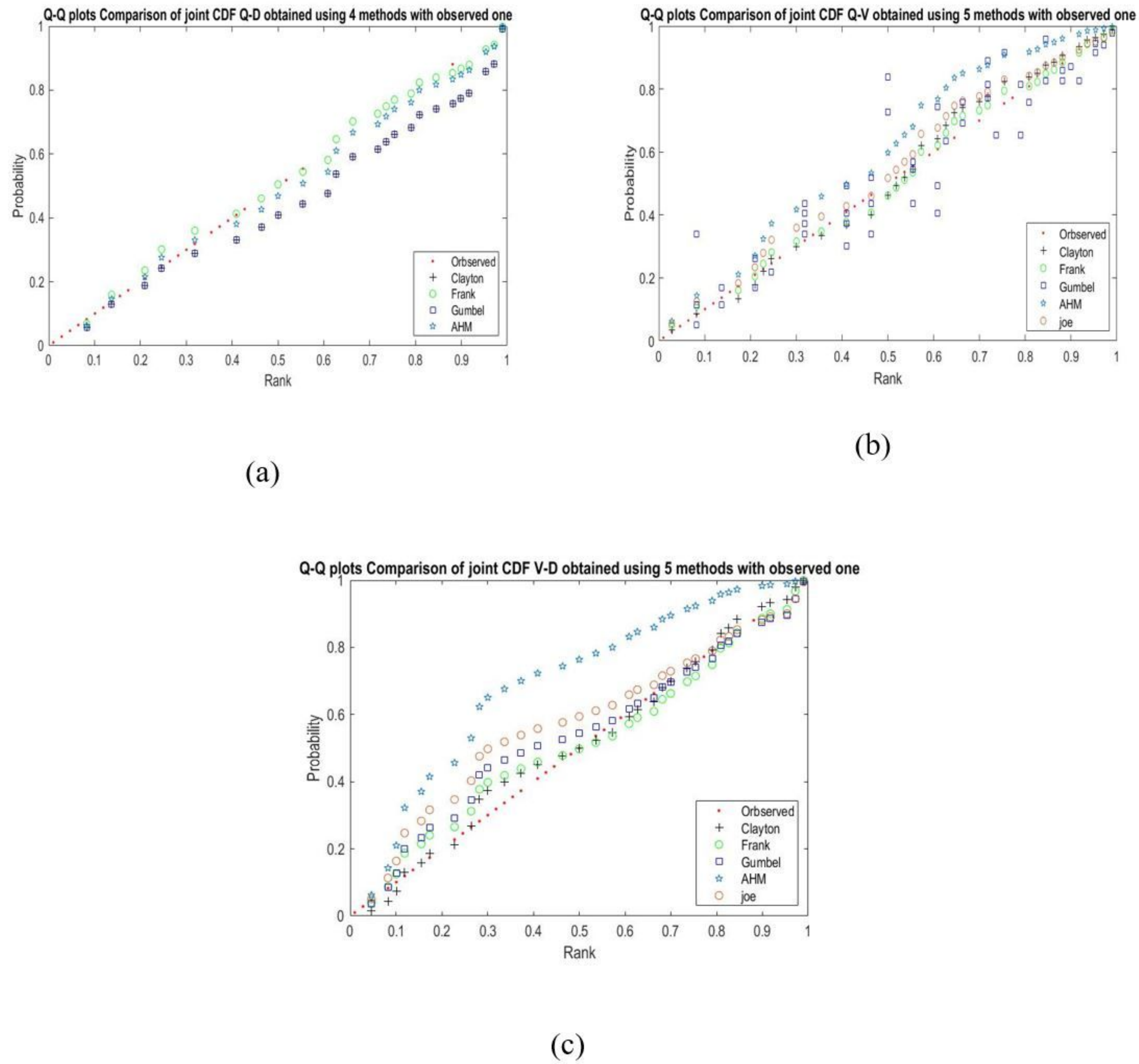


Figure 3

Curve Q-Q plot. (a) between flood peak and flood volume. (b) between flood peak and flood duration. (c) between the volume and duration of the flood

