University of Birjand

Water Harvesting Research Vol. 5, No. 1, Spring & Summer 2022, Original Paper p. 131-143

10.22077/JWHR.2023.5944.1076

Frequency Analysis and Joint Simulation of Qualitative Variables of River Flow Using Copula Functions

Gholamreza Ahangi[a], Keivan Khalili[b&*], Mohammad Nazeri Tahroudi[c]

[a]Ms.C Student, Department of Civil Engineering, Saba Institute of Higher Education, Urmia, Iran.
[b]Associate Professor, Department of Water Engineering, Urmia University, Urmia, Iran.
[c]Postdoctoral Researcher, Department of Water Engineering, Shahrekord University, Shahrekord, Iran

[*]Corresponding Author, E-mail address: k.khalili@urmia.ac.ir

Received: 27 December 2022/ **Revised**: 13 January 2023/ **Accepted**: 16 January 2023

Abstract

In this study, in order to analyze the frequency of surface water quality variables (EC-Cl and EC-SO4) and to simulate the dependent variables, the applicability of copula-based functions was addressed. Since meteorological and hydrological variables are dependent on other variables in their surroundings, their analysis using single variable models is not able to estimate the desired results. In this regard, EC, Cl and SO4 data were used in the period of 1971-2017 at Bitas, Gardeyagoub and Kotar stations in the Mahabadchai sub-basin in the south of Lake Urmia. By choosing log logistic, generalized extreme values and also log normal marginal functions, the superior copulas regarding the mentioned pair-variables were investigated. While confirming the accepted correlation between the investigated pair-variables in Bitas stations (EC-Cl=0.39 and EC-SO4=0.38), Gardeyagoub (EC-Cl=0.81 and EC-SO4=0.78) and Kotar (EC-Cl= 0.54 and EC-SO4=0.51) and also based on RMSE, MAE and NSE criteria, Galambos copula was chosen as the best copula in all stations. The joint analysis of the mentioned pair-variables using the Galambos copula led to the presentation of typical curves regarding the estimation of Cl and SO4 values corresponding to the specific EC unit with different probabilities in the studied stations. Given that the presented curves are based on the statistical distribution of data, they are specific to the studied station.

Keywords: Bivariate Copula, Conditional density, Electrical conductivity, Water Quality.

1. Introduction

Qualitative parameters of the river flow should be investigated and analyzed in a multivariate form due to the existence of dependence with other qualitative parameters. Therefore, univariate analysis cannot provide a complete and accurate assessment of these parameters and the correlation structure between them. The use of univariate methods in the analysis of the frequency of different events can lead to low or high estimates (Nazeri Tahroudi et al., 2019; Tahroudi et al., 2020). In the modeling of random variables in hydrology and water resources, the concepts of correlation between variables are of great importance. Linear correlation coefficients are not able to measure these concepts completely. For this reason, recently, in the simulation of random variables, alternative methods have been developed to measure random dependence, which is known as the detailed function, which was actually introduced in 1959 by Sklar and has since been investigated in various researches, including (Favre et al., 2004; De Michele et al., 2005; Genest et al., 2007; Chen et al., 2011; Salvadori and De Michele, 2006; Kao and Govindaraju, 2008; Serinaldi, 2009; Salvadori and De Michele, 2004; Wang et al., 2009; Mirakbari et al., 2010; Mirabbasi et al., 2013; Abdi et al., 2017; Tahroudi et al., 2020; Nazeri Tahroudi et al., 2022). Salvadori and De Michele (2004) presented a new approach that used copula functions to study the return periods of hydrological events. Bardossi (2006) analyzed the range of groundwater quality values using copula functions. By combining statistical methods with copula functions, they introduced a new method in the field of geostatistics. Salvadori and De Michele (2010) explained that how new multivariate extreme value distributions can be easily constructed by taking advantage of recent theoretical developments in copula theory. They showed how to introduce an appropriate number of parameters, a feature not shared by conventional extreme value models. On the other hand, they also introduced a suitable new definition of multivariate return period. Tahroudi et al. (2020) investigated the conditional behavior of bivariate of the joint groundwater level shortage and rainfall shortage in Naqadeh sub-basin in the Lake Urmia Basin using copula functions. The results of the study of changes in two deficiency signals showed that the decrease of 90-135 mm in rainfall in the region has increased the level of access to underground water between 1.2 and 1.7 meters. Nazeri Tahroudi et al. (2022) presented a method based on the conditional density of Vine copulas for drought monitoring and predicting the rainfall shortage during a duration of 60-day at Dashband station, Lake Urmia Basin. Annual rainfall and rainfall shortage in periods of 10-, 30- and 60- day were considered as input variables. D-, C- and R-vine copulas were used to show the dependence between variables and the results showed that the results of D-vine are more accurate. They concluded that if the rainfall is lower than the long-term average in the region, the rainfall shortage for the near future can be estimated with reasonable accuracy.

By examining various studies in the field of copula functions application in hydrology and water resources, it was concluded that the use of these functions in studies of drought, flood, extreme values of rainfall, and sediment load has very rare frequency. In this study, the effectiveness of this method was investigated to accomplish the analysis of the frequency of water quality. Based on this, the main contribution is to estimate the frequency of EC-Cl and EC-SO4 in the study area using two-dimensional copulas.

2. Materials and Methods

2.1. A case Study

The study area in this research is the Mahabadchai sub-basin in the south of Lake Urmia. In this study, the qualitative values of EC, Cl and So4 were used as pair-variables in Kotar, Bitas and Gardeyagoub stations in order to analyze the frequency of the qualitative variables of the river in the Mahabadchai sub-basin in the south of Lake Urmia. Figure 1 shows the location of the studied hydrometric stations in the Mahabadchai sub-basin. The changes of the investigated values in the period of 1977-2017 were presented in figures 2 to 4.

