Water Resources Management (2020) 34:3533–3549
https://doi.org/10.1007/s11269-020-02634-0

# Flood Frequency Analysis of Interconnected Rivers by Copulas

**Esmaeel Dodangeh**[1] **· Vijay P. Singh**[2] **· Binh Thai Pham**[3] **· Jiabo Yin**[4] **· Guang Yang**[5] **· Amirhosein Mosavi**[6,7]

Received: 21 December 2019 / Accepted: 21 July 2020 /
Published online: 29 July 2020


**Abstract**
Flood frequency analysis (FFA) considering the confluence of interconnected rivers is important for hydraulic structures (such as dams or diversions) design, but it has received little attention. This study develops a copula-based method for FFA and quantile estimation considering the confluence of two interconnected rivers, along with the uncertainty estimation by a nonparametric bootstrapping algorithm. Flood probability distribution and return periods are estimated for the two rivers by mapping from bivariate to univariate peak flow quantile estimation. The methodology is applied to the case study of Qezel Ozan and Shahrud Rivers which merge to one of the largest reservoir dams in Iran: Sefidrud (Manjil) dam. According to the results from Peak flow records from Gilvan station (GPF) at Qezel Ozan River and from Loshan station (LCF) at Shahrud River, Gaussian copula with Weibull and gamma margins fits best. Also, it shows that some peak flow quantiles with the same magnitudes have a different probability of occurrences at the confluence of the rivers, and the bivariate estimation uncertainty usually plays an important role in FFA. These findings suggest the use of bivariate instead of univariate distributions to the peak flows at the confluence of interconnected rivers, in which the sampling uncertainty should be considered.

**Keywords** Interconnected rivers · Copulas · Gaussian · Weibull · Return period · Sefidrud

✉ Esmaeel Dodangeh
smaeel.dodangeh@gmail.com

✉ Binh Thai Pham
phamthaibinh2@duytan.edu.vn

✉ Amirhosein Mosavi
amirhosein.mosavi@tdtu.edu.vn

Extended author information available on the last page of the article

## 1 Introduction

Floods are among the most devastating natural disasters that cause mortalities and financial losses in the affected areas. They are important not only for their destructive effects at the time of occurrence but also for their post-disaster damaging effects persisting for a long time. Global Assessment Report on Disaster Risk Reduction (UNISDR 2015) indicates that flood mortality rate is increasing in developing countries such as Iran. The flood mortality rate has increased by 11% in the Middle East and North Africa compared to the last century (Ghomian and Yousefian 2017). Over the past few years, floods have been the most serious natural disaster in Iran. The recent flood of March 2019 affected many parts of the country with a loss of 76 people and damages of 476 million US dollars.

Knowledge of flood magnitudes and frequencies is critical for flood preparedness by adopting preventive measures to mitigate their destructive effects. Flood frequency analysis (FFA) is employed for estimating flood magnitudes for given return periods; economically safe designs of hydraulic structures such as dams, culverts, bridge, and spillways and it can be also used to determine flood insurance premiums under potential flood damages.

Single-site FFA is performed by fitting appropriate univariate probability distributions to peak stream flow data (Modarres 2008). In the case of interconnected rivers, where two mainstreams with different drainage areas, land-use/land-cover schemes and consequently different probability distributions join together, FFA can however be challenging, especially when the two mainstreams marge at a dam where the historical streamflow data are usually not available. Fitting univariate probability distributions to the sums of flows may not be appropriate, because different pairs of flood quantiles which have the same probability of occurrence may have different severities of impact. Also, the stochastic nature of interconnected rivers with different drainage areas and hydro-geomorphic characteristics may differ from each other. Thus, the use of bivariate probability distributions will not guarantee reliable flood quantile estimation, as they assume that the flood records of the joining rivers share the same probability distributions. To our best knowledge, little literature provides the solution to the estimation of design flood at the confluence of the interconnected rivers, especially where two rivers merge at a reservoir dam.

In this study, an innovative application of copula functions (Sklar 1959) for design flood estimation at the confluence of interconnected rivers is presented by mapping from bivariate to univariate peak flow quantile estimation. Copulas have the capability of preserving the dependence structure of random variables having similar or different marginal probability distributions (Zhang and Singh 2019). Copulas have been applied to drought frequency analysis (Shiau and Modarres 2009; Mirakbari et al. 2010, 2012; Janga Reddy and Ganguli 2012; Madadgar and Moradkhani 2012; Lee et al. 2013; Saghafian and Mehdikhani 2014; Dodangeh et al. 2017), rainfall frequency analysis (Kao and Govindaraju 2007, 2008; Zhang and Singh 2007; Zhang et al. 2013), flood frequency analysis (Grimaldi and Serinaldi 2006; Shiau et al. 2006; Zhang and Singh 2006; Fu and Butler 2014), rainfall-runoff frequency analysis (Golian et al. 2012; Dodangeh et al. 2019), water quality analysis and other analyses (Zhang and Singh 2019). The copula model was applied to peak flow magnitudes estimation at the Sefidrud dam (where the Qezel Ozan and Shahrud Rivers merge). Various combinations of copula functions and univariate probability distributions were assessed to select the best combination to fit the joint cumulative distribution functions (CDFs) of annual peak flows of Qezel Ozan River and corresponding flows of Shahrud River. Then, the joint flood quantile and the uncertainty of joint flood quantile was estimated using a non-parametric bootstrapping method. Figure 1 demonstrates the general flowchart of the methodology used.