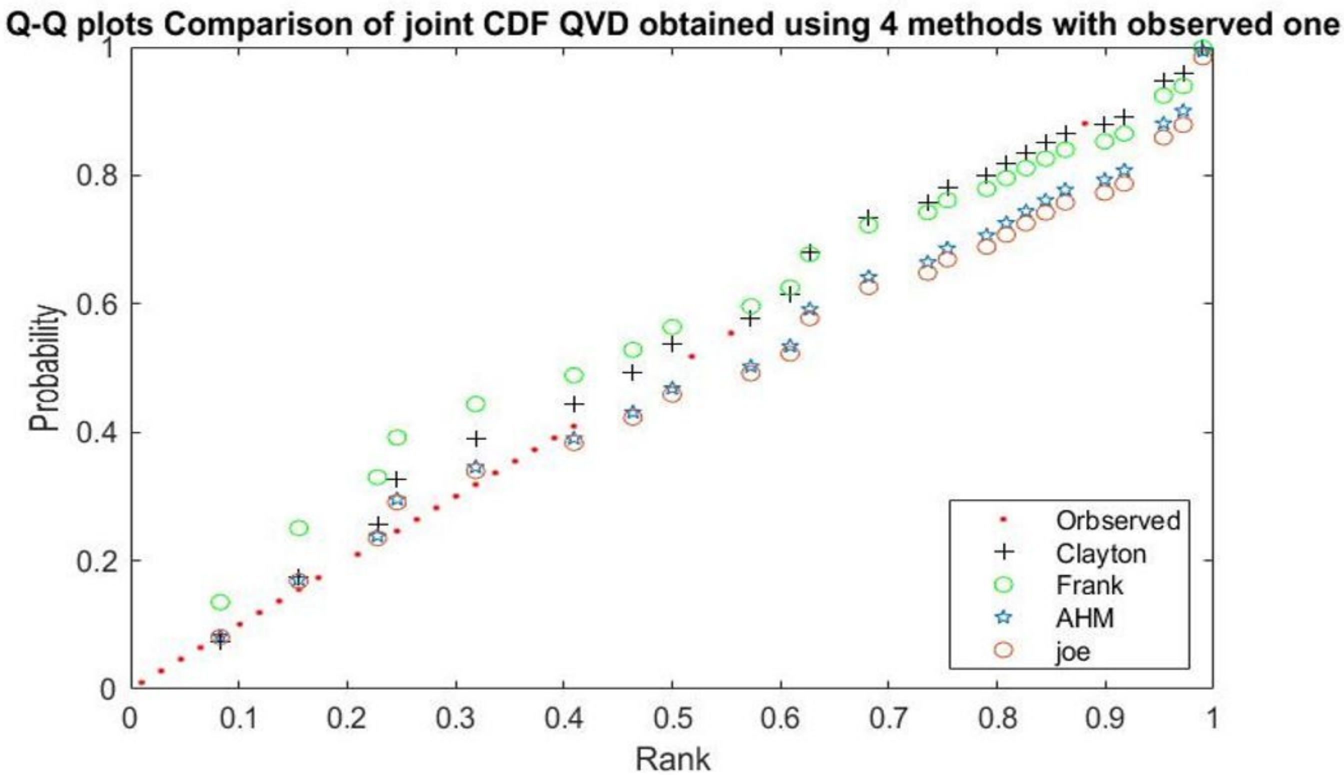


Figure 4

Curve Q_Q plot for three variables

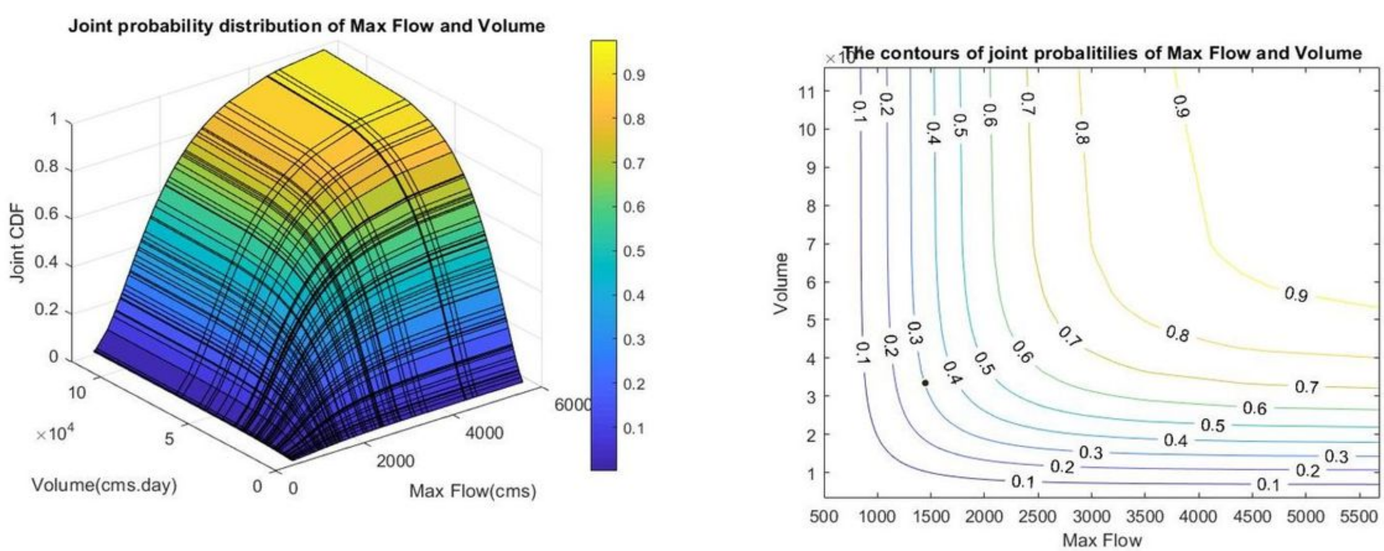


Figure 5

Joint Probability and contour lines of Frank distribution between flood peak and flood volume

