

Research Article

Copula-Based Joint Flood Frequency Analysis: The Case of Guder River, Upper Blue Nile Basin, Ethiopia

Mesfin Mamo Haile ⁽⁾,¹ Rakesh Khosa,² Asnake Kassahun Abebe,³ Ayansa Teshome Gelalcha,⁴ and Abera Misgana Tolera¹

¹Department of Hydraulic and Water Resources Engineering, Ambo University, Ambo, Ethiopia ²Department of Civil Engineering, Indian Institute of Technology Delhi, Hauz Khas, New Delhi 110016, India ³Department of Surveying Engineering, Ambo University, Ambo, Ethiopia ⁴Department of Civil Engineering, Ambo University, Waliso, Ethiopia

Correspondence should be addressed to Mesfin Mamo Haile; mesfinmamo0355@gmail.com

Received 5 April 2023; Revised 23 August 2023; Accepted 25 August 2023; Published 7 September 2023

Academic Editor: Stefano Federico

Copyright © 2023 Mesfin Mamo Haile et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The univariate analysis of hydrological extremes is a well-established practice in developing countries such as Ethiopia. However, for the design of hydrological and hydraulic systems, it is essential to have a thorough understanding of flood event characteristics, including volumes, peaks, time of occurrence, and duration. This study utilizes copula functions for bivariate modeling of flood peak and volume characteristics, examining the performance of four Archimedean copulas in the Guder basin located in Ethiopia from 1987 to 2017. Flood peak and volume were extracted using the theory of runs for analysis of their joint characteristics with the truncation level chosen as equal to the lowest annual maximum event. Univariate distributions with the best fitness on both variables were determined, and results showed that gamma and GEV-fitted flood peaks and lognormal-fitted flood volumes are the most suitable. Four Archimedean copulas were evaluated, and the Gumbel-Hougaard copula was found to be the best fit for the data based on graphical and measurable tests. Bivariate probability and return period were computed in "AND" and "OR" states. The joint return period for flood peak (97.49 m³/s) and volume (77.35 m³/s) was found to be 15 years in the "AND" state and approximately 4 years in the "OR" state. The study also evaluates univariate and conditional return periods, comparing them with the primary one. The copula method was an effective method for distributing marginal variables, highlighting its potential as a valuable tool in flood management.

1. Introduction

The increase in water demand caused by population growth and industrial development has caused issues related to water resource development and optimal use to gain researchers' attention more than ever in recent decades. Floods are a recurrent natural disaster all over the world, causing substantial socioeconomic consequences and human victims [1, 2]. As a result, developing models that allow for accurate projections of the extent of such catastrophic occurrences is crucial for achieving the construction and management of hydraulic structures such as dams, as well as acceptable flood risk assessments. In practice, such magnitudes are typically assessed using univariate flood frequency studies, which are primarily concerned with peak flows [3], [4] [5]. In addition to the uncertainty involved with the occurrence in both space and time, these events may frequently have varied degrees of association. As a result, using a univariate flood frequency analysis may result in an incorrect evaluation of the risks associated with the event. However, a review of the literature suggests that the bulk of research uses a univariate approach [6–8] rather than a more realistic approach that acknowledges the multivariate nature of the underlying phenomenology. In order to more properly reflect the risk associated with such events, it is required to analyze their simultaneous occurrence in order to better grasp their probabilistic properties [9].

Recently, researchers have widely used a new method known as the copula function in the multivariate frequency analysis of hydrological phenomena [10, 11]. Copula functions are not constrained by the limitations of ordinary bivariate distribution functions and can be used to create multivariate distribution functions by connecting different univariate marginal distribution functions. Furthermore, copulas can more accurately describe dependence structures among variables ([12], [13–17]).

Copulas have been introduced as an efficient tool for quantifying the dependence structure between correlated quantities. Many studies have demonstrated the flexibility provided by copulas in forming joint distributions [15, 18–23]. Copulas were first used in hydrological studies by De Michele and Salvadori [24] in rainfall frequency analysis. Naz et al. [15] used copula functions to investigate the flood at Tarbela Dam, Pakistan, and developed a method for forecasting floods along the Indus River. Other studies have also been used in flood analysis [17, 25–27].

Siraj et al., [27] tested ten bivariate copulas from three different copula families on flood data from the Sava River in Slovenia. The Gumbel-Hougaard copula was found to be the most appropriate for the peak and volume of the flood, as well as the peak and duration factors, whereas the student's t copula was found to be the most appropriate for the peak and duration pair of factors. Ganguli and Reddy [28] used a trivariate copula to break down the three flood factors, peak, volume, and duration. Based on the goodness-of-fit and tail-dependence tests, as well as a graphical investigation, the best choice among Clayton, Gumbel-Hougaard, Frank, and the student's t copula has been made. It was deduced that the student's t copula was the best fit for their data. Recently, another system's entropy copula implied for a multivariate examination of commonly influencing factors was tested on three flood factors (P, V, and D) at two distinct Chinese stations [29].

Amini et al. [30] applied the vine copula structures for multivariate analysis of flood characteristics in the Armand watershed, Iran, using the hydrographs of 68 flood events. The study found Johnson SB, lognormal (3P), log-logistic, and lognormal (3P) best-fitted marginal distributions for flood duration, peak discharge, time to peak, and flood volumes, respectively. They found that the Frank copula has the best-fitting copula at most of the edges and nodes. Zhang and Singh [18] investigated the relationship between flood peak, volume, and duration using four families of Archimedean copulas: Ali-Mikhail-Hagh, Frank, Gumbel-Hougaard, and Cook-Johnson. The margins were examined using the extreme values distribution type I and the log-Pearson type III. Their results showed that there are significant associations between flood peak and volume, as well as flood volume and duration. They realized that the Gumbel-Hougaard family would be more appropriate to explain the variable dependence structure. The copula model was also used to calculate the conditional return period. [31] developed a new method for calculating bivariate design events using copulas that allow for simultaneous and nonsimultaneous occasions of the variables under consideration. Copulas were also used in flood frequency analysis by different researchers [1, 19, 20, 28, 29–33]. Stamatatou et al. [34] compared univariate and joint bivariate return periods of floods that rely on different probability concepts in the Yermasoyia watershed, Cyprus. The pair of peak discharge and corresponding volume was estimated and compared using annual maximum series (AMS) and partial duration series (PDS) approaches.

