

*Corresponding author: Mesfin Mamo Haile, Department of Hydraulic and Water Resources Engineering, Ambo University, Ambo, Ethiopia

E-mail: mesfinmamo0355@gmail.com

RESEARCH ARTICLE

Bivariate Flood Frequency Analysis: A case study of Ribb River, Upper Blue Nile Basin, Ethiopia

Mesfin Mamo Haile

Department of Hydraulic and Water Resources Engineering, Ambo University, Ambo, Ethiopia

Abstract: Hydrological design, planning and design of flood mitigation structures require detailed knowledge of the characteristics of the flood event, i.e. peaks, volumes, occurrence times and duration. In addition to the uncertainty associated with the occurrence in both space and time, these events may often have a correlation of varying strengths. The literature study interestingly reveals that the majority of studies are based on a univariate approach rather than a more realistic approach that recognizes the multivariate nature of the underlying phenomenology. Therefore, the main objective of this study is to revisit the topic of sizing of flood events in terms of their frequency of occurrence using the ‘Copula’ based bivariate approach to analyze the joint distributions of correlated flood variables with a special focus on Ribb sub basin, Upper Blue Nile (Ethiopia) as the case study. The methodology was applied to flood attributes, i.e. flood peaks and volume generation from partial duration series (PDS) by applying run theory. The Joint Cumulative Distribution Function, the Conditional Cumulative Distribution Function and the associated Return Periods can be easily achieved based on the bivariate distribution of the copula and compared to the univariate analysis.

Keywords: bivariate flood frequency, copula, probability distribution, ribb river, Ethiopia.

1. Introduction

Univariate flood frequency analysis of hydrological extremes such as floods and droughts is well developed. In many cases, however, flood events are described by a number of variables of interest, namely peaks, volumes, occurrence times and duration. In addition to the uncertainty associated with the occurrence in both space and time, these events may often have a correlation of varying strengths. Therefore, the application of a univariate flood frequency analysis may lead to an inappropriate assessment of the risks associated with the event. However, a literature study interestingly reveals that the majority of studies are based on a univariate approach (A. Ramachandra Rao and Khaled H. Hamed, 2000; Haan, 1977) rather than a more realistic approach that recognizes the multivariate nature of the underlying phenomenology. In order to better understand the probabilistic characteristics of such events, it is necessary to study their joint occurrence in order to more precisely represent the risk associated with the event.

Over the past three decades, attention has been paid to addressing the common behavior of hydrological events such as floods, droughts and storms. Ashkar and Rousselle, 1982, used the joint distribution of flood peaks, volumes and durations by assuming the triangular shape of flood hydrographs above a specific threshold. Krstanovic and Singh (1987)



derived the common distribution of flood peaks and volume by applying the bivariate distribution of Gaussian and Exponential according to the principle of maximum entropy. (K. Singh & Singh, 1991) derived and used exponential bivariate to represent common distributions of precipitation intensity and corresponding depths with exponential marginals. (N.K.Goel, 1998; Sackl & Bergmann, 1987; Sheng Yue, 1999) used bivariate normal distributions to represent common distributions of flood peaks and volume. Applying normal bivariate distribution means that the marginals are naturally normal or require a prior normalizing transformation (e.g. by using a box-cox transformation) or one in a class of two power transformations. (S Yue, Ouarda, Bobée, Legendre, & Bruneau, 1999) obtained the joint distribution of flood peaks and volume as well as flood volume and durations by using the Gumbel mixed distribution. This model can only be useful for the positively correlated random variables. (Sheng Yue, 2000) applied the bivariate Gumbel-logistic distribution to analyse the joint flood frequency analysis of flood variables by applying the Gumbel marginal distributions. Yue (2000) applied bivariate lognormal distribution to analyze the joint occurrence of flood peaks and volume as well as flood volume and duration. (Sheng Yue, 2001) used bivariate Gamma distribution to represent the joint behavior of flood variables by applying the Gamma marginals. (Sheng Yue & Rasmussen, 2002) derived the relationship between joint return periods and univariate return periods by applying a Gumbel-logistic distribution to represent the joint distribution of flood peak and volume.

In most of the above literatures, the marginal distribution of the flood variables need to be the same type, either as an innate population attribute or upon transformation. In most cases, however, flood variables are dependent do not have the same type of marginal distributions. This difficulty may seriously undermine the accuracy of the analysis and results that are obtained therefrom. To overcome these limitations (Chen & Fan, 2004; Favre, Adlouni, Perreault, Thiémonge, & Bobée, 2004; Genest & Rivest, 1993; Grimaldi & Serinaldi, 2006; Zhang & Singh, 2006) applied the concept of Copulas. Copula relaxes the limitation of the conventional approach by selecting marginal from different families or without resorting to any transformation. This study aims to revisit the topic of sizing of flood events in terms of their frequency of occurrence using the 'Copula' based bivariate approach to analyze the joint distributions of correlated flood peaks and volume. The bivariate analysis is then employed to determine joint distributions, conditional distributions and corresponding return periods, and is verified using daily flow data from Ribb River, Upper Blue Nile (Ethiopia).

2. Materials and Methods

2.1. Description of Study Area

Rib catchment is one of the largest sub-catchment found in the eastern part of the Tana basin in Amhara regional state, Ethiopia. The catchments may be described as a flat to gently sloping plain. The area lies between coordinate systems of 37° 54' 38.42" E to 11° 59' 31.501" N as shown in **Error! Reference source not found.** It is a vast plain with a majority of the catchment area reaching an elevation of 1800m above mean sea level. The total catchment area of Ribb is 1943.48 km². This catchment is drained by the Rib River, which originates from Guna Mountain and finally joins Lake Tana in the vicinity where the River causes flooding. Ribb River is 84 km long and has 34 other small tributaries.