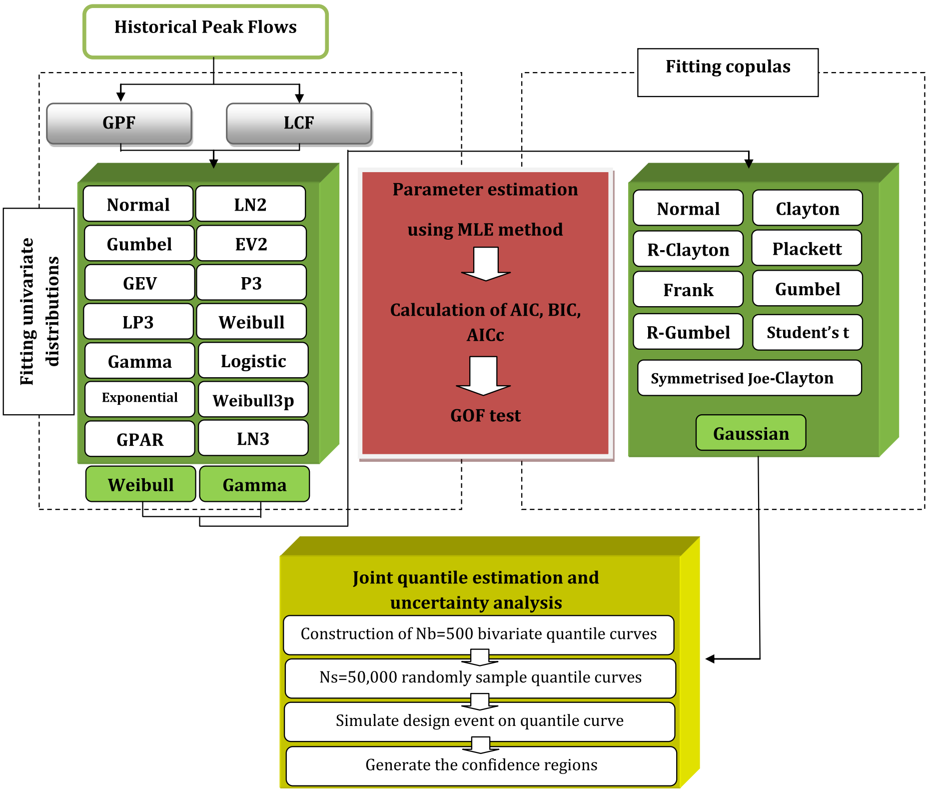


**Fig. 1** Flowchart of methodology

## 2 Description of Study Area

The study area is Sefidrud (Manjil) lake dam at the intersection of Qezel Ozan and Shahrud Rivers. The dam is located 75 km south of Rasht and is adjacent to the Manjil city. The main objective of the Sefidrud dam is to regulate inflows into the Sefidrud River to irrigate 189,832 ha of grasslands in Guilan and Foumanat Plain downstream of the dam. There are also secondary objectives, such as flood control, production of hydroelectric power with a nominal capacity of 87.5 MW, drinking water supply and urban industries of Central to and East of Guilan, and meeting the needs of fishery, aquaculture, and animal husbandry (Dodangeh 2010).

The drainage area of the Sefidrud lake is approximately 56,200 km$^2$ and the volume of the lake is 1756–10$^6$ m$^3$ which is supplied by two mainstreams, namely Qezel Ozan River in the west and Shahrud River in the east of the dam. The Shahrud River with a length of 180 km originates in the Taleghan, Alamkooh, Takht-e-Solaiman and central Alborz mountain ranges and it is the main river of the Qazvin province. The Qezel Ozan River with a length of 800 km is one of the longest rivers of the country that originates in the Chehel Cheshmeh mountain ranges in Kurdistan and East Azarbaijan province (Dodangeh et al. 2014). In the south of Guilan province, in the waters of Sefidrud dam, it joins the Shahrud River and forms the Sefidrud River.

Sefidrud dam is one of the most sediment-silting dams in the world and more than half of the reservoir capacity is filled with sediment (Kavian et al. 2016). Since the 1980s, Shas operation has been used to overcome siltation. However, considering the overflow discharge of about 5000 $m^3/s$, which was estimated at the time of dam construction, severe floods in Qezel Ozan and Shahrud Rivers may endanger the dam stability and the lives of residents of the areas adjacent to the lake (as illustrated in Fig. 2). Therefore, flood frequency and magnitude estimation are essential for determining the extent of flooding around the lake and for estimating potential damages caused by flooding. For this purpose, historical data for a 49-year period (1963–2011) from the Gilvan (#17–033) (GPF) and Loshan (#17–041) (LCF) gauging stations, located on Qezel Ozan and Shahrud Rivers, respectively, were used.

## 3 Methods

### 3.1 Copulas

For bivariate cases, the joint CDF, $F_{X,\,Y}(x, y)$, of random variables X and Y can be expressed using the copula function as:

$$F_{X,Y}(x,y) = C(F_X(x), F_Y(y)) \tag{1}$$

where C is the copula function, and $F_X(x)$ and $F_Y(y)$ are the marginal cumulative distribution functions (CDFs) of random variable X and Y, estimated using the univariate probability distributions. For the cestimation of univariate CDFs of the X and Y variables, the probability distributions suggested by Rao and Hamed (2002) were used (Table 1). Parameters of the marginal distributions were estimated independently and then the copula parameter was estimated based on the dependence between the marginal variables. Equation 1 can be rewritten in the form:

$$F_{X,Y}(x,y,\theta) = C\left(F_X(x,\theta_x), F_Y\left(y,\theta_y\right),\theta_c\right) \tag{2}$$

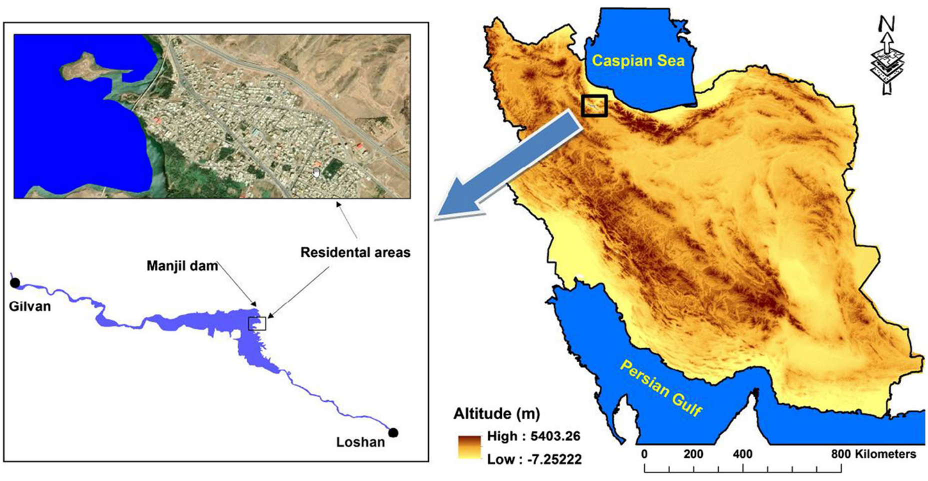


**Fig. 2** Location of the study area

**Table 1** Goodness of fit test of univariate probability distributions

| | AIC | AICc | BIC | KS test | |
|---|---|---|---|---|---|
| | | | | Z | *p* value |
| Gilvan | | | | | |
| Normal | 733 | 733 | 737 | 0.13 | 0.28 |
| LN2 | 732 | 732 | 735 | 0.14 | 0.21 |
| Gumbel | 726 | 726 | 730 | 0.07 | 0.92 |
| EV2 | 750 | 750 | 754 | 0.22 | 0.01 |
| GEV | 728 | 728 | 734 | 0.07 | 0.92 |
| P3 | 727 | 728 | 733 | 0.09 | 0.78 |
| LP3 | 727 | 727 | 732 | 0.08 | 0.82 |
| Weibull[a] | 724 | 725 | 725 | 0.10 | 0.66 |
| Gamma | 725 | 725 | 729 | 0.09 | 0.77 |
| Logistic | 732 | 732 | 735 | 0.14 | 0.23 |
| Exponential | 742 | 742 | 746 | 0.21 | 0.01 |
| Weibull3p | 726 | 726 | 732 | 0.09 | 0.73 |
| GPAR | 732 | 733 | 738 | 0.09 | 0.68 |
| LN3 | 734 | 734 | 739 | 0.09 | 0.90 |
| Loshan | | | | | |
| Normal | 555 | 555 | 559 | 0.09 | 0.76 |
| LN2 | 552 | 552 | 556 | 0.16 | 0.12 |
| Gumbel | 549 | 549 | 553 | 0.11 | 0.47 |
| EV2 | 568 | 567 | 572 | 0.21 | 0.02 |
| GEV | 551 | 551 | 556 | 0.10 | 0.64 |
| P3 | 551 | 552 | 557 | 0.13 | 0.27 |
| LP3 | 547 | 547 | 553 | 0.12 | 0.43 |
| Weibull | 545 | 546 | 549 | 0.12 | 0.40 |
| Gamma[a] | 544 | 544 | 549 | 0.09 | 0.72 |
| Logistic | 555 | 555 | 559 | 0.11 | 0.51 |
| Exponential | 559 | 559 | 563 | 0.17 | 0.08 |
| Weibul3p | 547 | 548 | 553 | 0.12 | 0.35 |
| GPAR | 546 | 547 | 550 | 0.12 | 0.35 |
| LN3 | 555 | 555 | 560 | 0.12 | 0.38 |