According to research, only one comprehensive study on this topic has been conducted previously, which was performed by Haile in 2022. Therefore, the novelty of this research in Ethiopia is that it provides valuable insights into a relatively new research area that can contribute to better flood risk management strategies. Multivariate flood frequency analysis can model the complex relationships between flood variables and provide a comprehensive understanding of the risk associated with floods, leading to accurate estimation of flood hazards, identification of sources of flood risk, and ultimately, better flood risk management strategies. Therefore, further research in this area has the potential to address the current data gap in Ethiopia and contribute to the development of effective flood risk management policies and practices. As a result, the aim of this study was to create a copula-based probability model for bivariate analysis of the Guder watershed's 30-year flow (1987-2017). Archimedean copulas such as the Clayton copula, Frank copula, Gumbel-Hougaard copula, and Ali-Mikhail-Haq copula have all been examined. Goodnessof-fit statistics, upper tail-dependence coefficients, and graphical analysis were used to compare the copulas. The primary, conditional, and Kendall return periods have also been computed for a better knowledge of the river flow in the study area in order to comprehend the danger of flood occurrences.

2. Materials and Methods

2.1. Description of Study Area. Guder watershed has a drainage area of 7011 km^2 and is located in central Ethiopia in the Blue Nile basin's south-eastern region. It may be found in the Oromia regional state between $7^{\circ}30$ to $9^{\circ}30$ N latitudes and $37^{\circ}00'$ to $39^{\circ}00'$ E longitudes. Tikur Incinis, Ambo, Cheliya, Dendi, Jima Rare, Mida Kegn, Gojo, Guduru, Liban Kutaye, Tokke Kutaye, and Ababo Guduru are the districts covered by the Guder watershed (Figure 1). It is bounded to the east by the Muger subbasin, to the south by the Awash basin, and to the west by the Fincha subbasin. The Guder River rises in the mountains to the south of the villages of Ambo and Guder. The river runs from south to north and empties into the Abbey River. The main tributaries of the Guder River include Huluka, Tarantar, and Debris Rivers [16, 35–39].

2.2. Methodology. The primary goal of this study is to apply the copula method and evaluate its results. Initially, approaches for specifying the marginal distribution functions for the various variables were used. This includes data



FIGURE 1: Map of the study area [16, 36, 37].

analysis and generation of flood characteristics [36, 40, 41], modeling extremes for the selected parameters using threshold peaks, and finally determining the appropriate distributions for each variable of interest. Following that, the return periods for each variable were calculated using the standard univariate approach. Copula represents the correlated variable's joint distributions. As a result, it is critical to examine the relationship between the variables. A scatterplot of the data, a scatterplot of the variables' standardized ranks, and a Chi-plot can be used to perform a graphical analysis of dependence. The Chi-plot is a plot of a rankbased measure of each observation's location versus a measure of the Chi-squared test of independence. If the transformed data are scattered above (below) the region defined by the Chi-plot's confidence interval [42, 43], the dependence is positive (negative).

To quantify the dependence between variables, dependence measures such as Kendall's rank correlation coefficient (or Kendall's tau), and Spearman's rho can be computed [16, 44]. Under the null hypothesis that the value of the association measures is 0, the p values associated with these measures can be calculated. If the p value is less than the significance level, statistical independence is rejected. Kendall's tau is regarded as the most important measure of association in the theory of copulas, particularly Archimedean copulas, where its relationship with the generating function is used for copula parameter estimation. One of the most common methods for estimating the parameter (θ) of the

Archimedean copula is the inversion of Kendall's tau. The relationship shown in Table 1 between Kendall's tau and the parameter is used to estimate in this method. Because this relationship is dependent on the copula functions generating function ϕ , different values of θ are obtained for different copula models for different copula models [14, 15, 17, 36].

$$\tau = 1 + \int_0^1 \frac{\phi(t)}{\phi'(t)} \mathrm{d}t,\tag{1}$$

where ϕ and ϕ' are the generating function and derivative of the Archimedean copula's generating function, respectively. Because this relationship is dependent on the generating function ϕ of the copula function, different values of θ are obtained for the various copula models listed in Table 1. Table 1 describes the key characteristics of the Archimedean copula families used in this study.

Following the selection of the best-performing copula approach, the bivariate distributions had to be created. According to Sklar's theorem [47], a copula is a joint distribution function of typical uniform random variables that may link marginal distribution functions with multivariate probability distributions. Let $F_{X,Y}$ be a joint distribution function of F_X and F_Y marginal distribution. The copula C can be described as follows:

$$F_{X,Y}(x, y) = C(F_X(x), F_Y(y)).$$
 (2)

If F_X , F_Y , and C are distribution functions, then $F_{X,Y}$ given by equation (2) is a joint distribution function with marginal F_X and F_Y .

Evaluation of the joint distributions is greatly simplified when copulas are employed to construct the bivariate distribution, which is demonstrated in the following. The joint distribution function, defined in equation (3), is the probability that both Q and V are less than or equal to specific thresholds of q and v, respectively; that is,

$$P(Q \le q, V \le v) = F(q, v) = C(F(q), F(v)) = C(u, v).$$
(3)

In practice, the probability of both Q and V exceeding their respective thresholds can also be defined by copulas:

$$P(Q > q, V > v) = 1 - F(q) - F(v) + C(F(q), F(v)) = 1 - u - v + C(u, v).$$
(4)

The conditional distribution of V given $Q \ge q$ and the conditional distribution of Q given $V \ge v$ can also be estimated as the following equations:

$$F(v \mid Q \ge q) = \frac{P(Q \ge q, V < v)}{P(Q \ge q)} = \frac{F(v) - C(F(q), F(v))}{1 - F(q)} = \frac{v - C(u, v)}{1 - u},$$
(5)

$$F(q \mid V \ge v) = \frac{P(Q < q, V \ge v)}{P(V \ge v)} = \frac{F(q) - C(F(q), F(v))}{1 - F(v)} = \frac{u - C(u, v)}{1 - v}.$$
(6)

The resulting joint probabilities of copula are expressed as a function of univariate marginal distributions. Therefore, for specified univariate marginal distributions, the joint distributions can be described simply through the use of copulas.

