RESEARCH ARTICLE | JUNE 21 2016

Bivariate copulas functions for flood frequency analysis FREE

Norizzati Salleh; Fadhilah Yusof; Zulkifli Yusop

AIP Conf. Proc. 1750, 060007 (2016)
https://doi.org/10.1063/1.4954612

Bivariate Copulas Functions for Flood Frequency Analysis

Norizzati Salleh [1, a)], Fadhilah Yusof [1, b)] and Zulkifli Yusop [2, c)]

[1]*Department of Mathematical Sciences, Faculty of Sciences, University of Technology Malaysia, 81310 Johor Bahru, Malaysia*
[2]*Water Resources Alliance,University of Technology Malaysia, 81310 Johor Bahru, Malaysia*

[a)] Corresponding author: *norizzatisalleh@gmail.com*
[b)]*fadhilahy@utm.my*
[c)]*zulyusop@utm.my*

Abstract. Bivariate flood frequency analysis offers improved understanding of the complex flood process and useful information in preparing flood mitigation measures. However, difficulties arise from limited bivariate distribution functions to jointly model major flood variables that are inter-correlated and each has different univariate marginal distribution. To overcome these difficulties, a Copula based methodology is presented in this study. Copula is functions that link univariate distribution functions to form bivariate distribution functions. Five Copula families namely Clayton, Gumbel, Frank, Gaussian and t Copulas were evaluated for modeling the joint dependence between peak flow-flood duration. The performance of four parameter estimation methods, namely inversion of Kendall's tau, inversion of Spearman's rho, maximum likelihood approach and inference function for margins for chosen copula's families are investigated. The analysis used 35 years hourly discharge data of Johor River from which the annual maximum were derived. Generalized Pareto and Generalized Extreme Value distribution were found to be the best to fit the flood variables based on the Kolmogorov-Smirnov goodness-of-fit test. Clayton Copula was chosen as the best fitted Copula function based on the Akaike Information Criterion goodness-of-fit test. It is found that, different methods of parameter estimation will give the same result on determining the best fit copula family. On performing a simulation based on a Cramer-von Mises as a test statistics to assess the performance of Copula distributions in modeling joint dependence structure of flood variables, it is found that Clayton Copula are well representing the flood variables. Thus, it is concluded that, the Clayton Copula based joint distribution function was found to be effective in preserving the dependency structure of flood variables. Thus, it is concluded that, the Clayton Copula based joint distribution function was found to be effective in preserving the dependency structure of flood variables.

Keywords: Flood variables; Parameter Estimation; Bivariate Copula

INTRODUCTION

Characterizing flood in the form of peak flow, volume and duration is crucial for designing hydraulic structures, planning and management of excess water. Flood variables are generally random in nature and mutually correlated. Flood frequency analysis defines the severity of a flood event by summarizing the characteristics of the flood, and by finding out their mutual dependence structure. A number of methodologies have been developed to perform univariate (see [1], [2], [3] and [4]) and multivariate flood frequency analysis by ([5] and [6]). However, according to [7], both univariate and multivariate techniques require several restrictive assumptions that all variables are independent, but the real phenomena are actually multivariate with significant interrelationships among the variables.

Thus, this study aims to use multivariate statistical tool called Copula functions to construct the dependency structure and joint probability distributions. The copula based method could provide reliable fitted marginal distribution of several variables, though they are not interpreted by the same family of probability distribution function [8]. Copula modeling was successfully enforced in other areas, particularly in survival analysis, actuarial science, and finance. In late times, the use of copulas has become popular for multivariate analysis in several fields, viz., in financial studies by ([9] and [10]), in hydrology by ([11], [12], [13], [14], [15],[16] and [17]) and for drought analysis by ([18] and [19]).

Advances in Industrial and Applied Mathematics
AIP Conf. Proc. 1750, 060007-1–060007-13; doi: 10.1063/1.4954612
Published by AIP Publishing. 978-0-7354-1407-5/$30.00

Following these successes, the concept of copula has been applied recently in flood frequency analysis to model the dependence structure among peak flow, volume and duration. Application of copula methods in modeling the joint distribution could relax the limitation of the traditional flood frequency analysis by selecting marginal from different families of probability distribution functions for flood characteristics. It is found that copula based flood frequency analysis performs better than conventional flood frequency analysis [20].

Thus the objective of this study was to analyze bivariate frequency of floods using copulas to model the relationships between peak flow and flood duration. In this study, we first determined separate univariate marginal distributions of peak flow and flood duration and then use copulas to construct the bivariate distribution. The performance of four parameter estimation methods in order to estimate the copula parameter, namely inversion of Kendall's tau, inversion of Spearman's rho, maximum likelihood approach and inference function for margins for chosen copula's families are investigated. Sungai Johor watershed was selected to illustrate the proposed methodology. The Sungai Johor watershed is one of the largest catchments in Johor with an area of about 2700 km^2. A large portion of this watershed is flood prone and is frequently affected by major floods. The watershed is also undergoing rapid development which has significantly evolved its demographic, industrial, transportation and social spheres. The peak flow is assumed to distribute as Generalized Pareto distribution while flood duration are assumed to follow Generalized Extreme Value, and tested by Kolmogorov-Smirnov (K-S) goodness-of-fit test. A total of five Copulas – Gaussian, t-Student, Clayton, Gumbel and Frank – are then used to link the predetermined univariate distribution to form the bivariate flood distribution.