[a] indicating the best fit univariate distribution

where $\theta_x$ and $\theta_y$ are the parameters of the univariate marginal distributions, θ is the parameter of bivariate distribution and $\theta_c$ is the copula parameter (Dung et al. 2015). Of the various copula functions used for multivariate frequency analysis of random variables (Joe 1997; Nelsen 2006; Sungur and Nelson 2006; Zhang and Singh 2019), Archimedean, Gaussian, Student's *t* and Plackett copulas were used to model the joint GPF-LCF pairs.

### 3.2 Parameter Estimation for Bivariate Distributions

Reliable parameter estimation is required to identify the best fit distributions (de Melo e Silva Accioly and Chiyoshi 2004). Both parametric and nonparametric approaches, such as Kendall's tau and Maximum likelihood estimation (MLE) method (Genest and Rivest 1993), have been employed. The non-parametric Kendall's tau estimation method is based on the association between Kendall's tau and copula parameters (Mirakbari et al. 2010). Kendall's tau estimation can be used for single-parameter copula families (Zhang and Singh 2006; Mirakbari et al. 2010; Janga Reddy and Ganguli 2012). However, for one- and two-parameter copula families such as Student's *t* copula, the maximum likelihood estimation method, which is more appropriate (Joe 1997), was used. For the estimation of copula parameters using the MLE method, the log likelihood (LL) can be expressed as:

$$L(\theta) = \sum_{i=1}^{n} \log[c_{\theta}\{F_X(x_i), F_Y(y_i)\}] \quad (3)$$

where F denotes the marginal distributions, $c_\theta$ denotes the copula density function with parameter θ, $X$ and $Y$ denote random variables and $x_i$ and $y_i$ denote historical observations (Dodangeh et al. 2019). Using the MLE method, the bivariate model parameter (θ) was estimated as the value for which the negative LL function is minimized.

### 3.3 Copula Model Selection Criteria

For identifying the copula which best fitted the GPF-LCF pairs, various performance measures and goodness-of-fit tests (GOF) were utilized. The Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are commonly used performance measures (Breymann et al. 2003). The AIC and BIC values were computed as follows:

$$AIC = 2L(\theta) + 2/(n^*k) \quad (4)$$

$$BIC = 2L(\theta) + ln(n)/(n^*k) \quad (5)$$

In Eqs. 4 and 5, k denotes to the number of estimated parameters, n is the data length, and L(θ) denotes to the log-likelihood value. For small data samples with n/k < 40, the second order variant of AIC, named $AIC_c$, was estimated as (Viglione 2008):

$$AIC_c = 2L(\theta) + 2/(n^*k)^*(n/(n{-}k{-}1)) \quad (6)$$

These performance measures are however only useful for comparative assessment between the models but are not capable of representing the degree-of-fit of the copula models (Genest and Rivest 1993). For verification of the copula model selection using the AIC, BIC and $AIC_c$ measures, the White test (White 1982) was employed to test if the copula belonged to the selected copula family.

### 3.4 Bivariate Quantile Estimation

The joint return period of a GPF-LCF quantile was estimated from the best fit copula with the joint CDFs of the GPF and LCF data:

$$T_{GPF-LCF} = T\left(GPF \geq GPF', LPF \geq LCF'\right) = \frac{E(L)}{1-F_{GPF}(GPF')-F_{LCF}(LCF') + C[F_{GPF}(GPF'), F_{LCF}(LCF')]} \quad (7)$$

Here C is the copula function, $T_{GPF\text{-}LCF}$ is the joint return period of GPF ≥ GPF′ and LCF ≥ LCF′. GPF′ and LCF′ are the specified values of the Gilvan peak flows and the corresponding Loshan flows. E(L) denoted the expected inter-arrival time of GPF ≥ GPF′ and LCF ≥ LCF′ events, that is calculated based on the historical data; $F_{GPF}$(GPF′) and $F_{LCF}$(LCF′) are CDFs of GPF and LCF.

### 3.5 Uncertainty Analysis of Bivariate Quantile Estimation

Bivariate quantile estimation has uncertainties associated with parameters of marginal distributions, parameter of copulas, and inappropriate univariate or copula model selection. There is an additional uncertainty for bivariate quantile associated with the quantile selection on the p-level curves. There are indefinite combinations of x and y (PF data points) on a given p-level curve with the same probability of occurrence. However, they have different implications for engineering applications (Salvadori and De Michele 2011). The point with the highest density on the p-level curve is the event with the highest likelihood and is considered as a design event in hydrologic engineering applications. For the estimation of this point a line that best fits the dependence between X and Y variables, is plotted. Intersection of this line with the p-level curve identifies the design event. For this purpose, the Acceptance-Rejection Sampling (ARS) algorithm (Martino and Míguez 2010) was employed to estimate the bivariate design event on the p-level curves. The following steps were followed to estimate the design event based on the ARS algorithm:

[1] Set the number of bootstrap samples at a sufficient large number $N_b$ ($N_b = 500$).
[2] Generate $N_b$ bootstrap samples by sampling with replacement from the observed *x-y* pairs. The size of all $N_b$ bootstrap samples should be identical to the size of the original sample of observed *x-y* pairs.
[3] Fit $N_b$ separate copulas to the generated $N_b$ bootstrap samples.
[4] Generate a random integer number in [1, $N_b$], and select the corresponding copula from the fitted copulas.
[5] Simulate a point along the level curve (at a specific $p$-level, for instance, $p = 0.95$) of the copula selected in step 4. Such a point is also referred to as the $p$-th quantile. Points along a level curve are simulated using the acceptance-rejection algorithm.
[6] Iterate steps 4 and 5 $N_s = 50{,}000$ times.
[7] Identify the confidence range around the p-level curves using the $N_s$ points.

## 4 Results and Discussion

### 4.1 Homogeneity and Randomness Tests

Randomness and homogeneity of the GPF and LCF data were examined before doing the joint FFA. The run test (Rodrigo et al. 1999) checks the homogeneity of the GPF and LCF data at the 95% confidence level and investigates the variation of data around the median (Modarres 2008). Results showed that both GPF and LCF were homogeneous at 95% confidence level with $p > 0.05$. Randomness of the GPF and LCF time series is examined with the Autocorrelation and Partial autocorrelation functions (ACF and PACF), i.e., autocorrelation between the unique data points is calculated at various time lags. If the ACF values were near zero, especially at the first lag, then the data set was considered random. The zero and near zero values of PACF at the first lag also indicated the randomness of data sets (see Fig. 3). It shows that the ACF coefficients are close to zero, which indicated that the streamflow data are randomly distributed over time.

