**Copula-Based Joint Distribution of Flood Peak and Volume for Improved Flood Risk Assessment: A Case Study of the Surma-Meghna River Basin, Bangladesh**

Nazmul Haque1, Khalidur Rahman1

1 Department of Statistics, Shahjalal University of Science & Technology, Sylhet-3114, Bangladesh

\*Corresponding Author: Khalidur Rahman; [khalid\_sust@yahoo.com](mailto:khalid-sta@sust.edu) ; Department of Statistics, Shahjalal University of Science and Technology, Sylhet 3114, Bangladesh.

**Abstract**

This study explores a copula-based multivariate approach for assessing flood risk, focusing on the interdependent behaviors of flood peak and volume in the Surma-Meghna River basin, Bangladesh. Traditional univariate flood frequency analyses often overlook the correlation among hydrological variables, which can lead to unreliable risk estimates. Utilizing 53 years of hydrological data, we evaluated several joint distribution models to accurately capture the dependence structure between these variables. For univariate frequency analysis of flood return periods, we employed the Weibull formula alongside its corresponding marginal distribution. The flood peak was modeled using a Log-Pearson Type III distribution, while the flood volume was best described by a Log-Normal distribution. We assessed various copula families—including Clayton, Frank, Gumbel, Ali-Mikhail-Haq (AMH), Farlie-Gumbel-Morgenstern (FGM), and Galambos—with the Clayton copula demonstrating the best fit, particularly for lower tail dependence, making it particularly effective for flood risk analysis in this region. Our findings indicate that using a copula framework provides a more detailed understanding of flood risks, which has significant implications for hydraulic design and floodplain management. This study underscores the necessity of accurately modeling joint distributions in flood frequency analysis to significantly improve predictions of flood occurrences and volumes. The copula approach shows promise in enhancing flood forecasting and assessing hydrological extremes, ultimately contributing to more effective flood risk management strategies.

**Keywords:** Copula, Flood Frequency, Joint Distribution, Surma-Meghna River, Multivariate Analysis, Flood Risk Estimation.

**Highlights:**

* The study employs copula models to analyze the dependence between flood peak and volume in the Surma-Meghna River system, Bangladesh.
* By modeling the joint behavior of flood peak and volume, the study provides an enhanced method for estimating flood risk.
* The Clayton copula is identified as the most suitable for modeling flood data, enhancing the precision of return period estimations for extreme flood events.
* This work advances multivariate hydrological modeling techniques, demonstrating their importance in flood risk assessment and mitigation strategies.

**1.Introduction**

The usage of copula functions as a technique in hydrology has been increased recently. Hydrological processes are frequently complex; therefore, it makes sense for analyses to take more than one variable into account. Copulas are frequently a useful substitute for the most popular univariate frequency analyses to facilitate correlated variables. Copula functions have been extensively applied in various hydrological studies, exhibiting their versatility and importance. They have been sufficiently utilized in rainfall analysis, drought modeling, dam overtopping risk estimation, hyetograph analysis, flood coincidence risk analysis, building geostatistical models, and flood frequency analysis. Numerous researchers have leveraged copula functions in these diverse applications, showcasing their significance in the field of hydrology. In rainfall analysis, Gargouri-Ellouze and Chebchoub (2008), Zhang and Singh (2007a), Ghosh (2010), Singh and Zhang (2007), Balistrocchi and Bacchi (2011), Vandenberghe et al. (2010) and Ariff et al. (2012) used copula function. Chen et al. (2012) used it for coincidence risk analysis. De Michele et al. (2005) used in the field of dam overtopping to estimate risk. Wong et al. (2013), Wong et al. (2010), Reddy and Ganguli (2012a), Liu et al. (2011), Lee et al. (2013), Ma et al. (2013), Kao and Govindaraju (2010) used it for draught analysis. Grimaldi and Serinaldi (2006b) used it in hyetograph analysis. Indeed, Grimaldi and Serinaldi (2006a), Zhang and Singh (2006), Renard and Lang (2007), Zhang and Singh (2007b), Karmakar and Simonovic (2009) used it for analyzing flood frequency. In previous years, researchers predominantly concentrated on the examination of univariate flood frequency, particularly on flood peaks due to their common pattern in flood control engineering. However, according to Goel (1998) and Yue S. (1999), flood flows show multivariate characteristics that can be characterized by peak, volume, and duration. This suggests that there is a correlation between several random variables. Therefore, it became most important to employ multivariate extreme models when examining the dependent structure that exists among environmental factors and the highest reservoir water level (Nadarajah S., 1993; Anderson CW, 1994). Additionally, researchers developed expressions for joint distribution functions to understand the highest flood peak and occurrence of time Gupta VK (1976) and Yue S. (1999) used the Gumbel mixed model to depict the joint probability distributions of flood peaks and volumes, as well as flood volumes and durations. This model included the traditional Gumbel marginal distributions. Additionally, Yue S. (2001) used standard Gumbel marginal distributions alongside with bivariate logistic distributions to represent the association between different El Ninʃo variables. De Michele and Salvadori (2003) published the first paper about copulas in hydrology. Favre (2004) used these bivariate distributions to represent the dependency of integrated hydrologic variables, which required precise marginal distributions that could be difficult to achieve in practice.

In many of the investigations mentioned above, two key assumptions were made. First, the flood characteristics have the same marginal probability distribution. Second, they are normally distributed. However, flood characteristics are not independent, do not follow the normal distribution in general unless modified, and do not have the same type of marginal distributions. Copulas were used to generate the bivariate distribution to solve these challenges. To our best knowledge, copula functions are a relatively new technique that has not been used in Bangladesh before for frequency flood research. This technique improves our understanding of flood risks by enabling more precise modeling of the dependence structure between various hydrological variables. In this study, our primary goal is to identify the most suitable copula-based joint distribution model for estimating extreme floods, thus contributing to better flood risk assessment and mitigation strategies.

The main goal of this study is to estimate different joint distribution models based on copulas and then use the most suitable one to effectively model the correlated flood peak and volume to assess the returning period of flood more accurately.

**2. Study Area and Data:**

The current study focuses on the Surma river basin, in Bangladesh. The Surma river is a part of the Surma Meghna river system. The Surma first evolved at the mountains of Shillong and Meghalaya in India. (25°34′56′′N91°53′40′′E). The Barak River is the main source, with a large catchment area in the Naga Manipurhills bordering Myanmar, which are distinguished by ridge and valley geography. Barak-Meghna is 950 km long, with 340 km in Bangladesh. Barak divides into the steep and highly lively Surma and Kushiyara rivers as it approaches the Bangladeshi border near Amalshid (24° 52' 33.18"N and 92° 29' 15.2"E) in Sylhet district. The maximum depth between Surma and Kushiyara is 550 feet, and the average depth is 282 feet (86 meters), (Tasnim, Z et al., 2022). For this study purpose the annual average discharge and maximum discharge level data of Sylhet streamflow station (SW267) (latitude 24.88794 and longitude 91.86801) in northern part of Bangladesh, are collected from Bangladesh Water Development Board (BWDB) and considered as flood peak and flood volume data, respectively for this analysis. A duration of 53 years of data, from 1964 to 2023, were taken into consideration for this study which is available from BWDB.

**3.METHODS**

**3.1 Dependencystructure:**

Pearson correlation is used for measuring linear relationships between flood peak and volume. Since in hydrology, flood data holds some nonlinear relationships between them hence spearman rank correlation and Kendall’s tau correlation is also measured for measuring the relationship between those.