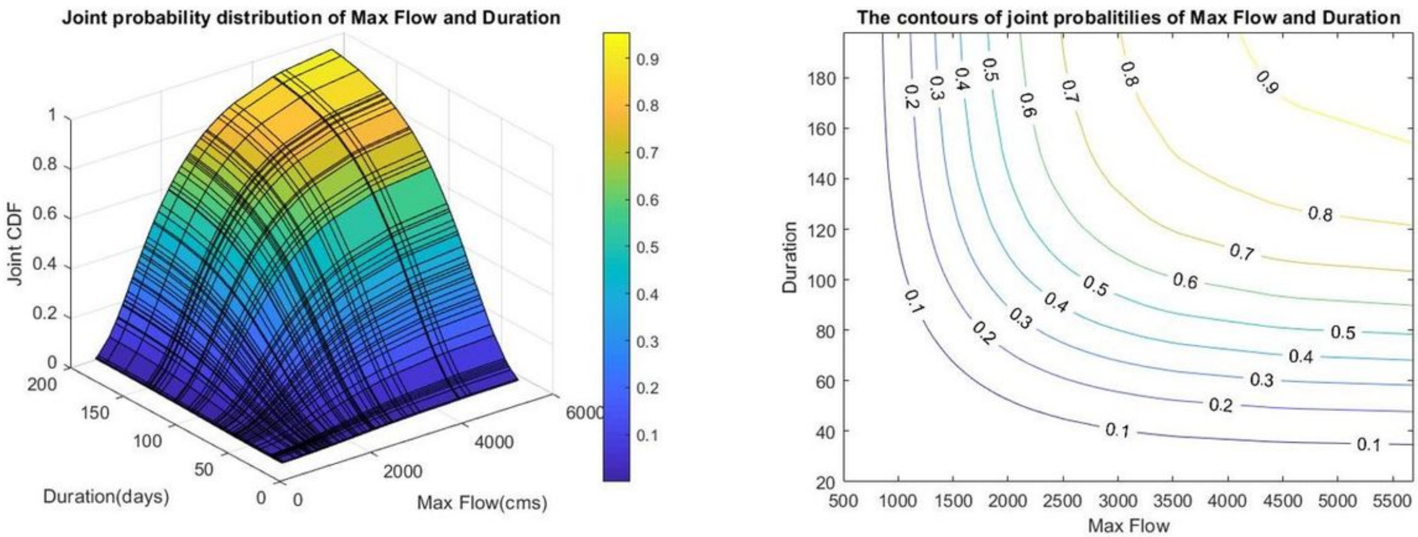


Figure 6

Joint Probability and contour lines of Frank distribution between flood peak and flood duration

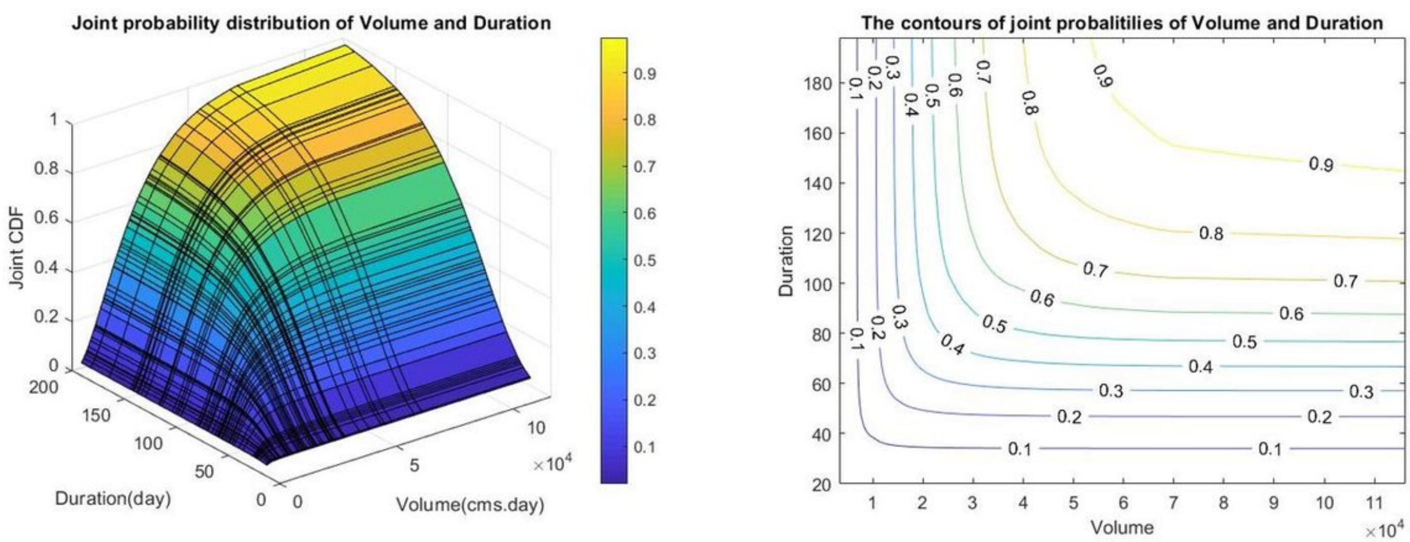


Figure 7

Joint Probability and contour lines of Clayton distribution between flood volume and flood duration

