University of Birjand
Water Harvesting Research Vol. 7, No. 2, Autumn & Winter 2024, Original paper p. 289-300
10.22077/JWHR.2024.8508.1160

Selection of the Best Copula Function in Bivariate Analysis of Water Resources Components (Case study: Siminehrood River Basin, Iran)

Fahimeh Sharifan[a], Yousef Ramezani[b&*], Mahdi Amirabadizadeh[b], Carlo De Michele[c]

[a]Ph.D. Student, Department of Water Engineering, University of Birjand, Birjand, Iran.
[b]Associate Professor, Department of Water Engineering, University of Birjand, Birjand, Iran.
[c]Professor, Department of Civil and Environmental Engineering, Politecnico di Milano, Milan, Italy.

*Corresponding Author, E-mail address: y.ramezani@birjand.ac.ir
Received: 02 September 2024/ **Revised**: 23 November 2024/ **Accepted**: 11 December 2024

Abstract
The copula function is a joint distribution of correlated random variables that are defined based on univariate marginal distributions. The aim of the present study is to select the best copula function to create joint probability distributions between the pair of parameters of precipitation- river discharge, river discharge- river salinity and river discharge- groundwater level in the Siminehrood River Basin. The necessity of using copula functions is the existence of correlation between the desired pair of parameters. For review the correlation between the pair of parameters, Kendall's tau statistic was used. Correlation between the precipitation- river discharge, river discharge- river salinity and river discharge- groundwater level were obtained 0.43, 0.64 and 0.44, respectively. After correlation evaluation, the marginal distribution of the parameters was investigated. Using the Kolmogorov-Smirnov and Anderson-Darling tests, statistical distribution functions for precipitation, river discharge, river salinity and groundwater level were obtained Lognormal, Gamma, Burr and Lognormal distributions, respectively. Then, by examining the dependence structure and the structure of copulas and using NSE, RMSE and BIAS evaluation criteria, Clayton's copula function was selected for all three pair of parameters, which was used to create a joint probability distribution between the pair of parameters in the Siminehrood River Basin.
Keywords: Clayton, Copula function, Correlation, Joint distribution, Marginal distribution.

1. Introduction
The latest approach introduced for the multivariate analysis of hydrological phenomena is the copula function, initially developed by Sklar in 1959. The utilization of copula functions in hydrological modeling signifies a significant advancement over traditional bivariate distribution functions. As noted by Salvadori and De Michele (2007), copulas facilitate the construction of multivariate distribution functions through the integration of various marginal distribution functions, allowing for a more accurate representation of dependence structures among multiple variables. Recently, copulas have emerged as a powerful tool for quantifying the dependence patterns between related variables.

The advantages of copulas in generating joint distributions are manifold:

1- Flexibility: Copulas allow for the selection of arbitrary marginal distributions, enabling researchers to model a wide variety of data types while adequately capturing the intrinsic characteristics of each variable.

2- Multi-Dimensional Capability: Unlike traditional approaches that are often limited to bivariate frameworks, copulas can seamlessly extend to three or more variables, making them particularly useful in complex systems where multiple interdependencies exist.

3- Separation of Margins and Dependence: Copulas provide the flexibility to analyze marginal distributions independently from the dependence structure.

This separation allows for a more focused analysis on each aspect, enhancing the overall modeling process (Serinaldi et al., 2009). The joint distributions derived from copulas provide a more accurate description of hydrological phenomena, eliminating reliance on subjective judgments or unfounded assumptions. By selecting an appropriate copula function model and accurately estimating key hydrological and meteorological features, it becomes possible to predict and simulate these phenomena more effectively. The results can then be applied to the design and management of water resource systems with greater precision. Previous research indicates that using univariate methods may result in either underestimating or overestimating of such phenomena and the associated damages (Raynal-Villasenor, 1987; Yuo et al., 2001).

Over the past two decades, copula functions have become increasingly popular among researchers for conducting multivariate analyses of hydro-climatological phenomena. Favre et al. (2004) employed bivariate copulas to clarify the relationship and dependence between discharge and flood volume. Their findings underscored the limitations of univariate analyses, highlighting the enhanced insights gained through copula approaches.

Shiau et al. (2006) investigated the two-dimensional joint frequency of flood peak discharge and flood volume using copula functions, further demonstrating the effectiveness of these tools in capturing complex relationships in hydrology. Zhang and Singh (2006) applied Archimedean copula functions to analyze the two distributions of various flood-related variables, including flood peak discharge and flood volume, flood peak discharge and flood duration, as well as flood volume and flood duration. Grimaldi and Sernaldi (2006a) constructed three distributions of flood event variables using a combination of symmetric and asymmetric mixed Archimedean functions, conducting extensive simulations to demonstrate the differences between these and traditional symmetric Archimedean correlations.

Salvadori and De Micele (2007) reviewed advancements in hydrological modeling that incorporate copula functions, such as the computation of conditional probabilities, bivariate simulations, return period of bivariate events under both conditional and unconditional scenarios, level curves of joint distributions, and the calculation of complexity variance. Additionally, Zhang and Singh (2007a) utilized the Gamble-Hoggard copula function to generate a trivariate distribution based on flood peak discharge, flood volume, and flood duration.

Ellouze-Gargouri and Bargaoui (2012) applied a geomorphological instantaneous unit hydrograph in conjunction with copula functions to analyze the relationships between the infiltration index parameter and various rainfall characteristics—including duration, depth, maximum intensity, and average intensity—in the Tunisia Basin. Initially, they employed the Nash model, which utilizes a gamma bivariate distribution, to compute the geomorphological instantaneous unit hydrograph. Following this, they derived the marginal distributions for the rainfall variables mentioned. To assess the infiltration index, they utilized copula functions, ultimately identifying the Frank copula as the most suitable for their analysis.