Equation of Kendall’s tau is:

where:

Cn = the total number of concordant pairs

Dn = the total number of discordant pairs

**3.2 Additional Graphical Tools for Detecting Dependence:**

**3.2.1 Chi-Plots:**

The chi-plot was first proposed by Fisher and Switzer (1985) and further they fully illustrated it in 2001. It was based on control charts and chi-square statistics for independence. the chi-statistics:

and the lambda-statistics:

where  ​, and ​ are the empirical distribution functions of the uniform random variables 𝑈1 and *U*2 and of (𝑈1, 𝑈2) respectively.

**3.2.2 K Plot:**

K plot is also a graphical tool for visualizing dependence structures. The suggestion was made by Genest and Boies (2003). It is based on a Q-Q plot. It can be expressed as

where k0=corresponding density. i in the range {1, 2… n}.

**3.3 Potential marginal distribution:**

Several candidate distributions, such as Normal, Logistic ,Weibull, Cauchy, Gamma, Log normal, Gumbel, Asymmetric Laplace Distribution (ALD), Pareto, Generalize extreme value (GEV) distribution, Lognormal 3, Weibull 3 has been used for flood volume and several extreme value distribution, such as Pearson type III, Log Pearson type III, Log normal, 3 -parameter Log normal, Weibull, 3-parameter Weibull, ALD, GEV, Gumbel has been used for flood peak. Assessment of the distribution has been measured by three statistical test i.e Kolmogorov-Smirnov test, Chi-squared fit statistics and Anderson Darling test. The null hypothesis of all of the tests assume that the sample follows a specified distribution.

**3.4 Copula:**

Copula is a joint Cumulative Density Function (CDF) of multiple random variables with marginal uniform (0,1) distribution that allows to construct a bivariate or multivariate probability function in terms of marginal distribution. The Sklar (1959) theorem, which asserts that any multivariate joint distribution can possibly be expressed in terms of univariate marginal distribution functions, served as the foundation for its composition. Let (X, Y) be any randomly distributed pair. Sklar's theorem allows it to be expressed as:

where the "copula" is a mapping function C = [0,1] × [0,1] and the marginal probability distributions F(x) and G(y). indicates that if the three parts (C, F, and G) are selected from the following parametric families: F (x, α), G (y, β), and C (u, v, θ), a suitable probabilistic model for (X, Y) can be generated. Here θ is the parameter vector for the dependency structure and α and β are the parameter vectors of the marginal distribution. The quantiles of the uniformity distribution variables, U = F(X) and V = G(Y), are denoted by the variables u and v, respectively.

There are numerous forms of copulas that are grouped into four classes that is Archimedean family, the Extreme value family, elliptical family, and other miscellaneous class. Copulas may be characterized as single, or vector parameter depending on comprehensiveness with which dependency structure can be described.

Some copula functions that are considered for this study are:

Clayton family:

Frank family:

Gumbel-Hougaard (GH) family:

Galambos family:

Ali-Mikhail-Haq (AMH) family:

FGM family:

**3.5 Parameter estimation:**

There are various methods of estimating the copula dependence structure. such as (1) moment-like method (MOM) based on the non-parametric dependence measures, i,e Kendall’s tau, spearman rho. (2) method of maximum pseudo-likelihood (MPL) and (3) method of exact maximum likelihood (EML).

**3.5.1 Moment-like method (MOM) based on the non-parametric dependence measures:**

|  |  |  |  |
| --- | --- | --- | --- |
| Copula | Generator | Kendall’s tau | Spearman rho |
| Clayton |  |  | No close form |
| Frank |  |  |  |
| Galambos | No close form | No close form | No close form |
| AMH |  |  |  |
| FGM | No close form |  |  |
| Gumbel |  |  | No close form |

Table 1: Formula of estimating parameters based on MOM method.

In the above table 1,

and the Debey function of order k defined by

For numerical computation, this series can be written as:

where is the Riemann zeta function (for***,*** note that .

Nelson (2006) or Genest and Fevre (2007) reported a form to obtain non-parametric estimates of θ which is derived from Kendall’s tau, τ, and Spearman’s rho, ρ. It can be obtained from

The above table shows some close from which are generated by these two formulae. Using this, it is possible to obtain the copula parameter of θ. For example, for clayton formula

For AMH copula τ requires the value between -0.1817 and 0.3333 and for FGM copula the admissible dependence space require between -2/9 to 2/9 (-0.222 to 0.222).

**3.5.2 Maximum pseudo-likelihood (MPL) method:**

In maximum pseudo likelihood method dependence structure assumed to be independent of the margin. It depicts non-parametrically by the respective scale ranks. In this process only dependence parameters are determined by maximizing the likelihood function. Parameter of marginal distribution cannot be obtained by maximizing the likelihood function, they can obtain based on ranks. The log-likelihood function is written as:

Where are non-parametric marginal probability based on ranks.

**3.5.3 Exact maximum likelihood method:**

In this methos all parameters can be determined by maximizing the log-likelihood function. The log-likelihood function is

Where, are the parameters vector of marginal distribution of F(x) and G(y) respectively. These parameters are estimated simultaneously. There is another an approach to estimating this parameter named “Inference From Margins” (IFM) method. In this method univariate marginal parameters are obtained first by maximizing the log-likelihood function of univariate marginals then the dependence parameter θ is obtained by maximizing the likelihood function. The log-likelihood function can be expressed as

Where indicates the margins having parameter that are obtained on a univariate basis. IFM method is barely used for estimating the parameters. It is useful in larger dimension where classical approach becomes more complex and computationally unwieldy.

**3.6 Goodness of fit test:**

After using several types of copulas, it is necessary to test the goodness of fit to obtain which copula fits well. This can be obtained in three ways. (a) graphical methods, (b) error statistics, and (c) formal goodness of fit tests.

**3.6.1 Graphical approach:**

There are plenty of graphical approaches to determine the goodness of fit test. One of them is employed by Nelson (2006). He plotted a scatter plot of observed data and fairly large sample data that are obtained by the copulas using their estimated parameter. This picture indicates the comparison between observed and fitted copula data. But it is noted that both very small and very large data can lead to misleading comparisons. Another approach is comparison the Cumulative Distribution Function (CDF) between empirical copula and computed copula. However, the important note is this is better suited for bivariate case only. The other two graphical approaches are related to K-plots and Q-Q plot.

**3.6.2 Error statistics:**

A quantitative technique to evaluate the performance of various copulas can be used, such as maximizing log-likelihood, the Akaike information criterion (AIC), or the Bayesian Information Criterion (BIC). Other error statistics include root mean squared error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), error (ERR), mean error (MERR), and mean normalized error (MNERR).

**3.6.3 Formal goodness of fit test:**

The availability of analytical goodness of fit tests for copulas and validity of parametric bootstrap procedures are proposed by Genest et al. (2009). Some Cramer-von Mises type of test statistics based on Rosenblatt’s transformation are recommended by wang and Wells (2000), Fermanian (2005), and Genest et al. (2006), among others. There are seven group of goodness of fit tests for copulas based on: (a) Empirical copula process; (b) Kendall’s process; (c) Rosenblatt integral transform; (d) Transformation for Archimedean copulas; (e) Kernel density; (f) White’s information matrix equality; and (g) Pseudo in-and-out-of-sample (PIOS) estimator.