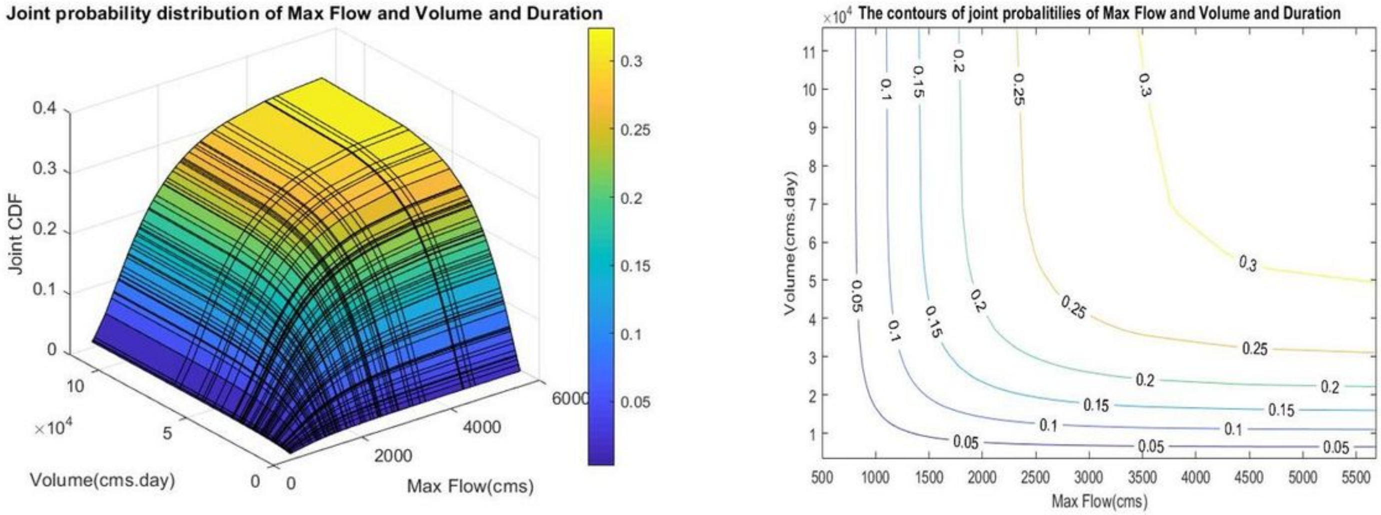


Figure 8

Joint Probability and contour lines of Clayton distribution between flood peak, volume, and duration