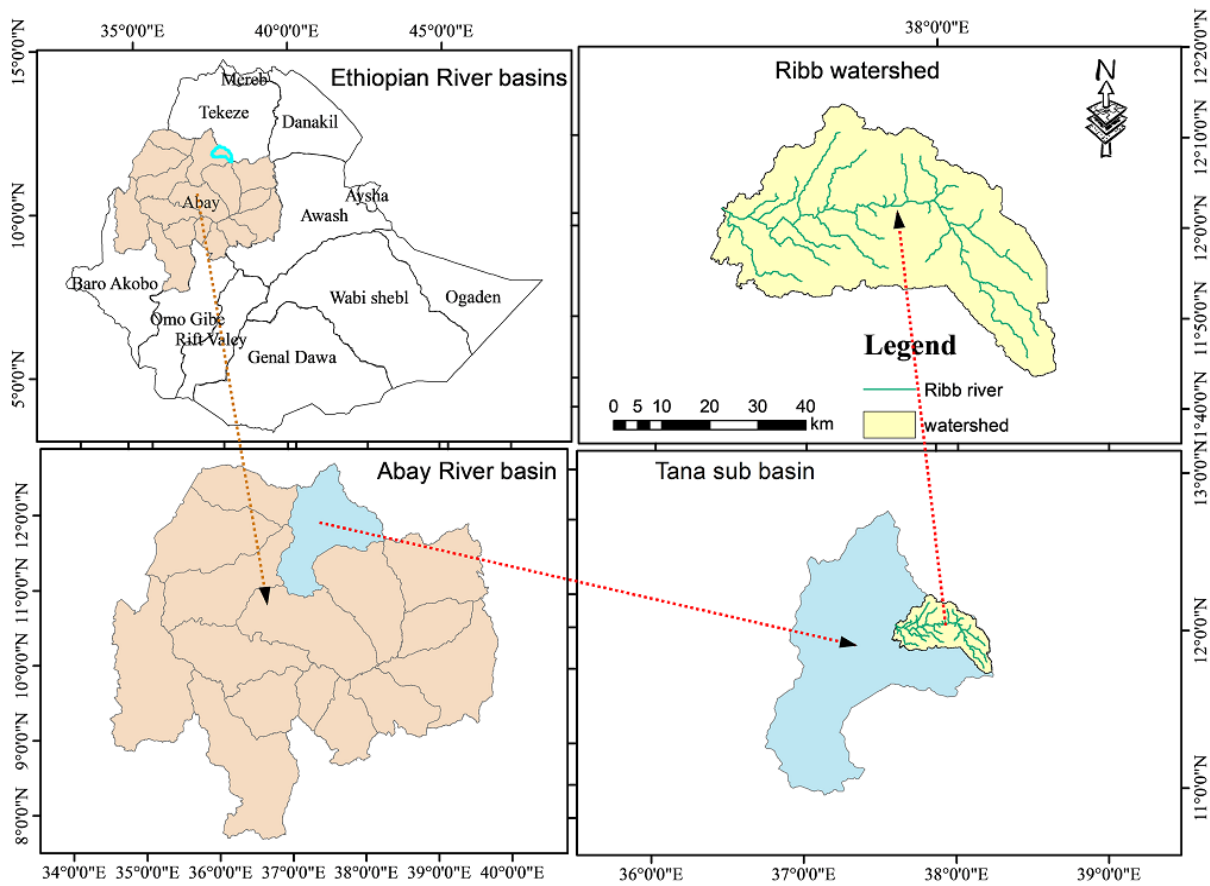


Figure 1: Map of the study area

2.2. Peaks over threshold series generation

Concept of “Theory of Runs” has been applied in this study to generate peaks and volume over threshold between a pair of up-crossing and its accompanying successive down-crossing constituting an analogous partial duration series (PDS).

In peaks over threshold model, a threshold discharge level Q_0 is fixed. This model replaces the continuous hydrographs of flows by a series of randomly spaced spikes on the time axis whose magnitudes are greater than Q_0 . The spikes themselves are of random size. Thus, in this model we have two random variables namely inter event times and peak magnitudes. The inter event times $(t_2 - t_1), (t_3 - t_2) \dots$ are random variables; having various statistical properties. In applications, only the average time between events, \bar{t} needs to be known.

2.3. Threshold selection

The selection of the threshold is a subjective process (figure 2). Madsen et al. (1993) recommended the use of the standard frequency factor k and data properties (mean value and standard deviation). The threshold value can be calculated using the following expression:

$$Q_0 = \bar{Q} + k\sigma_q \quad (1)$$

The frequency factor k with value 3 can be selected. The low k value means that a low threshold value is selected and more events are included in the Flood Frequency Analysis.

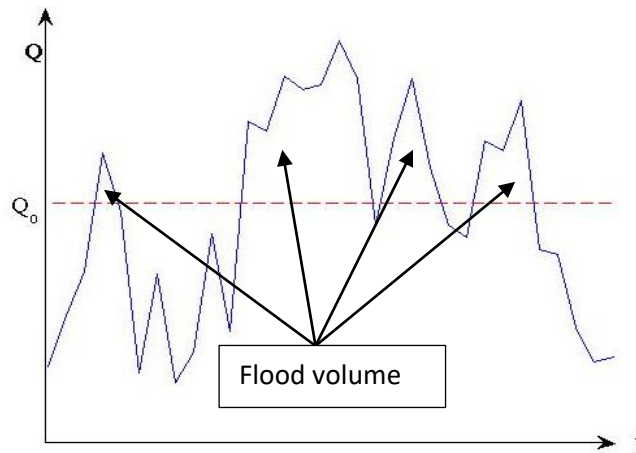


Figure 2: Actual hydrograph

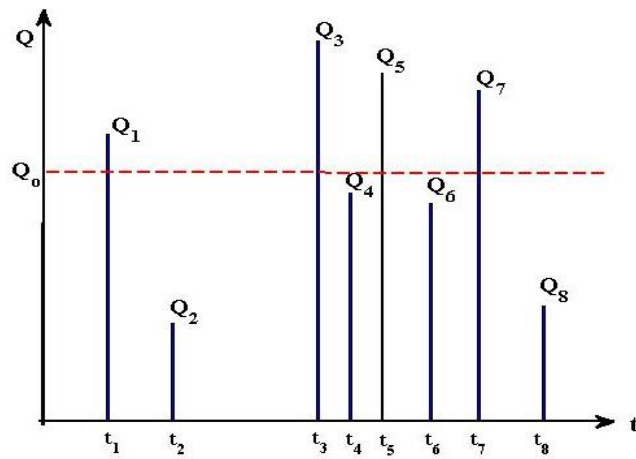


Figure 3: Generation of POT

2.4. Univariate Marginal Distributions

2.4.1. Empirical Non-Exceedance Probabilities

Empirical non-exceedance probabilities for observed values of the flood variables were estimated using

$$P_k = \frac{k - 0.4}{N + 0.2} \quad (2a)$$

$$P_k = \frac{k}{N + 1} \quad (2b)$$

Where: k is k^{th} smallest observation in the data set of a flood variable arranged in ascending order; P_k is probability of the k^{th} value; and N is number of observations of the variable or the sample size.

2.4.2. Lognormal distribution

The PDF of the lognormal distribution is expressed as:

$$f(x) = \frac{1}{x\sigma_y\sqrt{2\pi}} \exp\left[-\frac{(\ln x - \mu_y)^2}{2\sigma_y^2}\right] \quad (3)$$

The CDF of the lognormal distribution is expressed as:

$$F(x) = \frac{1}{\sigma_y\sqrt{2\pi}} \int_0^x \frac{1}{x} \exp\left[-\frac{(\ln x - \mu_y)^2}{2\sigma_y^2}\right] dx \sim \text{NOR}(\mu_y, \sigma_y^2), y = \ln x \quad (4)$$

where μ_y , and σ_y are the mean and standard deviation of natural logarithm of x .

2.4.3. Gamma (2) distribution

The PDF of the gamma (2) distribution is expressed as:

$$f(x) = \frac{1}{\alpha^\beta \Gamma(\beta)} x^{\beta-1} e^{-(x/\alpha)} \quad (5)$$

The CDF of the gamma (2) distribution is expressed as:

$$F(x) = \frac{1}{\alpha^\beta \Gamma(\beta)} \int_0^x x^{\beta-1} e^{-(x/\alpha)} dx \quad (6)$$

2.5. Copulas

Letting $F(x, y)$ be a joint cumulative distribution function with marginal distributions $F(x)$ and $F(y)$, then a copula C exists such that

$$F(x, y) = C(F(x), F(y)) = C(u, v) \quad (7)$$

Conversely, for any univariate marginal distributions $F(x)$ and $F(y)$ and any copula C , the function $F(x, y)$ given in Equation (7) is a joint distribution function with marginal distribution functions $F(x)$ and $F(y)$. where u and v denote two dependent cumulative distribution functions, $F(x)$ and $F(y)$, and range between 0 and 1.