**3.7 Returning period of flood:**

**3.7.1 For univariate frequency analysis:**

There are several methods for carrying out the returning period of flood using frequency analysis. For example, California formula, Hazen formula, Weibull formula, Chegodayev formula, Blom formula, Gringorten formula etc. Among them Weibull formula is frequently used for calculating returning period of flood. (Selaman et al., 2007). The Weibull formula is:

Where T denotes returning period of flood, R is the rank value of frequency and N is the number of total frequencies.

There is another way to calculate the returning period of flood using its marginal distribution. (Brunner et al., 2016). The return period (𝑥) of the event {𝑋≥𝑥} can be written as:

Where, is denoted as survival function and is the probability distribution function of random variable X at point x. is the mean inter-arrival time between two successive events.

**3.7.2 For bivariate frequency:**

**Joint return period:** There are four conditions that can happen.

Quadrant I: Pr [𝑋>𝑥, 𝑌>𝑦] =1−(𝑥)−𝐹𝑌(𝑦)+ 𝐹𝑋𝑌 (𝑥, 𝑦) = 𝑆𝑋𝑌(𝑥, 𝑦)

Quadrant II: Pr [𝑋≤𝑥, 𝑌>𝑦]

Quadrant III: Pr [𝑋≤𝑥, 𝑌≤𝑦] =𝐹𝑋𝑌 (𝑥, 𝑦)

Quadrant IV: Pr [𝑋>𝑥, 𝑌≤𝑦]



Figure 1: Quadrant of bivariate return period

We can deal with events in Quadrant I, where 𝑋 exceeds x and Y exceeds y, or we can work with events in Quadrants II and IV, where 𝑌 exceeds y or X exceeds x, when performing a flood frequency study. These potential combined occurrences using the OR and AND operators which is depicted in figure 1.

The joint OR and AND return periods of these two events using a copula can be calculated as follows:

and

where 𝑈 stands for 𝐹𝑋(𝑋), the peak discharge transformed via the probability integral transform, and 𝑉 stands for 𝐹𝑌(𝑌), is the mean inter-arrival time between two successive events and is the copula function. (Brunner et al., 2016).

**4.Result**

**4.1 Potential marginal distribution:**

Several candidate distributions are employed to test the potential marginal distribution. Kolmogorov-Smirnov (KS) test, Anderson darling test (AD) and Chi-squared statistic test have been used to find out the suitable fitted distribution of both flood peak. Considering the p value of those tests the lognormal distribution has been selected as a marginal distribution for flood volume and log-Pearson type III distribution has been found to be logical to select for flood peak for this data. For further information see supplementary file 1.

**4.2 Distribution and parameters estimation:**

Log Pearson type III distribution function:

where,

α=scale parameter

β= shape parameter

c= location parameter

Log Normal Distribution:

where,

μ = location parameter

σ= scale para

The parameter has been estimated by maximum likelihood method:

Parameter of Log Pearson type III and Log normal Distribution:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Log Pearson type III | | | Log Normal | |
| Shape | Scale | Location | Location | Scale |
| 8.765369 | -0.051623 | 7.976211 | 6.3357121 | 0.2180266 |

Table 2: Estimated parameters of Log-Pearson type III distribution and Log normal Distribution.

From table 2 we can see that the shape parameter of Log-Pearson type III distribution is 8.765 and the scale and location parameters are -0.0516, 7.9762 respectively. For Log normal distribution, the location parameters indicate the logarithmic value of mean that is 6.335 and the scale parameter indicates the logarithmic value of standard deviation that is 0.218.

**4.3 Dependence structure and copula test space:**

A graph with numbers and lines

Description automatically generatedFigure 2: Scatter plot of Flood Peak and flood Volume.

Figure 2 indicates a positive association between flood peak and flood volume which supports a positive relationship between those two variables.

For find out the dependence of flood peak and volume, Pearson correlation is first used. The value of correlation coefficient is 0.36, (table 3). It indicates that the linear relationship between flood peak and volume and its estimated coefficient value is 0.36. This depicts a relative lower linear relationship between them. In hydrology, flood data may show some nonlinear relationships between them that are not determined by the Pearson correlation coefficient. For checking nonlinear correlation, spearman rank correlation is used which shows the value of 0.43 that is pretty much higher than Pearson correlation coefficient. Further Kendall’s tau correlation coefficient has been calculated as 0.303 for this dataset.

|  |  |  |
| --- | --- | --- |
| Pearson correlation | Spearman Rank correlation | Kendall’s Tau |
| 0.3673136 | 0.4325807 | 0.303 |

Table 3: Value of several correlation coefficients.

A significant positive dependence is also indicated by both Chi- and K-plots in figure 7. Considering data exclusive from the lower left and upper right quadrants, as suggested by Abberger (2005).

A graph of a function

Description automatically generated with medium confidenceFigure 3: Chi-plot and K plot

A comparison of a graph

Description automatically generated with medium confidenceFigure 4: Chi plot characterization by lower tail and upper tail.

Seeing figure 4 we can conclude that some points near λ = 1 exhibit significance, indicating lower tail dependence. More importantly, it confirms upper tail independence, as points in the end zone are within control bounds (p-value = 0.95). Based on the sample Kendall's tau value of 0.30 and the characteristics of both upper and lower tail dependence, two Archimedean copulas, Clayton and Frank, and two extreme value copulas, Galambos and Gumbel, have been chosen. In fact, Gumbel is also an Archimedean copula. To further understand the issues that arise because of misspecification, two additional copulas, AMH and FGM, are shortlisted, with the limitation that the sample dependence exceeds a permitted range of these copulas. Although more copulas could have been considered in the initial screening stage, only these six are included in the copula test space, to keep the selection process shorter.

**4.4 Estimation of copula parameter:**

Among the many different estimation methods, pseudo maximum likelihood estimation and Kendall's tau are employed to estimate the dependence parameter of various copulas. Tau must have a value between -2/9 and 2/9 for FGM, and between -0.333 and 0.333 for AMH. FGM is not appropriate for estimating the parameter using this approach, as our tau is 0.303. Although AMH is used, this tau value does not find out the value of parameter for the limitation of convergence of parameter.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Estimation using MPL | | | | | | |
|  | Clayton | Frank | Galambos | Gumbel | AMH | FGM |
| Estimate | 1.054 | 3.128 | 0.494 | 1.278 | 0.999 | 1.000 |
| Standard Error | 0.327 | 0.859 | 0.079 | 0.083 | 0.005 | 0.277 |
| Likelihood | 10.610 | 5.735 | 1.813 | 2.335 | 10.591 | 4.637 |
| LCI | 0.412 | 1.444 | 0.338 | 1.115 | 0.989 | 0.457 |
| UCI | 1.696 | 4.811 | 0.649 | 1.440 | 1.010 | 1.543 |

Table 4a: Estimated parameters of several copulas’ family using MPL method

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Estimation using Tau | | | | | |
|  | Clayton | Frank | Galambos | Gumbel | AMH |
| Estimate | 0.867 | 2.946 | 0.705 | 1.434 | 0.948 |
| Standard Error | 0.381 | 0.201 | 0.200 | 0.190 | - |
| LCI | 0.121 | 2.552 | 0.312 | 1.061 | - |
| UCI | 1.614 | 3.341 | 1.098 | 1.807 | - |

Table 4b: Estimated parameters of several copulas’ family using the tau method.