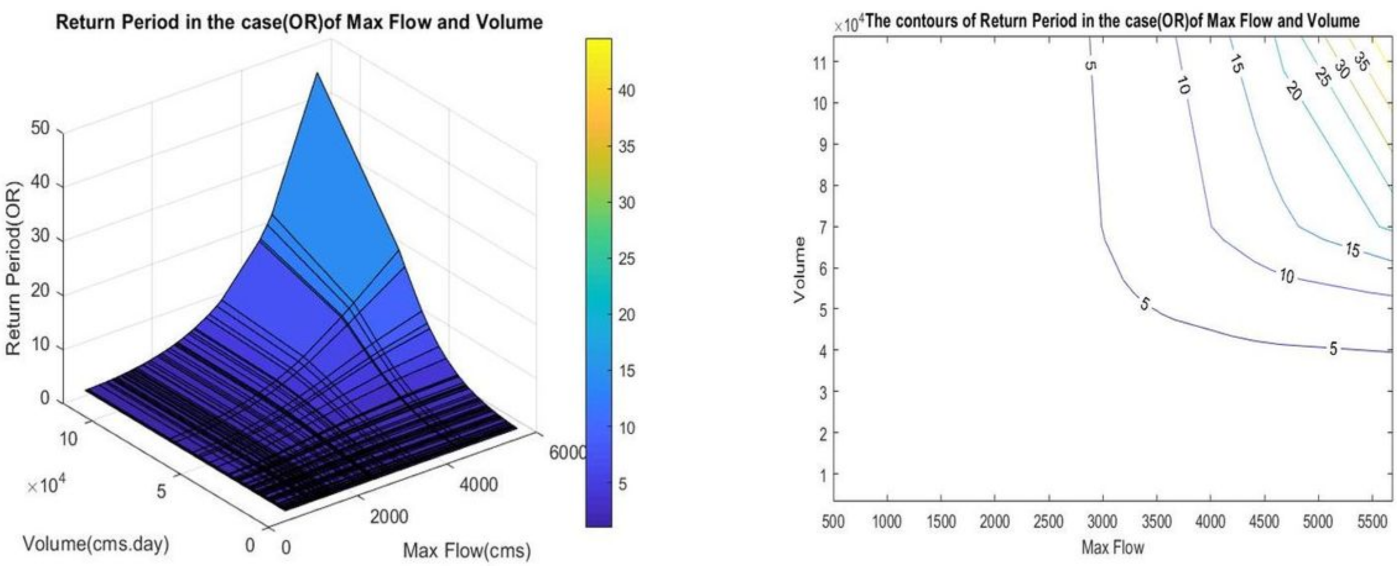


Figure 9

Return period and contour lines of the return period obtained by bivariate Frank function of type "or" between peak flow and flood volume

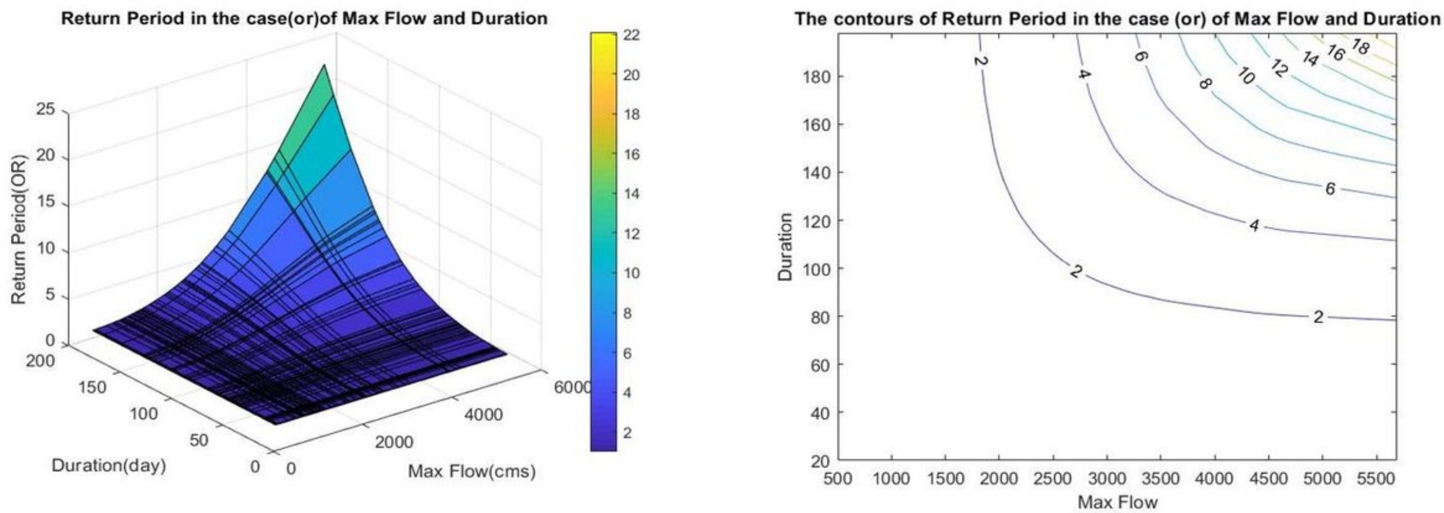


Figure 10

Return period and contour lines of the return period obtained by bivariate Frank function of type "or" between peak flow and duration