Schoelzel and Friederichs (2008) examined the dependence between precipitation and wind speed through the application of copula functions. They explored several types of copulas, including Archimedean, semi-elliptical, and normal copulas, concluding that copula functions serve as valuable tools in climatological research for capturing the dependencies among different meteorological variables.

Yoo and Cho (2019) employed copula functions to analyze the relationships among rainfall duration, total rainfall amount, and rainfall intensity. Findings a strong correlation structure between rainfall duration and depth, emphasizing the usefulness of copulas in computing joint probabilities and enhancing flood risk assessments.

Kao and Govindaraju (2008) explored a non-Archimedean copula function derived from the Plackett distribution family to model the temporal distribution of extreme rainfall events. This approach illustrated the flexibility of copulas in accommodating different types of dependence structures.

Additionally, Serinaldi et al. (2009) applied copula functions to conduct a probabilistic analysis of drought characteristics. This study demonstrates how copulas can be effectively utilized for modeling the joint behavior of drought-related variables, aiding in water resource management and planning.

Copula functions have been extensively employed in hydrology and water resources for various applications, including: rainfall frequency analysis (Grimaldi and Serinaldi, 2006b; Kao and Govindaraju, 2007; Zhang and Singh, 2007a; Kuhn et al., 2007; Keef et al., 2009; Kwon and Lall, 2016; Gao et al., 2018; Wei and Song, 2019), flood frequency analysis (Favre et al., 2006; Zhang and Singh, 2006; 2007; Xiao et al., 2016; Chen and Guo, 2019), drought frequency analysis (Shiau, 2010; et al., 2017; Hangshing and Dabral, 2018), analysis of marine storms (De Michele et al., 2007), flow simulation (Chen et al., 2015), theoretical analyses of multivariate boundary events (Salvadori and De Michele, 2010).

Based on the background of the research, it can be concluded that copula functions are a very efficient and effective tool for multivariate frequency analysis and simulation of hydrological events.

De Michele and Salvadori (2003) were pioneers in utilizing copula functions for rainfall frequency analysis within hydrological studies. Their research focused on hourly rainfall data from two rain gauge stations in La Presa, Italy, covering the statistical period from 1990 to 1996. Their findings revealed that the generalized Pareto distribution was the most suitable marginal distribution for the data, while the Frank copula function effectively captured the dependence structure between rainfall duration and intensity.

Lazoglou and Anagnostopoulou (2019) investigated the joint distribution of temperature and precipitation in Mediterranean region using both Archimedean and elliptical copula families. Their results highlighted Frank's copula as the most suitable for modeling temperature across most stations, while Gumbel's copula was preferred for precipitation. This study exemplifies the application of copulas in climatological research, facilitating better understanding of climate interactions.

Sobkowiak et al (2020) analyzed river flow in the upper of the Indus River Basin in Pakistan, utilizing copula functions to evaluate the joint dependence of river flow data. Their use of Pearson's coefficient to assess correlation among variables, alongside the Gumbel copula's performance reflects the effectiveness of copulas in hydrological data analysis and dependency modeling.

Janga Reddy and Ganguli (2012) assessed variability risks in precipitation and groundwater levels in the Manjara Basin, India, employing Archimedean copulas. This application underscores the ability of copulas to capture the interdependencies between groundwater level and precipitation, which is crucial for sustainable water resource management.

Nazeri Tahroudi et al. (2020), investigated the interrelationships between hydrological and meteorological droughts in the Zarinehrood Basin over a span of twenty-one years. By employing seven different copula functions, they identified Frank's copula as the most suitable for capturing the joint dependence between drought severity and duration.

Nazeri Tahroudi et al. (2022), focused on the Siminehrood River, where they aimed to simulate and predict rainfall and river flow deficiencies from 1992 to 2013. They determined that Clayton's copula function was the most effective for modeling the joint distribution of rainfall deficiency and river flow deficiency parameters. This study also revealed that river flow deficiency could be predicted with high accuracy based on rainfall deficiency,

Copula functions are a very efficient and useful tool for multivariable hydrological evaluation and simulation. Therefore, copula functions have been used in multivariate evaluation of different parameters. In this study, the best copula function between the pairs of precipitation- river discharge, river discharge- river salinity and river discharge-groundwater level in Siminehrood River Basin has been investigated and determined. Also, the parameters of the copula functions between these pair of parameters have been determined. The innovation of this research is the use of Archimedean, extreme-value, FGM and Plackett copulas, as well as the use of copula functions between the pairs of precipitation- river discharge, river discharge- river salinity and river discharge- groundwater level.

2. Materials and methods

2.1. Study area

The area examined in this research is the Siminehrood River Basin, located in the northwest region of Iran. This water basin spans from 45 degrees, 30 minutes, and 22 seconds east longitude to 46 degrees, 25 minutes, and 37 seconds east longitude, and from 36 degrees, 10 minutes, and 35 seconds north latitude to 37 degrees, 12 minutes, and 53 seconds north latitude. This region is part of the Lake Urmia Basin. The geographical location of this river is illustrated in Figure 1.

In this research, daily data on precipitation, river discharge, river salinity, and groundwater levels in the Siminehrood River Basin were utilized. Within the target basin, a rain gauge station was selected for measuring precipitation, a hydrometric station was established for monitoring river discharge and salinity values, and a piezometric well was chosen to detect changes in groundwater levels.