METHODS

Modeling the Peak flow and Flood Duration

This study used the Generalized Pareto (GP), Pearson Exponential, Beta and Generalized Extreme Value (GEV) distributions to model the flood variables' distribution. The Probability Distribution Function (PDF) is defined as:

$$P(a \leq X \leq b) = \int_{a}^{b} p(x)dx \tag{1}$$

The theoretical PDF is displayed as a continuous curve. The empirical PDF is denoted by:

$$P(X = x) = p(x) \tag{2}$$

A cumulative distribution function (CDF), $F(x)$ accumulates all of the probability less than or equal to x. The distribution function is therefore related to a continuous probability density function $F(x)$ demonstrated as below:

$$F(x) = P[X \leq x] = \int_{-\infty}^{x} f(x)dx \tag{3}$$

Whereas, for empirical CDF, it is described as:

$$F_n(x) = \frac{1}{n}[Number\ of\ observations \leq x] \tag{4}$$

where x is the random variable of peak flow, flood duration, or flood volume.

The following describes the Generalized Pareto (GP) and Generalized Extreme Value (GEV) models of the PDF and CDF.

Generalized Pareto (GP)

The PDF and CDF of the Generalized Pareto distribution with continuous shape parameter (k), continuous scale parameter ($\sigma > 0$) and continuous location parameter (μ) are shown below respectively:

$$f(x) = \begin{cases} \frac{1}{\sigma}\left(1+\kappa\frac{x-\mu}{\sigma}\right)^{-1-\frac{1}{k}} & k \neq 0 \\ \frac{1}{\sigma} exp\left(-\frac{x-\mu}{\sigma}\right) & k = 0 \end{cases} \tag{5}$$

$$F(x) = \begin{cases} 1-(1+\kappa\frac{x-\mu}{\sigma})^{-\frac{1}{k}} & k \neq 0 \\ 1-exp\left(-\frac{x-\mu}{\sigma}\right) & k = 0 \end{cases} \tag{6}$$

where

$$\mu \leq x < +\infty \quad for\ k \geq 0$$
$$\mu \leq x \leq \mu - {}^{\sigma}/_{k} \quad for\ k < 0$$

Generalized Extreme Value (GEV)

The PDF and CDF of the general extreme value with continuous shape parameter (k), continuous scale parameter (σ) and continuous location parameter (μ) are:

$$f(x) = \begin{cases} \frac{1}{\sigma}\exp\left(-(1+kz)^{-1/k}\right)(1+kz)^{-1-1/k} & k \neq 0 \\ \frac{1}{\sigma}\exp\left(-z-\exp(-z)\right) & k = 0 \end{cases} \tag{7}$$

$$F(x) = \begin{cases} \exp\left(-(1+kz)^{-1/k}\right) & k \neq 0 \\ \exp\left(-\exp(-z)\right) & k = 0 \end{cases} \tag{8}$$

where

$$z \equiv \frac{x-\mu}{\sigma}$$

$$1 + k\frac{(x-\mu)}{\sigma} > 0 \quad for\ k \neq 0$$

$$-\infty < x < +\infty \quad for\ k = 0$$

The goodness of Fit (GOF) of a statistical model describes how well it fits into a set of observations and it measures how perfectly selected distribution can accommodate the data. In this study, the GOF tests were performed at 5% level of significance.

Kolmogorov- Smirnov (K-S) Test

The Kolmogorov- Smirnov test is defined by:

H_0	:	The data follow a specified distribution
H_1	:	The data do not follow the specified distribution
Test Statistics	:	The Kolmogorov-Smirnov test statistic is defined as

$$D = \max_{1 \leq i \leq N}\left(F(Y_i) - \frac{i-1}{N}, \frac{i}{N} - F(Y_i)\right)$$

where F is the theoretical cumulative distribution of the distribution being tested which must be a continuous distribution and it must be fully specified (i.e., the location, scale, and shape parameters cannot be estimated from the data).

Significance Level : α

Critical Values : The hypothesis regarding the distributional form is rejected if the test statistic D is greater than the critical value obtained from a table. There are several variations of these tables in the literature that use somewhat different scaling for the K-S test statistic and critical regions. These alternative formulations should be equivalent, but it is necessary to assure that the test statistic is counted in a manner that is consistent with how the critical values were tabulated. Nevertheless, we do not provide the K-S tables in this thesis since software programs that perform a K-S test will provide the relevant critical values.