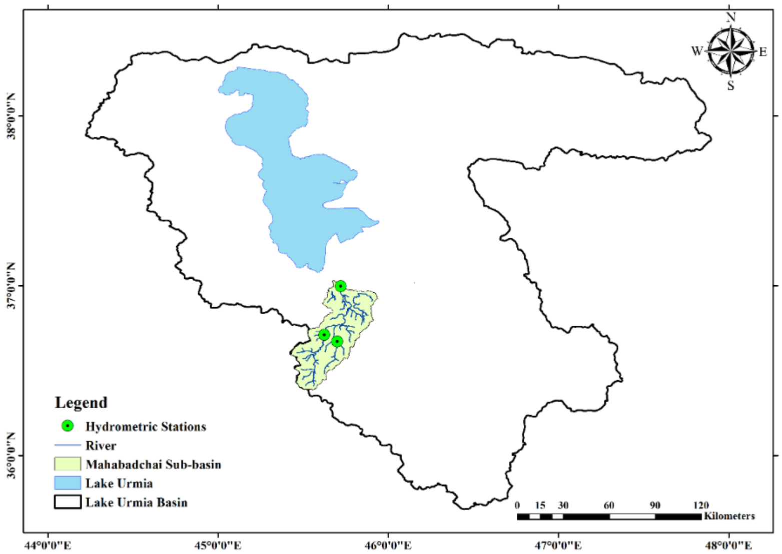

Fig. 1. The location of the stations and the studied area located in the northwest of Iran

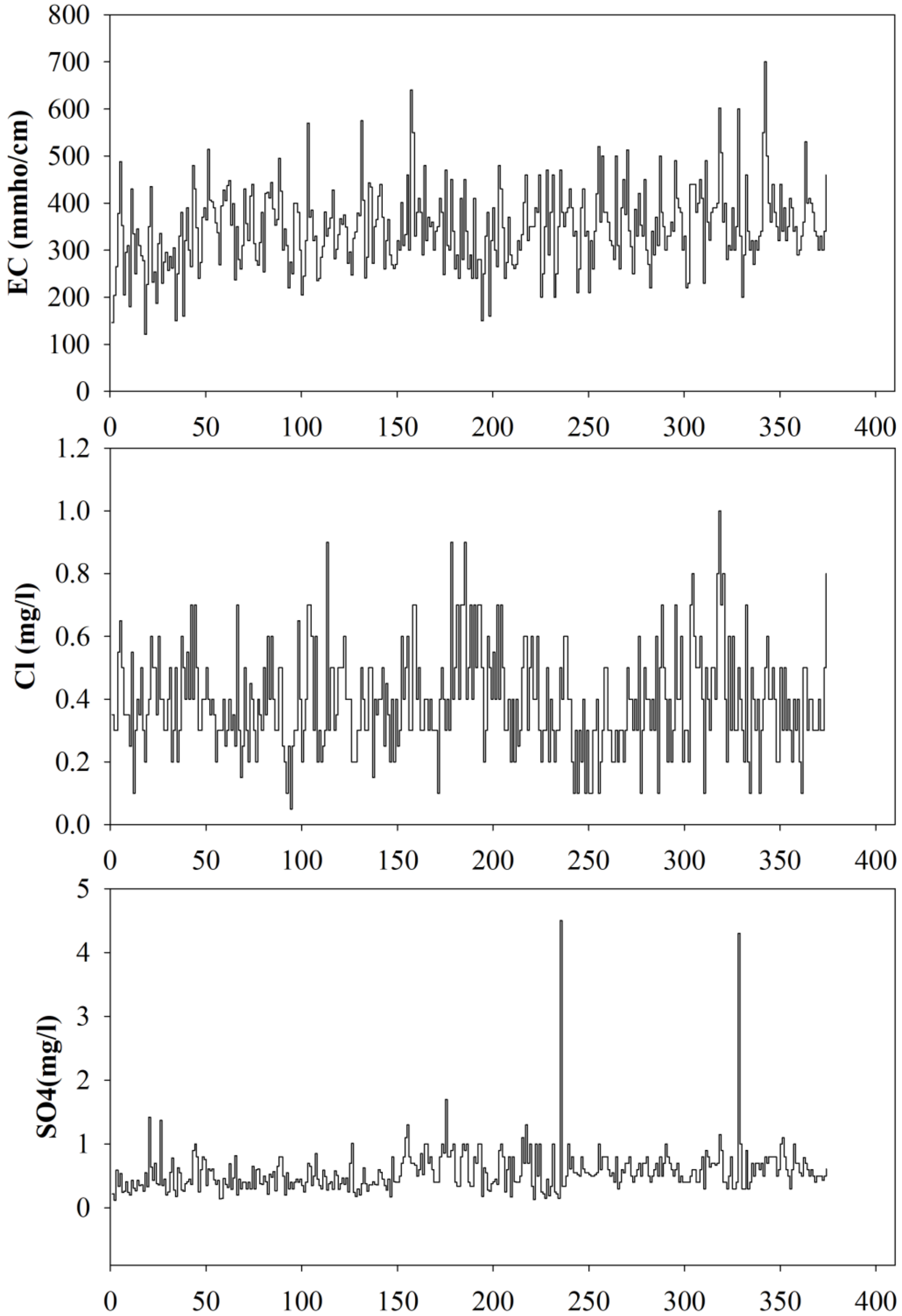

Fig. 2. Initial changes of investigated qualitative values at Bitas hydrometric station in the period of 1977-2017