The parameter theta for clayton has a value of 1.054 and a maximum likelihood value of 10.610, as seen in the above table 4a. It is 3.128 for Frank. For this family, 5.735 is the greatest likelihood value. The AMH copula family has the second-largest likelihood value, 10.591. The AMH's calculated parameter is 0.999. From table 4b we can interpret that the calculated parameter for Clayton using Kendall's tau parameter is 0.867 with a standard error of 0.381. It is 2.946 for Frank, with a standard error of 0.201. Gumbel and Galambos exhibit a lower standard error of 0.19 and 0.20, respectively, than both Clayton and Frank. The parameter value is 1.43 for Gumbel and 0.705 for Galambos.

**4.5 Checking Goodness of Fit:**

The assessment of copula has been carried out using three methods. (a) graphical approach. (b) using error statistic (c) using formal goodness of fit test.

**4.5.1 Graphical approach:**

The observed data is compared to a wide set of randomly produced samples. For this investigation, a set of 500 random samples is created for each of the six copula families under consideration, using MPL method-based parameters and the methods stated by Nelsen (2006). Figure 4.6 and figure 4.7 depict a comparison of observed data with randomly generated samples. These charts show that the general distribution of observed data is like that of random samples. The first graphic (Figure 5) is based on the uniformly distributed values of the observed data and the associated copula. The blue point indicates the observed value, while the black point represents the calculated value. It demonstrates that all members of the copula family effectively capture data.

We next plotted the real value of the observed data and calculated copula sample values in Figure 6. This graph illustrates that the generated sample of Gumbel and Galambos has upper tail dependence, whereas our observed data does not. The generated sample of FGM and Frank does not show lowered dependence, whereas Clayton and AMH represent an optimal scenario with real data. It is important to note that the observed data shows very high flows with moderate volumes, which are not reproduced by the simulation set. The other four copulas' simulated sets are adequate, although the AMH and Clayton copulas appear to do better in the lower tail.

A diagram of a number of dots

Description automatically generated with medium confidence  
Figure 5: Comparison between uniformly distributed observed data and their corresponding calculated copulas sample.

A group of blue dots

Description automatically generated  
Figure 6: Comparison between real observed data and their corresponding calculated copulas sample.

Secondly comparison of CDF between observed and generated sample is shown in figure 7.

It shows that, the matching by the Clayton and Gumbel is better than the other copulas, with the differences being minimal for the clayton copula. The Clayton copula had the best graphical match out of the six copulas tested using four different approaches. Formal test results can provide additional evidence of relative superiority.

A graph of a group of people

Description automatically generated with medium confidence

Figure 7: Comparison of CDF between observed and generated sample.

**4.5.2 Using error statistic:**

| Estimation using MPL | | | | | | | Estimation using Tau | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Name | AMH | Clayton | Frank | Galambos | Gumbel | FGM | Clayton | Frank | Galambos | Gumbel |
| RMSE | 0.38 | 0.45 | 0.40 | 0.41 | 0.39 | 0.42 | 0.41 | 0.40 | 0.43 | 0.40 |
| MAE | 0.31 | 0.39 | 0.33 | 0.35 | 0.31 | 0.35 | 0.34 | 0.33 | 0.35 | 0.33 |
| MAPE | 483.9 | 536.8 | 564.5 | 405.3 | 667.1 | 761.7 | 741.3 | 809.7 | 566.5 | 234.9 |
| ERR | 0.62 | 0.77 | 0.66 | 0.70 | 0.61 | 0.71 | 0.67 | 0.66 | 0.71 | 0.67 |
| MERR | 0.00 | 0.00 | -0.01 | -0.03 | -0.02 | 0.01 | -0.03 | 0.09 | 0.06 | -0.07 |
| MNERR | 0.62 | 0.77 | 0.66 | 0.70 | 0.61 | 0.71 | 0.67 | 0.66 | 0.71 | 0.67 |

Table 5: Error statistics of MPL and tau parameters.

The above table 5 shows the various error statistics for two estimation methods. In the MPL approach, the root mean square error for the AMH copula is 0.38, which is the lowest RMSE of any copula. Clayton, Frank, Galambos, Gumbel, and GFM had scores of 0.45, 0.40, 0.41, 0.39, and 0.42, respectively. This is fairly like all copulas. AMH, Clayton, Frank, Galambos, Gumbel, and Frank had mean absolute errors of 0.31, 0.39, 0.33, 0.35, 0.31, and 0.35, respectively. The mean normalized error for AMH, Clayton, Frank, Galambos, Gumbel, and Frank is 0.62, 0.77, 0.66, 0.70, 0.61, and 0.71, respectively. It is difficult to conclude from the error statistics which copula is optimal for our dataset. We can find a similar outcome using the tau estimation method. Clayton, Frank, Galambos, and Gumbel demonstrate identical error statistics in different error statistics calculations. Clayton, Frank, Galambos, and Gumbel had an RMSE of 0.41, 0.40, 0.43, and 0.40, respectively. As usual, Clayton, Frank, Galambos, and Gumbel have normalized mean square errors of 0.67, 0.66, 0.71, and 0.67, respectively. As a result, in order to reach our final evaluation, we have to perform a formal goodness of fit test.

**4.5.3 Formal goodness of fit test:**

To conduct formal goodness-of-fit testing, the Cramer-von Mises type statistics (CMn, Sn, and T n) for four copulas (Clayton, Frank, Galambos, and Gumbel) are evaluated using Rosenblatt transform. The parametric bootstrap process simulates random samples of sizes 100, 1000, 10,000, and 100,000. We compute the four statistics' values, p-values, and critical values at a 5% significance level using different ties method. Tables 4.8, 4.9, 4.10, 4.11 present the results for MPL parameter estimation methods used in this investigation. To reduce computational complexity, just 100 and 1000 random samples of Galambos copula were simulated using the parametric bootstrap procedure. Here The multiplier approach (Rn) is used for all copulas of simulated random samples of sizes 100, 1000, 10,000, and 100,000. SnB and SnC do not perform for it due to their limitations. Note that, In table Tables 6a, 6b, 6c, 6d, Sn indicates Cramer von mise type statistic that focus on overall fit across the entire distribution, SnB indicates another variation of Cramer-von\_mises type statistics that focus on L2 norm and kernel-based assessment, SnC indicates Kolmogorv-Samirnov statistics that focus on the Maximum deviation at any point in the distribution, and Rn is based on ranks and is designed to capture discrepancies in the tails of the distribution more effectively. It involves comparing the ranks of the empirical and theoretical copulas.