Theoretical Aspect of Copula

Copulas are functions developed to model the formation of multivariate uniform distribution function based on their univariate distribution functions respectively [21]. The application of copulas contributes to association determination for flood characteristics independently with respect to their marginal distribution. In a simple form of understanding Copula, the two-dimensional (bivariate) Copula is a function of $C: [0,1]^2 \rightarrow [0,1]$, with the following properties:

1. When C is grounded: for all $u, v \in [0,1]$, $C(u,0) = 0$ and $C(0,v) = 0$. It means if either one of the marginal functions is zero then the joint distribution is zero.
2. When C is 2-increasing: $u_1, u_2, v_1, v_2 \in [0,1]$ such that $u_1 \leq u_2$ and $v_1 \leq v_2$, then $C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0$
3. For all $u, v \in [0,1]$, $C(u,1) = u$ and $C(1,v) = v$. If either one of the marginal is 1, the joint distribution behave as univariate of opposite variable.

According to Sklar (1959), for d-dimensional continuous random variables $\{X_1, \ldots X_d\}$ with marginal cumulative distribution functions (CDFs); $u_j = F_{X_j}(x_j), j = 1, \ldots d$ there exists one unique d-copula $C_{u_1,\ldots u_d}(u_1, \ldots u_d)$ such that:

$$C_{u_1,\ldots u_d}(u_1, \ldots u_d) = H_{x_1,\ldots x_d}(x_1, \ldots x_d) \tag{9}$$

where u_j is the jth marginal and $H_{x_1,\ldots x_d}$ is the joint-CDF of $\{X_1, \ldots X_d\}$. Since the marginals are nondecreasing from 0 to 1 for continuous random variables, copulas C can be regarded as a transformation of $H_{x_1,\ldots x_d}$ from $[-\infty, \infty]^d$ to $[0,1]^d$.

This transformation segregates the marginal distributions into $H_{x_1,\ldots x_d}$. Thus, $C_{u_1,\ldots u_d}$ is only suitable to describe the relationship between variables; it is also gives a look on dependence structure. So, in this research 2-D Copula (bivariate) was used. If H is the joint distribution function with margins U_1 and U_2, the Copula exists for x_1, x_2, in $\bar{R}([-\infty, +\infty])$ as shown in Equation 10 below:

$$C(U_1, U_2) = H(x_1, x_2) \tag{10}$$

Archimedean Copula

The Archimedean copulas are widely applied in hydrology, because they can be easily generated and are capable of capturing wide range of dependence structure with desirable properties such as, symmetry and associativity. The general equation of the Archimedean (symmetric) Copula is shown in Equation (11).

$$C(u_1, \ldots, u_n) = \varphi^{-1}[\varphi(u_1) + \cdots + \varphi(u_n)] \tag{11}$$

where $\varphi(t)$ is copula generator and $\varphi^{-1}(s)$ is the inverse of generator φ.

In this study, the following commonly used copulas from Archimedean are used to investigate the relationship between flood variables.

1. **Gumbel Copula**
   The Gumbel-Archimedean Copula was first introduced by [22]. The mathematical expression for Gumbel Copula is uniquely defined by its generator $\varphi(t) = (-lnt)^{\theta}$, where t varies from 0 to 1 regardless of whether it is equal to u_1 or u_2, with $\theta \geq 1$. The formulation of this Copula family is shown in Equation (12) and the joint density function follows Equation (13).

$$C(u_1, u_2; \theta) = C_\theta\left(F_{x_1}(x_1), F_{x_2}(x_2)\right) = H(x_1, x_2) \quad (12)$$

$$= \exp\left(-\left((-lnu_1)^{\theta} + (-lnu_2)^{\theta}\right)^{\frac{1}{\theta}}\right) \quad (13)$$

2. **Clayton Copula**
   Clayton Copula was introduced by [23] and developed by [24]. The formulation of this copula and the mathematical expression of its joint density are shown in Equation (14) as well as (15) and (16).

$$C(u_1, u_2; \theta) = C_\theta\left(F_{x_1}(x_1), F_{x_2}(x_2)\right) = H(x_1, x_2) \quad (14)$$

$$= (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta} \quad (15)$$

$$= (u_1 u_2)^{-\theta-1}(\theta + 1)(u_1{}^{-\theta} + u_2{}^{-\theta} - 1)^{-\frac{1}{\theta}-2} \quad (16)$$

   where θ is a parameter of the generating function $\varphi(t) = \frac{[(t)^{-\theta}-1]}{\theta}$ and $\theta \in [0, \infty]$. As a uniformly distributed random variable, t varies from 0 to 1 regardless of whether it is equal to u_1 or u_2

3. **Frank Copula**
   Frank Copula was developed by [25]. The formulation of this bivariate Copula family is expressed in Equation (17) and its joint density function is shown in Equation (18) and (19).

$$C(u_1, u_2; \theta) = C_\theta\left(F_{x_1}(x_1), F_{x_2}(x_2)\right) = H(x_1, x_2) \quad (17)$$

$$= -\frac{1}{\theta}n\left[1 + \frac{\left(e^{-\theta u} - 1\right)\left(e^{-\theta v} - 1\right)}{(e^{-\theta} - 1)}\right] \quad (18)$$

$$= \frac{\theta e^{-\theta(u+v)}(e^{-\theta} - 1)}{(e^{-\theta(u+v)} - e^{-\theta u} - e^{-\theta v}e^{-\theta})} \quad (19)$$

where θ is a parameter of the generating function $-\ln\left[(e^{-\theta t} - 1)/(e^{-\theta} - 1)\right]$. As a uniformly distributed random variable, t varies from 0 to 1 regardless of whether it is equal to u_1 or u_2.