(Nelsen, 2006; Shiau, Feng, & Nadarajah, 2007; Zhang & Singh, 2006) summarize several types of copulas. In this study, the following commonly used copulas and associated density functions are employed to investigate the relationship between flood peak and volume.

Table 1: Important copulas in practical application

Copula	Formulae
Ali-Mikhail-Haq	$C(u, v) = \frac{uv}{1 - \theta(1-u)(1-v)}, -1 \leq \theta \leq 1$
Clayton	$C(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-\frac{1}{\theta}}, \theta \geq 0$
Frank	$C(u, v) = -\frac{1}{\theta} \ln \left[1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1} \right], \theta \neq 1$
Gumbel-Hougaard	$C(u, v) = \exp \left\{ -\left[(-\ln u)^\theta + (-\ln v)^\theta \right]^{\frac{1}{\theta}} \right\}$

Where: θ is a parameter used to measure the degree of association between u and v . A copula is a joint distribution function expressed in terms of univariate marginal distribution functions; the following elementary properties of copulas are thus easily tested as $C(u, 0) = C(0, v) = 0$ and $C(u, 1) = u$ and $C(1, v) = v$

2.6. Parameter Estimation

In this study, we used approach based on the rank correlation to estimate parameter of copulas for bivariate flood frequency analysis of flood peaks and volume. This approach is based on the relationship between the dependence parameter θ and the rank correlation coefficient. For random variable X and Y , let $\{x_1, x_2, \dots, x_n\}$ and $\{y_1, y_2, \dots, y_n\}$ denote n observations of X and Y , respectively, then

$$\tau = \binom{n}{c}^{-1} (n_c - n_d) \tag{8}$$

$$\tau = \binom{n}{c}^{-1} \sum_{i < j} \text{sign}[(x_i - x_j)(y_i - y_j)], i, j \in [1, n]$$

Where: n_c and n_d are the number of concordance pairs and discordance pairs, respectively, and n is the number of observations. (Nelsen, 2006) expresses the Kendall's τ in terms of copulas (Table 2).

Table 2: Four Copula types for application

Copula type	Kendall's tau
Ali-Mikhail-Haq	$\tau = \frac{3\theta - 1}{3\theta} - \frac{2}{3} \left(1 - \frac{1}{\theta}\right)^2 \ln(1 - \theta)$
Clayton	$\tau = \frac{\theta}{\theta + 2}$
Frank	$\tau = 1 - \frac{4}{\theta} [1 - D_1(\theta)]$
Gumbel-Hougaard	$\tau = \frac{1}{1 - \theta}$

Note: $D_k(x)$ is the Debye function, for any positive integer k ,

$$D_k(x) = \frac{k}{x^2} \int_0^x \frac{\ln t}{e^t - 1} dt$$

For Archimedean Copulas, Kendall's τ takes the form of $\tau = 1 - 4 \int_0^1 \frac{\varphi(t)}{\varphi'(t)} dt$, where $\varphi(t)$ is the generator function. Based on the relationship between Kendall's τ and the copulas, the dependence parameter can be estimated.

2.6.1. Dependence of Flood Variables

The dependence of flood variables was ascertained using Pearson's linear correlation coefficient ρ , which can be expressed as

$$\rho = \frac{E[(x - \mu_x)(y - \mu_y)]}{\sigma_x \sigma_y} = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{N \sqrt{S_x^2} \sqrt{S_y^2}} \tag{9}$$

Where: N is sample size, \bar{x} and \bar{y} = sample means of flood variables X and Y ; S_x^2 and S_y^2 = sample variances of X and Y .



2.7. Identification of Archimedean Copula

(Genest & Rivest, 1993) proposed a methodology for selection between bivariate Archimedean copulas. For random variables X and Y of size n , with CDF $F_X(x)$ and $F_Y(y)$, respectively, the corresponding copula is $C(u, v)$, where u, v are the CDF of X and Y , respectively. The following steps are involved in Genest and Rivest method

- 1) Let the random variable $Z = Z(x, y)$ which had the property $K(z) = P(Z \leq z)$ where $K(z)$ is defined as $K(z) = z - \frac{\varphi(z)}{\varphi'(z)}$ where $\varphi(z)$ is the generator function and $\varphi'(z)$ denotes the derivative of $\varphi(z)$ with respect to z . The appropriate generator function needs to be identified so as to identify the appropriate copula.

- 2) calculate the empirical copula, $\hat{K}(z)$
 Define the variable

$$z_i = \frac{\text{Number of } (x_j, y_j) \text{ such that } x_j < x_i \text{ and } y_j < y_i}{n - 1} \quad i, j = 1, 2, \dots, n \quad (10)$$

- 3) Set the estimate of K such that $\hat{K}(z) = \text{the proportion of } z_i \text{'s } \leq z$
- 4) Calculate Kendall's tau by

$$\tau = \binom{n}{2}^{-1} \sum_{i < j} \text{sign}[(x_i - x_j)(y_i - y_j)], i, j \in [1, n]$$

Calculate the dependence parameter θ according to the relationship between θ and Kendall's tau, and then the generator function $\varphi(z)$ corresponding to each copula is obtained for a specified copula.

- 5) Using $\varphi(z)$, calculate a parametric estimate of K by $K(z) = z - \frac{\varphi(z)}{\varphi'(z)}$ corresponding to each generator function.
- 6) Compare the $K_\varphi(z)$ with the nonparametric estimate $\hat{K}(z)$, and choose the generator function that has the closest difference between $K_\varphi(z)$ and $\hat{K}(z)$, as the appropriate one. This can be determined by employing Q-Q plot or by minimizing the distance function $\int K_\varphi(z) - \hat{K}(z) dK(z)$

2.8. Joint Probability distributions

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 (7) is the probability that both X and Y are less than or equal to specific thresholds of x and y , respectively; that is

$$P(X \leq x, Y \leq y) = F(x, y) = C(F(x), F(y)) = C(u, v) \quad (11)$$

Under the condition that X and Y are mutually independent, the joint probability $F(x, y)$ becomes the product of individual probabilities of $F(x)$ and $F(y)$; that is

$$F(x, y) = F(x)F(y) \quad (12)$$

Conversely, if X and Y are completely dependent, that is, X becomes a function of Y and vice versa, the joint cumulative distribution is then reduced to a univariate probability such that

$$F(x, y) = F(x) \text{ or } F(x, y) = F(y) \quad (13)$$

In practice, the probability of both X and Y exceeding their respective thresholds can also be defined by copulas.

$$P(X > x, Y > y) = 1 - F(x) - F(y) + C(F(x), F(y)) \quad (14)$$

$$= 1 - u - v + C(u, v)$$

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 use of copulas.