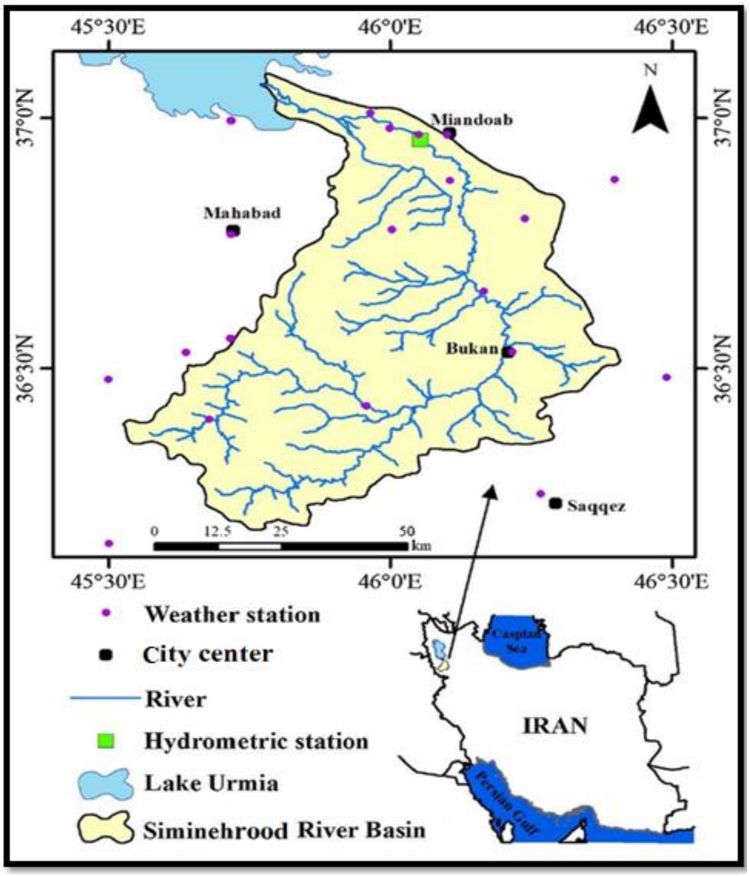

Fig. 1. Geographical location of the Siminehrood River Basin

In this research, precipitation data in the statistical period of 1988-2018, river discharge data in the statistical period of 1966-2019, river salinity data in the statistical period of 1966-2019, and groundwater levels in the statistical period of 2002-2018 have been used.

Dashband Bukan hydrometric station is one of the oldest and most complete hydrometric stations in the region. For this reason, it was chosen as a hydrometer station that represents the river discharge of the total basin (Nazeri Tahroudi et al., 2019). Dashband Bukan rain gauge station and Mirabad well were selected using the entropy theory to represent the precipitation and the groundwater level of the total basin, respectively (Nazeri Tahroudi et al., 2019). The statistical characteristics of the desired parameters are given in Table 1.

Table 1. Statistical characteristics of the evaluated data on an annual scale

Hydrometric Station	Dashband Bukan
Rain Gauge Station	Dashband Bukan
Piezometer	Mirabad
Mean Precipitation (mm)	296.1
Mean Discharge (m^3/s)	17.85
Average Salinity (μS/cm)	368.5
Average Groundwater Level (m)	5.82
Standard Deviation of Precipitation (mm)	20.33
Standard Deviation of Discharge (m^3/s)	18.84
Standard Deviation of Salinity (μS/cm)	158.5
Standard Deviation of Groundwater Level (m)	0.67

2.2. Selection of marginal distributions

First, the statistical distributions of the desired water resources components should be estimated. In this study, EasyFit 5.6 software was used to estimate statistical distributions and also their parameters. After fitting the statistical distributions on the desired parameters, the goodness of fit of the distributions should be calculated using the Kolmogorov-Smirnov and Anderson-Darling tests. If the results are significant at the level of 5%, that distribution will be confirmed for the desired parameters (Nazeri Tahroudi et al., 2019).

2.3. Copula functions

Copula functions are mathematical constructs that enable the combination of marginal distributions of various random variables to form a multivariate distribution. The multivariate cumulative distribution function (CDF) can be expressed in terms of its univariate marginal CDFs along with the copula function. A multivariate copula is essentially the joint CDF of standard uniform random variables, capturing the dependence structure between the variables. For a copula$C: [0,1]^2 \rightarrow [0,1]$, it must satisfy specific properties, which are as follows:

Boundary Conditions:

$$C(u,0) = C(0,v) = 0\,, C(1,v) = v\,, C(u,1) = u \quad (1)$$

Associativity Condition:

If $u_1 \geq v_1$ and $u_2 \geq v_2$ and $u_1, u_2, v_1, v_2 \in [0,1]$

$$C(u_1,u_2) + C(v_1,v_2) - C(u_1,u_2) - C(v_1,v_2) \geq 0 \quad (2)$$

Sklar's theorem serves as a fundamental principle in copula theory, stating that any n-dimensional distribution function F can be expressed using a copula C as follows:

$$F(x_1,\dots,x_n) = C(F_1(x_1),\dots,(F_n(x_n)) \quad (3)$$

where $F_1,\dots, F_n$ represent the multivariate distribution functions. If these multivariate distributions are continuous, the copula function C is unique and can be expressed in the following way:

$$(u_1,\dots,u_n) = F\left(F_1^{-1}(u_1),\dots,F_n^{-1}(u_n)\right) \quad (4)$$

where the quantile F_{k-1} is defined as $F_{k-1} = \inf\{x \in R | F_k(x) \geq u_k)\}$. It can also be stated that if C is a copula function and $F_1,\dots,F_n$ are univariate distribution functions, then the function F is an n-dimensional distribution function with margins $F_1,\dots,F_n$. The copulas are divided into different groups such as Archimedean, elliptic and extreme-value, each of these groups includes different copulas.