Elliptical Copula

Elliptical Copula does not have a closed expression, however it able to present a pair of marginal's relationship. Gaussian (normal) and t Copula are commonly used for elliptical Copula families. Based on Sklar's Theorem, the general equation of elliptical copula is shown in Equation (20) while for Gaussian and t-Copula in Equation (21)

$$C(u, v) = \Phi_\theta(\Phi_\theta{}^{-1}(u), \Phi_\theta{}^{-1}(v)) \quad (20)$$

where $\Phi_\theta(.,.)$ is the bivariate standard normal CDF and $\Phi_\theta{}^{-1}(.)$ is the inversed univariate normal univariate CDF.

$$C(u, v) = t_\theta(t_\theta{}^{-1}(u), t_\theta{}^{-1}(v)) \quad (21)$$

where $t_\theta(.,.)$ is the bivariate standard cumulative t-distribution and $t_\theta{}^{-1}(.)$ is the inverse of the univariate t-distribution univariate CDF.

1. Gaussian (Normal) Copula

According to [26], the general equation of Gaussian Copula is shown in Equation (22) and can be connected to the specific joint density function in Equation (23).

$$C_G(u,v;\theta) = \Phi_G(\Phi^{-1}(u),\Phi^{-1}(v);\theta) \tag{22}$$

$$= \int_{-\infty}^{\emptyset^{-1}(u)}\int_{-\infty}^{\emptyset^{-1}(v)} \frac{1}{2\pi(1-\theta^2)^{1/2}} \exp\left\{-\frac{x^2-2\theta xy+y^2}{2(1-\theta^2)}\right\} dxdy \tag{23}$$

where $\Phi^{-1}(.)$ is the inverse function of the (CDF) $\Phi(.)$ and θ is the linear correlation coefficient between $\Phi^{-1}(u)$ and $\Phi^{-1}(v)$ restricted to interval $(-1,1)$.

2. t (student) Copula

The construction steps for t-Copula can be found in [27]. Its association with a bivariate t-distribution can be illustrated by stating two dependent variables with υ degrees of freedom and a correlation linear coefficient, θ (Lian et al. 2013).

$$C^t(u,v;\theta) = \Phi_G(\Phi^{-1}(u),\Phi^{-1}(v);\theta) \tag{24}$$

$$= \int_{-\infty}^{t_v^{-1}(u)}\int_{-\infty}^{t_v^{-1}(v)} \frac{1}{2\pi(1-\theta^2)^{1/2}} \left\{1+\frac{x^2-2\theta xy+y^2}{\upsilon(1-\theta^2)}\right\}^{-(v+2)/2} dydx \tag{25}$$

where $t_v^{-1}(.)$ denotes the inverse function of the CDF $t_v(.)$ with υ as the degree of freedom, and θ is between $t_v^{-1}(u)$ and $t_v^{-1}(v)$ and x and y are dependent variables.

Parameter Estimation of Copulas

In this study, there are several methods used in order to estimate the Copula parameter (θ), which is a parameter used to measure the degree of association between two univariate CDFs. The methods are: Approached Based on the Rank Correlation, Maximum Likelihood Approach (ML), and Inference Function for Margin (IFM). All these methods will be discussed in the following subsection.

Approached Based on the Rank Correlation

If there exists one-to-one correspondence between copula parameter $\hat{\theta}$ and rank correlation, then by substituting the empirical values of the rank correlation into the relation $\hat{\theta} = f(\hat{\tau})$ will yield the estimate of copula parameter. Table 1 below shows the summary of Copula families, parameter and relationship of Kendall's τ in different distributions.

TABLE 1. Summary of Copula Families, Parameter and Relationship Of Kendall's Tau

Copula	Gumbel	Clayton	Frank	t	Gaussian
Parameter $\boldsymbol{\theta}$	$\theta \geq 1$	$\theta \geq 1, \theta \neq 0$	$\theta \neq 0$	$-1 < \theta < 1$	$-1 < \theta < 1$
Kendall's	$(1-\theta^{-1})$	$\frac{\theta}{(\theta+2)}$	$1-\frac{4}{\theta}(1-D_1(\theta))$	$\frac{2}{\pi}\arcsin(\theta)$	$\frac{2}{\pi}\arcsin(\theta)$
Kendall's $\boldsymbol{\tau = 1 + 4\int_0^1 \frac{\phi(t)}{\phi'(t)} dt}$					
Notation first Debye function $\boldsymbol{D_1(\theta) = {}^1/_\theta \int_0^\theta \frac{t}{(e^t-1)} dt}$					

Maximum Likelihood Approach (ML)