The copula-based return periods were computed after demonstrating the joint distribution. The joint (primary) return periods known as "OR" mode and "AND" mode [31, 48, 49] were computed in this study and are defined as follows:

$$T_{\rm OR} = \frac{E(t)}{1 - C(u, v)},$$

$$T_{\rm AND} = \frac{E(t)}{1 - u - v + C(u, v)},$$
(7)

where u and v are uniformly distributed U(0, 1). The U represents $F_X(x)$, and V represents $F_Y(y)$, and they were created by applying the probability integral transform to X and Y, a transformation that allowed us to simplify our work by using an equivalent set of values that follow the standard uniform distribution.

In addition, the return period can also be defined by the event for Q given $V \ge v$ or the event for V given $Q \ge v$ that is called the conditional return period for Q given $V \ge v$ and the conditional return period for V given $Q \ge q$, respectively:

$$T_{Q|V \ge \nu} = \frac{E(t)}{[1 - \nu] [1 - u - \nu + C(u, \nu)]},$$

$$T_{V|Q \ge q} = \frac{E(t)}{[1 - u] [1 - u - \nu + C(u, \nu)]}.$$
(8)

On the other hand, a secondary return period, also known as Kendall's return period, is defined as follows:

$$T_{K} = \frac{E(t)}{\overline{K_{C}}(t)} = \frac{E(t)}{1 - K_{C}(t)},$$
(9)

where K_C is Kendall's distribution function for the theoretical copula function.

As a reminder, $u = F_Q(q)$, and $v = F_V(v)$; hence, the (u, v) pairs that satisfy the above equations form the associated bivariate return period curve for a given return period value. In other words, for a given copula value C(uv) = t, the (u, v) pairs located in the copula level curve equals t are associated with the return period value.

3. Results and Discussion

The technique described in the methodology section was used to evaluate the design variables that could be used in the analysis of hydraulic or hydrologic problems on the Guder River. In this case, both the flood peak and the flood volume are important; thus, assuming the two variables are related, the copula approach would be more significant than the single-variable approach. The threshold was chosen in accordance with the partial duration series methodology in order to generate a larger sample of extreme events than in the previous analysis [31, 40]. The flood volume V was then estimated by locating the flood volume that corresponded to the chosen flood peak events. More specifically, the truncation line is taken as 40 and 60 m³/s for station-1 and station-2, respectively; i.e., flood events are defined as a daily stream flow equal to or greater than 40 and 60 m³/s. Eightyseven and hundred fifty-four flood data are abstracted from the recorded daily stream flow data (1987-2017) for the two

TABLE 1: Copula function,	generating function	$\phi(t)$, and the	functional	relationship of	of Kendall's	tau with	the copula	parameter f	for
Archimedean copulas [10, 4	5] [46].			-				-	

Copula methods	C(u, v)	$\phi(t)$	Parameter estimation
Ali-Mikhail-Haq	$uv/1 - \theta(1-u)(1-v), \theta \in [-1,1]$	$\phi(t) = \ln\left(1 - \theta t\right)/t$	$\tau = 3\theta - 1/3\theta - 2/3(1 - 1/\theta)^2 \ln(1 - \theta)$
Clayton	$(u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}, \theta \ge 0$	$t^{- heta} - 1/ heta$	au = heta/ heta + 2
Frank	$-1/\theta \ln[1 + (e^{-\theta u} - 1)(e^{\theta v} - 1)/e^{-\theta} - 1], \theta \neq 0$	$e^{-\theta t} - 1/e^{-\theta} - 1$	$\tau = 1 - 4/\theta [1 - D_1(\theta)]$
Gumbel-Hougaard	$\exp\left\{-\left[\left(-\ln u\right)^{\theta}+\left(-\ln v\right)^{\theta}\right]\right\}^{-1/\theta}, \theta \ge 1$	$(-\ln t)^{\theta}$	au = 1/1 - heta

Note. $D_k(x)$ is the Debye function, for any positive integer k, and determined by $D_k(x) = (k/x^2) \int_0^x (t^k/e^t - 1) dt$.

stations, respectively. The observed number of pairs (n) is 87 for station 1, consisting of peak discharge (Q) and volume (V), with a mean interarrival time e(t) of 0.35 years. On the other hand, for station 2, there are 154 pairs with a mean interarrival time of 0.20 years. The location and scale parameters for both variables are depicted in Table 2 at the two stations.

3.1. Univariate Analysis. A univariate flood frequency analysis was performed after the extreme events were chosen. The datasets were subjected to various probability models that were best fitted by gamma and generalized extreme value (GEV) distributions for peak discharge at stations 1 and 2, respectively, and lognormal for flood volumes at both stations. The parameters of the distribution were estimated using the maximum likelihood method, which will also be used in the estimation of copula parameters [16, 48]. Following that, the Kolmogorov-Smirnov (K-S) goodness-of-fit and graphical tests were developed to identify the distributions that produced an adequate fit to the data. After performing graphical and goodness-of-fit (GOF) tests and computing K-S values for each accepted model, the best-fit distributions were chosen (Figure 2). The maximum likelihood (ML) method was used to estimate model parameters (Table 3). Finally, once a suitable model had been identified, the univariate return periods for 2, 5, 10, 25, 50, and 100 years were calculated and are shown in Table 4. In the lower return periods, the estimated univariate return values differ only for the peak discharge variable.