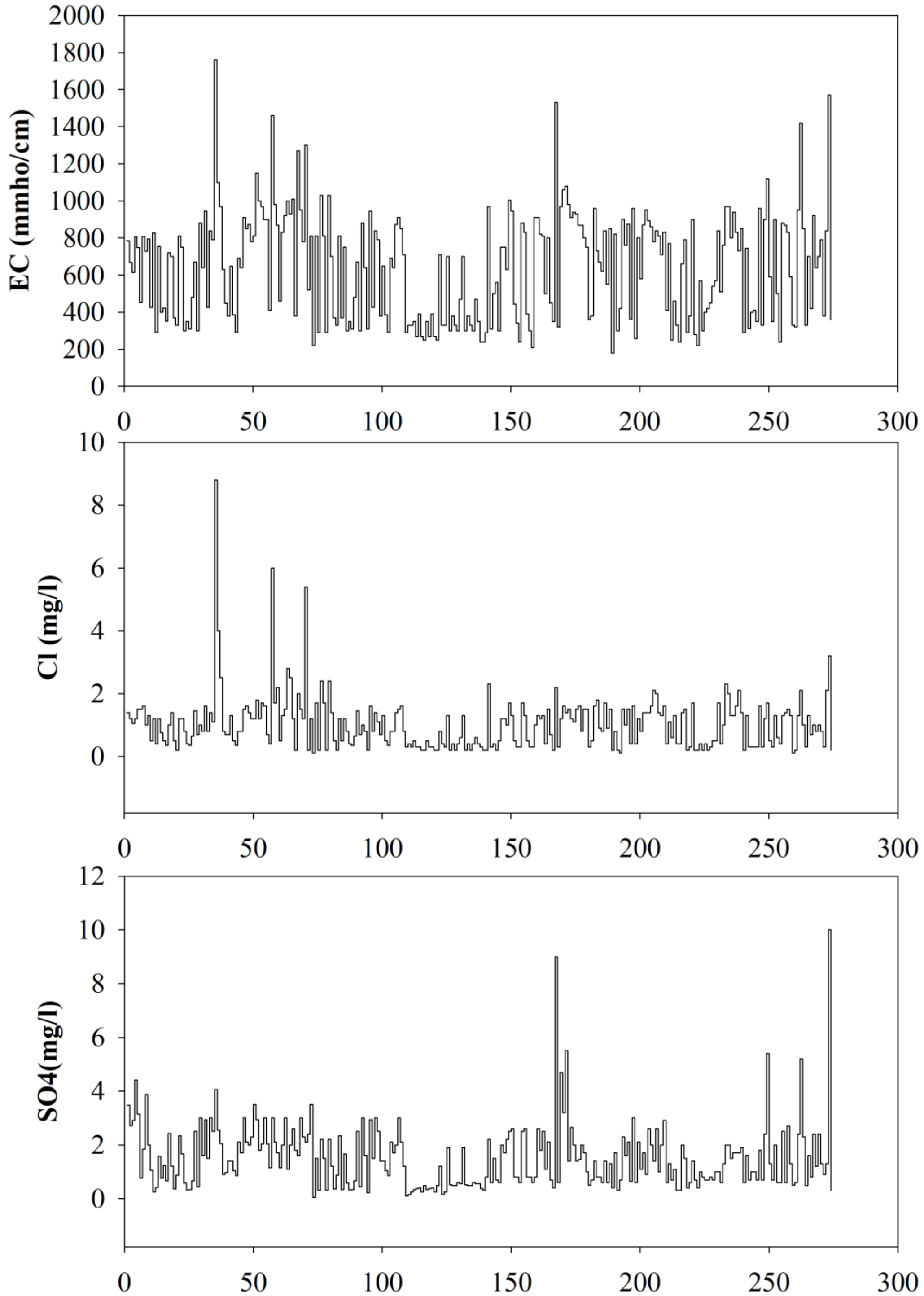

Fig. 3. Initial changes of investigated qualitative values at Gerdeyagoub hydrometric station in the period of 1977-2017

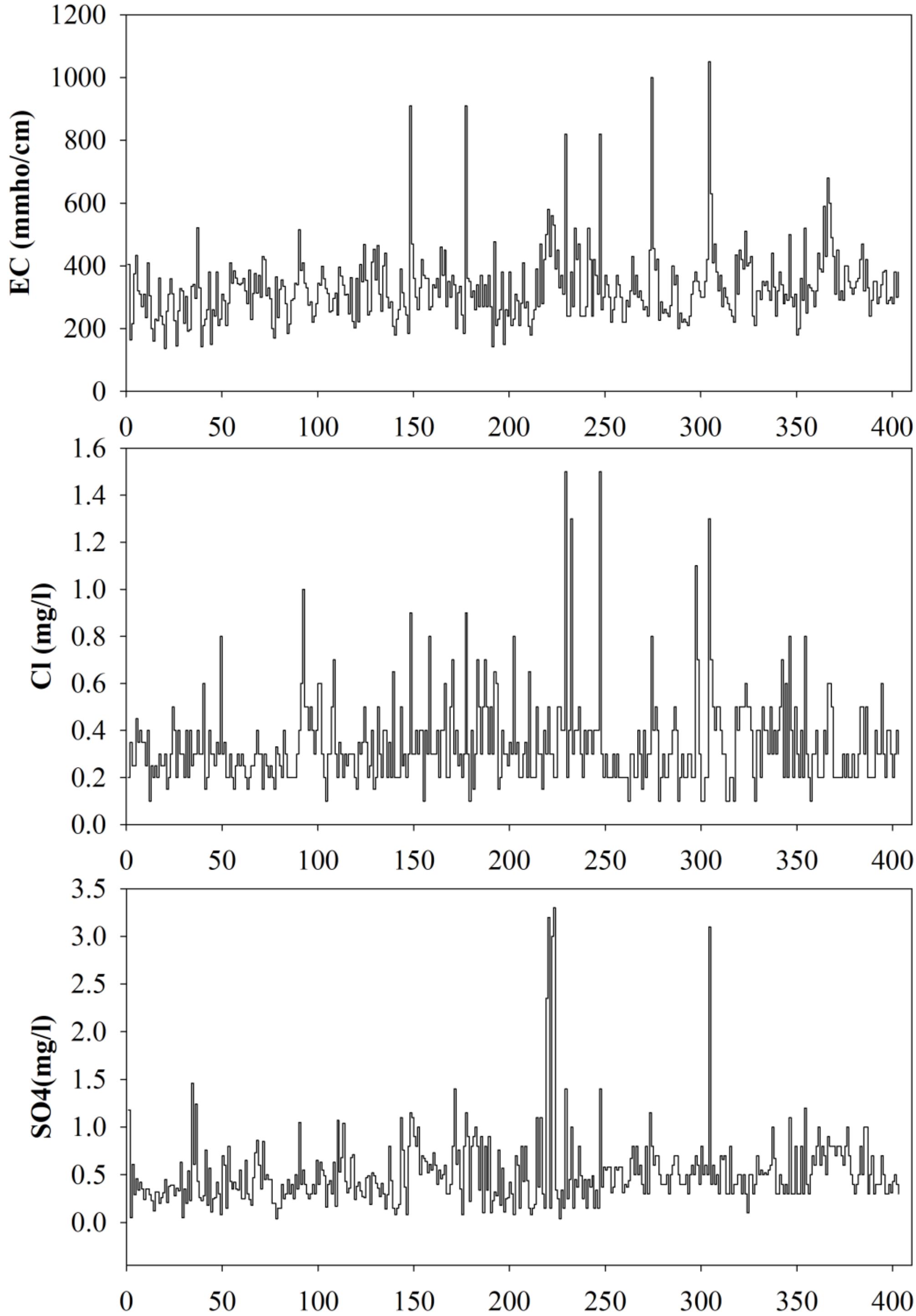

Fig. 4. Initial changes of investigated qualitative values at Kotar hydrometric station in the period of 1977-2017