2.9. Joint Return Periods

The joint return period of the event (X, Y) may be defined as

$$T(x, y) = \frac{1}{1 - C(F(x), F(y))} \quad C(F(x), F(y)) = \Pr[(X \leq x, Y \leq y)] \quad (15)$$

This event represents the case that either x or y or both are exceeded ($X > x$, or $Y > y$, or $X > x$ and $Y > y$). The joint return period of X and Y associated with the event ($X > x$ and $Y > y$, i.e. both x and y are exceeded) may be defined as

$$T'(x, y) = \frac{1}{C'(F(x), F(y))} = \frac{1}{1 - F(x) - F(y) - C(F(x), F(y))} \quad (16)$$

In addition, the return period can also be defined by the event for X given $Y \geq y$ or event for Y given $X \geq x$ that are called the conditional return period for X given $Y \geq y$ and the conditional return period for Y given $X \geq x$, respectively.

$$T_{X|Y \geq y} = \frac{1}{[1 - F(y)][1 - F(x) - F(y) + C(F(x), F(y))]} \quad (17a)$$

$$T_{Y|X \geq x} = \frac{1}{[1 - F(x)][1 - F(x) - F(y) + F(x, y)]} \quad (17b)$$

The conditional distribution of Y given $X \geq x$ and the conditional distribution of X given $Y \geq y$ can also be estimated as the following.

$$F(y|X \geq x) = \frac{P(X \geq x, Y < y)}{P(X \geq x)} = \frac{F(y) - C(F(x), F(y))}{1 - F(x)} \quad (18a)$$

$$F(x|Y \geq y) = \frac{P(X < x, Y \geq y)}{P(Y \geq y)} = \frac{F(x) - C(F(x), F(y))}{1 - F(y)} \quad (18b)$$

2.10. Goodness-of-Fit Statistics

Goodness-of-fit statistics are most popularly used for statistical model selection. In this study, the root mean square error (RMSE), bias, maximum likelihood (ML), Akaike Information Criterion (AIC), K-S test, Chi-Squared and Anderson-Darling(AD) test were applied in both univariate and bivariate cases. RMSE can be used to measure the goodness of fit of the distribution. Accordingly, the best model is the one which has the minimum RMSE. RMSE can be expressed through the mean square error (MSE) as:

$$MSE = E(x_c - x_o) = \frac{1}{n-1} \sum (x_c(i) - x_o(i))^2 \quad (19)$$

Then

$$RMSE = \sqrt{MSE} \quad (20)$$

Where $E[\cdot]$ is the expectation of $[\cdot]$, x_c denotes the computed value, n denotes the sample size, and x_o denotes the observed value.

Bias is a measure of the deviation of the estimated quantity from the observed (or true) value. The best model is the one with smallest bias. Bias is expressed by:

$$Bias = \sum_{i=1}^n \frac{x_o(i) - x_c(i)}{x_o(i)} \quad (21)$$

Where $x_o(i)$ denotes the i^{th} observed value, and $x_c(i)$ denotes the i^{th} computed value.

The AIC criterion, developed by (Akaike, 1974), consists of two parts: lack-of-fit of the model and the unreliability of the model due to the number of model parameters. The AIC criterion can be expressed through two approaches: maximized likelihood and MSE of the model as:

$$AIC = -2\log(\text{maximized likelihood for model}) + 2(\text{no. of fitted parameters}) \quad (22)$$

2.11. Chi-square test

In the test, the data are divided into k class intervals (k is recommended to be more than 5). The statistic Chi-square (χ^2) is given by

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \quad (23)$$

Where O_i is the observed number of events in the class interval i , E_i is the number of events that would be expected from the summed theoretical distribution and k is an arbitrary number of classes to which the observed data are divided.

2.12. Kolmogorov-Smirnov test

The test statistic is the maximum vertical distance between the empirical and hypothetical cumulative distribution function (CDF), which is defined as follows,

$$D_n = \sup_x |F(x) - F_n(x)| \quad (24a)$$

Where $F(x)$ are the estimated values by the proposed theoretical distribution, and $F_n(x)$ is denoted by:

$$F_n(x) = \begin{cases} 0, & x < x_1 \\ k/n, & x_k < x < x_{k+1} \\ 1, & x \geq x_n \end{cases} \quad (24b)$$

Where $x_1, x_2, \dots, x_k, \dots, x_n$ are the values of the increasingly ordered sample data and n is the sample size.

Table 3: Critical D value for K-S test (Kanji, 1993)

α	0.10	0.05	0.25	0.01	0.005	0.001
Critical value, D_n^α	1.22	1.36	1.48	1.63	1.73	1.95

2.13. Anderson-Darling Test

(T. W. Anderson and D. A. Darling, 1954) defined the statistic for this test as

$$W_n^2 = n \int_{-\infty}^{\infty} [F_n(x) - F^*(x)]^2 \psi(F^*(X)) dF^*(x) \quad (25a)$$

Where ψ is a nonnegative weight function which can be computed by, $\psi = [F^*(x)(1 - F^*(x))]^{-1}$ In order to make the computation of this statistic easier, (Arshad, Rasool, & Ahmad, 2003) applied the following formula,

$$W_n^2 = -n - \frac{1}{n} \sum (2i - 1) \{ \log F^*(X) + \log(1 - F^*(X_{n+1-i})) \} \quad (25b)$$

Where, $F^*(x_i)$ is the cumulative distribution function of specified distributions, x_i 's are the ordered data and n is the sample size. This study used the following modified AD statistic which takes into accounts the sample size n,

$$W_n^{2*} = W_n^2 (1 + 0.75/n + 2.25/n^2) \quad (25c)$$

3. Results and Discussion

To illustrate the above concepts, as well as the joint distribution of flood peak and volume and the corresponding return period of an actual basin, the Ribb River, is analyzed. The basin has an area of 1500 km², located at the Upper Blue Nile, Ethiopia. In the basin, annual flood events occur in the summer due to the contribution of summer precipitation to river runoff. There are 21 years of daily streamflow data available from 1987 to 2007 at the gauging station Addiszemen, near the outlet of the basin.

3.1. Flood peak and volume generation

Two properties, flood volume (V) and flood peak (Q) are used to characterize the extreme flood events. The flood peak is defined as the maximum daily flow during the flood period, while the flood volume is defined as the cumulative flow volume during the flood period above some truncation levels.

In this study a method suggested by (Madsen, Pearson, & Rosbjerg, 1997) was used to determine truncation level. The threshold value was calculated using equation (1). The frequency factor k with value 3 is selected. Therefore, peak thresholds are considered at

40m³/s in this study, i.e. flood events are defined as a daily stream flow equal to or greater than 40m³/s. 87 peaks are abstracted from the recorded daily stream flow data.

3.2. Univariate Flood Frequency Analyses

3.2.1. Fitting Marginal Probability Distributions to Flood peak and volume

The first step in constructing bivariate distribution is to determine the univariate marginal probability distributions. Eight univariate probability distributions are the candidates for fitting flood peak and volume. The goodness-of-fit statistics, i.e., Kolmogrov-Smirnov (K-S) test (eq. 28a), Anderson-Darling (eq.29c) and Chi-square test (eq.27), were applied for selection of distribution. The goodness-of-fit statistics for each fitted distribution are given in *Table* for flood peaks. Thus, based on the goodness-of-fit statistics, the observed flood peaks are fitted by a Gamma distribution. Empirical non-exceedance probabilities were estimated using the Gringorten plotting-position formula (eq.2a) for flood peak and Weibull plotting-position (2b) for flood volume given in (Table 4).