In the context of bivariate copula modeling, two correlated random variables X and Y are considered, which have respective marginal probability density functions $f_x\left(x; a_1, a_2, \dots, a_p\right)$ and $f_y\,(y; \lambda_1, \lambda_2, \dots, \lambda_r)$. Here, $a_1, a_2, \dots, a_p$ are parameters associated with the distribution of $f_x\,(x)$, and $\lambda_1, \lambda_2, \dots, \lambda_r$are parameters associated with the distribution of $f_y\,(y)$.

To estimate the parameters from n independent pairs of observations, the log likelihood function for X and Y is $lnL_x\,(x; a_1, a_2, \dots, a_p)$ and $lnL_y\,(y; \lambda_1, \lambda_2, \dots, \lambda_r)$ are maximized individually in order to estimate the parameters.

$a_1, a_2, \dots, a_p$ and $\lambda_1, \lambda_2, \dots, \lambda_r$ are the estimated parameters. The log likelihood function of the joint probability density function of functions $f_{X\,Y}\,(x\,,y)$ is defined as follows:

$$\begin{aligned} \ln L(x,y; \hat{a_1}, \hat{a_2}, \dots, \hat{a_p}, \hat{\lambda_1}, \hat{\lambda_2}, \dots, \hat{\lambda}, \theta) \\ = \ln L_c(x,y; F_X(x), F_Y(y), \theta) \\ + \ln L_X(\ln L(x; \hat{a_1}, \hat{a_2}, \dots, \hat{a_p}) \\ + ln\;\; L_Y(y; \hat{\lambda_1}, \hat{\lambda_2}, \dots, \hat{\lambda_r}) \end{aligned} \quad (5)$$

In equation (5), the log likelihood function of the copula density function is lnL_c. By placing the estimated values for $\lambda_1, \lambda_2, \dots, \lambda_r$ and $a_1, a_2, \dots, a_p$ in equation 5, the log likelihood function lnL to calculate the estimated copula parameter $\widehat{\theta}$ is maximized.

In this research, used of fitting 10 different copula functions to desired pair of parameters, aiming to create a suitable bivariate distribution that first must calculated the correlation between these parameters. To quantify the correlation between the parameters, the Kendall's tau statistic is employed. This non-parametric measure of correlation is particularly useful in the context of copulas. This statistic (equation 6) is used in all studies of copula functions (De Michele and Salvadori, 2003; Czado, 2010; Ramezani et al., 2019).

$$\tau = \binom{N}{2} \sum_{i<j} sign\,[(x_i - x_j)(y_i - y_j)]) \quad (6)$$

In equation (6), *N* is the number of data, *sign()* is the sign function, and *x* and *y* are the values of the pair of desired parameters.

To determine the best copula, the parameters of the copula function are obtained using the IFM method (Mirabbasi et al., 2012; Nazeri Tahroudi et al., 2020). The IFM method consists of two steps: first: marginal distributions are calculated from the observed values; and second: the copula dependence parameter (θ) is estimated by maximizing the log-likelihood function (Mirabbasi et al., 2012). The log-likelihood function is as follows:

$$l_a = \sum_{k=1}^{n} \ln c(F_1(x_1^i; a_1), \dots, F_p(x_p^i; a_p); \theta) + \sum_{i=1}^{n}\sum_{j=1}^{n} \ln f_j(x_j^i; a_j) \quad (7)$$

First, *n* separate estimates can be obtained for each univariate marginal distribution, for example:

$$\bar{a_j} = argmax \sum_{i=1}^{n} f_j(x_j^i; a_j) \quad (8)$$

Then the estimation of θ with previous marginal functions will be as follows:

$$\bar{\theta} = argmax \sum_{k=1}^{n} \ln c(F_1(x_1^i; a_1), \dots, F_p(x_p^i; a_p); \theta) \quad (9)$$

The optimal copula is then selected by comparing the results of each copula with those of the empirical copula.

For a joint two-dimensional copula, the empirical copula of the measured data (u_i, v_i) is defined as follows:

$$C_e(u_i, v_i) = \frac{1}{n}\sum_{i=1}^{n} I\left(\frac{Q_i}{n+1} \le u_i, \frac{P_i}{n+1} \le v_i\right) \quad (10)$$

In this equation, C_e is the empirical copula, n is the number of measured data and $I(A)$ is the indicator variable of the logic expression A. If the expression A is true, it will take the value of one and if it is false, it will take the value of zero. Q_i and P_i are the ranks of the *i-th* measured data corresponding to the desired pair of copulas.

Table 2. Copula functions used in this research

Family	C(u,v)	The domain of $\boldsymbol{\theta}$
Ali-Mikhail-Haq (AMH)	$\frac{uv}{1-\theta(1-u)(1-v)}$	$-1 \le \theta \le 1$
Clayton	$(u^{-\theta}+v^{-\theta}-1)^{-1/\theta}$	$\theta \ge 0$
Frank	$-\frac{1}{\theta}\ln[1 + \frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{e^{-\theta}-1}]$	$\theta \ne 0$
Galambos	$uv\exp\{[(-lnu)^{-\theta} + (-lnv)^{-\theta}]^{-\frac{1}{\theta}}\}$	$\theta \ge 0$
Gumbel-Hougaard (GH)	$\exp\{-[(-lnu)^{\theta} + (-lnv)^{\theta}]^{\frac{1}{\theta}}\}$	$\theta \ge 1$
Plackett	$\exp\{-[(-lnu)^{\theta} + (-lnv)^{\theta}]^{1/\theta}\}$	$\theta \ge 0$
Farlie-Gumbel-Morgenstern FGM)	$\frac{1}{2}\frac{1}{\theta-1}\{1+(\theta-1)(u+v) -[(1+(\theta-1)(u+v))^2 -4\theta(\theta-1)uv]^{\frac{1}{2}}\}$	$-1 \le \theta \le 1$
Gumbel-Barnett (GB)	$uv\,exp\{-\theta(logu)(logv)\}$	$0 < \theta \le 1$
A12	$\frac{1}{1+[(\frac{1}{u}-1)^{\theta}+(\frac{1}{v}-1)^{\theta}]^{1/\theta}}$	$\theta \ge 1$
Joe	$1-[(1-u)^{\theta}+(1-v)^{\theta} -(1-u)^{\theta}(1-v)^{\theta}]^{1/\theta}$	$\theta > 1$