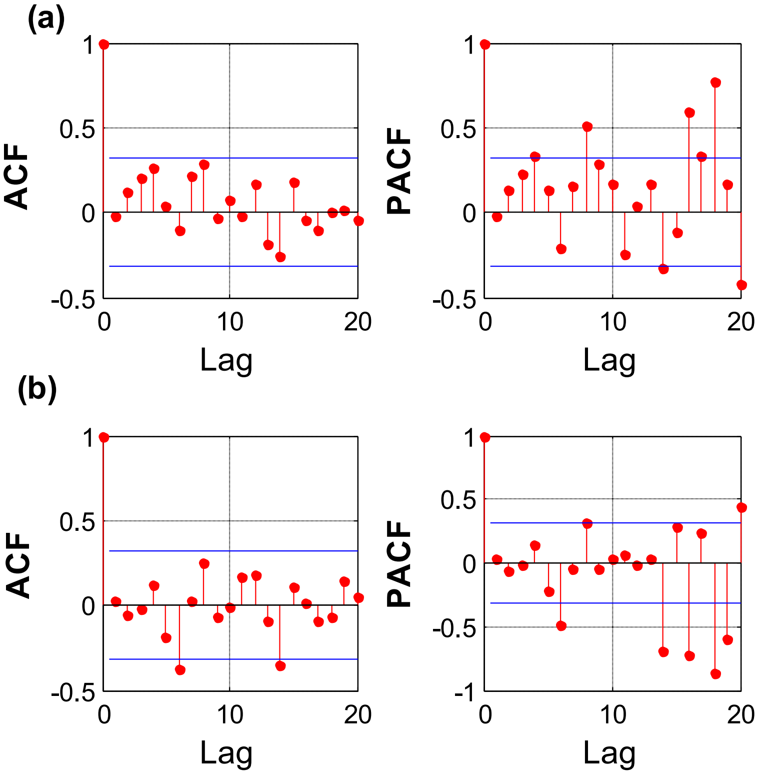


**Fig. 3** ACF and PACF values of the GPF (**a**) and LCF (**b**) data series

## 4.2 Univariate GPF and LCF Quantile Estimation

Univariate FFA is performed by fitting the probability distributions to the GPF and LCF data. Fourteen probability distributions, including the normal, 2-parameter log-normal (LN2), Gumbel, Extreme values type II (EV2), Generalized extreme values (GEV), Pearson III (P3), Log-Pearson III (LP3), Weibull, Gamma, Logistic, exponential, 3-parameter Weibull (Weibull3p), generalized Pareto (GPAR), and 3-parameter log-normal (LN3) (Rao and Hamed 2002), are considered. Parameters of these distributions are estimated using the MLE method and their goodness-of-fit was evaluated based on the AIC, AICc, BIC criteria and Kolmogorove-Smirnove (KS) test. Results of the best fit distributions are given in Table 1. The best fit distribution is that with the least AIc, AICc and BIC values. The Weibull and gamma distributions best fitted the GPF (AIC =724, AICc = 725, BIC = 725) and LCF (AIC =544, AICc = 544, BIC = 549) peak flow data. The KS test also supported the selection of the Weibull and gamma distributions to fit the GPF ($p$ value = 0.66) and LCF data (p value = 0.72). Past studies also found the Weibull distribution fitting annual maximum rainfall and river discharge (Aksoy 2000; Burn and Goel 2001; Chandrasekhar 2002; Ghorbani et al. 2010). The gamma distribution has also been used for flood frequency analyses (Haan 1977; Yevjevich and Obeysekera 1984; Singh 1998). Figure 4 illustrates the satisfactory fit of Weibull and gamma distributions to univariate GPF and LCF data.

**Table 2** Performance evaluation and goodness of fit test of copula functions

| | Copula | AIC | AICc | BIC | LL | Parameter values | White test | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | Statistic | *p* value |
| 1 | Gaussian[a] | −17.84 | −17.83 | −17.80 | −8.94 | 0.55 | 0.01 | 0.8 |
| 2 | Clayton | −17.45 | −17.43 | −17.41 | −8.74 | 0.90 | 0.02 | 0.79 |
| 3 | Rotated Clayton | −9.80 | −9.79 | −9.76 | −4.92 | 0.69 | 0.47 | 0.2 |
| 4 | Plackett | −16.73 | −16.71 | −16.70 | −8.38 | 5.54 | 0.01 | 0.95 |
| 5 | Frank | −16.45 | −16.43 | −16.41 | −8.24 | 3.82 | 0.01 | 0.92 |
| 6 | Gumbel | −13.50 | −13.49 | −13.47 | −6.77 | 1.50 | 0.13 | 0.6 |
| 7 | Rotated Gumbel | −17.64 | −17.63 | −17.61 | −8.34 | 1.54 | 0.18 | 0.49 |
| 8 | Student's t | −17.78 | −17.81 | −17.71 | −8.93 | 0.55, 100 | 1.44 | 0.64 |
| 9 | Symmetrised Joe-Clayton | −17 | −17.03 | −16.93 | −8.04 | 0.15,0.44 | 0.44 | 0.25 |

[a] indicating the best fit copula

### 4.3 Joint GPF-LCF Quantile Estimation

For the joint quantile estimation of the GPF-LCF, bivariate distributions based on nine copula functions, as given in Table 2, were built. According to the White test, all of the copulas had an acceptable fit (at 95% significant level) to the joint CDFs of the GPF-LCF data. The Gaussian copula was identified as the best fit copula based on the AIC, AICc and BIC criteria (AIC = −17.84, AICc = −17.83, BIC = −17.80). Visual inspection of the degree of fit of the Gaussian