Table 4: Ranking of distribution fitted to flood peaks

Distribution	K-S	Rank	AD	Rank	Chi-square	Rank
Exponential	0.23498	8	10.414	8	21.338	3
Gamma	0.09099	1	0.88436	1	15.178	1
GEV	0.0946	3	4.7504	7	N/A	
Gumbel	0.12912	7	1.9277	6	24.181	7
Log-Logistic	0.12617	6	1.5527	5	24.148	6
Log-Pearson 3	0.09263	2	1.0144	3	20.89	2
Lognormal	0.10545	5	1.1818	4	23.318	5
Normal	0.09547	4	0.98282	2	21.52	4

Therefore, there is insufficient evidence to reject the null hypothesis that the observed flood peaks are drawn from a Gamma distribution at significant level of 5 percent. Parameters of the distribution fitted to flood peaks are estimated using the MOM, MLE, and LMM as shown in Table 5.

Table 5: Parameters of Gamma distributions for flood peaks.

Distribution	Method	Parameters	
		β	α
Gamma (β, α)	MOM	16.98	4.82
	MLE	17.40	4.71
	LMM	17.62	4.65

The goodness-of-fit techniques, i.e., MSE (eq. 23), RMSE (eq. 24), BIAS (eq. 25), AIC (eq. 26) and K-S test (eq.28a) values are calculated as given in Table to identify best parameter estimation method. These indicate that four goodness-of-fit statistics show the same tendency, i.e., when a smaller MSE is obtained, smaller RMSE, BIAS and AIC are obtained correspondingly. Thus, based on the goodness-of-fit statistics, the chosen best fitted distributions for flood peaks are identified, with the best fitted cumulative density plots shown Figure 4 and **Error! Reference source not found.**

Table 6: Goodness-of-fit statistics for flood peak parameter estimation

Method	MOM	MLE	LMM
MSE	0.0028	0.0021*	0.0085
RMSE	0.0526	0.0462*	0.092
BIAS	6.23	0.0606*	24.64
AIC	-216.5	-226.4*	-174.4
KS	0.1096	0.093*	0.19

Note: * represents the chosen method

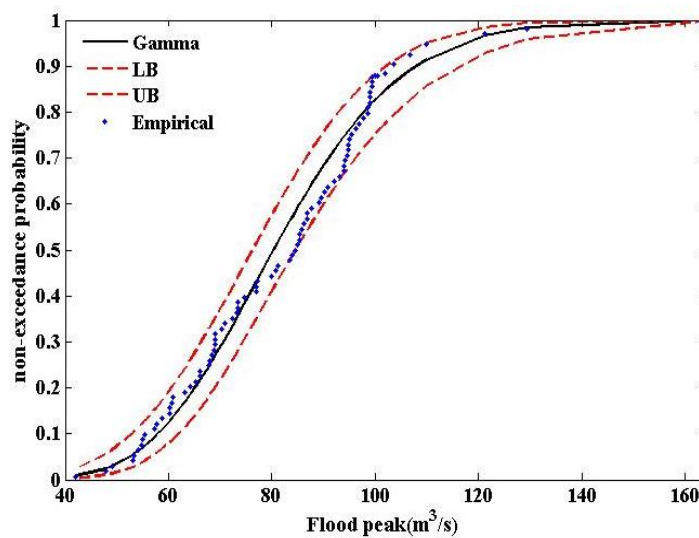


Figure 4: Gamma distribution fitting flood peaks.

On the other hand, flood volumes are fitted by the two-parameter lognormal distribution. The goodness-of-fit statistics for each fitted distribution are given in Table for flood volume. Thus, based on the goodness-of-fit statistics, the observed flood volumes are fitted by a lognormal distribution. Parameters of the distribution fitted to flood volume are estimated using the method of moments (MOM), maximum likelihood estimation (MLE), and L-moment method (LMM) as shown in Table .

The goodness-of-fit statistics, i.e., MSE, RMSE, BIAS, AIC and K-S test values are computed as given in Table to identify best parameter estimation method. Thus, based on the goodness-of-fit statistics, the chosen best fitted distributions for flood volume are given in Table , with the best fitted cumulative probability plots shown Figure 5.

Table 7: Ranking of distribution fitted to flood volume.

Distribution	K-S	Rank	AD	Rank	Chi-square	Rank
Exponential	0.26585	7	12.978	8	37.317	7
Gamma	0.11592	4	1.662	3	9.1037	3
GEV	0.1455	5	2.3956	4	10.352	4
Gumbel	0.27543	8	7.4955	6	12.727	5
Log-Logistic	0.06695	3	0.35331	2	2.6487	2

Distribution	K-S	Rank	AD	Rank	Chi-square	Rank
Log-Pearson 3	0.0659	2	4.2548	5	N/A	
Lognormal	0.061301	1	0.34645	1	2.6254	1
Normal	0.26386	6	10.354	7	24.296	6

Table 8: Parameters of fitted probability distributions for flood volume

Distribution	Method	μ_y	σ_y
Log-normal (μ_y, σ_y)	MOM	16.23198	1.6345
	MLE	16.23198	1.643959
	LMM	16.23204	1.643927

Table 9: Goodness-of-fit statistics for flood volume

Method	MOM	MLE	LMM
MSE	0.0008	0.0007	0.0008
RMSE	0.028	0.0295	0.0295
BIAS	2.172	5.87	5.87
AIC	-267.4	-262.25	-262.25
KS	0.059	0.0741	0.0741

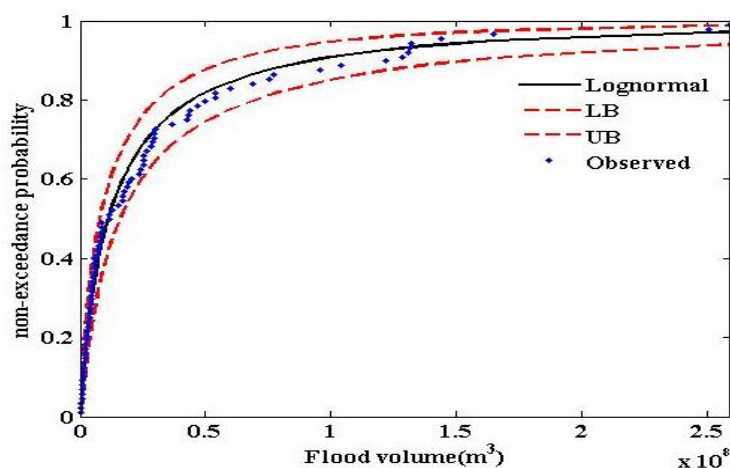


Figure 5: Flood Volume fitted by lognormal distribution

3.3. Joint Flood Frequency Analysis using Copula

3.3.1. Correlation between Flood Peaks and Volumes

For this study, Pearson’s linear coefficient (ρ) (eq.13) and Kendall’s tau rank coefficient (τ) eq. (8) were used to determine correlation between flood peaks and volume. The estimated values of ρ and τ are given in s.



which shows that there is positive correlation between flood peaks and volumes.

