Research article

Trivariate distribution modelling of flood characteristics using copula function—A case study for Kelantan River basin in Malaysia

Shahid Latif* and Firuza Mustafa

Department of Geography, University of Malaya, Kuala Lumpur 50603, Malaysia

* Correspondence: Email: macet.shahid@gmail.com; Tel: +60166316364.

Abstract: Water resources operational planning, managements or either flood defence infrastructure designs often demand the estimations of flow exceedance probability for visualizing the risk of flood episodes. Numerous literature often incorporated copulas for the development of bivariate joint dependence structure among the flood characteristics, flood peak flow, volume and duration, but it could be more realistic and comprehensive if we focus all the three mutually correlated flood characteristics simultaneously. Actually, the inclusion of more flood characteristics could provide better and much justifiable information in correlation and dependency modelling. In this study, trivariate copulas are incorporated and applied to a case study to analyse flood episodes in the Kelantan River basin at Gulliemard bridge gauge station in Malaysia. Firstly, for describing best-fitted bivariate copulas for establishing the joint dependence structure of each flood attribute pairs, the Gaussian copula is recognized most justifiable model for peak-volume pair and the Frank copula for peak-duration and volume-duration pairs. After that, the trivariate joint distribution is modelled using one Archimedean copula, the Frank copula and one Elliptical copula, the Gaussian copula. Based on Cramer-von-Mises-type statistics, Sn and p-value, the Gaussian copula best representing the trivariate dependence structure of flood and which further employed in the deriving of trivariate joint and conditional return periods.

1. Introduction

Water-related operational planning, managements or either flood defence infrastructure designs often demands accurate estimations of the flood exceedance probability for assessing the hydrologic risk [1–3]. The probabilistic assessment often provides a flexible way to inference and extrapolates long term historical streamflow characteristics by fitting the most justifiable probability distribution functions and estimating their specified flood exceedance probability or return periods. The flood frequency analysis or FFA is an approach to relate flood design quantiles and their frequency of occurrence or non-exceedance probability by fitting the probability distribution functions [4,5]. The unreliability of univariate FFA are already highlighted through numerous studies (e.g. [4,6]), which cannot sufficiently characterize the full structure of flood hydrograph and might reveals the underestimation or overestimation of associated risk of correlated flood characteristics. In actuality, the flood is a multidimensional random consequence usually characterized completely through its trivariate mutual correlated vector such as flood peak discharge flow, volume and duration of flood hydrograph [7,8]. The multivariate distribution modelling often facilitates an effective approach in the prediction of hydrologic risk through visualizing the mutual dependencies among its multiple intercorrelated characteristics based on the joint probability density functions or JPDFs and joint cumulative distribution functions or JCDFs [5,9] also, to demonstrate the uncertainties interlinked with these hydrologic events. More especially, from the hydraulic designing perspective where, the accountability of multivariate design variables is often an insightful strategies [10,11]. The necessity of estimating flood design hydrograph instead of the single variable flood modelling or FFA i.e., either flood peak /or volume /or duration as a function with non-exceedance probabilities motivated numerous demonstrations (e.g. [6,12,13]), towards the incorporation of distinguished varieties of traditional bivariate or trivariate distribution functions for establishing the joint relationship among flood characteristics.

All the above distribution-based flood modelling approaches often surrounded with several statistical limitations such as (a) each flood characteristics must assume to have gaussian (or normal) distributions (b) statistical parameter of univariate marginal structure is often employed to model their joint dependence structure (c) limited space are available for the justification of their joint dependence structure under the traditional probability functions (d) complexity in the mathematical formulation as the number of random variable got increases [2,7,15]. Therefore, De Michele and Salvadori (e.g., [16]) firstly introduced copulas function for establishing the joint dependence structure between storm intensity and duration series. After that, extended literatures incorporated bivariate or few trivariate copulas distribution as a model risk for tackling different hydrological extremes (i.e., [15,17,18]). In actual, the copula function perform the modelling of individual univariate distribution and their joint dependence structure separately into two different stages, which allow higher degree of flexibility in the selection best fitted marginal distributions and their joint structure to capture a wider extent of linear and non-linear dependencies alongwith their preservation in their mutual dependence structure [19–21].