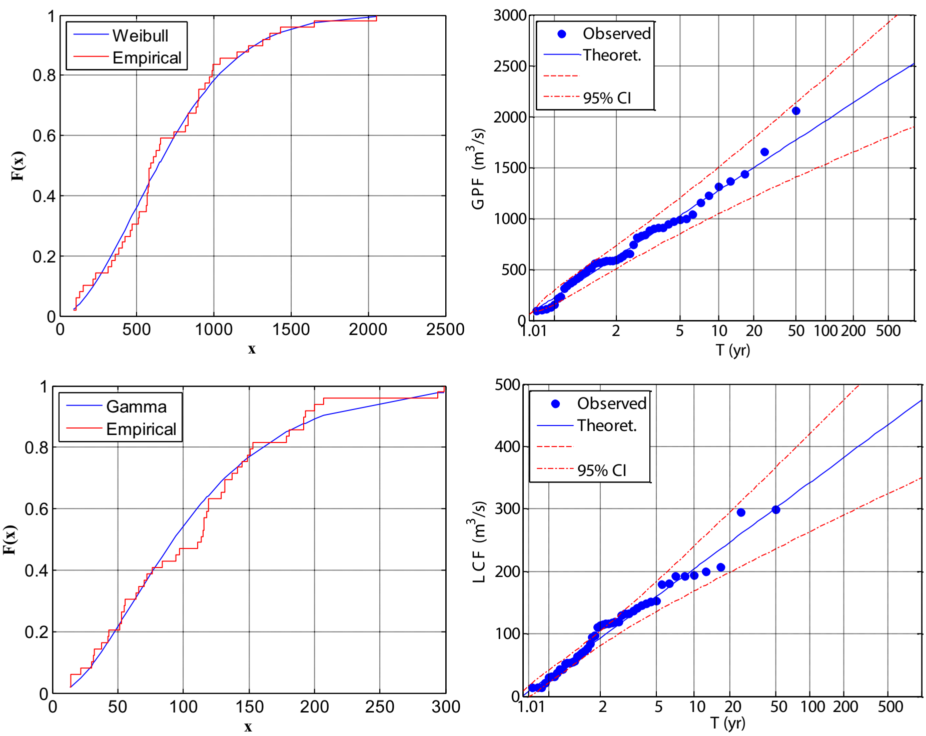


**Fig. 4** Univariate frequency analysis of GPF and LCF using Weibull and Gamma distributions

copula was also considered using the Q-Q plot. Figure 5 illustrates the empirical CDFs (Kn(t)) of the GPF-LCF pairs against that obtained by the Gaussian copula (K(t)). Clearly, empirical points fit will with Gaussian copula, which has a symmetric structure without tail dependency. It is selected because of its weaker dependencies of the most peak GPF and LCF events which may be attributed to the different GPF and LCF generation mechanisms.

For the joint quantile estimation of GPF-LCF using the Gaussian copula, the univariate CDFs derived by Weibull and gamma distributions were used (see the joint probabilities of occurrence and return periods of the GPF-LCF pairs in Figs. 6 and 7). It shows that most of GPF-LCF events (95%) occurs with a return period less than 100 -years, indicating the high risk of concurrent flooding in the Qezel Ozan and Shahrud Rivers. The comparison of the univariate and joint quantile estimates of the GPF and LCF in Figs. 4 and 7 indicates that univariate GPF and LCF quantiles occurred with shorter return periods than joint GPF-LCF quantiles. For example, the GPF quantile equal to 2000 $m^3/s$ and LCF quantile equal to 300 $m^3/s$ occurred with a 50-yr return period. However, the joint quantile of GPF-LCF equal to 2000–300 $m^3/s$ exceeded the 100-yr return period. For clarification, we constructed the conditional probability and return periods of the LCF quantiles given the various quantiles of GPF. Figure 8a, b shows the conditional return period and the probability of occurrence of LCF for different GPF quantiles and it demonstrates that the return period and probability of occurrence of LCF varies with GPF levels. The return period of LCF ≈ 330 $m^3/s$ is 50-yr return period for GPF ≥ 313 $m^3/s$, while the return period of the same LCF quantile is 200 years for GPF ≥ 895 $m^3/s$. Because the Sefidroud lake on the Sefidroud River is formed from two rivers, the univariate flood frequency analysis of GPF and LCF is not sufficient for optimal reservoir management and flood risk analysis around the lake. Thus the results of joint GPF-LCF quantile estimates can be helpful.

### 4.4 Evaluation of Uncertainty of the Joint GPF-LCF Quantiles Estimates

The joint quantile estimates of GPF-LCF have uncertainties stemming from sample size, and inappropriate univariate and copula model selection and their associated parameter

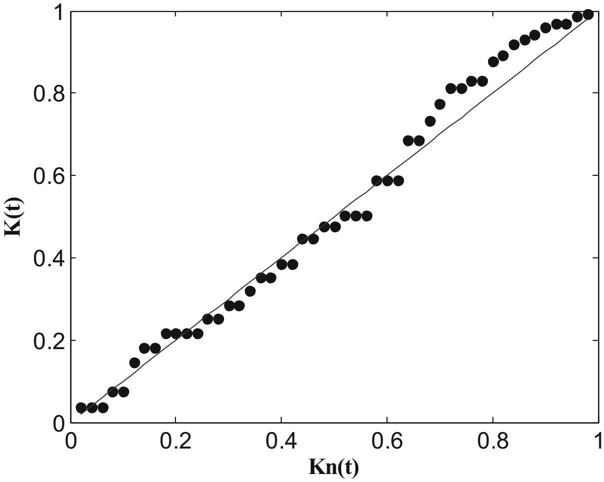


**Fig. 5** Q-Q plot of the Gaussian copula

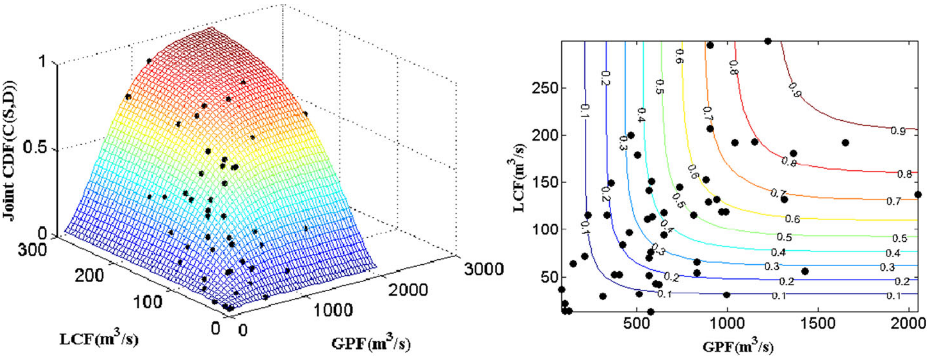


Fig. 6 Joint probability of occurrence of the GPF and LCF data

estimations. The non-parametric bootstrapping algorithm proposed in section 3.5 was used to estimate the uncertainties of the joint GPF-LCF quantiles. Figure 9 shows the uncertainty of bivariate GPF-LCF quantiles at various probability levels of $p = 0.9$ (≈10-yr return period), $p = 0.95$ (≈ 20-yr return period), $p = 0.98$ (≈ 50-yr return period) and $p = 0.99$ (≈ 100-yr return period). The 25–95% confidence regions (CRs) display the uncertainties of GPF-LCF quantile estimates. The parallel extent of CRs with p-level curves indicates the uncertainties caused by inefficient univariate distribution and copula function selection and associated parameter estimations (PU). Besides, the extent of CRs along the bottom left to top right direction demonstrates the sampling uncertainty (SU). The figures show that the dimension of CRs along the bottom left to top right direction increased for larger *p*-level curves. That implies that the GPF-LCF quantile estimates at larger hazard levels were more sensitive to sample size.

### 4.5 Mapping from Bivariate to Univariate FFA

Finally, we estimate the peak flows generated at the confluence of two rivers by mapping from the bivariate to univariate FFA. For this purpose, the return period of the sum of the GPF and LCF was derived from the fitted copula. Figure 10 illustrates the return periods of peak flows of Qezel Ozn and Shahrud Rivers against the sum of the two river peak flows.