3.3.2. Identification of Best Copula Method

Nonparametric estimation method is used to estimate copula parameters and for identification of copula based bivariate flood frequency analysis of flood peaks and volumes. The four commonly used Archimedean copula families are considered as candidates for joint flood frequency analysis. These are Ali-Mikhail-Haq, Clayton, and Gumbel-Hougaard copula. The parameter estimated by Kendall's τ is given in Table for each copula method.

The goodness of fit of observed data to the theoretical bivariate distribution obtained by using copula functions are tested by AIC and KS methods. Table 12 shows that, AIC and KS values for bivariate distributions obtained by using different copula function for flood peaks and volumes. It can be concluded from Table 12 that Clayton copula is appropriate to characterize joint distribution of flood peaks and volumes for this study. Based on this, joint CDF, conditional probability and their joint return periods for different combinations of flood peaks and volumes are calculated.

Table 10: Correlation of flood peaks and volume

Kendall's tau (τ)	0.593
Pearson rho (ρ)	0.683

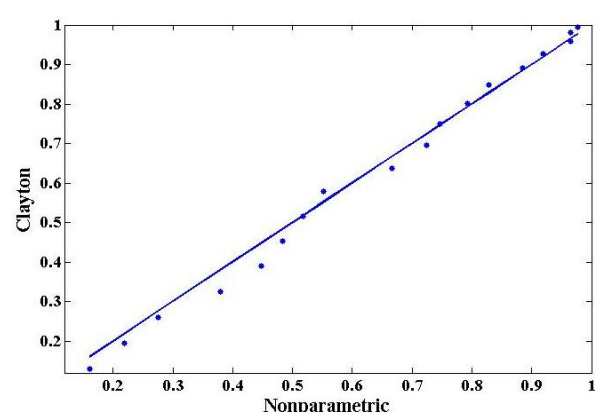
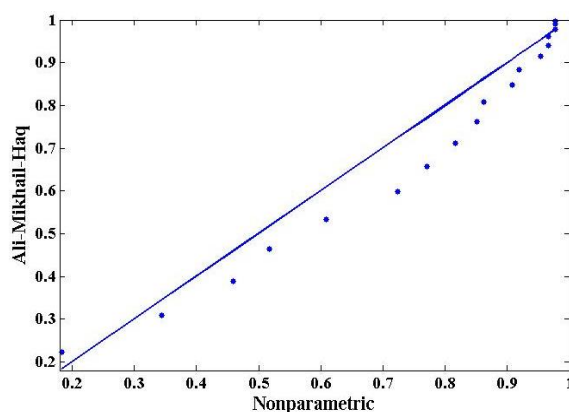
Table 11: Parameters estimated for copulas.

Copula	Gumbel-Hougaard	Clayton	Ali-Mikhail-Haq	Frank
Theta	2.62	3.24	0.897	8.56

Table 12: Goodness-of-fit statistics for copulas

Method	Gumbel-Hougaard	Clayton	Ali-Mikhail-Haq	Frank
AIC	-117.46	-129.53*	- 84.59	-75.18
KS	0.0628	0.0575*	0.1608	0.1735

Note: * represents the chosen copula.



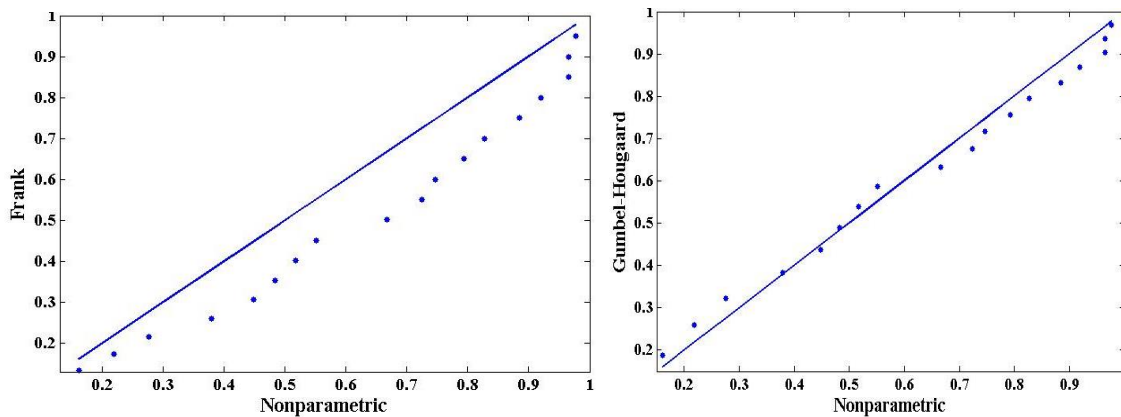


Figure 6: Comparison of nonparametric and parametric $K(z)$

a. Joint cumulative density function (JCDF)

The combined CDF of flood peaks and volumes is calculated based on Table 1 of Clayton copula method, and is showed in *Figure 7*. In which u and v were estimated by eq.6 and 4 respectively.

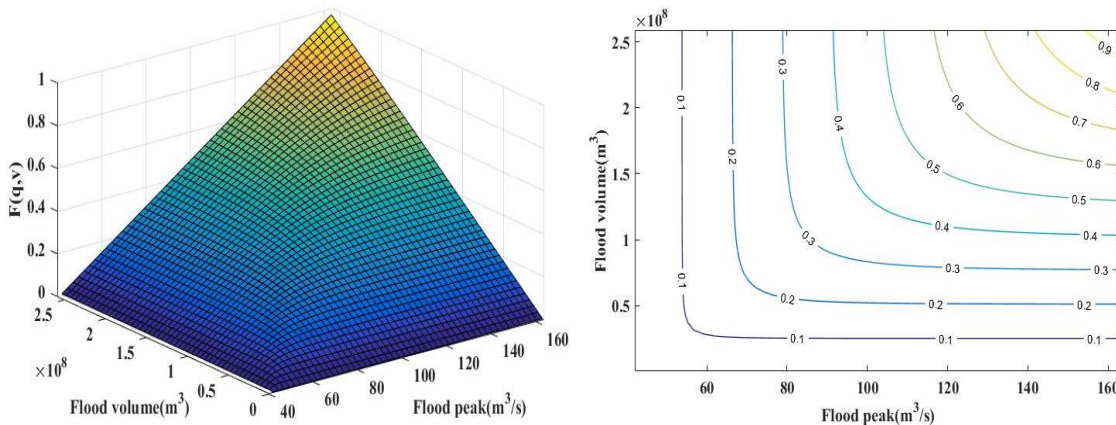


Figure 7: JCDF of flood peaks and volumes.

b. Joint and Conditional Return periods

The average inter-arrival time defined as stream flow equaling to or exceeding $40\text{m}^3/\text{s}$, for the Addiszemen gauge station is 88.1days, namely 0.241years. Based on this information estimated from the observed floods, the joint return period for flood volume equal to or greater than a given value or flood peak equal to or greater than another given value, i.e., $V \geq v$ or $Q \geq q$ is expressed as:

$$T_{QV} = \frac{E(t)}{1 - F(q, v)} \tag{26}$$

Where $E(t) = 0.241$

The combined return period for flood peaks equal to or greater than a given value and flood volume equal to or greater than another given value, $V \geq v$ and $Q \geq q$; is expressed as

$$T'_{QV} = \frac{E(t)}{F'(q, v)} \tag{27}$$

$$F'(q, v) = 1 - F(q) - F(v) + F(q, v) \tag{28}$$

The joint return period $T(q, v)$ and their contour lines are displayed in Figure 8.