Existing copula distribution modelling frequently focused towards the establishment of bivariate joint analysis of the flood attribute pairs such as between peak flow and volume series and/or volume and duration series, and/or peak flow and -duration series(i.e., [9,22]). But the more

comprehensive flood risk analysis can be achieved through accounting all the trivariate random vector simultaneously by introducing the trivariate copula distribution modelling. Multiple relevant vectors of the specified hydrological episodes could likely depends upon the potential damage also, the ignorance of spatial dependency among these uncertain flood characteristics may responsible for the underestimation of uncertainty [11,23]. Therefore, the consideration of multiple flood relevant random vectors could provide better demonstration of their correlation or mutual dependence structure. Few existing incorporations such as Grimaldi and Serinaldi (i.e., [17]), performed flood distribution modelling by adapting different trivariate functions such as the mono-parametric and fully nested structure of Frank functions, Gumbel logistic distributions and pointed the significance of Frank function under FNA structure. Similarly, Serinaldi and Grimaldi, (i.e., [24]), derived trivariate flood dependence structure using the same fully nested structure. Genest et al., (i.e., [25]), modelled the annual spring flood analysis over Romaine River in Canada using the meta-elliptical copulas and their results revealed that such incorporation facilities an effective modelling environment for the analysis of multi-dimensional observations alongwith the preservation of the pair-wise dependencies among multiple random vectors through the correlation matrix but exhibited some modelling limitation such as might be ineffective under the low probabilities unless the asymptotic properties of data will be justified through the strong arguments. Similarly, Reddy and Ganguli (i.e., [3]), applied the fully nested Archimedean or FNA class copula and Student's t copula (Elliptical class copula) for the annual flood characteristics and examined the significance of multidimensional designs events by comparing univariate, bivariate and trivariate return periods and thus revealed that it could be an essential effort to demonstrate the joint and conditional flood occurrence in the light of trivariate return periods. Similarly, Fan and Zheng (i.e., [26]), adopted the entropy copula based on the Gibbs sampling procedure along with the Gaussian and the Archimedean copula for simulation of trivariate flood characteristics and revealed that using the entropy copula one can easily projected into higher dimensional frame directly just like as the Gaussian copula.

The Kelantan River basin is often affected by the most intensive monsoonal flooding in Malaysia and perceiving for increasing in term of their frequency and magnitude [27–29]. Few historical extremes happening such as intense and prolonged precipitation in the year 2002 caused flooding of a total area of 1640 km^2 and affected the population of 714,287 or either in the early month of December 2014, much heavy precipitation occurred for many of days triggered the flood event in most of the part of eastern coast of Kelantan river basin and it was the worst flood ever recorded in history and affected more than 200,000 people [29]. The cause of frequent failure of flood defence infrastructure in Malaysia due to the impact of moderately severe of flood episodes might be responsible due to the lack of complete flood hydrograph or in other words, where only flood peak discharge samples often targeted in deriving flood frequency curve during the structural development. Therefore, multivariate probabilistic assessments of flood characteristics and their associated return periods could be a comprehensive way for making a defensive risk-based decision making in the various basin perspective water-related issues. In this study, the copulas distribution modelling is incorporated for establishing trivariate joint dependence structure of flood peak, volume and duration series. The probabilistic model is implemented on the block (annual) maxima based flood sampling procedure, also called at-site event-based methodology, in which the daily basis

streamflow discharge records from period 1961–2016 are collected for the Kelantan River Basin at the Gulliemard Bridge gauge station in Malaysia. Both the Archimedean class and Elliptical class copula function are introduced and their adequacy are tested in the establishment of trivariate joint dependency simulations of flood characteristics. For the trivariate cases, joint primary return period in both "OR" and "AND" cases (for annual flood analysis) are estimated and also compared with the bivariate and univariate return periods. Also, trivariate conditional distribution and their associated return periods are investigated and compared with the bivariate cases.

2. Theoretical framework

2.1. Trivariate distribution using the copula function

Let us consider, if the flood peak flow, P, volume, V and duration, D series be the three intercorrelated flood characteristics then the joint probability distribution, F, can join the probabilities of these random variables and can be expressed as [8,30];

$$F(p, v, q) = P'(P \le p, V \le v, D \le d) = \int_0^d \int_0^v \int_0^p f(p, v, d) \, dp \, dv \, dd \tag{1}$$

where p, v, q = values of flood characteristics P, V and D; and P'= Non-exceedance probability.