|  | 100 | | | 1000 | | 10000 | | 1000000 | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | parameter | statistic | P value | statistic | P value | statistic | P value | statistic | P value |
| Clayton | 1.0234 | 0.0194 | 0.7178 | 0.0194 | 0.6948 | 0.0194 | 0.6961 | 0.0194 | 0.6934 |
| 0.0215 | 0.5099 | 0.0215 | 0.5360 | 0.0215 | 0.5519 | 0.0215 | 0.5514 |
| 0.0230 | 0.5000 | 0.0230 | 0.6139 | 0.0230 | 0.5972 | 0.0230 | 0.5993 |
| 0.0206 | 0.599 | 0.0198 | 0.6369 | 0.0198 | 0.6498 | 0.0207 | 0.6097 |
| Frank | 3.0946 | 0.0321 | 0.2129 | 0.0321 | 0.2163 | 0.0321 | 0.2017 | 0.0321 | 0.2034 |
| 0.0321 | 0.2624 | 0.0321 | 0.2173 | 0.0321 | 0.2072 | 0.0321 | 0.2013 |
| 0.0321 | 0.1733 | 0.0321 | 0.2043 | 0.0321 | 0.2033 | 0.0321 | 0.2045 |
| 0.0321 | 0.203 | 0.0321 | 0.1763 | 0.0321 | 0.2011 | 0.0321 | 0.2015 |
| Gumbel | 1.3352 | 0.0680 | 0.0050 | 0.0680 | 0.0005 | 0.0680 | 0.0010 | 0.0680 | 0.0011 |
| 0.0680 | 0.0050 | 0.0680 | 0.0015 | 0.0680 | 0.0015 | 0.0680 | 0.0012 |
| 0.0680 | 0.0149 | 0.0680 | 0.0005 | 0.0680 | 0.0009 | 0.0680 | 0.0010 |
| 0.0680 | 0.005 | 0.0680 | 0.0025 | 0.0680 | 0.0013 | 0.0680 | 0.0012 |
| Galambos | 0.57184 | 0.0738 | 0.0049 | 0.0738 | 0.0044 | - | - | - | - |
| 0.0754 | 0.0049 | 0.0754 | 0.0034 | - | - | - | - |
| 0.0772 | 0.0049 | 0.0772 | 0.0034 | - | - | - | - |
| 0.0752 | 0.0049 | 0.0764 | 0.0014 | - | - | - | - |

Table 6a: Formal goodness of fit test result of 100,1000,10000,100000 generated sample based on Sn method.

The above table 6a demonstrates that Gumbel and Galambos copula have significant p-values less than 0.05. That is, in the Sn approach, Gumbel and Galambos copula do not fit well with our data set. Rather than these two, Clayton and Frank are insignificant , that’s mean, for this data those two-copula family capture the data well.

|  | 100 | | | 1000 | | 10000 | | 1000000 | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | parameter | statistic | P value | statistic | P value | statistic | P value | statistic | P value |
| Clayton | 1.0234 | 0.0396 | 0.2822 | 0.0396 | 0.2952 | 0.0396 | NA | 0.0396 | NA |
| 0.0405 | 0.2525 | 0.0405 | 0.2642 | 0.0405 | 0.2801 | 0.0405 | NA |
| 0.0407 | 0.3614 | 0.0407 | 0.2782 | 0.0407 | NA | 0.0407 | NA |
| 0.0399 | 0.2624 | 0.0403 | 0.2662 | 0.0398 | NA | 0.0398 | NA |
| Frank | 3.0946 | 0.0349 | 0.4406 | 0.0349 | 0.3711 | 0.0349 | 0.3654 | 0.0349 | 0.3651 |
| 0.0349 | 0.4010 | 0.0349 | 0.3731 | 0.0349 | 0.3621 | 0.0349 | 0.3618 |
| 0.0349 | 0.3416 | 0.0349 | 0.3372 | 0.0349 | 0.3607 | 0.0349 | 0.3600 |
| 0.0349 | 0.4802 | 0.0349 | 0.3422 | 0.0349 | 0.3657 | 0.0349 | 0.3634 |
| Gumbel | 1.3352 | 0.0488 | 0.0248 | 0.0488 | 0.0824 | 0.0488 | 0.0872 | 0.0488 | 0.0895 |
| 0.0488 | 0.0446 | 0.0488 | 0.0884 | 0.0488 | 0.0881 | 0.0488 | 0.0895 |
| 0.0488 | 0.0347 | 0.0488 | 0.0924 | 0.0488 | 0.0860 | 0.0488 | 0.0878 |
| 0.0488 | 0.0842 | 0.0488 | 0.0914 | 0.0488 | 0.0884 | 0.0488 | 0.0879 |

Table 6b: Formal goodness of fit test result of 100,1000,10000,100000 generated sample based on SnB method.

In this table 6b after simulating 1000000 times Gumbel also shows insignificant with , Clayton copula has been failed to calculate their corresponding p value for its convergence Issues. Frank demonstrates that it has fitted , well in this data.

|  | 100 | | | 1000 | | 10000 | | 1000000 | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | parameter | statistic | P value | statistic | P value | statistic | P value | statistic | P value |
| Clayton | 1.0234 | 0.0652 | 0.1139 | 0.0652 | 0.1823 | 0.0652 | NA | 0.0652 | NA |
| 0.0744 | 0.0941 | 0.0744 | 0.0924 | 0.0744 | 0.0863 | 0.0744 | NA |
| 0.0754 | 0.1139 | 0.0754 | 0.1144 | 0.0754 | 0.1181 | 0.0754 | NA |
| 0.0678 | 0.1535 | 0.0675 | 0.1374 | 0.0651 | NA | 0.0684 | NA |
| Frank | 3.0946 | 0.0760 | 0.0743 | 0.0760 | 0.0534 | 0.0760 | 0.0497 | 0.0760 | 0.0517 |
| 0.0760 | 0.0743 | 0.0760 | 0.0455 | 0.0760 | 0.0497 | 0.0760 | 0.0513 |
| 0.0760 | 0.0149 | 0.0760 | 0.0385 | 0.0760 | 0.0527 | 0.0760 | 0.0511 |
| 0.0760 | 0.0743 | 0.0760 | 0.0485 | 0.0760 | 0.0509 | 0.0760 | 0.0516 |
| Gumbel | 1.3352 | 0.0959 | 0.0050 | 0.0959 | 0.0045 | 0.0959 | 0.0093 | 0.0959 | 0.0086 |
| 0.0959 | 0.0050 | 0.0959 | 0.0085 | 0.0959 | 0.0067 | 0.0959 | 0.0089 |
| 0.0959 | 0.0149 | 0.0959 | 0.0085 | 0.0959 | 0.0096 | 0.0959 | 0.0078 |
| 0.0959 | 0.0050 | 0.0959 | 0.0075 | 0.0959 | 0.0087 | 0.0959 | 0.0091 |

Table 6c: Formal goodness of fit test result of 100,1000,10000,100000 generated sample based on SnC method.

Like the SnB approach mentioned above, Clayton's convergence problem prevents it from calculating the p value even after 100,000 simulations. Galambos is significant (p ≤0.05), indicating that this family is not appropriate for this data according to the SnC method, which is depicted in table 6c, and Frank is insignificant (p ≥0.05), indicating that it has been fitted for this data set.