The JCDF $F'(q, v)$ of the flood peaks and volumes and their return period $T'(q, v)$ are estimated by eq. (27) and (32) respectively. And the result is depicted in the Figure 9 below. From Figure 7 and Figure 9, it can be understood that the two JCDFs are not similar. It is also distinguished that the two types of joint return period are also different.

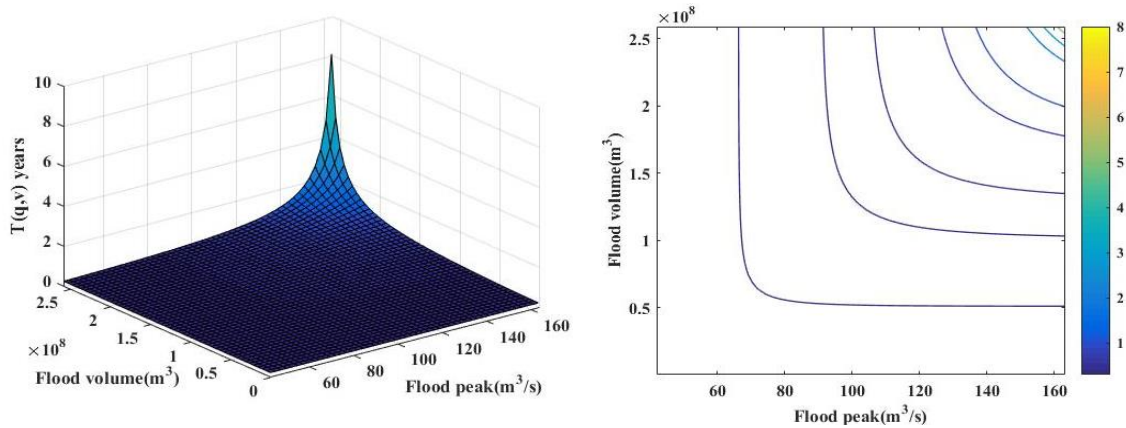


Figure 8: Joint return period.

Contrasting the return period defined by one variable, the specific bivariate return periods can be obtained using various combinations of the flood peaks and volume. Hence, the joint return period for flood peaks and flood volume must be showed using the contour lines. However, the joint return periods $T(q, v)$ and $T'(q, v)$ show different characteristics. For the same value of q and v , $T'(q, v)$ is greater than $T(q, v)$ according to Eqs. (30) and (31). For example, $q = 129m^3/s$ and $v = 250Mm^3$, $T(q, v) = 5$ years, and $T'(q, v) = 97$ years. One variable flood frequency analysis reveals that the return period of such a flood peak is 15 years, and the return period of flood volume is 8.30 years. Bivariate return periods estimated by T'_{QV} become very large when flood peak and volume exceed large given values at the same time; that is, such events seldom occur in the long term.

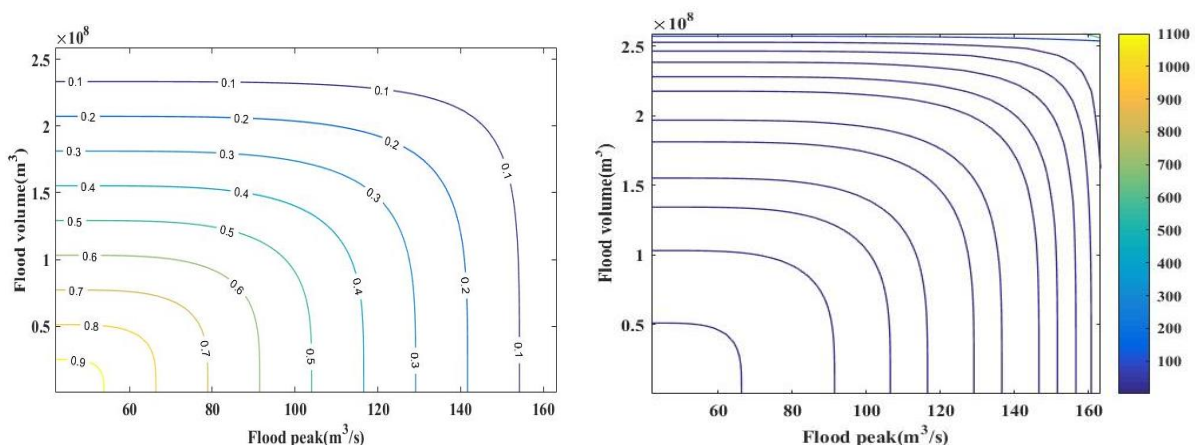


Figure 9: $F'(q, v)$ and $T'(q, v)$

The conditional probability $F_{q|v}$ and their return period $T_{q|v}$ of flood peak Q given flood volume V is estimated based on Equation (18b) and (18a), and is plotted in Figure 10. Likewise, the conditional probability $F_{v|q}$ and their return period $T_{v|q}$ of flood volume V

given flood peak Q is estimated based on Equation (17a) and (17b), and is plotted in **Error!**
Reference source not found..

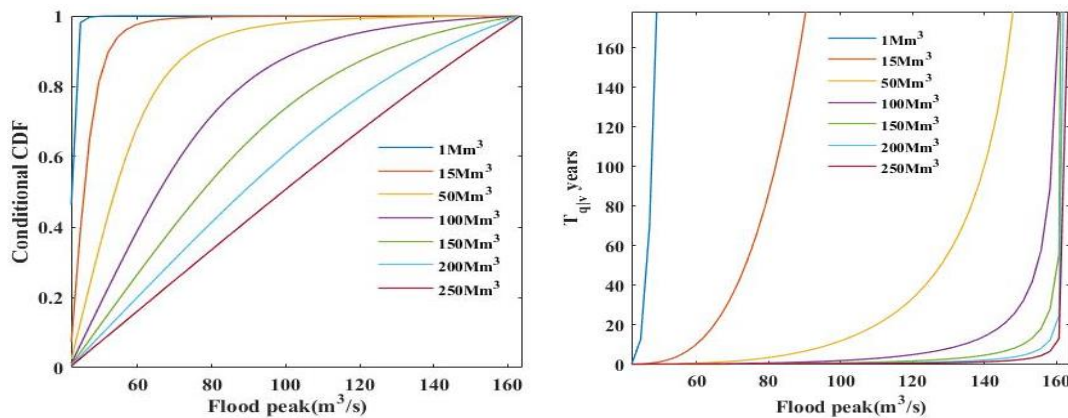


Figure 10: $F_{q|v}$ and $T_{q|v}$ of flood peak given flood volume.

As shown on Figure 10 and 11, anyone can determine the return period of flood peak above a certain magnitude given flood volume equaling or exceeding any specific value. For example, given a 5-year flood event defined uniquely by the flood volume, i.e., flood volume exceeding 77.35Mm³, the return period of flood peaks equal to or greater than 97.49m³/s is 5 years. Undoubtedly, this evidence cannot be gained from the one variable frequency analysis.