2.4. Goodness of fit tests for copulas

In this research, Nash-Sutcliffe efficiency (NSE), root mean square error (RMSE) and BIAS evaluation criteria were used to select the best copula function.

$$NSE = 1 - \frac{\sum_{i=1}^{n}|O_i - p_i|^2}{\sum_{i=1}^{n}|O_i - \bar{O}_i|^2} \quad (11)$$

$$RMSE = \sqrt{\frac{(p_i - O_i)^2}{n}} \quad (12)$$

$$BIAS = \frac{\sum_{i=1}^{n}(p_i - O_i)}{n} \quad (13)$$

In these equations, $\bar{p_i}$ and $\bar{O_i}$ are the average of the predicted and measured parameters in the i-th month and *n* is the number of data in each parameter, respectively.

The best result for the selection of copula functions will be created when RMSE and BIAS tend to zero and NSE to one.

3. Results and Discussion

3.1. Investigating the correlation between the pair of parameters

As mentioned before, the correlation between the pair of desired parameters has been calculated using the Kendall's tau statistic. These results are given in Table 3.

Table 3. The Kendall's tau values for the pair of parameters

The pair of parameter	Kendall's Tau
Precipitation - River discharge	0.43
River discharge - River salinity	0.64
River discharge - Groundwater level	0.44

According to Table 3, for the pairs of precipitation- river discharge, river discharge- river salinity and river discharge- groundwater level, the correlation coefficient is more than 0.4 and is within the acceptable range (Tahroudi et al., 2020). This coefficient between river discharge and river salinity parameters has a high value.

In order to use copula functions, the existence of correlation between the pair of parameters is required, which is confirmed in this research according to the Kendall's tau correlation coefficients, the use of copula functions and its bivariate calculations.

3.2. Selection of marginal distributions

Using the EasyFit software, the desired parameters were fitted to various statistical distribution functions. The best fit among these distribution functions was selected based on the Kolmogorov-Smirnov and Anderson-Darling tests. If the fit of a distribution function is confirmed at a significance level of 5% by these tests, that statistical distribution is accepted (Wen et al., 2022).

The results obtained from estimating the statistical distribution of precipitation, river discharge, river salinity, and groundwater level parameters using Kolmogorov-Smirnov and Anderson-Darling tests showed that the statistical distributions lognormal, Wakeby, and Fatigue Life for the precipitation parameter, Gamma, Log-Pearson 3, Wakeby, and Fatigue Life for the river discharge parameter, Burr, Pearson 5, Wakeby and Log Logistic for the river salinity parameter, and Lognormal, Johnson SU and Dagum for the groundwater level parameter were selected as fitted statistical distributions with parameters, and the Lognormal, Gamma, Burr and Lognormal statistical distributions were selected as the best distribution by taking rank 1 for both tests in EasyFit software for the precipitation, river discharge, river salinity, and groundwater level parameters, respectively.

After selecting and determining the marginal distribution according to the parameters, using the copula functions in Table 2, the joint distribution function was constructed. As mentioned before, in order to use the copula function, there must be a correlation between the pair of parameters. After confirming the existence of correlation between the pair of parameters, the copula dependence parameter was calculated using the IFM method.

To select the best copula, NSE, RMSE and BIAS evaluation criteria were used to compare the probability values obtained from theoretical copula functions and empirical copula functions. The results of the goodness-of-fit tests for the copula functions for each desired parameter are presented in Tables 4, 5, and 6.

According to Tables 4, 5, and 6, it can be seen that Clayton's copula has a better performance in all three pairs of precipitation- river discharge, river discharge- river salinity and river discharge- groundwater level, with the least error.

The selection of copulas that correspond to the parameters is influenced by the range of dependence levels they can effectively describe. For example, the Gumbel-Hougaard (GH) copula is applicable only for positive correlations. In contrast, the Clayton and Frank copulas can accommodate both positive and negative dependencies. AMH copula is suitable for weak dependence, specifically within the range of $-0.1807 < \tau < 0.3333$, while the Farlie-Gumbel-Morgenstern (FGM) copula is appropriate for the range $-0.22 < \tau < 0.22$ (Nelsen, 2006).

The results obtained from the correlation of the parameters show the positive dependence between the parameters. Also, due to the better performance of Clayton's copula compared to other copulas, Clayton's copula is selected for all three pair of investigated parameters to create joint distributions. The results of this research are consistent with the following studies:

Table 4. The results of the goodness of fit test of copula functions for precipitation (mm) and river discharge (m^3/s)

Evaluation criteria	AMH	Clayton	Frank	Galambos	GH	Plackett	FGM	GB	A12	Joe
NSE	0.82	0.9	0.53	0.82	0.8	0.84	0.7	0.58	0.82	0.85
RMSE	0.12	0.09	0.19	0.12	0.12	0.11	0.13	0.18	0.12	0.1
BIAS	0.11	0.08	0.18	0.11	0.12	0.1	0.13	0.17	0.11	0.1
Teta	0.99	2.17	-0.91	0.73	1.4	5.23	1	0.1	1	2.08

Table 5. The results of the goodness-of-fit test of copula functions for river discharge (m^3/s) and river salinity (μS/cm)