|  | 100 | | | 1000 | | 10000 | | 1000000 | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | parameter | statistic | P value | statistic | P value | statistic | P value | statistic | P value |
| Clayton | 1.0234 | 0.1153 | 0.6584 | 0.1153 | 0.7048 | 0.1153 | 0.6903 | 0.1153 | 0.6892 |
| 0.1242 | 0.6386 | 0.1242 | 0.6349 | 0.1242 | 0.6293 | 0.1242 | 0.6310 |
| 0.1312 | 0.5495 | 0.1312 | 0.5619 | 0.1312 | 0.5861 | 0.1312 | 0.5895 |
| 0.1178 | 0.6683 | 0.1199 | 0.6259 | 0.1145 | 0.6973 | 0.1221 | 0.6457 |
| Frank | 3.0946 | 0.6517 | 0.2327 | 0.6517 | 0.2023 | 0.6517 | 0.1876 | 0.6517 | 0.1894 |
| 0.6517 | 0.2228 | 0.6517 | 0.1733 | 0.6517 | 0.1878 | 0.6517 | 0.1920 |
| 0.6517 | 0.2525 | 0.6517 | 0.1793 | 0.6517 | 0.1907 | 0.6517 | 0.1923 |
| 0.6517 | 0.1931 | 0.6517 | 0.1813 | 0.6517 | 0.1900 | 0.6517 | 0.1907 |
| Gumbel | 1.3352 | 0.9461 | 0.0941 | 0.9461 | 0.0834 | 0.9461 | 0.0820 | 0.9461 | 0.0825 |
| 0.9461 | 0.1040 | 0.9461 | 0.0644 | 0.9461 | 0.0863 | 0.9461 | 0.0822 |
| 0.9461 | 0.1040 | 0.9461 | 0.0934 | 0.9461 | 0.0839 | 0.9461 | 0.0821 |
| 0.9461 | 0.0347 | 0.9461 | 0.0834 | 0.9461 | 0.0849 | 0.9461 | 0.0826 |
| Galambos | 0.57184 | 1.0523 | 0.0643 | 1.0523 | 0.0944 | 1.0523 | 0.0781 | 1.0523 | 0.0771 |
| 1.059 | 0.104 | 1.059 | 0.0734 | 1.059 | 0.0720 | 1.059 | 0.0745 |
| 1.0669 | 0.0940 | 1.0669 | 0.0724 | 1.0669 | 0.0723 | 1.0669 | 0.0734 |
| 1.0463 | 0.104 | 1.0596 | 0.0664 | 1.0479 | 0.0724 | 1.0471 | 0.0770 |

Table 6d: Formal goodness of fit test result of 100,1000,10000,100000 generated sample based on Rn method.

Using the multiplier method we have obtained the result that is depicted in table 6d, every member of the copula family exhibits insignificance (p ≥0.05), with Clayton having the highest P value. Even though every model for this dataset had an optimal fit, we can verify this by displaying the figure 8 that follows.

A chart of a number of different types of dna

Description automatically generated with medium confidence

Figure 8: P values of bootstrapping sample using in formal goodness of fit test.

We only present the p-values of the three copulas (Clayton, Frank, and Gumbel) because the majority of the tests show that the data are inappropriate for the Galambos family. The greatest p value in the preceding figure indicates a good fit between the data and the Clayton family of copula, which fits the data set well.

**4.6 Return period comparison between univariate frequency analysis and bivariate frequency analysis:**

For univariate frequency analysis of returning period of flood has been calculated using flood peak. Weibull formula and its corresponding marginal distribution (log-Pearson type III) has been employed to assess the returning period and their corresponding percentage value. The mean inter-arrival time between two successive events is 1. In bivariate frequency analysis, for calculating returning period of flood, OR and AND conditions have been used. The following table shows 53 years analysis of our study. For full information see supplementary file 2.

|  |  | | Univariate frequency analysis | | | | Bivariate frequency analysis | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Year | Flood  Peak | Flood  Volume | Weibull method | Percentage | Marginal distribution | Percentage | OR method | Percentage | AND method | Percentage |
| 1,964 | 2,120.00 | 856.94 | 5.40 | 18.52 | 5.27 | 18.99 | 4.81 | 20.80 | 104.05 | 0.96 |
| 1,974 | 2,120.00 | 617.15 | 4.50 | 22.22 | 5.27 | 18.99 | 2.34 | 42.76 | 9.70 | 10.31 |
| 1,976 | 2,480.00 | 562.97 | 54.00 | 1.85 | 162.06 | 0.62 | 1.98 | 50.60 | 212.62 | 0.47 |
| 1,987 | 1,990.00 | 520.38 | 3.18 | 31.48 | 2.86 | 34.99 | 1.43 | 70.00 | 3.39 | 29.51 |
| 2010 | 1,869.30 | 935.91 | 2.00 | 50.00 | 1.92 | 51.99 | 1.92 | 52.21 | 126.60 | 0.79 |
| 2,022 | 2,235.11 | 611.55 | 9.00 | 11.11 | 11.20 | 8.93 | 2.54 | 39.38 | 19.29 | 5.18 |

Table 7: Comparison of univariate frequency analysis and bivariate frequency analysis.

The OR method calculates the return period based on the probability that at least one of the two events exceeds a certain threshold. This method typically results in more conservative (shorter) return periods. The AND method calculates the return period based on the probability that both variables simultaneously exceed their respective thresholds. This often leads to longer return periods as it considers the joint exceedance.

According to table 7 in 1964 the return period of flood using univariate Weibull formula is 5.40 and using marginal distribution is 5.27. It means that the flood peak is expected to exceed 2,120.00 cm3 once every 5.40 or 5.27 years on average. That is there is approximately an 18.52% or 18.99% chance that the flood peak will exceed the specified threshold in any given year. In bivariate frequency analysis, return period​ is 4.8 years, this means that either the peak discharge exceeds 2,120.00 cm3 or the volume exceeds 856.94 cm3 once every 4.8 years on average. Return period using AND method is 104.05 years, which means that both peak discharge and volume exceed 2,120.00 cm3 and 856.94 cm3 once every 104.05 years on average. The chance being exceeds peak discharge or the volume is an approximately 59.95% in any given year and the chance being exceed both is 0.96%.

In 1974, the return period for the Weibull technique was 4.50, which implies that the average flood peak value exceeded 2,120.00 cm3 once every 4.5 years, however the bivariate method reveals that the flood peak exceeded 2120 cm3 or the flood volume exceeded 617.15 cm3 once every 2.34 years. According to univariate frequency analysis, there is a 22.22% possibility of exceeding the flood peak threshold in any given year, however bivariate frequency analysis shows a 42.67% risk as well as 10.31% risk of being exceed the threshold of both flood peak and volume.

Similarly, according to 1987, using the univariate marginal distribution technique, there is an approximately 34.99% chance of exceeding the flood peak of 1990.00 cm3 in any given year, compared to approximately 31.48% using the Weibull method. According to bivariate frequency analysis, the chances of either flood peak exceeding 1990.00 cm3 or flood volume exceeding 520.38 cm3 are approximately 70%, and the chances of both exceeding the threshold value are approximately 29.51% in any given year.

In the same way, according to 2022, the likelihood of exceeding the flood peak of 2,235.11 cm3 in any given year is about 8.93% when using the univariate marginal distribution technique, compared to approximately 11.11% when using the Weibull method. Return period of flood is 8.93 and 9 respectively. But bivariate frequency analysis indicates that in any given year, there is a 39.38% chance of either the flood peak exceeding 2,235.11 cm3 or the flood volume exceeding 611.55 cm3, and a 5.18% chance of both exceeding the threshold value. The average return period of flood for either rises above one is 2.54, meaning that once every 2.54 years, either the peak discharge exceeds 2,235.11 cm3 or the volume exceeds 611.55 cm3.

**5. Discussion:**

This study highlights the usefulness of applying copula function to investigate the returning period of flood when hydrological data are correlated to each other. We conducted a bivariate frequency analysis of flood events to study the relationship between flood peak and flood volume, as univariate frequency analysis is insufficient to capture all information due to the correlated behavior of hydrological events (Goel, 1998 and Yue S., 1999). The main objective was to obtain a thorough understanding of the combined dependence structure of these two important hydrological variables, as well as to estimate the return periods of catastrophic flood events using advanced statistical approaches.