2.2. Copula-based frequency analysis

Due to the multi-variable nature of meteorological and hydrological phenomena, the use of univariate models does not have the necessary efficiency to provide quantitative analysis results. The copula function is a joint distribution of correlated random variables, which is a function of the univariate marginal distributions. In other words, the copula function is a function that connects univariate marginal functions to create a multivariate distribution function. By using copulas, the effect of dependence is separated from the effects of marginal distributions, which indicates that the use of the copula method does not require the similarity of the type of marginal distributions. This advantage provides more flexibility for choosing univariate marginal distributions.

For N-dimensional continuous random variables $X_1, X_2, ..., X_N$ that have marginal distributions $F(x_i) = P_{x_i}(X_i \le x_i)$, the joint distribution of X is defined as follows (Sklar, 1959).

$$H_{X_1,...,X_N}(x_1, x_2, ..., x_N) = P\left[X_1 \le x_1, X_2 \le x_2, ..., X_N \le x_N\right] \tag{1}$$

In Sklar theory, the multivariate probability distribution H can be expressed using marginal distributions as well as the dependence structure by the copula function C:

$$C(F_{X_1}(x_1), F_{X_2}(x_2), ..., F_{X_N}(x_N)) = H_{X_1,...,X_N}(x_1, x_2, ..., x_N) \tag{2}$$

Where $F_{X_i}(x_i)$ is equal to the ith value of the continuous marginal distribution. $H_{X_1,...,X_N}$ is also a cumulative distribution of $X_1, X_2, ..., X_N$. Considering that for continuous random variables, copula C can be considered as a transformation $H_{X_1,...,X_N}$ from $\left[-\infty, +\infty\right]^N$ to $\left[0,1\right]^N$ (Nelsen, 2007). In the two-dimensional case, let H be the joint distribution of variables X_1 and X_2 with cumulative distributions of $u_1 = F_{X_1}(x_1)$ and $u_2 = F_{X_1}(x_1)$. Now there is a two-dimensional copula in the set of real numbers, which is as follows.

$$H(x_1, x_2) = C(u_1, u_2) = C(F_{X_1}(x_1), F_{X_2}(x_2)) \tag{3}$$

In general, two-dimensional copula functions have the following properties:

1) For each u_1 and u_2 in the interval $\left[0,1\right]$:

$$C(u_1, 0) = C(0, u_2) = 0 \tag{4}$$

$$C(u_1, 1) = C(1, u_2) = 1 \tag{5}$$

This feature is also called the boundary conditions of the two-dimensional copula function. According to these boundary conditions, it is inferred that if one of the marginal distributions has a value equal to zero, then the value of the copula function (joint distribution) will also be equal to zero. In the same way, a similar conclusion can be presented.

2) For the values u_{11}, u_{12} $(u_{11} \le u_{12})$ in U_1 and u_{21}, u_{22} $(u_{21} \le u_{22})$ in U_2 the following inequality holds:

$$C(u_{12}, u_{22}) - C(u_{12}, u_{11}) - C(u_{11}, u_{22}) + C(u_{11}, u_{21}) \ge 0 \tag{6}$$

3. Results and Discussion

As mentioned, in this study, two-dimensional copula functions were used to analyze the qualitative values of the river flow at selected stations in the Mahabadchai sub-basin in the period of 1977-2017. In this study, the pair-variables of variables (EC-Cl) and (EC-SO4) were considered for all three sub-basins. The main goal of this research is the two-dimensional analysis and finally the simulation of qualitative values compared to the value of EC in each station. In this regard, the marginal distribution corresponding to each time series was examined first. The results of examining the marginal distributions at the Kotar station showed that the log logistic, generalized extreme values and normal distributions have a good fit with the studied qualitative data. In the case of Gardeyagoub station, the results of fit the marginal distributions on the qualitative values of the river flow showed that the Weibull marginal distribution had the best fit, and for the SO4 values, the log normal marginal distribution and the generalized extreme values had the best fit. In the case of Cl values, the log logistic distribution, gamma and generalized extreme values, and in the case of EC values, the logistic distribution was the best fit. In the next step, in order to use copula functions, first, the correlation between the investigated values was checked using Kendall's tau statistic. The results of the correlation between the investigated qualitative pair-variables were presented in Figures 5 to 7.

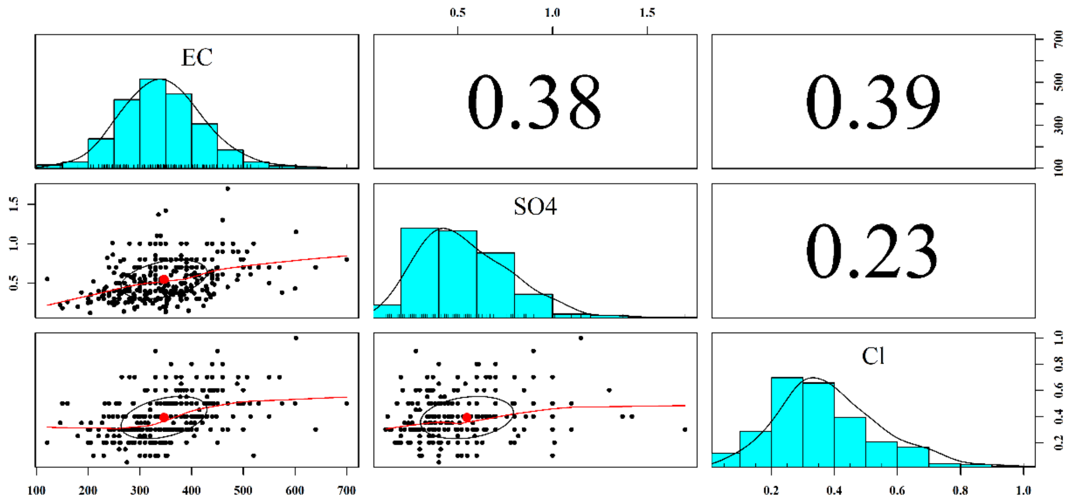

Fig. 5. The results of correlation analysis of studied qualitative pair-variables at Bitas station