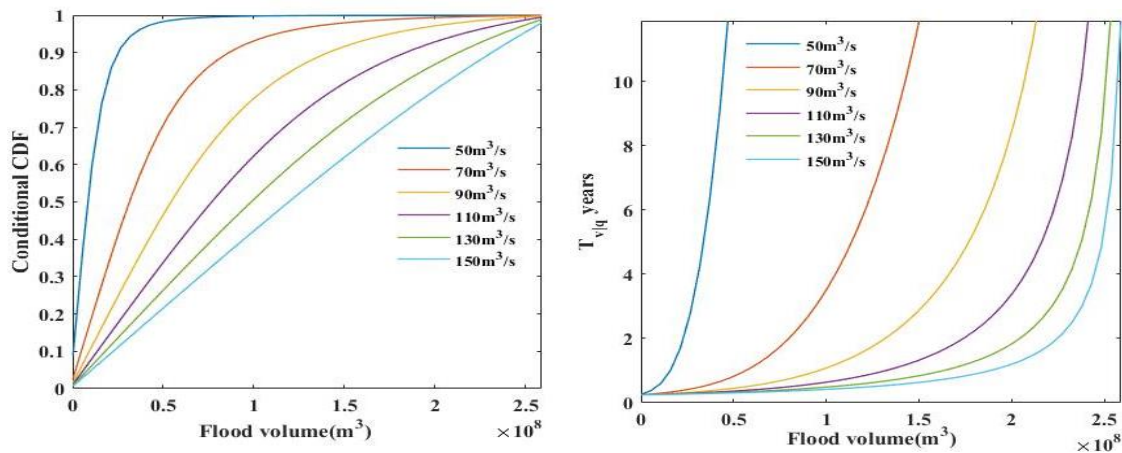


Figure 11: $F_{v|q}$ and $T_{v|q}$ of flood, volume given flood peak.

4. Conclusion

Damaging floods occur when high flood peak sustain for a longer period or huge volume of floodwater inundates an area for a longer time. This emphasizes the need to study the joint distribution of flood peak and volume together. The flood peak is the maximum daily flow during the flood period, while the flood volume is the cumulative flow volume during the flood period the truncation levels. The flood parameters were further fitted using the concept of Copula to construct joint bivariate distribution function of peak flow-volume. Based on Kolmogorov-Smirnov (KS) and Akaike Information Center (AIC) test the Clayton copula fits flood peak and volume. The correlation among flood variables indicates that there is a good agreement between flood peak and volume. The bivariate copula provides all information, which cannot be obtained by single-variable flood frequency analysis. This study should be useful for practitioners to analyze multivariate hydrological extreme events such as floods and droughts. These results can be useful for hydrological

structure design studies and expected that same methodology can be applied to any station over Ethiopia. The overall conclusion is that univariate frequency analysis cannot provide a sufficient probabilistic assessment of correlated multivariate hydrological events and may lead to overestimation or underestimation of the severity of these events.

References

- A. Ramachandra Rao and Khaled H. Hamed. (2000). Flood Frequency Analysis. 356.
- Akaike, H. (1974). A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, 19(6), 716-723.
- Arshad, Rasool, & Ahmad. (2003). Anderson Darling and Modified Anderson Darling Tests for Generalized Pareto distribution. 85–88.
- Chen, X., & Fan, Y. (2004). Estimation of Copula-Based Semiparametric Time Series Models. (615).
- Favre, A., Adlouni, S. El, Perreault, L., Thiémonge, N., & Bobée, B. (2004). Multivariate hydrological frequency analysis using copulas. *Water Resources Research*, 40(1).
<https://doi.org/10.1029/2003WR002456>
- Genest, C., & Rivest, L.-P. (1993). Statistical Inference Procedures for Bivariate Archimedean Copulas. *Journal of the American Statistical Association*, 88(423), 1034–1043.
<https://doi.org/10.2307/2290796>
- Grimaldi, S., & Serinaldi, F. (2006). Asymmetric copula in multivariate flood frequency analysis. *Advances in Water Resources*, 29(8), 1155–1167.
- Haan, C. T. (1977). *Statistical Methods*.
- Madsen, H., Pearson, C. P., & Rosbjerg, D. (1997). Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events: 2. Regional modeling. *Water Resources Research*, 33(4), 759.
- N.K.Goel, S. M. S. & S. C. (1998). MULTIVARIATE MODELING OF FLOOD FLOWS. 124(2), 146–155.
- Nelsen, R. B. (2006). *An Introduction to copulas* second edition.
- Sackl, B., & Bergmann, H. (1987). A Bivariate Flood Model and Its Application BT - Hydrologic Frequency Modeling: Proceedings of the International Symposium on Flood Frequency and Risk Analyses, 14–17 May 1986, Louisiana State University, Baton Rouge, U.S.A. (V. P. Singh, Ed.).
https://doi.org/10.1007/978-94-009-3953-0_40
- Shiau, J. T., Feng, S., & Nadarajah, S. (2007). Assessment of hydrological droughts for the Yellow River, China, using copulas. *Hydrological Processes*, 21(16), 2157–2163.
<https://doi.org/10.1002/hyp.6400>
- Singh, K., & Singh, V. P. (1991). Derivation of bivariate probability density functions with exponential marginals. *Stochastic Hydrology and Hydraulics*, 5(1), 55–68.
<https://doi.org/10.1007/BF01544178>
- T. W. Anderson and D. A. Darling. (1954). A Test of Goodness of Fit. 49(268), 765–769.
- Yue, S, Ouarda, T. B. M. ., Bobée, B., Legendre, P., & Bruneau, P. (1999). The Gumbel mixed model for flood frequency analysis. *Journal of Hydrology*, 226(1–2), 88–100.
[https://doi.org/10.1016/S0022-1694\(99\)00168-7](https://doi.org/10.1016/S0022-1694(99)00168-7)
- Yue, Sheng. (1999). Applying Bivariate Normal Distribution to Flood Frequency Analysis. *Water International*, 24(3), 248–254. <https://doi.org/10.1080/02508069908692168>
- Yue, Sheng. (2000). The gumbel logistic model for representing a multivariate storm event. *Advances in Water Resources*, 24(2), 179–185. [https://doi.org/10.1016/S0309-1708\(00\)00039-7](https://doi.org/10.1016/S0309-1708(00)00039-7)
- Yue, Sheng. (2001). A bivariate gamma distribution for use in multivariate flood frequency analysis. *Hydrological Processes*, 15(6), 1033–1045. <https://doi.org/10.1002/hyp.259>

- Yue, Sheng, & Rasmussen, P. (2002). Bivariate frequency analysis: Discussion of some useful concepts in hydrological application. *Hydrological Processes*, 16(14), 2881–2898. <https://doi.org/10.1002/hyp.1185>
- Zhang, L., & Singh, V. (2006). Bivariate Flood Frequency Analysis Using the Copula Method. *Journal of Hydrologic Engineering*, 11(2), 150–164. [https://doi.org/10.1061/\(ASCE\)1084-0699\(2006\)11:2\(150\)](https://doi.org/10.1061/(ASCE)1084-0699(2006)11:2(150))