Maximum Likelihood Approach (ML) was used to determine the Copula parameter (θ) by using copula package in R software. Basically, this method involves several steps below:

1. For two variables identified as X and Y, the PDFs are $f_X(x;\alpha)$ and $f_Y(y;\beta)$ respectively where α and β are the corresponding parameters of $f_X(x)$ and $f_Y(y)$ that have $\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_j, \ldots, \alpha_m, \quad i \epsilon [1, m]$ and $\beta_1, \beta_2, \beta_3, \ldots, \beta_j, \ldots, \beta_n, \quad j \epsilon [1, n]$ respectively. m and n are the number of the parameters. In this step, only the type of marginal need to be identified, not the parameters α and β.
2. An assumed copula expressed as $C_\theta(u, v, \alpha, \beta, \theta)$ where θ is the dependency parameter, u and v are the corresponding CDF of X and Y is then selected.
3. The copula density function is derived as:

$$\begin{aligned} f_{X,Y}(x, y; \alpha, \beta, \theta) &= c_\theta(u, v; \alpha, \beta, \theta) \frac{\partial u}{\partial x} \frac{\partial v}{\partial y} \\ &= c_\theta(u, v; \alpha, \beta, \theta) f_X(x; \alpha) f_Y(y; \beta) \end{aligned} \tag{26}$$

where

$$c_\theta(u, v; \alpha, \beta, \theta) = \frac{\partial^2 C(u, v)}{\partial u \partial v} \tag{27}$$

4. The log-likelihood function is thus written as:

$$\begin{aligned} &\ln L[\, f_{X,Y}(x, y; \alpha, \beta, \theta)] \\ &= \sum_{k=1}^{K} ln c_\theta(F_X(x_k; \alpha), F_Y(y_k; \beta); \theta) + \sum_{k=1}^{K} [ln\, f_X(x; \alpha) \\ &+ f_Y(y; \beta)] \end{aligned} \tag{28}$$

where K is the number of the observations.

5. Let $\Omega = (\alpha, \beta, \theta)$ and thus Equation (2.4.3.4) can be solved as:

$$\frac{\partial lnL[f_{X,Y}(x, y; \alpha, \beta, \theta)]}{\partial \Omega} = 0 \tag{29}$$

Inference Function for Margin (IFM)

The Inference function for margin (IFM) is one of method in estimating the parameter of Copula. Basically this method has two steps:

1. In the first step the parameters for the models of the marginal are estimated:
   Based on the log-likelihood functions of two margins, the parameters α and β are respectively estimated for the PDF of $f_X(x;\alpha)$ and $f_Y(y;\beta)$ where α and β are the parameters of $f_X(x)$ and $f_Y(y)$ that have $a_1, a_2, \ldots a_i, \ldots a_m, \; i \in [1, m]$ and $\beta_1, \beta_2, \ldots \beta_j, \ldots \beta_n, \; j \in [1, n]$.
2. In the second step the parameter of the copula model are estimated:
   By using the estimated α and β, the full log-likelihood function is solved as:

$$\begin{aligned} &lnL\big[f_{X,Y}(x, y; \alpha, \beta, \theta)\big] \\ &= \sum_{k=1}^{k} lnC_\theta(F_X(x_k; \alpha), F_Y(y_k; \beta); \theta) \\ &+ \sum_{k=1}^{k} [lnf_X(x_k; \alpha) + f_Y(y_k; \beta)] \end{aligned} \tag{30}$$

where θ is dependency parameter.

Goodness-of-Fit Statistics for Copula

Akaike Information Criteria (AIC)

The AIC which was developed by [28] was used in this study for determine the best fitted Copula family for flood frequency analysis. AIC can be obtained by calculating the maximum likelihood of the model. Thus AIC can be expressed as:

$$AIC = -2\log(maximized\ likelihood\ of\ model) + 2m \tag{31}$$

where m is the number of estimated fitted parameter dependent on the type of univariate marginal distribution and Copula.

RESULT AND DISCUSSION

Univariate Marginal Distributions of Flood Variables

The statistical summary of the flood parameters are shown in Table 2 below.

TABLE 2. Basic Statistics of Flood Variables

	Peakflow (m^3/sec)	Duration (Hr)
maximum	724.73	600
minimum	76.9	144
average	248.2276471	349.4117647
SD	163.8648299	125.5016884
CV(%)	66.01393192	35.91799164

Table 2 shows the average peak flow and flood duration were $248.2276471 m^3/sec$ and 349.4117647Hr respectively. It also shows that the peak flow has the highest variation with coefficient of variation (CV) of 66.01393192% whereas flood duration has the less variation with CV of 35.91799164%. The coefficient of variation represents the ratio of the standard deviation to the mean, and it is a useful statistic for comparing the degree of variation from one data series to another, even if the means are drastically different from each other.

While, Table 3 presents (k), continuous scale parameter (σ), and continuous location parameter (μ) of various distributions under study. Based on the (K-S) GOF tests, it is found that the Generalized Pareto distribution fit well with peak flow variable and observed peak flow that was obtained are shown in Fig. 1. On the other hand, Generalized Extreme Value (GEV) distribution is found to be the best fit distribution for flood duration and observed flood duration and presented in Fig. 2.