To begin with, we fitted the marginal distributions of the individual variables. The flood peak data was found to follow a Log-Pearson Type III distribution, a widely accepted distribution in hydrology due to its flexibility in modeling skewed flood data. This choice was supported by goodness-of-fit tests and visual inspections of density plots. Similarly, the flood volume data was best described by a Log-Normal distribution. The Log-Normal distribution was appropriate given the nature of the data, which often demonstrates a right-skewed distribution with many low values and few very high values. Noted that the U. S water Resources Council (WRC) recommended a uniform technique for estimating flood flow frequencies, using log-Pearson type III distribution (Thomas Jr, W. O. 1985). Zhang, T et al. (2018) found lognormal distribution performed best and a tend to describe well the nonstationary behavior of both flood peak and volume. Hemant Chowdhary et al. (2011) tested several candidate distributions and finally chosen Pearson type III and three-parameter Weibull distribution for their flood peak and flood volume.

We used copula functions to capture the dependence structure between flood peak and flood volume due to the limited form of the joint distribution function ever made. When variables have unique distributions and are mutually correlated, only copula functions provide the opportunity to determine their dependence and construct their joint distribution function. Among the various copulas examined, the Clayton copula has been found the best fit to our data, with a parameter of 1.05. The Clayton copula is particularly suitable for modeling the lower tail dependence, which is crucial in flood analysis as it captures the tendency of extreme low values in one variable to be associated with extreme low values in the other. This dependence structure is vital for accurate risk assessment and return period estimation in flood frequency analysis. George Papaioannou et al. (2016) found all the Archimedean class copula families do better among the suitable and applicable copula families. Hemant Chowdhary et al. (2011) and Salleh et al. (2016) found the clayton copula as the best to their study.

For the univariate joint distribution, we employed the Weibull method to the flood peak data. The Weibull plotting position formula is a non-parametric method commonly used in hydrological studies to estimate the return periods of flood events. This method provided a reliable basis for comparing and validating the results obtained from the copula-based approach (Selaman et al., 2007)

To estimate the return period of flood events, we used the bivariate OR and AND methods. The OR approach considers the joint probability of exceedance of either flood peak or flood volume, providing a more comprehensive risk assessment compared to univariate methods and AND approach considers the joint probability of exceedance of both flood peak and volume. Thus, it considers an extreme case of flood situation. Since OR method used closed values more effectively in producing return periods, it provided more consistent return periods than the univariate approach. That is, while the univariate Weibull method analysis yielded varying return durations, the OR approach made more consistent and trustworthy flood frequency estimations. Our study used the AND technique, which requires multiple variables to achieve high thresholds simultaneously, to identify unusual and extreme flood episodes. As a result of the statistically lower likelihood of such events, this technique frequently yields longer return durations such as 100 years or more. In our study this situation has been observed in years 1964, 1976 and 2010. The rarity of these extreme conditions is reflected in these extended return durations, highlighting the significance of taking combined occurrences into account when assessing flood risk more accurately. Our findings coincide with the earlier studies, because the bivariate return period offers a more realistic scenario for flood risk management by considering the simultaneous occurrence of extreme events in both variables (Brunner et al., 2016).

The findings from our study have significant implications for flood risk management and infrastructure design in the catchment area. By understanding the joint behavior of flood peak and flood volume, we can better predict the likelihood of extreme flood events and design more resilient hydraulic structures. The use of copulas, particularly the Clayton copula in this case, enhances the accuracy of joint probability estimates, which is critical for making more informative decisions in flood-prone areas.

Moreover, the methodology employed in this research can be extended to other hydrological studies involving multiple correlated variables. The approach of combining marginal distributions with copulas and employing advanced methods for return period calculation represents a robust framework for bivariate frequency analysis.

**6. Conclusion:**

The copula approach is a powerful method that enables us to apply the insights gained from single-variable hydrological frequency analysis to multivariate research. This approach does not need to impose specific distributional forms or data transformation procedures. We have also employed a range of graphical and formal tests to evaluate the goodness-of-fit and determine the most optimal copula for the data that have been analyzed. When these tests were conducted, the Clayton copula was identified as the favored choice.

In our study, after employing several graphical and formal tests to examine the goodness-of-fit to identify suitable copula structures for the specific set of maximum discharge and average discharge data, the Clayton copula was found to be the most suitable among the considered other copula family.

Furthermore, our analytical tests strongly rejected the validity of the copula family of extreme value such as Gumbel and Galambos for the targeted dataset. We found that there is no reliance on the upper tail in 53 years of maximum discharge and average discharge data. Our findings suggest that mis-specifying the copula of an extreme value as a suitable copula for the dataset could lead to significant overestimation in hydrologic designs related to maximum discharge and average discharge data.

Moreover, it is to be noted that our interpretation contradicts the findings of previous studies conducting by De Michele et al. (2005), Zhang & Singh (2006), Genest & Favre (2007), Poulin et al. (2007), and Karmakar & Simonovic (2009), which indicated that the Gumbel copula of extreme value family was superior to other copulas; although they indicated that the Clayton family is relatively well fitted. This inconsistency could be attributed to the unique characteristics of the specific data set under examination. We recommend further examination of additional cases to explore variations in inference more thoroughly.

In summary, our study has demonstrated that the application of bivariate frequency analysis enhances the assessment of flood risk compared to univariate flood frequency analysis, and the use of the copula of Clayton family is an optimal model for the considered maximum discharge and average discharge data that provides valuable insights for hydrological events.

**Acknowledgements:** not applicable.

**Author contributions:**

**Nazmul Haque:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Data curation, Conceptualization.

**Khalidur Rahman**: Supervision, Writing – review & editing, Visualization, Resources, Validation, Methodology, Investigation, Conceptualization.

**Funding sources:** We would like to thank the University Research Center, Shahjalal University of Science and Technology, Sylhet, for funding this research under the project PS/2024/1/37.

**References:**

Anderson CW, D. I. (1994). Maximum reservoir water level. In Reservoir Safety and Environment, pp. 200–213.

Abberger, K. (2005). A simple graphical method to explore tail-dependence in stock-return pairs. *Applied Financial Economics*, *15*(1), 43-51.

Ariff NM, Jemain AA, Ibrahim K, Zin WZW. 2012. IDF relationships using bivariate copula for storm events in Peninsular Malaysia. Journal of Hydrology, 470: 158-171. DOI: 10.1016/j.jhydrol.2012.08.045.

Balistrocchi M, Bacchi B. 2011. Modelling the statistical dependence of rainfall event variables through copula functions. Hydrology and Earth System Sciences, 15: 1959-1977. DOI: 10.5194/hess-15-1959-2011.

Brunner, M. I., Seibert, J., & Favre, A. C. (2016). Bivariate return periods and their importance for flood peak and volume estimation. Wiley Interdisciplinary Reviews: Water, 3(6), 819-833.

Chowdhary, H., Escobar, L. A., & Singh, V. P. (2011). Identification of suitable copulas for bivariate frequency analysis of flood peak and flood volume data. *Hydrology Research*, *42*(2-3), 193-216.

BWDB. https://bwdb.portal.gov.bd/

Chen L, Singh VP, Guo SL, Hao ZC, Li TY. 2012. Flood Coincidence Risk Analysis Using Multivariate Copula Functions. Journal of Hydrologic Engineering, 17: 742-755. DOI: 10.1061/(asce)he.1943-5584.0000504.

De Michele C, Salvadori G. 2003. A Generalized Pareto intensity-duration model of storm rainfall exploiting 2-Copulas. Journal of Geophysical Research-Atmospheres, 108. DOI: 10.1029/2002jd002534.