Evaluation criteria	**AMH**	**Clayton**	**Frank**	**Galambos**	**GH**	**Plackett**	**FGM**	**GB**	**A12**	**Joe**
NSE	0.72	0.93	0.43	0.92	0.8	0.87	0.64	0.42	0.85	**0.89**
RMSE	0.15	0.07	0.21	0.08	0.13	0.1	0.17	0.21	0.1	**0.09**
BIAS	0.13	0.06	0.19	0.07	0.11	0.09	0.15	0.19	0.09	**0.08**
Teta	1	6.93	-0.33	5.32	1.86	20	1	0.1	1.6	**5.11**

Table 6. The results of the goodness of fit test of copula functions for river discharge (m^3/s) and groundwater level (m)

Evaluation criteria	**AMH**	**Clayton**	**Frank**	**Galambos**	**GH**	**Plackett**	**FGM**	**GB**	**A12**	**Joe**
NSE	0.73	0.94	0.47	0.92	0.82	0.88	0.66	0.43	0.9	**0.93**
RMSE	0.15	0.07	0.21	0.08	0.12	0.09	0.16	0.21	0.07	**0.07**
BIAS	0.13	0.06	0.18	0.07	0.11	0.09	0.15	0.19	0.08	**0.06**
Teta	0.99	6.7	-0.025	20	1.95	20	1	0.1	8.75	**7.1**

Salleh et al (2016) evaluated of the bivariate flood frequency for the Johor River over a 35-year period and demonstrated the applicability of copula functions for capturing the relationship between flood peak discharge and flood duration. By selecting Clayton's copula, they showed that it effectively maintains the dependence structure of these flood variables, which is crucial for reliable flood risk assessment and management.

Also, Ayantobo and Song (2019) focused on the evaluation of multivariate drought in China from 1961 to 2013, where Clayton's copula function again emerged as the best performer.

The use of trivariate copulas highlights the importance of understanding the interactions between multiple drought-related variables, which can enhance the accuracy of drought frequency estimation. In another research, the application of copula functions to analyze the relationship between salinity and groundwater levels in the Kahriz and Tajsoo Lake Urmia sub-basins further confirms the robustness of Clayton's copula.

This study illustrates how copulas can facilitate an understanding of complex environmental interactions, particularly in regions where salinity and groundwater levels are critical for agricultural and ecological sustainability (Zavareh et al., 2023).

Figures 2, 3 and 4 show the scatter plots of the empirical copula values against the Clayton copula for three pairs of precipitation- river discharge, river discharge-river salinity and river discharge-groundwater level compared to the 45° axis.

According to Figures 2, 3, and 4, it is evident that for all three pair of parameters, the values predicted by the Clayton copula function are generally lower than those of the empirical copula. In some instances, the predicted values align closely with the 45-degree line. However, the proximity of the predicted values to this line suggests that the performance of the Clayton copula is acceptable.

Therefore, it is suitable for the joint analysis of the pairs of precipitation- river discharge, river discharge- river salinity, and river discharge- groundwater level in the Siminehrood River Basin.

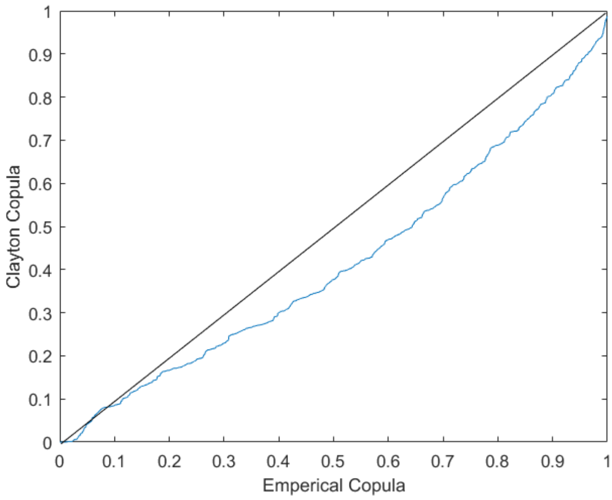

Fig. 2. Scatter plot of the values of the empirical copula against the Clayton copula for the precipitation-river flow

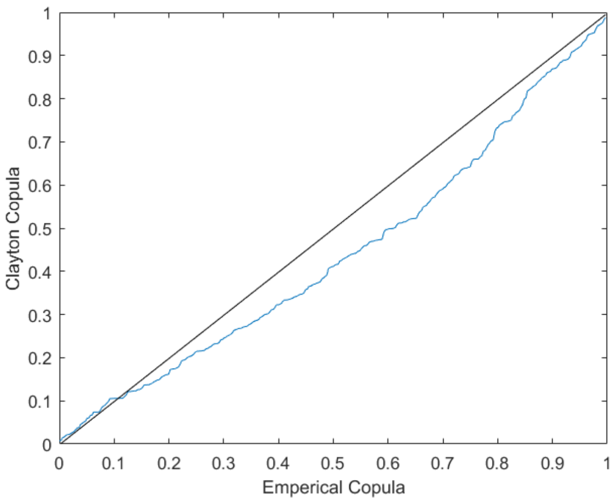

Fig. 3. Scatter plot of the values of the empirical copula against the Clayton copula for the river discharge-river salinity

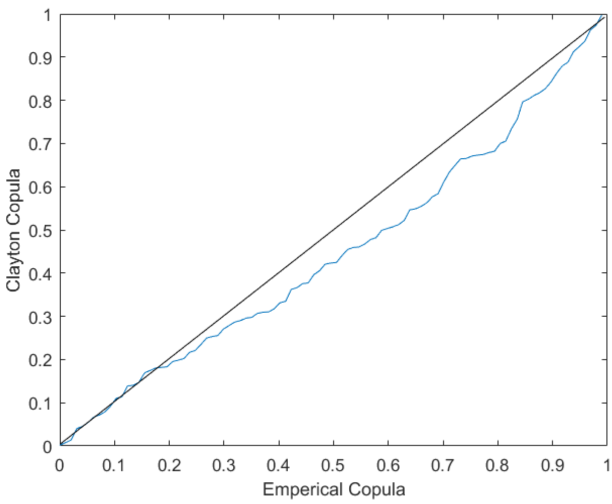

Fig. 4. Scatter plot of the values of the empirical copula against the Clayton copula for the river flow-groundwater level

4. Conclusion

In most of the hydrological researches, univariate evaluation and analysis and independence between parameters have been used due to their ease. While one of the problems of hydrological researchers is to investigate the nonlinear behavior of random variables and their multivariate probability distribution. For this reason, realistic methods should be used.