TABLE 3. Fitting Result Parameters for Various Distributions of Flood Variables

Flood Variables	Best fitted distribution	Parameters
Peakflow (P)	Gen. pareto	$k = -0.03906$ $\sigma = 184.8$ $\mu = 70.686$
Duration (D)	Gen. extreme value (GEV)	$k = -0.20041$ $\sigma = 122.45$ $\mu = 299.35$

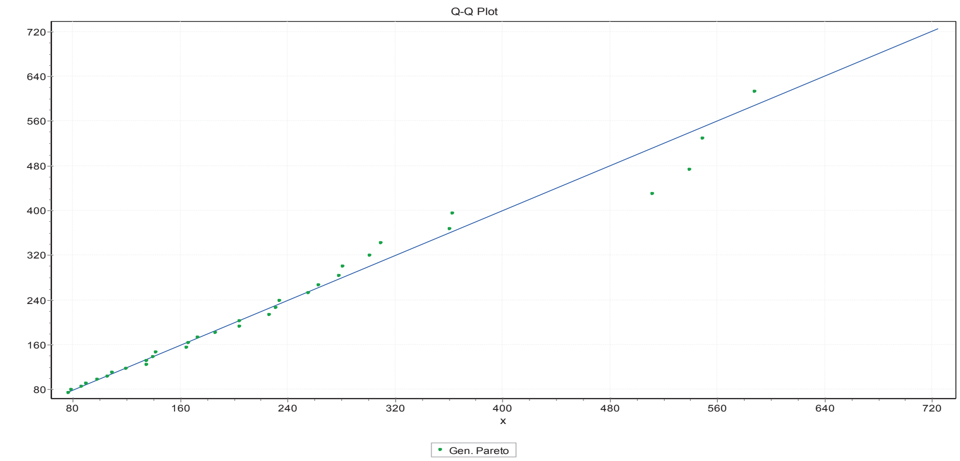

FIGURE 1. Comparison of the Observed Peak Flow and the Fitted Gen. Pareto Distribution

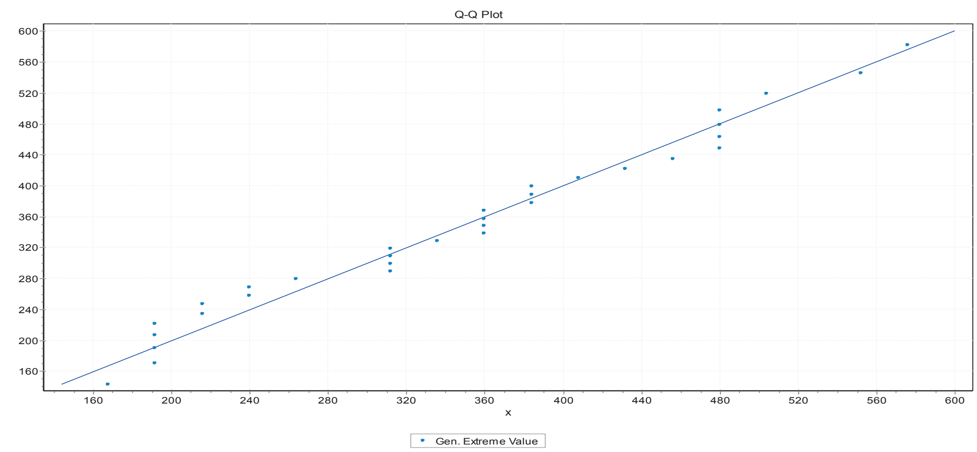

FIGURE 2. Comparison of the Observed Flood Duration and the Fitted GEV Distribution

Copula-Based Bivariate Flood Distribution

Construction of the bivariate flood distribution is achieved in this study using copulas including Gaussian, Student-t, Clayton, Gumbel and Frank. Besides that, several methods are used to estimate the copula parameter that is rank correlation method including Kendall's tau and Spearman's rho, Maximum Likelihood Approach (ML) and Inference Functions for margins (IFM). Table 4 presents a summary of the parameters derived from all the methods stated before for peakflow- flood duration paired of flood parameters.

TABLE 4. Copula Linear Correlation Parameter, θ Using Method of (a) Inversion of Kendall's Tau (b) Inversion of Spearman's Rho (c) Maximum Likelihood Approach (d) Inference Functions for Margins

TABLE 4(a)

Copula family	Peak flow-Duration
t	0.03982
Gaussian	0.02300034
Frank	0.1318168
Clayton	0.02972278
Gumbel	1.014861

The degree of freedom for t-copula	2.97818

TABLE 4(b)

Copula family	Peak flow-Duration
t	0.04108
Gaussian	0.02922
Frank	0.1675
Clayton	0.03776
Gumbel	1.0204
The degree of freedom for t-copula	2.9961

TABLE 4(c)