3.2. Bivariate Analysis. Following the univariate analysis, Kendall and Spearman's correlation coefficient was used to perform a formal assessment of the dependence between the pairs of variables under consideration. The first station yielded Kendall and Spearman's correlation coefficient values of 0.60 and 0.79, while the second station yielded a value of 0.67 and 0.85, respectively, highlighting the strong correlation between the two variables (Table 5).

The scatter plot of the *u* and *v* is displayed in Figures 3(a) and 3(b) for both station-1 and station-2, respectively, where a strong correlation between the two variables is readily apparent. The scatter plot of the pseudo-observations is plotted in Figures 3(c) and 3(d) for station-1 and station-2, respectively; this shows a positive relation of dependence between variables. The values of Kendall's τ and Spearman's ρ_s rank-based nonparametric measures of dependence validate the results provided by the graphical information and

TABLE 2: Location and scale parameters for variables flood peak (Q) and volume (V).

Stations	Variables	Location parameters	Scale parameters
1	Q	81.93	19.88
	$V (10^{6})$	33.75	51.78
2	Q	89.21	17.03
2	$V(10^{6})$	12.19	25.38

also supported by the Chi-plot (Figure 4). The Chi-plot shows that the majority of the values are located above the region defined by the confidence intervals, indicating that the positive dependency between the two variables (peak and volume) is strong.

Nonparametric estimation method is used to estimate copula parameters and to identify copula-based bivariate flood frequency analysis of flood peaks and volumes [44, 50]. The four commonly used Archimedean copula families are considered as candidates for joint flood frequency analysis. These are Ali-Mikhail-Haq, Clayton, Frank, and Gumbel-Hougaard copula. The feasible copulas were checked by testing the admissible range of dependence supported by each one via Kendall's τ value. As a result, the Ali-Mikhail-Haq ($\tau \in [0.1817, 1/3]$) was eliminated (Figures 5 and 6).

The parameter estimated by Kendall's τ is given in Table 6 for each copula method. The goodness of fit of observed data to the theoretical bivariate distribution obtained by using copula functions is tested by AIC [51] and KS methods. Table 7 shows that AIC and KS values for bivariate distributions are obtained by using different copula functions for flood peaks and volumes. Considering both tests, the Gumbel-Hougaard copula was selected as the best copula model for station-1 and station-2. Based on this, joint cumulative density function (JCDF), conditional probability, and corresponding return periods for different combinations of flood peaks and volumes are calculated.

In other definition, if the plot is in agreement with a straight line passing through the origin at a 45-degree angle, then the generating function is satisfactory [46], [52]. The 45-degree line indicates that the quantiles are equal. Otherwise, the copula function needs to be reidentified.

3.2.1. Joint and Conditional Density Function

(1) Joint Cumulative Density Function (JCDF). The JCDF of flood peaks and volumes is calculated based on Gumbel-Hougaard copula methods for station-1 and station-2,



FIGURE 2: Comparison of the observed flood variable and their fitted distributions: (a) gamma distribution for flood peak of station-1, (b) lognormal distribution for flood volume of station-1, (c) gamma distribution for flood peak of station-2, and (d) lognormal distribution for flood volume of station-2.

TABLE 3: Estimated parameter for fitted distribution and g	goodness	of fit
--	----------	--------

Station	Flood peak	KS	Flood volume	KS
1	Gamma distribution ($\beta = 17.4$, $\alpha = 4.71$)	0.0909	Lognormal distribution ($\mu = 16.23$, $\sigma = 1.64$)	0.0613
2	GEV ($u = 77.4437$, $\alpha = 11.3452$, $k = 0.0546$)	0.0973	Lognormal distribution ($\mu = 14.8152$, $\sigma = 1.538$)	0.0617

respectively, in which u and v were estimated by corresponding marginal distributions of flood peak and volume. The contour lines demonstrate that, given a fixed occurrence probability of a flood event, one can obtain various occurrence combinations

of flood peaks and volumes, and vice versa, which cannot be obtained by single-variable flood frequency analysis (Figure 7). These results should be useful for hydrological structure design studies.

		-	-			
Return period (years)		Station 1	Station 2			
	Flood peak	Flood volume (10 ⁶ m ³)	Flood peak	Flood volume (10 ⁶ m ³)		
5	97.49	77.35	93.78	25.92		
10	108.09	100.14	101.47	34.75		
25	120.46	124.42	110.74	44.16		
50	129.06	140.11	117.31	50.23		
100	137.20	154.2	123.60	55.69		

TABLE 4: Results of univariate return periods 5, 10, 25, 50, and 100 for flood peak and volume.

TABLE 5: Value and p value of the rank-based nonparametric measures of dependence: Kendall's τ and Spearman's ρ .

Dependence measure	Station-1	Station-2
Kendall's τ	0.60	0.670
Spearman's ρ	0.79	0.85



FIGURE 3: (a, b) Scatter plot of (u, v); (c, d) scatter plot of pseudo-observations derived from the observed data for station-1 and station-2, respectively.

(2) Conditional Density Function. The conditional probabilities $F_{q/\nu}$ and $F_{q/\nu}$ are estimated based on (5) and (6) and are depicted in Figure 8 for both stations.

(3) Joint Return Periods. The T_{OR} (equation (7)), T_{AND} (equation (7)), and T_K (equation (9)) joint return periods

associated with the theoretical events with a peak equal to q_T and volume equal to v_T for univariate return periods *T* equal to 5, 10, 50, and 100 years are estimated for both the Station-1 and Station-2, being q_T and v_T the quantiles obtained from the fitted marginal distributions. The results presented in Table 8 indicate that although T_{OR} values linked to the



FIGURE 4: Chi-plot for station-1 and station-2.



FIGURE 5: Comparison of nonparametric and parametric K(z) for station-1.