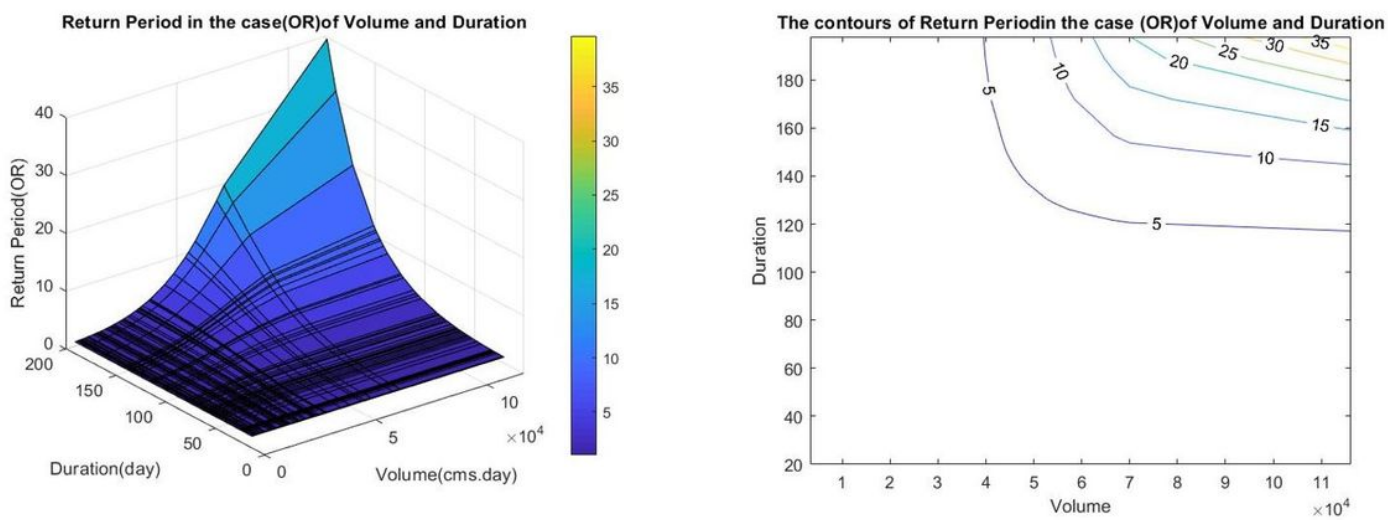


Figure 11

Return period and contour lines of the return period obtained by bivariate Clayton function of type "or" between volume and duration

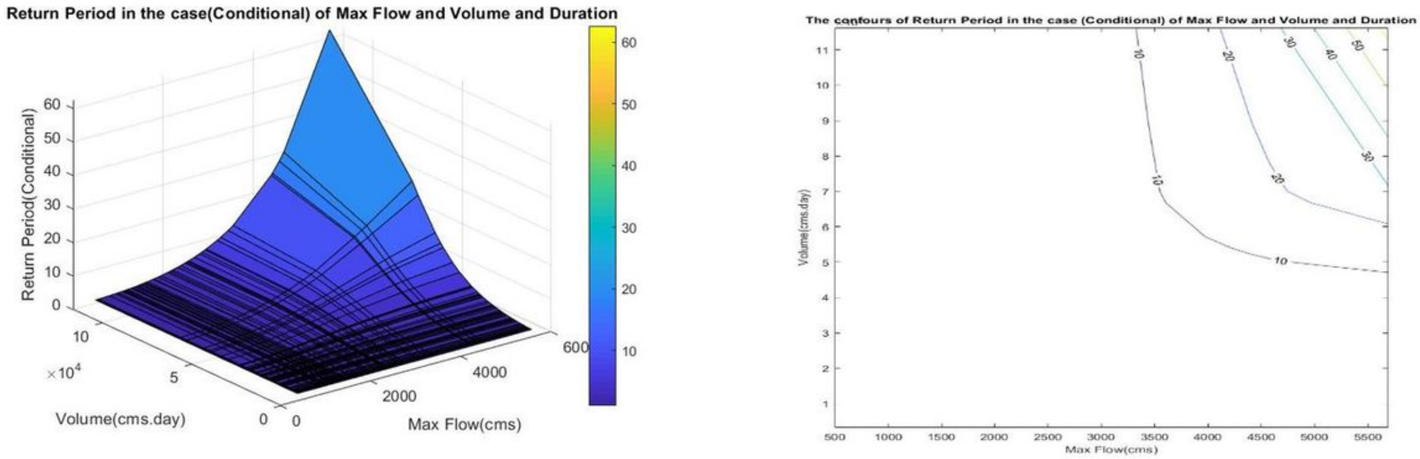


Figure 12

Conditional return period of three variables and contour lines for the 60-day flood duration