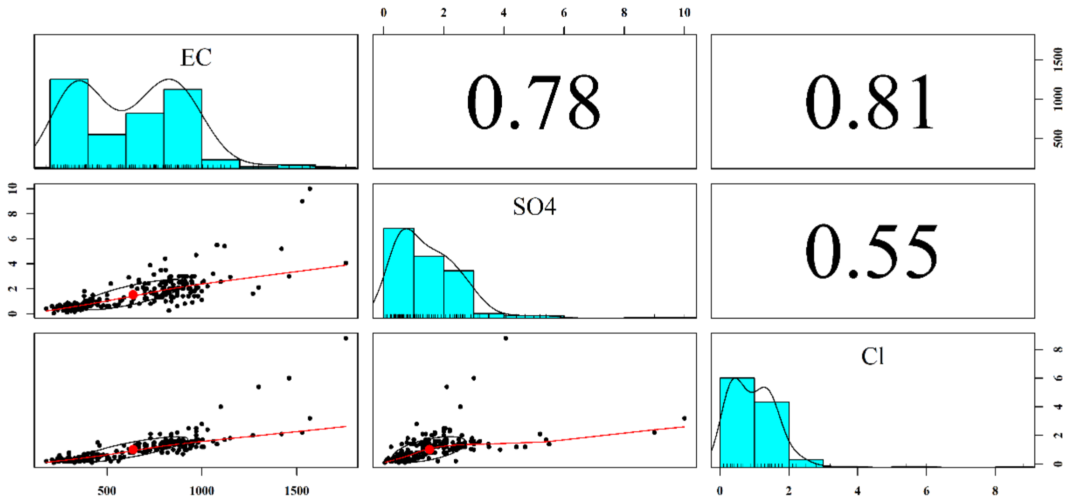

Fig. 6. The results of correlation analysis of studied qualitative pair-variables at Gerdeyagoub station

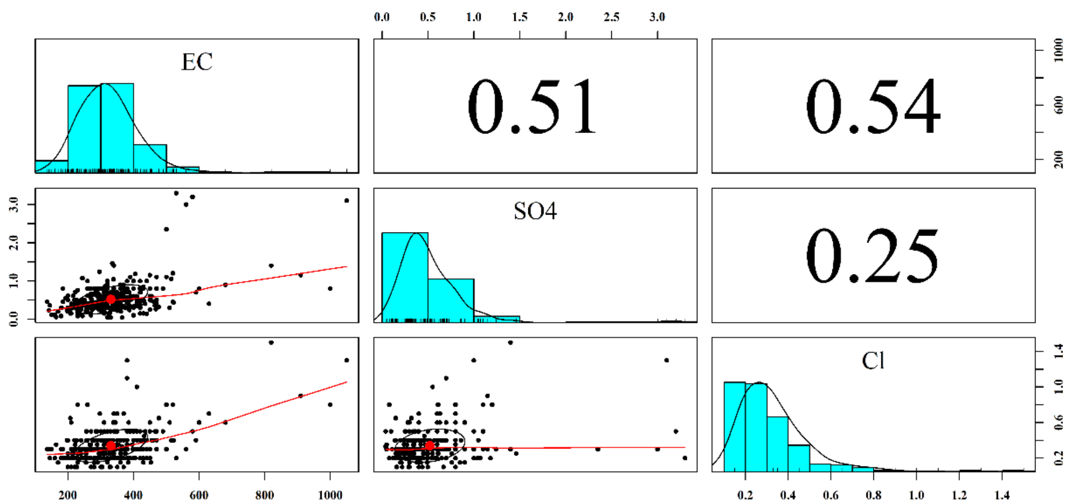

Fig. 7. The results of correlation analysis of studied qualitative pair-variables at Kotar station

According to the correlation presented in Figures 5 to 7, it can be seen that there is an acceptable correlation between the pair of EC-Cl and EC-SO4 series. According to Figure 5, it can be seen that in Bitas hydrometric station there is a correlation of 0.38 and 0.39 between the pair variable of EC-SO4 and EC-Cl, respectively. These coefficients are equal to 0.78 and 0.81 for pair variables of EC-SO4 and EC-Cl in Gardeyagoub station. Kendall's Tau coefficients for the pair variables of EC-SO4 and EC-Cl at Koter station are equal to 0.51 and 0.54, respectively. These coefficients show the existence of correlation between the mentioned pair-variables and indicate the necessity of using the copula function. In order to select the best copula function, various evaluation statistics such as MAE, NSE and RMSE were used to compare the probability values obtained from the experimental copulas and the theoretical copulas in all stations. Table 1 shows the results of investigation and evaluation of various copulas in the fitting of EC-SO4 and EC-Cl variables based on RMSE, MAE and NSE statistics at the studied stations.

Table 1- The results of the examination of various copula functions corresponding to the pair-variables of EC-Cl and EC-SO4 at the studied stations

Station	Pair-variable	NSE	MAE	RMSE	Copula Parameter	Copula Function
Kotar	EC-Cl	0.97	0.04	0.05	7.25	Galambos
	EC-SO4	0.99	0.02	0.02	15.86	
Bitas	EC-Cl	0.97	0.04	0.05	6.83	
	EC-SO4	0.99	0.03	0.03	7.21	
Gerdeyagoub	EC-Cl	0.95	0.05	0.06	8.66	
	EC-SO4	0.96	0.04	0.05	8.24	

The results presented in Table 1 showed that Galambos copula is the best copula function based on NSE, RMSE and MAE statistics to investigate and analyze the frequency of pair-variables of EC-Cl and EC-SO4 at the studied stations. This function has been selected at the studied stations with NSE more than 95%.