FIGURE 6: Comparison of nonparametric and parametric K(z) for station-2.

TABLE 6: Estimated value of the copula parameter (θ) .

Copula	$\hat{\theta}$ (station-1)	$\hat{\theta}$ (station-2)
Ali-Mikhail-Haq	Not applicable	Not applicable
Clayton	3.24	4.37
Gumbel-Hougard	2.62	3.19
Frank	8.56	10.92

TABLE 7: AIC and KS test values for copulas.

Station	Station Test Gumbel-Hougaard		Clayton	Frank
Station-1	AIC	-129.53*	-117.4652	-75.1834
	KS	0.0575*	0.0628	0.1735
Station 2	AIC	-141.076^{*}	-131.47	-125.43
Station-2	KS	0.0413*	0.0684	0.052

Note. * represents the chosen copula.

Gumbel copula are greater than those linked to the Clayton copula, T_{AND} and T_K values are much higher. It can also be seen that the greater the return period, the larger the difference between joint return periods related to each copula. Therefore, as expected for not being an extreme-value

copula, the Clayton copula underestimates the risk associated with the T_{AND} and T_K joint return period. Hence, this analysis supports the fact that not taking into consideration the upper tail dependence in joint extreme events modeling can lead to underestimate the risk [53] [46], [13–15, 17, 31, 38].

Shiau et al. [52] predicted, in reference to "OR" and "AND" cases and flood peak and volume, that "the use of $T_{\rm OR}$ or $T_{\rm AND}$ as the design criterion depends on what situations will destroy the structure. $T_{\rm OR}$ can be used to calculate the average recurrence interval when either the flood peak or flood volume exceeds a certain magnitude. $T_{\rm AND}$, on the other hand, is used when the flood volume and flood peak must exceed a certain magnitude that will cause damage." T_K was proposed by Salvadori [24] and is thus widely used (e.g., [48, 50, 54–56]). Based on the following arguments, the idea behind T_K is to overcome an apparent shortcoming of T_{OR} and T_{AND} . Different pairs of (U, V), e.g., (u, v), (u, v), and (u^{v}, v^{v}) , lying on the same level curve of a bivariate joint distribution, have the same joint probability, i.e., $P[Q \le q \cap V \le v] = P[Q \le q' \cap V \le v'] = P[Q \le q'' \cap V \le v''],$ but they define different and partially overlapping P(Q > q, V > v) critical regions. As a result, we have infinite



FIGURE 7: JCDF of peak and volume: (a) $P(Q \le q, V \le v)$ for station-1, (b) P(Q > q, V > v) for station-1, (c) $P(Q \le q, V \le v)$ for station-2, and (d) P(Q > q, V > v) for station-2.

OR (AND) critical regions with the same joint probability, making a choice between them impossible (e.g., [27, 54, 57]). Because the lack of correspondence between each $T_{OR}(T_{AND})$ value and a unique critical region is incorrect from a measure theoretic standpoint, Salvadori [24] introduced T_K , which is based on the Kendall distribution (or measure) KC and measures the chance of observing an event in one of the two distinct subregions defined by a level curve characterized by a distinct value of joint probability. The contours for joint return periods T_{OR} and T_{AND} are depicted in Figure 9.

Based on the contours of the joint return periods, given a return period, one can obtain various combinations of flood peaks and volumes, and vice versa. These various scenarios can be very useful for hydrologic engineers to carry out effective hydrologic engineering management and designs, such as spillways and flood control reservoirs, in which a design flood hydrograph (DFH) is needed [58]. The DFH is characterized by its peak, volume, duration, and shape, as addressed in previous studies [59–61] [62]. Various pairs of flood peak and volume values associated with a given return period, say 100 years, provide one more possible choice on which the DFH should be selected. The variability of design flood events under bivariate flood frequency analysis allows a better selection of the most crucial case according to a specific water resources planning, management, or design problem, which cannot be



FIGURE 8: Conditional probability for Guder River: (a) $F(q | V \ge v)$ for station-1, (b) $F(v | Q \ge q)$ for station-1, (c) $F(q | V \ge v)$ for station-2, and (d) $F(v | Q \ge q)$ for station-2.

achieved via single-variable frequency analysis. The conditional joint return periods were computed using the copula-based distributions selected for each set of correlated flood variables. The conditional joint returns periods $T_{Q|V \ge v}$ and $T_{V|Q \ge q}$ of flood peak given flood volume and flood volume given peak based on the conditions $Q \ge q$ and $V \ge v$ were computed using equation (8).

Figure 10 shows the conditional joint return periods for the Guder River at Station-1 and Station-2 for the abovementioned conditions. The figures show that the conditional return period is significantly affected by the nature of conditioning. Taking the conditional return period of the Guder River Station-1 and 2 as an example, the return period of a specified value (say, flood peak) when the conditioning is in terms of a fixed value (say, flood volume) is found to be lower than the return period when the conditioning is in terms of a value less than or equal to the fixed value. On the one hand, for low specified flood volume values, no

Station	Copula	$T=T_Q=T_V$	q_T (m ³ /s)	$(10^6 m^3)$	t	$T_{\rm OR}$	$T_{\rm AND}$	$K_{c}(t)$	T_K
		5	97.49	77.35	0.7437	4	15	0.9382	16
		10	108.09	100.14	0.8397	6	34	0.9614	26
1 Gumbel-Ho	Gumbel-Hougaard	25	120.46	124.42	0.9020	10	107	0.9764	42
	C C	50	129.06	140.11	0.9255	13	258	0.9820	56
		100	137.20	154.20	0.9392	16	625	0.9853	68
-		5	93.78	25.92	0.7688	4	24	0.9686	32
		10	101.47	34.75	0.8995	10	92	0.9863	73
2	Gumbel-Hougaard	25	110.74	44.16	0.9271	14	176	0.9901	101
	C C	50	117.31	50.23	0.9538	22	462	0.9937	159
		100	123.60	55.69	0.9805	51	3418	0.9973	376

TABLE 8: Comparison between joint return periods associated with the theoretical events.