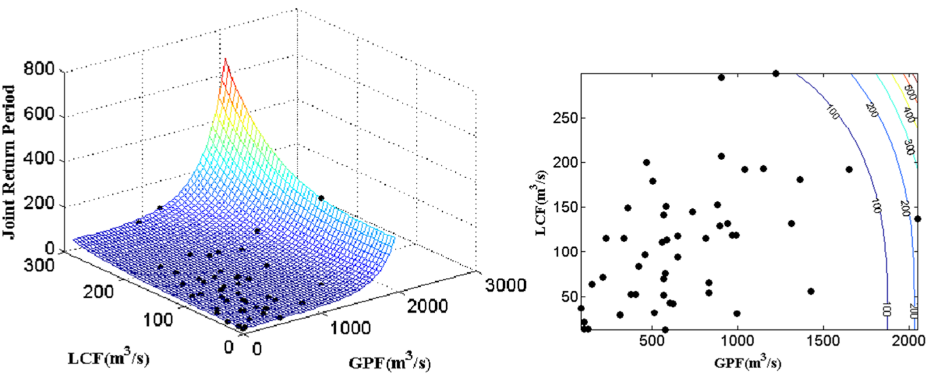


Fig. 7 Joint return period (RT) of the GPF and LCF

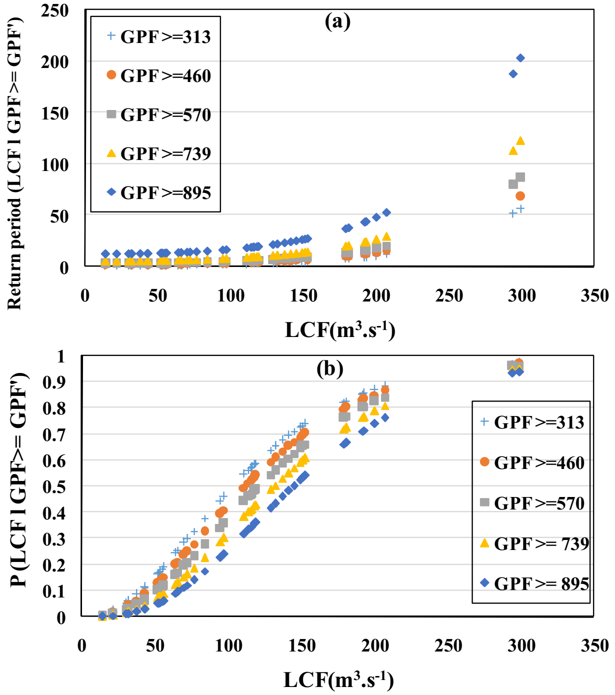


**Fig. 8** Conditional RT (**a**) and probability (**b**) of LCF given GPF

It shows that a significant number of historical peak flows >1000 $m^3$/s at the confluence of two rivers occurs with return periods less than 10 -years. This should be attached great importance from policymakers for flood control around the dam lake. Also, the same quantile of peak flows could occur with different return periods, which is caused by different combinations of GPF-LCF pairs and marginal probabilities. For example, both of 470–200 and 560–111 GPF-LCF events generated approximately the same inflow of 670 $m^3$/s to the Sefidrud reservoir (Fig. 10), however, they lead to different occurrence probabilities of marginal events and affect flood control differently. This finding shows that simply fitting univariate distributions to the annual maxima of the sums of peak flows does not satisfy the requirement of FFA for interconnected rivers. Thus, this study provides a simple guide for statistical properties analysis of interconnected rivers using copulas, which does not seem to have received much attention.

The flood dynamics and characteristics might be impacted by anthropogenic climate changes mainly due to increasing greenhouse gas emissions. As governed by the Clausius-Clapeyron relationship, the holding vapor capacity of atmosphere is increasing at rates with about 7% per degree warming, and thus lead to a substantial intensification of extreme precipitations and flooding in most areas of the globe. The change to flood characteristics is of great importance for water resources planning and management (Yin et al. 2018), which is not considered in this study due to lack of space. Future works can be devoted to investigate the bivariate flood quantiles under changing environments. Moreover, our developed methods can also be extended to estimate design floods at the

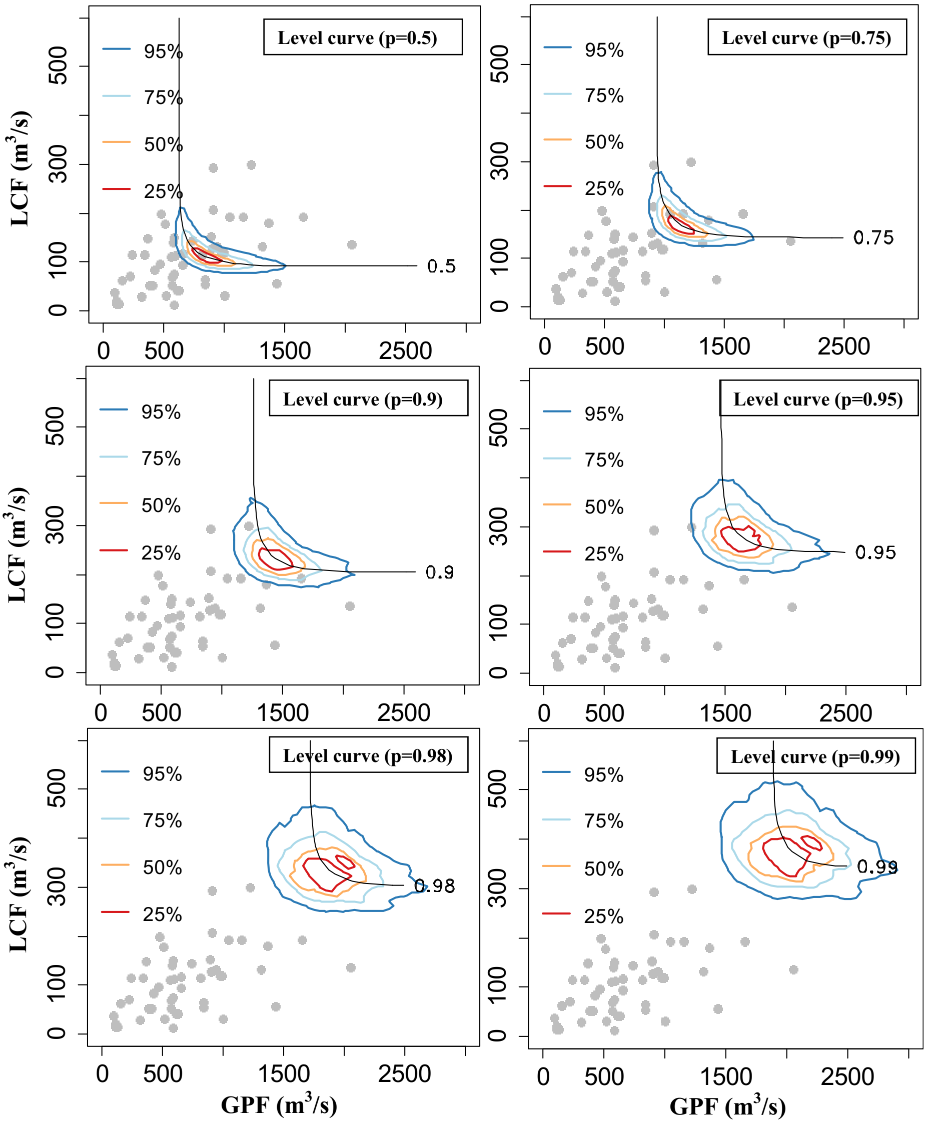


**Fig. 9** Uncertainty estimation of GPF-LCF quantiles at various probability levels

confluence of more than two rivers, in which a trivariate or higher-dimensional copulas can be further employed.