3.1. Estimating the probability of occurrence of EC-SO4 pair-variable at the studied stations

According to the presented figures, SO4 values can be estimated according to the changes of EC values with different possibilities. Figure 8 shows the probability contours of the joint distribution function for EC and SO4 data with 10 to 90% probabilities at Bitas station. This figure shows how SO4 and EC values can be determined simultaneously at a station using joint probabilities. This can provide users and researchers with very useful information related to the probable behavior of investigated qualitative variables. For example, for an EC value equal to 550 µmho/cm in Bitas station according to Figure 8, with 50, 70 and 90% probabilities, the SO4 values in this station under the condition of occurrence of selected EC can be about 0.5, 75, and 1 mg/l respectively. In Gardeyagoub station according to Figure 9, if there is an EC value equal to 1000 µmho/cm with 50, 70 and 90% probabilities, it is possible to estimate SO4 values in this station to be about 1.4, 2 and 3 mg/l, respectively.

These numbers regarding SO4 values are estimated to be 0.4, 0.6 and 1 mg/l, respectively, under the condition of occurrence of EC equal to 500 µmho/cm according to Figure 10 at Kotar station. The mentioned figures can be used as a typical curve to identify and estimate different amounts of SO4 with different probabilities in the mentioned stations. This information can be very useful in analyzing the quality of the river.

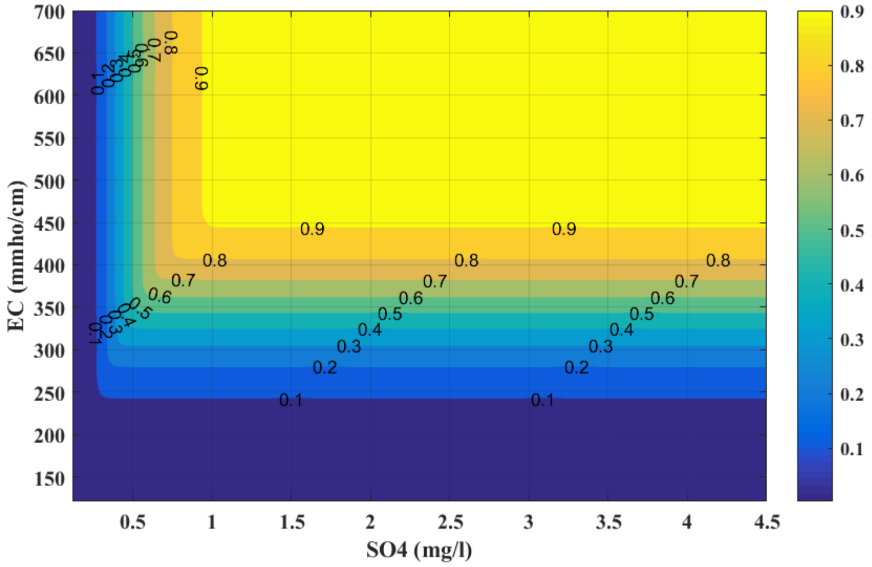

Fig. 8. Joint probability analysis of EC-SO4 pair-variable in Bitas station using the best copula

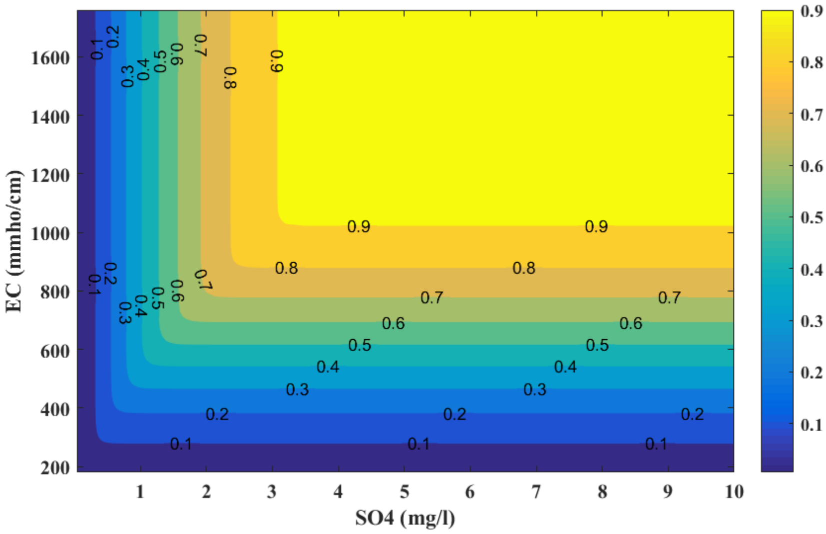

Fig. 9. Joint probability analysis of EC-SO4 pair-variable in Gerdeyagoub station using the best copula

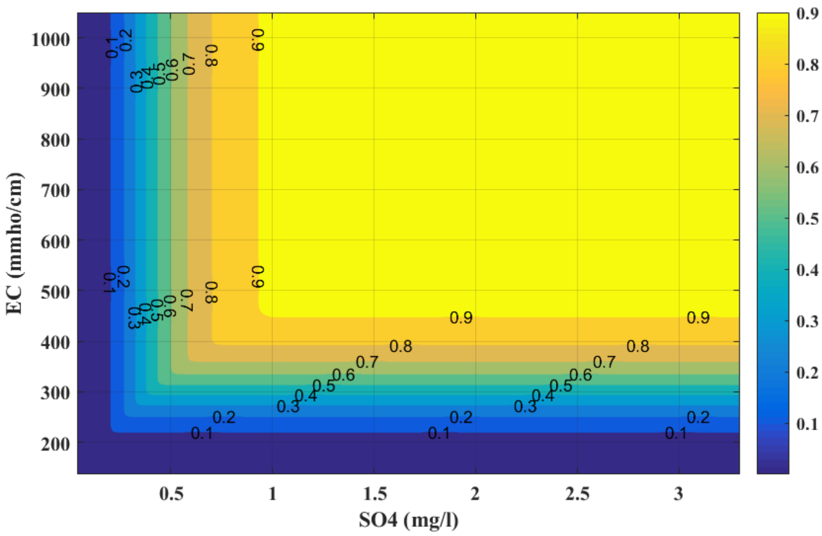

Fig. 10. Joint probability analysis of EC-SO4 pair-variable in Kotar station using the best copula

3.2. Estimating the probability of occurrence of EC-Cl pair-variable at the studied stations

The frequency of occurrence of EC-Cl pair-variable was estimated at the studied stations and the results were presented in figures 11 to 13 for Bitas, Gardeyagoub and Kotar stations, respectively.