FIGURE 9: Joint return period of flood peak and volume: (a) T_{OR} for station-1, (b) T_{AND} for station-1, (c) T_{OR} for station-2, and (d) T_{AND} for station-2.

significant differences were found under these two conditions, i.e., in terms of a fixed volume value and a volume less than or equal to the fixed volume value. On the other hand, significant differences were found under these two conditions if a specified peak discharge value was high. Furthermore, if the return period of a larger specified value is desired, then larger differences are obtained under the two conditions. A decreasing trend of the return period is



FIGURE 10: Conditional return period for Guder River: (a) $F(q | V \ge v)$ for station-1, (b) $F(v | Q \ge q)$ for station-1, (c) $F(q | V \ge v)$ for station-2, and (d) $F(v | Q \ge q)$ for station-2.

obtained for the same specified value under different conditions. This can be explained by the correlation structure of flood peak and volume, which are positively correlated. The positive correlation results in a less likely occurrence of high specified discharge for low conditioning volume than would be the case for high specified discharge under high conditioning volume.

The conditional return period was always higher than the primary return period T_{AND} and T_{OR} . The secondary return

period was higher than the primary return period T_{OR} and lower than the return period T_{AND} . These relationships are in accordance with the different findings ([46], [13–15, 17, 38]).

From Table 8, it is evident that the return period for the AND operation is consistently larger than the return period for the OR operation. Additionally, the univariate return period values always fall between these two primary return period notations. This finding aligns with the results reported by [46]. The study's results highlight the utility of copulas as a valuable mathematical tool for modeling various multidimensional hydrological processes, including floods and sediment transport-related processes.

4. Conclusions

In this study, a bivariate flood frequency analysis was carried out using a large number of bivariate copulas as well as various statistical and graphical tests. The annual maxima series and peaks of threshold approaches were used to collect the data samples, and the corresponding univariate and bivariate return periods were evaluated and compared as the next step. Overall, the results agreed with [46], [13–15, 17, 31, 38] who demonstrated that the relationship between univariate and joint return periods can be written as $T_{\rm OR} < T_{\rm UN} < T_{\rm AND}$.

The examination of correlation in the two samples of data confirmed that the hydrological variables were significantly dependent on one another. It is worth noting that, while the association pattern changed when different sampling methods were used, the return period estimates did not differ significantly. Overall, the presence of dependence among hydrological variables, as well as the demands of the common problems at hand, points to the need for multivariate distributions to be constructed, particularly when dealing with design values. As a result, more research should be conducted to investigate the significance of copula application in hydrological analysis, specifically in return period estimation.

These findings could be useful for hydrological system design studies, and the same approach could be extended to any station in Ethiopia. The general conclusion is that univariate frequency analysis cannot offer an adequate probabilistic evaluation of correlated multivariate hydrological events, which can result in overestimation or underestimation of their magnitude.

Data Availability

The datasets generated during and/or analysed during this study are available from the corresponding author on reasonable request.

Consent

Informed consent to participate and to publish was obtained from all individual participants included in the study.

Disclosure

A preprint of this article has previously been published [63].

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

The authors were responsible for the study conception and design, data collection, analysis and interpretation of results, and manuscript preparation.

References

- [1] A. Candela, G. Brigandì, and G. Aronica, "Estimation of synthetic flood design hydrographs using a distributed rainfall-runoff model coupled with a copula-based single storm rainfall generator," *Natural Hazards and Earth System Sciences*, vol. 14, pp. 1819–1833, 2014.
- [2] L. Yan, S. Yin-He, Y. Qian, S. Zhi-Yu, W. Chun-Zi, and L. Zi-Yun, "Method of reaching consensus on probability of food safety based on the integration of finite credible data on block chain," *IEEE Access*, vol. 9, pp. 123764–123776, 2021.
- [3] C. Cunnane, "Statistical distributions for flood frequency analysis," Operational Hydrology Report, pp. 23–33, World Meteorological Organization, Geneva, Switzerland, 1989.
- [4] J. R. Stedinger, R. M. Vogel, and E. Foufoula-Georgiou, "Frequency analysis of extreme events," in *Handbook of Hydrology*, D. R. Maidment, Ed., McGraw-Hill, New York, NY, USApp. 18.1–18.66, 1993.
- [5] Y. Liu, K. Zhang, Z. Li, Z. Liu, J. Wang, and P. Huang, "A hybrid runoff generation modelling framework based on spatial combination of three runoff generation schemes for semi-humid and semi-arid watersheds," *Journal of Hydrology*, vol. 590, Article ID 125440, 2020.
- [6] A. R. Rao and K. H. Hamed, *Flood Frequency Analysis*, CRC Press, Boca Raton, FL, USA, 1999.
- [7] C. T. Haan, Statistical Methods, ISO Central Secretariat, Geneva, Switzerland, 1977.
- [8] G. Wang, B. Zhao, B. Wu et al., "Research on the macromesoscopic response mechanism of multisphere approximated heteromorphic tailing particles," *Lithosphere*, vol. 2022, no. 10, Article ID 1977890, 2022.
- [9] C. Gao, M. Hao, J. Chen, and C. Gu, "Simulation and design of joint distribution of rainfall and tide level in Wuchengxiyu Region, China," *Urban Climate*, vol. 40, Article ID 101005, 2021.
- [10] R. B. Nelsen, An Introduction to Copulas: Lecture Notes in Statistics, vol. 139, Springer, New York, NY, USA, 2006.
- [11] S. Grimaldi and F. Serinaldi, "Asymmetric copula in multivariate flood frequency analysis," *Advances in Water Resources*, vol. 29, no. 8, pp. 1155–1167, 2006.
- [12] G. Salvadori and C. De Michele, "On the use of copulas in hydrology: theory and practice," *Journal of Hydrologic En*gineering, vol. 12, no. 4, pp. 369–380, 2007.
- [13] F. Ahmadi, F. Radmaneh, M. R. Sharifi, and R. Mirabbasi, "Bivariate frequency analysis of low flow using copula functions (case study: dez River Basin, Iran)," *Environmental Earth Sciences*, vol. 77, no. 18, p. 643, 2018.
- [14] S. Abdollahi, A. Akhoond-Ali, R. Mirabbasi, and J. F. Adamowski, "Probabilistic event based rainfall-runoff modeling using copula functions," *Water Resources Management*, vol. 33, no. 11, pp. 3799–3814, 2019.
- [15] S. Naz, M. Ahsanuddin, S. Inayatullah, T. Siddiqi, and M. Imtiaz, "Copula-based bivariate flood risk assessment on Tarbela dam, Pakistan," *Hydrology*, vol. 6, no. 3, p. 79, 2019.
- [16] M. M. Haile, "Application of geospatial technology coupled with fuzzy logic operations in mapping soil-erosion critical zones: the case of guder sub-basin, upper blue nile basin, Ethiopia," *Applied Water Science*, vol. 12, no. 8, p. 195, 2022a.
- [17] S. Moradzadeh Rahmatabadi, M. Irandoust, and R. Mirabbasi, "Multivariate analysis of rainfall-runoff characteristics using copulas," *Journal of Earth System Science*, vol. 132, no. 2, p. 93, 2023.