## 5 Conclusions

Flood frequency analysis considering the confluence of interconnected rivers is of great importance in water resources engineering, but it has not yet been addressed in the literature. In this study we used potential of copulas for peak flow quantile estimation at

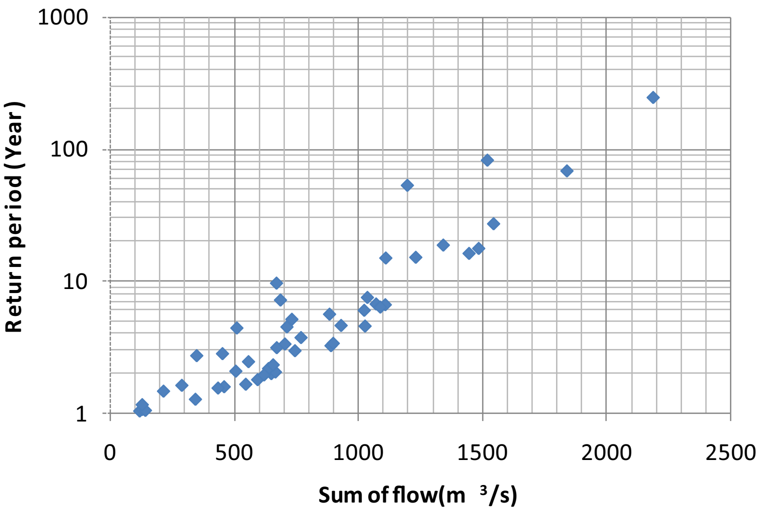


**Fig. 10** Univariate peak inflow estimation at the Sefidrud dam location

the confluence of two interconnected rivers through mapping from bivariate to univariate peak flow quantile estimation.

The following conclusions can be drawn from this study:

(1) The Gaussian copula with Weibull and gamma margins fits best for the estimation of the joint CDFs of the GPF-LCF data, based on its performance in terms of AIC, AICc, BIC. The efficiency of the copula function was also ascertained by White GOF test.

(2) The joint probability distribution of the GPF-LCF pairs was derived for flood frequency analysis at the confluence of two rivers.

(3) Bivariate quantile estimation uncertainties increased with increasing probability of hazard levels. The joint GPF-LCF quantiles at higher probability levels lead to higher uncertainties.

(4) For FFA at the confluence of the two rivers, mapping from the bivariate to univariate quantile estimation was done by plotting the joint return period of GPF-LCF pairs against their sum. Floods of greater than 1000 $m^3/s$ occurred with return periods less than 10 years at the confluence of two rivers.

(5) Flood quantiles with the same magnitudes and different impacts occurred with different return periods which cannot be isolated in univariate FFA of summed GPF-LCF pairs. As there is one-one mapping between the probability and quantiles in univariate FFA, the marginal probabilities of GPF and LCF are not taken into account by simply fitting the univariate probability distributions to the summed GPF-LCF pairs. Thus all of the summed GPF-LCF pairs with the same magnitudes are assigned by the same return period which is not the case in practice. By using the copula-based method, we are able to isolate the same magnitudes of peak flows with different impacts at the confluence of the rivers.

The results of this study can be used to determine the extent of flooding with different return periods around the dam lake to prepare dam safety plans and to adopt preventive measures for reducing flood damages especially in the residential areas adjacent to the lake dam. As there is no direct literature addressing FFA for interconnected rivers, this study can be a good reference for future works for appropriate FFA at the confluence of the interconnected rivers.

**Acknowledgements** We are very grateful to the editor and anonymous reviewers for their valuable comments and constructive suggestions that helped us to greatly improve the manuscript. In this study, Jiabo is supported by the Post-doctoral Innovative Talent Support Program of China (BX20200257), and partly funded by the Fundamental Research Funds for the Central Universities (2042020kf0003).

**Data Availability** Some or all data, models, or code generated or used during the study are available from the corresponding authors by request.

## References

Aksoy H (2000) Use of gamma distribution in hydrological analysis. Turkish J Eng Environ Sci 24:419–428
Breymann W, Dias A, Embrechts P (2003) Dependence structures for multivariate high-frequency data in finance. Quant Financ 3:1–14. https://doi.org/10.1080/713666155
Burn DH, Goel NK (2001) Flood frequency analysis for the Red River at Winnipeg. Can J Civ Eng 28:355–362. https://doi.org/10.1139/l00-122
Chandrasekhar S (2002) Stochastic problems in physics and astronomy. Rev Mod Phys 15:1–89. https://doi.org/10.1103/revmodphys.15.1
de Melo e Silva Accioly R, Chiyoshi FY (2004) Modeling dependence with copulas: a useful tool for field development decision process. J Pet Sci Eng 44:83–91. https://doi.org/10.1016/j.petrol.2004.02.007
Dodangeh E (2010) Regional low flow frequency analysis in Sefid-Roud basin. Dissertation, Isfahan University of Technology (IUT)
Dodangeh E, Soltani S, Sarhadi A, Shiau JT (2014) Application of L-moments and Bayesian inference for low-flow regionalization in Sefidroud basin, Iran. Hydrol Process 28:1663–1676. https://doi.org/10.1002/hyp.9711
Dodangeh E, Shahedi K, Shiau JT, Mirakbari M (2017) Spatial hydrological drought characteristics in Karkheh river basin, Southwest Iran using copulas. J Earth Syst Sci 126. https://doi.org/10.1007/s12040-017-0863-6
Dodangeh E, Shahedi K, Solaimani K et al (2019) Data-based bivariate uncertainty assessment of extreme rainfall-runoff using copulas: comparison between annual maximum series (AMS) and peaks over threshold (POT). Environ Monit Assess 191. https://doi.org/10.1007/s10661-019-7202-0
Dung NV, Merz B, Bárdossy A, Apel H (2015) Handling uncertainty in bivariate quantile estimation - an application to flood hazard analysis in the Mekong Delta. J Hydrol 527:704–717. https://doi.org/10.1016/j.jhydrol.2015.05.033
Fu G, Butler D (2014) Copula-based frequency analysis of overflow and flooding in urban drainage systems. J Hydrol 510:49–58. https://doi.org/10.1016/j.jhydrol.2013.12.006
Genest C, Rivest LP (1993) Statistical inference procedures for bivariate Archimedean copulas. J Am Stat Assoc 88:1034–1043. https://doi.org/10.1080/01621459.1993.10476372
Ghomian Z, Yousefian S (2017) Natural disasters in the middle-east and North Africa with a focus on Iran: 1900 to 2015. Heal Emergencies Disasters Q. https://doi.org/10.18869/nrip.hdq.2.2.53
Ghorbani MA, Ruskeepää H, Singh VP, Sivakumar B (2010) Flood frequency analysis using Mathematica. Turk J Eng Environ Sci. https://doi.org/10.3906/muh-1002-2
Golian S, Saghafian B, Farokhnia A (2012) Copula-based interpretation of continuous rainfall–runoff simulations of a watershed in northern Iran. Can J Earth Sci 49:681–691. https://doi.org/10.1139/e2012-011
Grimaldi S, Serinaldi F (2006) Asymmetric copula in multivariate flood frequency analysis. Adv Water Resour 29:1155–1167. https://doi.org/10.1016/j.advwatres.2005.09.005
Haan CT (1977) Statistical methods in hydrology. Stat Methods Hydrol 41:190–192. https://doi.org/10.1016/0022-1694(79)90123-9
Janga Reddy M, Ganguli P (2012) Application of copulas for derivation of drought severity-duration-frequency curves. Hydrol Process 26:1672–1685. https://doi.org/10.1002/hyp.8287
Joe H (1997) Multivariate models and dependence concepts. Monogr Stat Appl Probab 73:ALL
Kao SC, Govindaraju RS (2007) A bivariate frequency analysis of extreme rainfall with implications for design. J Geophys Res Atmos 112. https://doi.org/10.1029/2007JD008522