By having salinity information (EC) at the studied hydrometric stations, it is easy to estimate the amount of chlorine under the condition of the occurrence of a specific amount of salinity with different probabilities. For example, if the EC value in Bitas station is around 450 μmho/cm, the surface water chlorine in this station will be around 0.37, 0.46 and 0.6 mg/l with a probability of 50, 70 and 90%, respectively. In other words, with a probability of more than 90%, the range of changes caused by EC equal to 450 to 700 μmho/cm will bring the amount of chlorine between 0.6 and 1 mg/l. In Gardeyagoub station according to Figure 12, if the EC measured in this station is 1200 μmho/cm, with a probability of 60, 80 and 90%, the amount of chlorine in this station will be equal to 1, 1.5 and 2 mg/l.

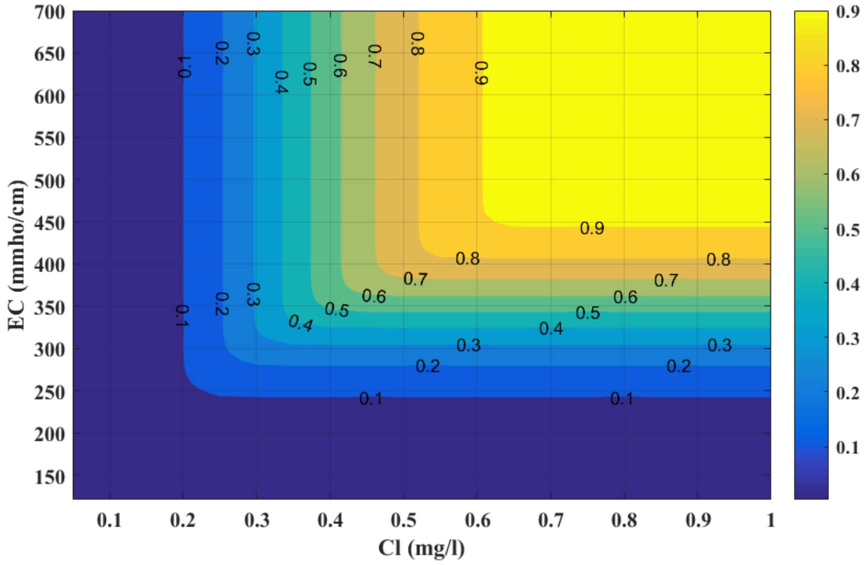

Fig. 11. Joint probability analysis of EC-Cl pair-variable in Bitas station using the best copula

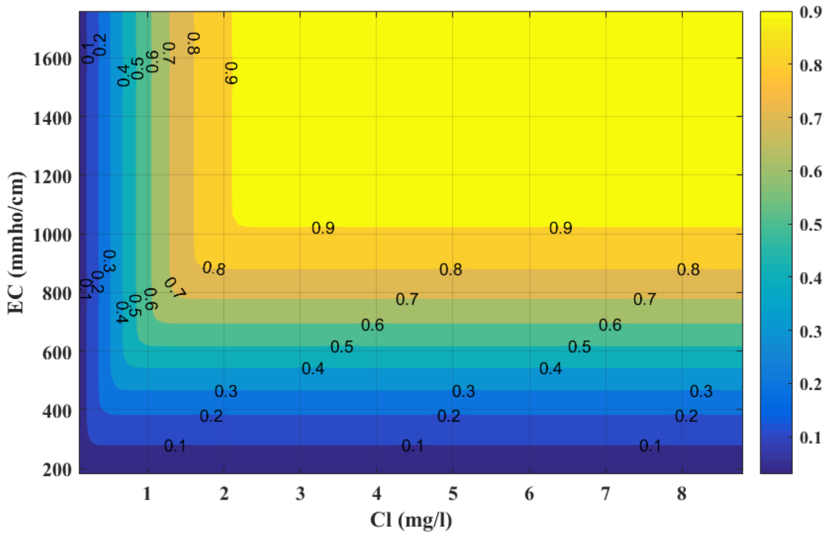

Fig. 12. Joint probability analysis of EC-Cl pair-variable in Gerdeyagoub station using the best copula

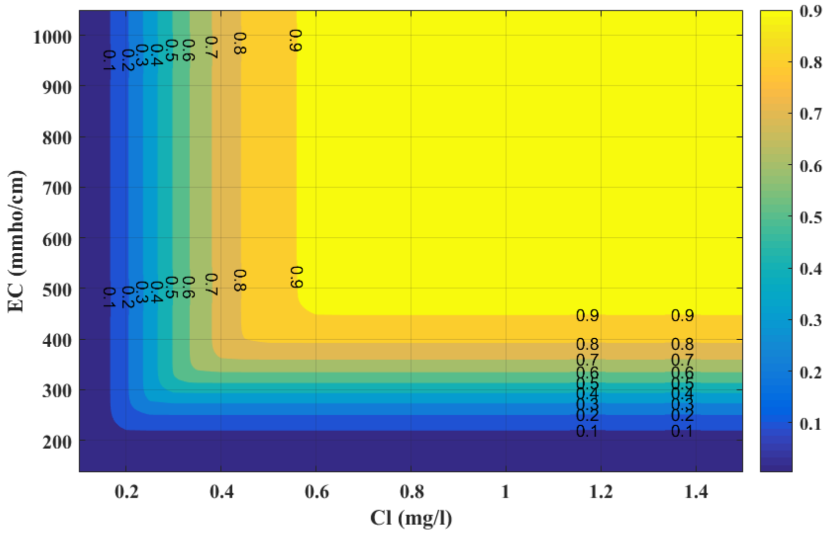

Fig. 13. Joint probability analysis of EC-Cl pair-variable in Kotar station using the bset copula

In this station, according to the range of changes of EC values at the studied statistical, with a probability of more than 90%, the EC values will be more than 1000 µmho/cm, and as a result, the amount of chlorine in the river flow in this station with a probability of 90% is more than 2 mg/l. According to Figure 13, it can be seen that if the salinity level (EC) in the Kotar hydrometric station is about 450 µmho/cm, with 60, 80 and 90% probabilities, the amount of chlorine in the surface water flow in this station will be about 0.35, 0.48 and 0.58 mg/l respectively. According to the analysis of the frequency of occurrence of the CE-Cl pair variable in Koter hydrometric station, it can be seen with a probability of more than 90% that the electrical conductivity in this station during the studied statistical period with this probability rate is more than 450 µmho/cm and chlorine in surface water was more than 0.58 mg/l. According to the presented results, it can be seen that the proposed copula-based approach can well simulate Cl and SO4 values corresponding to EC values with different probabilities. Khashei et al. (2022) in their research on two-variable simulation of reference evapotranspiration in eastern Iran, described the performance of copula-based simulation model as satisfactory. Also, Pronoos Sedighi et al. (2022) in simulating the amount of suspended sediment load based on the flood discharge, considered the accuracy of the copula-based simulation model to be acceptable. The results of this research are in complete agreement with the studies of Khashei et al. (2022) and Pronoos Sedighi et al. (2022).