- [18] L. Zhang and V. P. Singh, "Bivariate flood frequency analysis using the copula method," *Journal of Hydrologic Engineering*, vol. 11, no. 2, pp. 150–164, 2006.
- [19] S. Madadgar and H. Moradkhani, "Improved Bayesian multimodeling: integration of copulas and Bayesian model averaging," *Water Resources Research*, vol. 50, no. 12, pp. 9586–9603, 2014.
- [20] S. Haider, R. Rahman, S. Ghosh, and R. Pal, "A copula based approach for design of multivariate random forests for drug sensitivity prediction," *PLoS One*, vol. 10, no. 12, Article ID e0144490, 2015.
- [21] A. Abdi, Y. Hassanzadeh, S. Talatahari, A. Fakheri-Fard, and R. Mirabbasi, "Parameter estimation of copula functions using an optimization-based method," *Theoretical and Applied Climatology*, vol. 129, no. 1-2, pp. 21–32, 2017.
- [22] S. Amini, R. Z. Bidaki, R. Mirabbasi, and M. Shafaei, "Flood risk analysis based on nested copula structure in Armand Basin, Iran," *Acta Geophysica*, vol. 70, no. 3, pp. 1385–1399, 2022a.
- [23] S. Amini, R. Z. Bidaki, R. Mirabbasi, and M. Shafaei, "Multivariate analysis of flood characteristics in Armand Watershed, Iran using vine copulas," *Arabian Journal of Geosciences*, vol. 16, no. 1, p. 11, 2022b.
- [24] G. Salvadori and C. De Michele, "Frequency analysis via copulas: theoretical aspects and applications to hydrological events," *Water Resources Research*, vol. 40, no. 12, pp. 1–17, 2004.
- [25] H. Chowdhary, L. A. Escobar, and V. P. Singh, "Identification of suitable copulas for bivariate frequency analysis of flood peak and flood volume data," *Hydrology Research*, vol. 42, no. 2-3, pp. 193–216, 2011.
- [26] A. I. Requena, L. Mediero, and L. Garrote, "A bivariate return period based on copulas for hydrologic dam design: accounting for reservoir routing in risk estimation," *Hydrology* and Earth System Sciences, vol. 17, no. 8, pp. 3023–3038, 2013.
- [27] M. Sraj, N. Bezak, and M. Brilly, "Bivariate flood frequency analysis using the copula function: a case study of the Litija station on the Sava River," *Hydrological Processes*, vol. 29, no. 2, pp. 225–238, 2015.
- [28] P. Ganguli and M. J. Reddy, "Probabilistic assessment of flood risks using trivariate copulas," *Theoretical and Applied Climatology*, vol. 111, no. 1-2, pp. 341–360, 2013.
- [29] F. Li and Q. Zheng, "Probabilistic modelling of flood events using the entropy copula," *Advances in Water Resources*, vol. 97, pp. 233–240, 2016.
- [30] S. Amini, R. Z. Bidaki, R. Mirabbasi, and M. Shafaei, "Multivariate analysis of flood characteristics in Armand Watershed, Iran using vine copulas," *Arabian Journal of Geosciences*, vol. 16, no. 1, p. 11, 2023.
- [31] J. Bender, T. Wahl, A. Müller, and J. Jensen, "A multivariate design framework for river confluences," *Hydrological Sciences Journal*, vol. 61, no. 3, pp. 471–482, 2016.
- [32] G. Salvadori, F. Durante, C. De Michele, M. Bernardi, and L. Petrella, "A multivariate copula-based framework for dealing with hazard scenarios and failure probabilities," *Water Resources Research*, vol. 52, no. 5, pp. 3701–3721, 2016.
- [33] M. Shafaei, A. Fakheri-Fard, Y. Dinpashoh, R. Mirabbasi, and C. De Michele, "Modeling flood event characteristics using Dvine structures," *Theoretical and Applied Climatology*, vol. 130, no. 3-4, pp. 713–724, 2017.
- [34] N. Stamatatou, L. Vasiliades, and A. Loukas, "Bivariate flood frequency analysis using copulas," *Proceedings*, vol. 2, no. 11, p. 635, 2018.