Copula family	Peak flow-Duration
t	0.01451
Gaussian	0.09824
Frank	0.05146
Clayton	0.2988
Gumbel	1.0327
The degree of freedom for t-copula	2.0001

TABLE (4d)

Copula family	Peak flow-Duration
t	0.06106986
Gaussian	0.08687449
Frank	0.1820346
Clayton	0.45074
Gumbel	1.0001
The degree of freedom for t-copula	2.671197

With the Copula linear correlation parameter θ and the degree of freedom parameter from Table 4, the Copula-based bivariate joint CDFs for five different Copula families are constructed using all values of Copula parameter form various methods stated before. Based on the Akaike information criterion (AIC) method, a good model must has low AIC value. Based on the AIC value obtained presented in Table 5, it shown that different method of parameter estimation will give the same result in order to find the best fit Copula family for flood variables. As we can see, the lowest AIC value for all the methods used for copula parameter estimation goes to Clayton Copula for peak flow- duration paired of flood variables.

TABLE 5. AIC Value for Five Copula Functions for Flood Variables Using (a) Inversion of Kendall's Tau (b) Inversion of Spearman's Rho (c) Maximum Likelihood Approach (d) Inference Functions for Margins; as A Method of Copula Parameter Estimation

TABLE (5a)

Copula family	Peak flow-Duration
t	1.6978
Gaussian	1.910855
Frank	2.002824
Clayton	1.464596
Gumbel	1.972412

TABLE (5b)

Copula family	Peak flow-Duration
t	1.5871
Gaussian	1.890757
Frank	2.00802
Clayton	1.340241
Gumbel	1.966362

TABLE (5c)

Copula family	Peak flow-Duration
t	-0.1137811
Gaussian	1.783037
Frank	1.998037
Clayton	-0.2263732
Gumbel	1.96092

TABLE (5d)

Copula family	Peak flow-Duration
t	0.2001551
Gaussian	1.786017
Frank	2.011068
Clayton	0.141844
Gumbel	1.999758

In this paper, to illustrate the bivariate frequency analysis of floods, we selected only the Clayton copula to model the dependent structure between peak flow- duration since Clayton Copula is found to be the best fitted family of Copula for flood variables. Whereas since the different method of copula parameter estimation give the same result in order determining best fit copula family, thus we choose value of copula parameter obtained based on Maximum Likelihood Approach (ML) method because (ML) approach is the most efficient in determining the Copula parameter because it is consistent and asymptotically normal under the usual regularity conditions for the multivariate model and for each of its margins.

The Clayton based bivariate cumulative distribution function of peak flow- duration is demonstrated in Equation 32 below.

$$F(x,y) = \left[\left(1-\left(1-0.03905\,\frac{x-70.684}{184.48}\right)^{\frac{1}{0.03905}}\right)^{-0.2988} + \exp\left(-\left(1-0.20041\left(\frac{y-299.35}{1222.45}\right)\right)^{\frac{1}{0.20041}}\right)^{-0.2988} - 1\right]^{-\frac{1}{0.2988}} \quad (32)$$

The joint cumulative distribution function of the peak flow- duration based on Clayton Copula is illustrated in Fig. 3.

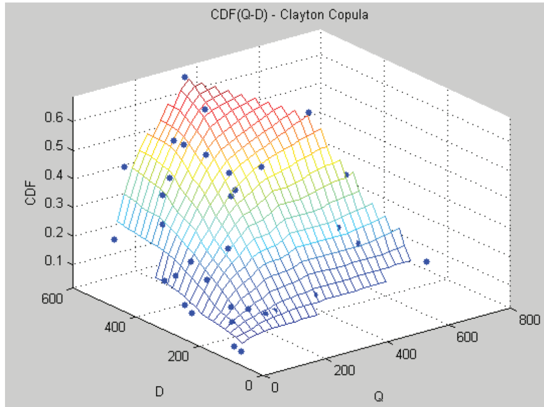

FIGURE 3. Joint cumulative distributions function of peak flow and flood duration

Figure 3 illustrates the contours of the Clayton based joint cumulative distribution $F_{Q,D}(q,d)$ with respect to univariate probabilities $F_Q(q)$ and $F_D(d)$. For each cumulative distribution contour, there is an inverse relationship between peak flow and duration. It means that if flood peak is high, the flood duration will be low or vice versa. Also with higher cumulative distribution contours allow higher peak flow and flood duration. Therefore, Figure 3 refers to chance or probabilities of two condition peak flow and flood duration occurring at the same time. Whereas, Fig. 4 shows joint distribution contours for peak flow and flood duration at 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9. The contour line marked 0.1 means that if peak discharge is more than probabilities of 0.82, the flood duration will be less than probabilities of 0.949.