The advantages of using copula functions to generate joint distributions are flexibility in choosing the desired marginal distribution and dependence structure, extension to more than two variables, and analysis of the marginal distribution and dependence structure separately, while its disadvantages are the complexity of calculations, especially in dimensions higher than 2, and the need for correlation between the variables under study.

This study employed multivariate analysis of copula functions to determine the best copula that can create joint probability distributions for the aforementioned parameters in the Siminehrood River Basin. After confirming the existence of correlation between the pair of parameters using Kendall's tau statistic, the marginal distributions were established through the Kolmogorov-Smirnov and Anderson-Darling goodness-of-fit tests. Upon evaluating the dependence between parameters and copula structures, the Clayton copula was selected as the best choice in the Siminehrood River Basin for the pair of parameters (precipitation- river discharge, river discharge- river salinity, and river discharge-groundwater level).

Considering that the Siminehrood River is a permanent river at the Dashband Bukan hydrometric station, and since the river discharge is correlated with precipitation, river salinity, and groundwater level, it can be said that using two-dimensional copula functions, river discharge, river salinity, and groundwater level can be simulated and predicted. Also, the results of comparing different copula functions with the empirical copula function indicate the good and appropriate efficiency of copula functions in predicting and simulating hydrological variables. Simultaneous evaluation of precipitation- river discharge and river discharge- river salinity and river discharge-groundwater level due to their correlation can be very useful in managing the Siminehrood River Basin.

According to the importance of water resources, the use of copula functions is a very effective tool for bivariate modeling of water resources components, because copula functions maintain the dependence structure of water resources components effectively.

5. Disclosure Statement

The authors reported no potential conflict of interest.

6. References

Ayantobo, O. O., Li, Y., & Song, S. (2019). Copula-based trivariate drought frequency analysis approach in seven climatic sub-regions of mainland China over 1961–2013. *Theoretical and Applied Climatology, 137*, 2217-2237.

Chen, L., Guo, S., Chen, L., & Guo, S. (2019). Copula-based flood frequency analysis. *Copulas and Its Application in Hydrology and Water Resources*, 39-71.

Chen, L., Singh, V. P., Guo, S., Zhou, J., & Zhang, J. (2015). Copula-based method for multisite monthly and daily streamflow simulation. *Journal of Hydrology, 528*, 369-384.

Czado, C. (2010, May). Pair-copula constructions of multivariate copulas. In Copula Theory and Its Applications: Proceedings of the Workshop Held in Warsaw, 25-26 September 2009 (pp. 93-109). Berlin, Heidelberg: Springer Berlin Heidelberg.

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., Passoni, G., & Vezzoli, R. (2007). A multivariate model of sea storms using copulas. *Coastal Engineering, 54*(10), 734-751.

Ellouze-Gargouri, E., & Bargaoui, Z. (2012). Runoff estimation for an ungauged catchment using geomorphological instantaneous unit hydrograph (GIUH) and copulas. *Water resources management, 26*(6), 1615-1638.

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

Genest, C., & Favre, A. C. (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of hydrologic engineering, 12*(4), 347-368.

Grimaldi, S., & Serinaldi, F. (2006b). Asymmetric copula in multivariate flood frequency analysis. *Advances in Water Resources, 29*(8), 1155-1167.

Guo, E., Zhang, J., Si, H., Dong, Z., Cao, T., & Lan, W. (2017). Temporal and spatial characteristics of extreme precipitation events in the Midwest of Jilin Province based on multifractal detrended fluctuation analysis method and copula functions. *Theoretical and Applied Climatology, 130*, 597-607.

Hangshing, L., & Dabral, P. P. (2018). Multivariate frequency analysis of meteorological drought using copula. *Water resources management, 32*, 1741-1758.

Hao, C., Zhang, J., & Yao, F. (2017). Multivariate drought frequency estimation using copula method in Southwest China. *Theoretical and Applied Climatology, 127*, 977-991.

Janga Reddy, M., & Ganguli, P. (2012). Risk assessment of hydroclimatic variability on groundwater levels in the Manjara basin aquifer in India using Archimedean copulas. *Journal of Hydrologic Engineering, 17*(12), 1345-1357.

Kao, S. C., & Govindaraju, R. S. (2007). A bivariate frequency analysis of extreme rainfall with implications for design. *Journal of Geophysical Research: Atmospheres, 112*(D13).

Kao, S. C., & Govindaraju, R. S. (2008). Trivariate statistical analysis of extreme rainfall events via the Plackett family of copulas. *Water Resources Research, 44*(2).

Kao, S. C., & Govindaraju, R. S. (2010). A copula-based joint deficit index for droughts. *Journal of Hydrology, 380*(1-2), 121-134.

Keef, C., Svensson, C., & Tawn, J. A. (2009). Spatial dependence in extreme river flows and precipitation for Great Britain. *Journal of Hydrology, 378*(3-4), 240-252.