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: 50-57. DOI: 10.1061/(asce)1084-0699(2005)10:1(50).

Favre, A. -C. (2004). Multivariate Hydrological Frequency Analysis Using Copulas. Water Resources Research 40:WOllOl. doi:10.1029/2003WR002456.

Fisher, Nicholas & Switzer, Paul. (1985). Chi-Plots for Assessing Dependence. Biometrika. 72. 253-265. 10.1093/biomet/72.2.253.

Fermanian, J.-D. 2005 Goodness-of-fit tests for copulas. J. Multivariate Anal. 95, 119–152.

Genest, Christian & J-C, Boies. (2003). Detecting Dependence With Kendall Plots. The American Statistician. 57. 275-284. 10.1198/0003130032431.

Goel, N. K., Seth, S. M., and Chandra, S. 1998. “Multivariate modeling of flood flows.” J. Hydraul. Eng., 1242, 146–155.

Gupta VK, D. L. (1976). On the joint distribution of the largest flood and its time of occurrence. Water Resour. Res., 12(2): 295–304.

Gargouri-Ellouze E, Chebchoub A. 2008. Modelling the dependence structure of rainfall depth and duration by Gumbel's copula. Hydrological Sciences Journal-Journal Des Sciences Hydrologiques, 53: 802-817. DOI: 10.1623/hysj.53.4.802.

Ghosh S. 2010. Modelling bivariate rainfall distribution and generating bivariate correlated rainfall data in neighbouring meteorological subdivisions using copula. Hydrological Processes, 24: 3558 3567. DOI: 10.1002/hyp.7785.

Genest, C., Quessy, J.-F., & Rémillard, B. (2006). Goodness-of-Fit Procedures for Copula Models Based on the Probability Integral Transformation. Scandinavian Journal of Statistics, 33(2), 337–366. http://www.jstor.org/stable/4616928

Genest, C. & Favre, A.- C. 2007 Everything you always wanted to know about copula modeling but were afraid to ask. J. Hydrol. Engng 12(4), 347–368.

Genest, C., Rémillard, B., Beaudoin, D., 2009. Goodness-of-fit tests for copulas: A review and a power study. Insurance Math. Econom. 44 (2), 199–213

Grimaldi S, Serinaldi F. 2006a. Asymmetric copula in multivariate flood frequency analysis. Advances in Water Resources, 29: 1155-1167. DOI: 10.1016/j.advwatres.2005.09.005.

Grimaldi, S., & Serinaldi, F. (2006b). Design hyetograph analysis with 3-copula function. *Hydrological Sciences Journal*, *51*(2), 223-238.

Kao SC, Govindaraju RS. 2010. A copula-based joint deficit index for droughts. Journal of Hydrology, 380: 121-134. DOI: 10.1016/j.jhydrol.2009.10.029.

Karmakar, S., & Simonovic, S. P. (2009). Bivariate flood frequency analysis. Part 2: A copula‐based approach with mixed marginal distributions. Journal of Flood Risk Management, 2(1), 32-44.

Lee T, Modarres R, Ouarda T. 2013. Data-based analysis of bivariate copula tail dependence for drought duration and severity. Hydrological Processes, 27: 1454-1463. DOI: 10.1002/hyp.9233.

Liu CL, Zhang Q, Singh VP, Cui Y. 2011. Copula-based evaluations of drought variations in Guangdong, South China. Natural Hazards, 59: 1533-1546. DOI: 10.1007/s11069-011-9850-4.

Ma MW, Song SB, Ren LL, Jiang SH, Song JL. 2013. Multivariate drought characteristics using trivariate Gaussian and Student t copulas. Hydrological Processes, 27: 1175-1190. DOI: 10.1002/hyp.8432.

Nelsen, R. B. 2006 An Introduction to Copulas. Springer, Berlin.

Nadarajah S, W. C. (1993). Modeling dependency between climate extremes for New Zealand. World Resour. Rev., 13(4): 526–539.

Reddy MJ, Ganguli P. 2012a. Application of copulas for derivation of drought severity-duration frequency curves. Hydrological Processes, 26: 1672-1685. DOI: 10.1002/hyp.8287.

Renard B, Lang M. 2007. Use of a Gaussian copula for multivariate extreme value analysis: Some case studies in hydrology. Advances in Water Resources, 30: 897-912. DOI: 10.1016/j.advwatres.2006.08.001.

Selaman, O. S., Said, S., & Putuhena, F. J. (2007). Flood frequency analysis for Sarawak using Weibull, Gringorten and L-moments formula. J. Inst. Eng, 68, 43-52.

Singh VP, Zhang L. 2007. IDF curves using the Frank Archimedean copula. Journal of Hydrologic Engineering, 12: 651-662. DOI: 10.1061/(asce)1084-0699(2007)12:6(651).

Sklar, A. (1959) Functions de Repartition an Dimension Set Leursmarges. Publications de L’In-stitut de Statistique de L’Universite de Paris.

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

Thomas Jr, W. O. (1985). A uniform technique for flood frequency analysis. *Journal of Water Resources Planning and Management*, *111*(3), 321-337.

Tasnim, Z., & Raju, K. M. A. H. (2022). HYDROLOGICAL SIMULATION OF UPPER REACH OF SURMA RIVER BASIN AT SYLHET STATION USING SWAT.

Vandenberghe S, Verhoest NEC, De Baets B. 2010. Fitting bivariate copulas to the dependence structure between storm characteristics: A detailed analysis based on 105 year 10 min rainfall. Water Resources Research, 46. DOI: 10.1029/2009wr007857.

Wong G, Lambert MF, Leonard M, Metcalfe AV. 2010. Drought Analysis Using Trivariate Copulas Conditional on Climatic States. Journal of Hydrologic Engineering, 15: 129-141. DOI: 10.1061/(asce)he.1943-5584.0000169.

Wong G, van Lanen HAJ, Torfs P. 2013. Probabilistic analysis of hydrological drought characteristics using meteorological drought. Hydrological Sciences Journal-Journal Des Sciences Hydrologiques, 58: 253-270. DOI: 10.1080/02626667.2012.753147.

Wang, W. & Wells, M. T. 2000 Model selection and semiparametric inference for bivariate failure-time data. J. Am. Stat. Assoc. 95(449), 62–72.

Yue, S., Ouarda, T. B. M. J., Bobée, B., Legendre, P., and Bruneau, P. 1999. “The gumbel mixed model for flood frequency analysis.” J. Hydrol., 226, 88–100.

Yue, S. (2001). A statistical measure of severity of El Nin˜o events. Stochast. Envir. Res. and Risk Assess., 15(2): 153–172.

Zhang L, Singh VP. 2007a. Bivariate rainfall frequency distributions using Archimedean copulas. Journal of Hydrology, 332: 93-109. DOI: 10.1016/j.jhydrol.2006.06.033.

Zhang L, Singh VP. 2007b. Trivariate flood frequency analysis using the Gumbel-Hougaard copula. Journal of Hydrologic Engineering, 12: 431-439. DOI: 10.1061/(asce)1084-0699(2007)12:4(431).

Zhang, L. S. V. P., & Singh, V. P. (2006). Bivariate flood frequency analysis using the copula method. Journal of hydrologic engineering, 11(2), 150-164.

Zhang, T., Wang, Y., Wang, B., Tan, S., & Feng, P. (2018). Nonstationary flood frequency analysis using univariate and bivariate time-varying models based on GAMLSS. *Water*, *10*(7), 819.