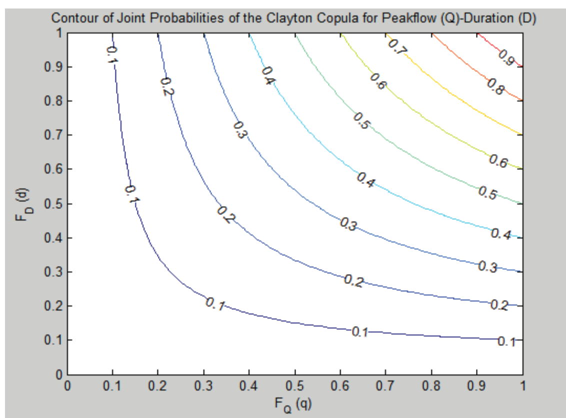

FIGURE 4. Contour of Joint Probabilities of the Clayton for Peak flow (Q) - Duration (D)

CONCLUSION

A methodology based on copulas is presented in this study to construct a bivariate flood frequency analysis. The advantages using Copulas is that sophisticated bivariate distribution modeling is reduced to investigate the relationship between flood variables under the condition of given univariate marginal distributions. Besides, various method of parameter estimation on estimating the copula parameter is used in this study. The following conclusion is drawn from this study: (1) different method of parameter estimation of copula parameter gives the same result in order to find the best fitted family for peakflow-duration paired of flood variables. (2) The Clayton Copula family is selected as the best fit copula for peakflow-duration paired of flood variables for case study at Sungai Johor. (3) The Copula method is found to be very effective tool for bivariate modeling of flood risks, as copula are effectively preserving the dependence structure of multiple flood characteristics.

ACKNOWLEDGEMENT

The authors would like to thanks Universiti Teknologi Malaysia (UTM) and MOHE grant vote number (4L832) for financial funding.

REFERENCES

1. C. D. Michele and G. Salvadori, Journal of Geophysical Research Atmosphere. 108(1), 1884-2012 (2003).
2. E. Goodarzi, M. Ziaei, and L.T. Shui, in *Introduction to Risk and Uncertainty in Hydrosystem Engineering, edited by* E. Goodarzi, M. Ziaei, and L.T. Shui, (Springer, Netherlands, 2013), pp. 123-141.
3. R. K. Linsley, Water Resources Research. 22(9S), 1595-1645 (1986).
4. L. Zhang, "Multivariate hydrological frequency analysis and risk mapping," Doctoral dissertation. Louisiana State Univesity. 2005.
5. K. Singh and V.P. Singh, Stochastic Hydrology and Hydraulics. 5(1), 55-68 (1991).
6. S. Yue, T. B. M. J. Ouarda and B. Bobee, Journal of Hydrology. 246(1), 1-18 (2001).
7. L. Zhang and V. P. Singh, Journal of Hydrologic Engineering. 11(2), 150-164 (2006).
8. R. B. Nelsen, *An introduction to copulas*. (Springer, New York, 1999).
9. U. Cherubini, E. Luciano and W. Vecchiato, in *Copula methods in finance, edited by* John Wiley and Sons Ltd., (The Atrium, England, 2004).
10. E.W. Frees and E.A. Valdez, American Actuarial Journal. 2(1), 1-25 (1998).
11. C. Genest, and A. C. Favre, Journal of Hydrologic Engineering. 12(4), 347-368 (2007).
12. H. Chowdhary, L.A. Escobar and V.P. Singh, Hydrology Research. 42(3), 193-216 (2011).
13. S. Grimaldi and F. Serinaldi, Advances in Water Resources. 29(8), 1155-1167 (2006).
14. C. Favre, S. El Adlouni, L. Perreault, N. Thiemonge, and B. Bobbe, Water Resources Research. 40(1), (2004).
15. C. Wang, N. B. Chang and G. T. Yeh, Hydrological Processes. 23(10), 1471-1486 (2009).
16. B. Klein, M. Pahlow, Y. Hundecha and A. Schumann, Journal of Hydrologic Engineering. 15(5), 360-369 (2010).
17. S. Karmakar and S. P. Simonovic, Journal of Flood Risk Management. 2(1), 32-44 (2009).
18. S. C. Kao and R. S. Govindaraju, Journal of Hydrology. 380(1), 121-134 (2010).
19. S. Song and V. P. Singh, Stochastic Environmental Research and Risk Assesment. 24(5), 783-805 (2010).
20. J. J. Lian, K. Xu and C. Ma, Hydrology and Earth System Sciences. 17(2), 679-689 (2013).
21. Sklar, Paris Institute of Statistics. 2(8), 229-231 (1959).
22. R. D. Cook and M. E. Johnson, Journal of Royal Statistical Society Series B(Methodological). 43(2), 210-218 (1981).
23. M. J. Frank, in Aequationes Math. 19(1), 194-226 (1979).
24. E. J. Gumbel, "Distributions des Valeurs Extremes en Plusieurs Dimensions," (Institute of Statistics, Paris, 1960).
25. G. H. Wong, "Drought Predictions: Applications in Australia," Doctoral dissertation, University of Adelaide (2010).
26. C. Cunnane, Journal of Hydrology. 100(1), 269-290 (1987).
27. D. G. Clayton, Biometrika. Volume 65(1), 141-151 (1978).
28. H. Akaike, Automatic Control, IEEE Transactions on. 19(6), 716-723 (1974).