4. Conclusion

In this study, by using copula functions in two-dimensional mode, the joint frequency analysis and simulation based on the conditional density of EC-Cl and EC-SO4 pair-variables has been done in Kotar, Bitas and Gardeyagoub stations in the Mahabadchai sub-basin, Lake Urmia Basin. In this study, an attempt has been made to simulate the values of Cl and SO4 in the mentioned sub-basins by using the proposed approach based on copula functions and its conditional density, under the condition of the occurrence of a certain amount of EC. At first, the marginal distribution of EC, SO4 and Cl variables was examined in the mentioned stations, and among the various examined distributions, generalized extreme values, log logistic, Weibull and log normal distributions were selected as the best distributions. These distributions were selected based on RMSE and NSE statistics. By choosing the best marginal distributions for each variable in each station, the best copula functions for each pair-variables were also estimated. Based on RMSE, NSE and MAE statistics, Galambos copula was selected as the best copula for all studied stations. The results of the correlation between EC-Cl and EC-SO4 pair-variables at the studied stations based on NSE statistics showed that there is an acceptable correlation between these two pair-variables. The highest correlation (EC-Cl=0.81

and EC-SO4=0.78) was observed in Gardeyagoub station and the lowest correlation (EC-Cl=0.39 and EC-SO4=0.38) was observed in Bitas station. By confirming the existence of correlation between the studied pair-variables, which requires the use of copula functions, the frequency analysis of the occurrence of the mentioned pair-variables was estimated. This work led to the presentation of typical curves regarding the estimation of Cl and SO4 values based on a specific value of the EC variable with different probabilities in each station. These curves estimate different values of Cl and SO4 in each station according to the changes of EC values in the same station. Considering the probability of more than 90%, we can have enough confidence in the obtained results, and we can estimate the values of Cl and SO4 values based on the EC values in the necessary and field cases.

In fact, by presenting the two-dimensional analysis curves of the mentioned pair-variables, by measuring the EC values of the river flow at the studied stations, it is possible to estimate the values of Cl and SO4 with different probabilities. With this method, it is possible to avoid excess costs regarding the measurement of different variables. This method can replace the measurement and calculation of quality parameters of surface water and there is no limitation in the implementation of this proposed method. Considering that this method relies on the marginal distribution of the investigated time series, there is no limitation in its implementation and climatic and geographical factors will not affect the accuracy of the model. The proposed approach and the provided typical curves are stable and reliable in the region as long as the marginal distribution of the used variables does not change.

5. Disclosure statement

No potential conflict of interest was reported by the authors

6. References

Abdi, A., Hassanzadeh, Y., Talatahari, S., Fakheri-Fard, A., & Mirabbasi, R. (2017). Regional bivariate modeling of droughts using L-comoments and copulas. *Stochastic Environmental Research and Risk Assessment, 31*(5), 1199-1210.

Bárdossy, A. (2006). Copula-based geostatistical models for groundwater quality parameters. *Water Resources Research*, *42*(11).

Chen, L., Singh, V. P., Shenglian, G., Hao, Z., & Li, T. (2011). Flood coincidence risk analysis using multivariate copula functions. *Journal of Hydrologic Engineering*, *17*(6), 742-755.

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.

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

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

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).

Mirabbasi, R., Anagnostou, E. N., Fakheri-Fard, A., Dinpashoh, Y., & Eslamian, S. (2013). Analysis of meteorological drought in northwest Iran using the Joint Deficit Index. *Journal of Hydrology*, *492*, 35-48.

Mirakbari, M., Ganji, A., & Fallah, S. R. (2010). Regional bivariate frequency analysis of meteorological droughts. *Journal of Hydrologic Engineering*, *15*(12), 985-1000.

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. *Journal of hydrology*, *10*(3), 282-290.

Nazeri Tahroudi, M., Khashei Siuki, A., & Ramezani, Y. (2019). Redesigning and monitoring groundwater quality and quantity networks by using the entropy theory. *Environmental monitoring and assessment*, *191*(4), 1-17.

Nazeri Tahroudi, M., Ramezani, Y., De Michele, C., & Mirabbasi, R. (2022). Multivariate analysis of rainfall and its deficiency signatures using vine copulas. *International Journal of Climatology*, *42*(4), 2005-2018.

Nelsen, R. B. (2006). Archimedean Copulas. *An Introduction to Copulas*, 109-155.

Pronoos Sedighi, M., Ramezani, Y., Nazeri Tahroudi, M. and Taghian, M., 2022. Joint frequency analysis of river flow rate and suspended sediment load using conditional density of copula functions. *Acta Geophysica*, pp.1-13.

Salvadori, G., & De Michele, C. (2004). Frequency analysis via copulas: Theoretical aspects and applications to hydrological events. *Water resources research*, *40*(12).

Khashei, A., Shahidi, A., Nazeri-Tahroudi, M., & Ramezani, Y. (2022). Bivariate simulation and joint analysis of reference evapotranspiration using copula functions. *Iranian Journal of Irrigation & Drainage*, *16*(3), 639-656.

Salvadori, G., & De Michele, C. (2006). Statistical characterization of temporal structure of storms. *Advances in Water Resources, 29*(6), 827-842.

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

Sklar, M. (1959). Fonctions de repartition an dimensions et leurs marges. *Publ. inst. statist. univ. Paris*, *8*, 229-231.

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.