- [35] M. M. Haile, "Approaches for delineating groundwater recharge potential zone using fuzzy logic model," *Geofluids*, vol. 2022, Article ID 3637230, 15 pages, 2022.
- [36] M. M. Haile, "Bivariate flood frequency analysis: a case study of rib river, upper blue Nile basin, Ethiopia," ARRUS Journal of Engineering and Technology, vol. 2, no. 2, pp. 46–64, 2022b.
- [37] M. M. Haile and A. K. Abebe, "GIS and fuzzy logic integration in land suitability assessment for surface irrigation: the case of Guder watershed, Upper Blue Nile Basin, Ethiopia," *Applied Water Science*, vol. 12, no. 10, p. 240, 2022.
- [38] M. M. Haile, J. Tafesse Yadeta, and L. Amena Gambe, "Delineation of groundwater potential zones using geo-spatial and AHP techniques: Ambo area, Blue Nile Basin, Ethiopia," *Sustainable Water Resources Management*, vol. 8, no. 1, p. 36, 2022.
- [39] A. M. Tolera, M. M. Haile, T. Merga, and G. A. Feyisa, "Assessment of land suitability for irrigation in West Shewa zone, Oromia, Ethiopia," *Applied Water Science*, vol. 13, no. 5, p. 112, 2023.
- [40] W. B. Langbein, "Annual floods and the partial-duration flood series," *Transactions- American Geophysical Union*, vol. 30, no. 6, pp. 879–881, 1949.
- [41] G. Wang, B. Zhao, R. Lan et al., "Experimental study on failure model of tailing dam overtopping under heavy rainfall," *Lithosphere*, vol. 2022, no. 10, Article ID 5922501, 2022.
- [42] N. I. Fisher and P. Switzer, "Chi-plots for assessing dependence," *Biometrika*, vol. 72, no. 2, pp. 253–265, 1985.
- [43] N. I. Fisher and P. Switzer, "Graphical assessment of dependence: is a picture worth 100 tests?" *The American Statistician*, vol. 55, no. 3, pp. 233–239, 2001.
- [44] C. Genest and L. Rivest, "Statistical inference procedures for bivariate Archimedean copulas," *Journal of the American Statistical Association*, vol. 88, no. 423, pp. 1034–1043, 1993.
- [45] R. B. Nelsen, An Introduction to Copulas, Springer, New York, NY, USA, 1999.
- [46] G. C. Salvadori, De Michele, N. Kottegoda, and R. Rosso, Extremes in Nature. An Approach Using Copulas, Water Science and Technology Library, Dordrecht, Netherlands, Springer, 2007.
- [47] A. Sklar, Fonctions de répartition à n dimensions et leurs marges, Université Paris, Paris, France, 1959.
- [48] G. Salvadori, F. Durante, G. R. Tomasicchio, and F. D'Alessandro, "Practical guidelines for the multivariate assessment of the structural risk in coastal and offshore engineering," *Coastal Engineering*, vol. 95, pp. 77–83, 2015.
- [49] B. Gräler, M. J. van den Berg, S. Vandenberghe et al., "Multivariate return periods in hydrology: a critical and practical review focusing on synthetic design hydrograph estimation," *Hydrology and Earth System Sciences*, vol. 17, no. 4, pp. 1281–1296, 2013.
- [50] S. Vandenberghe, N. E. C. Verhoest, and B. De Baets, "Fitting bivariate copulas to the dependence structure between storm characteristics: a detailed analysis based on 105 year 10 min rainfall," *Water Resources Research*, vol. 46, no. 1, Article ID W01512, 2010.
- [51] H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *Proceeding of the Second International Symposium on Information Theory*, B. N. Petrov and F. Csaki, Eds., pp. 267–281, Akademiai Kiado, Budapest, Hungary, September 1973.
- [52] J. Shiau, H. Wang, and C. Tsai, "Bivariate frequency analysis of floods using copulas," *Journal of the American Water Resources Association*, vol. 42, no. 6, pp. 1549–1564, 2006.

- [53] A. Poulin, D. Huard, A.-C. Favre, and S. Pugin, "Importance of tail dependence in bivariate frequency analysis," *Journal of Hydrologic Engineering*, vol. 12, no. 4, pp. 394–403, 2007.
- [54] G. Salvadori and C. De Michele, "Multivariate multiparameter extreme value models and return periods: a copula approach," *Water Resources Research*, vol. 46, no. 10, Article ID W10501, 2010.
- [55] F. Durante and G. Salvadori, "On the construction of multivariate extreme value models via copulas," *Environmetrics*, vol. 21, pp. 143–161, 2010.
- [56] G. Salvadori, C. De Michele, and F. Durante, "On the return period and design in a multivariate framework," *Hydrology* and Earth System Sciences, vol. 15, no. 11, pp. 3293–3305, 2011.
- [57] J. Li, Z. Wang, X. Wu, C. Xu, S. Guo, and X. Chen, "Toward monitoring short-term droughts using a novel daily scale, standardized antecedent precipitation evapotranspiration index," *Journal of Hydrometeorology*, vol. 21, no. 5, pp. 891– 908, 2020.
- [58] X. Zhu, Z. Xu, Z. Liu et al., "Impact of dam construction on precipitation: a regional perspective," *Marine and Freshwater Research*, vol. 74, no. 10, pp. 877–890, 2022.
- [59] S. Yue, T. B. M. J. Ouarda, B. Bobee, P. Legendre, and P. Bruneau, "The Gumbel mixed model for flood frequency analysis," *Journal of Hydrology*, vol. 226, no. 1-2, pp. 88–100, 1999.
- [60] S. Yue, "Applying bivariate normal distribution to flood frequency analysis," *Water International*, vol. 24, no. 3, pp. 248–254, 1999.
- [61] S. Yue and P. Rasmussen, "Bivariate frequency analysis: discussion of some useful concepts in hydrological application," *Hydrological Processes*, vol. 16, no. 14, pp. 2881–2898, 2002.
- [62] J. Zhou, L. Wang, X. Zhong et al., "Quantifying the major drivers for the expanding lakes in the interior Tibetan Plateau," *Science Bulletin*, vol. 67, no. 5, pp. 474–478, 2022.
- [63] M. M. Haile and K. Khosa, Copula Based Joint Flood Frequency Analysis: The Case of Guder River, Upper Blue Nile Basin, Ethiopia, 2023.