Kuhn, G., Khan, S., Ganguly, A. R., & Branstetter, M. L. (2007). Geospatial–temporal dependence among weekly precipitation extremes with applications to observations and climate model simulations in South America. *Advances in Water Resources, 30*(12), 2401-2423.

Kwon, H. H., & Lall, U. (2016). A copula-based nonstationary frequency analysis for the 2012–2015 drought in California. *Water Resources Research, 52*(7), 5662-5675.

Latif, S., & Mustafa, F. (2020). Copula-based multivariate flood probability construction: a review. *Arabian Journal of Geosciences, 13*(3), 132.

Lazoglou, G., & Anagnostopoulou, C. (2019). Joint distribution of temperature and precipitation in the Mediterranean, using the Copula method. *Theoretical and applied climatology, 135*, 1399-1411.

Mirabbasi, R., Fakheri-Fard, A., & Dinpashoh, Y. (2012). Bivariate drought frequency analysis using the copula method. *Theoretical and applied climatology, 108*, 191-206.

Nazeri Tahroudi, M., Ramezani, Y., De Michele, C., & Mirabbasi, R. (2020). A new method for joint frequency analysis of modified precipitation anomaly percentage and streamflow drought index based on the conditional density of copula functions. *Water Resources Management, 34*, 4217-4231.

Nazeri Tahroudi, M., Ramezani, Y., De Michele, C., & Mirabbasi, R. (2022). Application of copula functions for bivariate analysis of rainfall and river flow deficiencies in the siminehrood River Basin, Iran. *Journal of Hydrologic Engineering, 27*(11), 05022015.

Nazeri Tahroudi, M., Shahidi, A., Khashei Seyuki, A., & Ramezani, Y. (2019). Designing the monitoring network of rain gauge stations using chaos theory (case study: Lake Urmia basin). *Iran Irrigation and Drainage Journal, 13*(2), 296-308. [InPersian]

Nelsen, R. B. (2006). An introduction to copulas. Springer.

Ozga-Zielinski, B., Ciupak, M., Adamowski, J., Khalil, B., & Malard, J. (2016). Snow-melt flood frequency analysis by means of copula based 2D probability distributions for the Narew River in Poland. *Journal of Hydrology: Regional Studies, 6*, 26-51.

Qian, L., Wang, H., Dang, S., Wang, C., Jiao, Z., & Zhao, Y. (2018). Modelling bivariate extreme precipitation distribution for data-scarce regions using Gumbel–Hougaard copula with maximum entropy estimation. *Hydrological Processes, 32*(2), 212-227.

Ramezani, Y., Nazeri Tahroudi, M., & Ahmadi, F. (2019). Analyzing the droughts in Iran and its eastern neighboring countries using copula functions. *IDŐJÁRÁS/Quarterly Journal of the Hungarian Meteorological Service, 123*(4), 435-453.

Renard, B., & Lang, M. (2007). Use of a Gaussian copula for multivariate extreme value analysis: Some case studies in hydrology. *Advances in Water Resources, 30*(4), 897-912.

Salleh, N., Yusof, F., & Yusop, Z. (2016, June). *Bivariate copulas functions for flood frequency analysis.* In AIP conference proceedings (Vol. 1750, No. 1). AIP Publishing.

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).

Schoelzel, C., & Friederichs, P. (2008). Multivariate non-normally distributed random variables in climate research–introduction to the copula approach. *Nonlinear Processes in Geophysics, 15*(5), 761-772.

Serinaldi, F., Bonaccorso, B., Cancelliere, A., & Grimaldi, S. (2009). Probabilistic characterization of drought properties through copulas. *Physics and Chemistry of the Earth, 34*(10-12), 596-605.

Shiau, J. T. (2006). Fitting drought duration and severity with two-dimensional copulas. *Water resources management, 20*, 795-815.

Shiau, J. T., Wang, H. Y., & Tsai, C. T. (2006). Bivariate frequency analysis of floods using copulas1. JAWRA *Journal of the American Water Resources Association, 42*(6), 1549-1564.

Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges. In Annales de l'ISUP, 8(3), 229-231.

Sobkowiak, L., Perz, A., Wrzesiński, D., & Faiz, M. A. (2020). Estimation of the river flow synchronicity in the upper indus river basin using copula functions. *Sustainability*, *12*(12), 5122.

Tahroudi, M. N., Ramezani, Y., De Michele, C., & Mirabbasi, R. (2020). Analyzing the conditional behavior of rainfall deficiency and groundwater level deficiency signatures by using copula functions. *Hydrology Research, 51*(6), 1332-1348.

Wei, T., & Song, S. (2018). Copula-based composite likelihood approach for frequency analysis of short annual precipitation records. *Hydrology Research, 49*(5), 1498-1512.

Wen, Y., Yang, A., Kong, X., & Su, Y. (2022). A Bayesian-model-averaging copula method for bivariate hydrologic correlation analysis. *Frontiers in Environmental Science, 9*, 744462.

Xiao, Y., Guo, S., Liu, P., Yan, B., & Chen, L. (2009). Design flood hydrograph based on multicharacteristic synthesis index method. *Journal of Hydrologic Engineering, 14*(12), 1359-1364.

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. *Journal of hydrology, 566*, 23-42.

Yoo, C., & Cho, E. (2019). Effect of multicollinearity on the bivariate frequency analysis of annual maximum rainfall events. *Water, 11*(5), 905.

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.

Zavareh, M. M., Mahjouri, N., Rahimzadegan, M., & Rahimpour, M. (2023). A drought index based on groundwater quantity and quality: Application of multivariate copula analysis. *Journal of Cleaner Production, 417*, 137959.

Zhang, L., & Singh, V. P. (2007a). Bivariate rainfall frequency distributions using Archimedean copulas. *Journal of Hydrology, 332*(1-2), 93-109.