Kao SC, Govindaraju RS (2008) Trivariate statistical analysis of extreme rainfall events via the Plackett family of copulas. Water Resour Res 44. https://doi.org/10.1029/2007WR006261

Kavian A, Dodangeh S, Abdollahi Z (2016) Annual suspended sediment concentration frequency analysis in Sefidroud basin, Iran. Model Earth Syst Environ 2. https://doi.org/10.1007/s40808-016-0101-2

Lee T, Modarres R, Ouarda TBMJ (2013) Data-based analysis of bivariate copula tail dependence for drought duration and severity. Hydrol Process 27:1454–1463. https://doi.org/10.1002/hyp.9233

Madadgar S, Moradkhani H (2012) Drought analysis under climate change using copula. J Hydrol Eng 18:746–759. https://doi.org/10.1061/(asce)he.1943-5584.0000532

Martino L, Míguez J (2010) Generalized rejection sampling schemes and applications in signal processing. Signal Process 90:2981–2995. https://doi.org/10.1016/j.sigpro.2010.04.025

Mirakbari M, Ganji A, Fallah SR (2010) Regional bivariate frequency analysis of meteorological droughts. J Hydrol Eng 15:985–1000. https://doi.org/10.1061/(asce)he.1943-5584.0000271

Mirakbari M, Ganji A, Fallah SR (2012) Regional bivariate frequency analysis. J Hydrol Eng 15:985–1000. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000271

Modarres R (2008) Regional frequency distribution type of low flow in north of Iran by L-moments. Water Resour Manag 22:823–841. https://doi.org/10.1007/s11269-007-9194-8

Nelsen RB (2006) An introduction to copulas. Springer Science Business Media, Inc, New York

Rao AR, Hamed KH (2002) Flood frequency analysis

Rodrigo FS, Esteban-Parra MJ, Pozo-Vázquez D, Castro-Díez Y (1999) A 500-year precipitation record in southern Spain. Int J Climatol 19:1233–1253. https://doi.org/10.1002/(SICI)1097-0088(199909)19:11<1233::AID-JOC413>3.0.CO;2-L

Saghafian B, Mehdikhani H (2014) Drought characterization using a new copula-based trivariate approach. Nat Hazards 72:1391–1407. https://doi.org/10.1007/s11069-013-0921-6

Salvadori G, De Michele C (2011) Estimating strategies for multiparameter multivariate extreme value copulas. Hydrol Earth Syst Sci 15:141–150. https://doi.org/10.5194/hess-15-141-2011

Shiau JT, Modarres R (2009) Copula-based drought severity-duration-frequency analysis in Iran. Meteorol Appl 16:481–489. https://doi.org/10.1002/met.145

Shiau JT, Wang HY, Tsai CT (2006) Bivariate frequency analysis of floods using copulas. J Am Water Resour Assoc 42:1549–1564. https://doi.org/10.1111/j.1752-1688.2006.tb06020.x

Singh VP (1998) In: Singh VP (ed) Gamma distribution BT - entropy-based parameter estimation in hydrology. Springer Netherlands, Dordrecht, pp 202–230

Sklar A (1959) Fonctions de répartition à n dimensions et leurs marges. Publ Inst Statis, Univ Paris-VIII

Sungur EA, Nelson RB (2006) An introduction to copulas. J Am Stat Assoc 95:334. https://doi.org/10.2307/2669568

UNISDR (2015) Global assessment report on disaster risk reduction. International Stratergy for Disaster Reduction (ISDR). https://doi.org/9789211320282

Viglione A (2008) Model selection techniques for the frequency analysis of hydrological extremes: the MSClaio2008 R function. ftp3.ie.freebsd.org

White H (1982) Maximum likelihood estimation of misspecified models. Econometrica 50:1–25. https://doi.org/10.2307/1912526

Yevjevich V, Obeysekera JTB (1984) Estimation of Skewness of hydrologic variables. Water Resour Res 20:935–943. https://doi.org/10.1029/WR020i007p00935

Yin J, Guo S, He S, Guo J, Hong X, Liu Z (2018) A copula-based analysis of projected climate changes to bivariate flood quantiles. J Hydrol 566:23–42. https://doi.org/10.1016/j.jhydrol.2018.08.053

Zhang L, Singh VP (2006) Bivariate flood frequency analysis using the copula method. J Hydrol Eng 11:150–164. https://doi.org/10.1061/(asce)1084-0699(2006)11:2(150

Zhang L, Singh VP (2007) Bivariate rainfall frequency distributions using Archimedean copulas. J Hydrol 332:93–109. https://doi.org/10.1016/j.jhydrol.2006.06.033

Zhang L, Singh VP (eds) (2019) Copulas and their applications in water resources engineering. In: Copulas and their applications in water resources engineering. Cambridge University Press, Cambridge, pp i–ii

Zhang Q, Li J, Singh VP, Xu CY (2013) Copula-based spatio-temporal patterns of precipitation extremes in China. Int J Climatol 33:1140–1152. https://doi.org/10.1002/joc.3499

## Affiliations

**Esmaeel Dodangeh[1] • Vijay P. Singh[2] • Binh Thai Pham[3] • Jiabo Yin[4] • Guang Yang[5] • Amirhosein Mosavi[6,7]**

[1] Department of Watershed Management, Sari Agriculture Science and Natural Resources University, P.O. Box 737, Sari, Iran

[2] Department of Biological & Agricultural Engineering, Texas A & M University, 2117 TAMU, College Station, TX 77843-2117, USA

[3] Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam

[4] State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan 430072, China

[5] Department of Civil and Environmental Engineering, University of Wisconsin-Madison, 1415 Engineering Dr., Madison, WI 53706, USA

[6] Environmental Quality, Atmospheric Science and Climate Change Research Group, Ton Duc Thang University, Ho Chi Minh City, Vietnam

[7] Faculty of Environment and Labour Safety, Ton Duc Thang University, Ho Chi Minh City, Vietnam