According to Salvadori and De Michele (i.e., [31]), the multivariate joint return period can be derived from the Eq 1, as given below;

$$T(P \ge p, V \ge v, D \ge d) = \frac{\mu}{1 - (P'(P \le p, V \le v, D \le d) = \int_0^d \int_0^v \int_0^p f(p,v,d) \, dp \, dv \, dd)} \tag{2}$$

where $F(.)$ = joint CDF or JCDF; T = return period; μ = average inter-arrival time of sequential hydrologic or flood event =1.

The ideas of the copula method have been developed by Saklar (i.e., [19]). According to Nelsen (i.e., [20]), the copula are function that connects multivariate probability distributions to their univariate marginal functions. One of the major advantages of copula function is to modelling the dependence structure of the multiple intercorrelated univariate marginal distribution independently. Mathematically, let us consider the situation of bivariate random series, according to Sklar's theorem [20], if (X, Y) be the bivariate random variables with continuous marginal distributions $u_1 = F_X(x) = P(X \le x)$, and $u_2 = F_Y(y) = P(Y \le y)$, then it can be characterized uniquely by its associated dependence function called Copula or C which can be defined on the unit square, can be expressed as;

$$H_{X,Y}(x, y) = C[F_X(x), F_Y(y)] = C(u_1, u_2) \tag{3}$$

where, C = any type of bivariate copulas under consideration; $F_X(x) = F_Y(y)$ = CDF of univariate random variables "X" and "Y"; $H_{X,Y}(x, y)$ = bivariate joint probability distribution functions which can be expressed in terms of its univariate marginal functions and the associated dependence function C, as revealed from Eq 3. According to Shiau (i.e., [32]) and Zhang and Singh (i.e., [7]), the copula C must be unique if are continuous and thus can easily capture the wider

extent of dependencies among the random variables. Conversely, if $F_X(x)$, $F_Y(y)$ and the copula functions is given, then the above Eq 1 must define the bivariate joint distribution functions with its marginal distributions and Similarly, if $f_X(x)$ and $f_Y(y)$ are the PDF of variable X and Y, then the joint probability density of the two random variables can be expressed as;

$$f_{X,Y}(x,y) = c(F_X(x), F_Y(y)) \, f_X(x)f_Y(y) \tag{4}$$

where, c is the density function of bivariate copula C, can be defined as;

$$c(u,v) = \frac{\partial^2 c(u,v)}{\partial u \, \partial v} \tag{5}$$

in which, $u_1 = F_X(x)$ and $u_2 = F_Y(y)$.

Similarly, we consider the situation of trivariate distribution series where the joint distribution of random variables can be expresses as;

$$H_{X,Y,Z}(x,y,z) = C[F_X(x), F_Y(y), F_Z(z)] = C(u_1, u_2, u_3) \tag{6}$$

where $H_{X,Y,Z}(x,y,z)$ = trivariate joint distribution of random variables; $F(.)$ = marginal distribution; and C = trivariate copula function.

In this study, we introduced the Archimedean copula called Frank copula and elliptical copula called the Gaussian copula for establishing trivariate joint dependency of the annual basis (i.e., block (annual) maxima) flood characteristics i.e., flood peak flow, volume and duration series. The Archimedean copulas are widely accepted in numerous demonstration which exhibited a different varieties of families and also much effective and flexible to capture wider extent of joint dependencies [17,20]. On the other side, the elliptical family-based Gaussian copula is also introduced as a candidate model for testing their adequacy in the establishment of trivariate joint dependency simulations of flood characteristics. The Gaussian copula is an implicit copula which can be expressed as an integral over the density of X, and that can expressed mathematically for bivariate case as given below [33];

$$C_\theta(u,v) = \int_{-\infty}^{-\phi^{-1}(u_1)} \int_{-\infty}^{-\phi^{-1}(u_2)} \frac{1}{2\pi(1-\theta^2)^{1/2}} \exp\left[-\frac{s^2 - 2\theta st + t^2}{2(1-\theta^2)}\right] ds dt \tag{7}$$

The Gaussian copula shows almost no dependence in the tails and is mostly distributed around centre of the distribution but because of simple intuition as it is based on normal distribution, it is quite popular among the hydrologist and water practioner in extreme event modelling (i.e., [33–35]).

Mathematically, the two and three-dimensional Frank and Gaussian (or Normal) copula can be expressed as:

For the 3-dimension Frank copula;

$$C_\theta^3(u_1, u_2, u_3) = \frac{-1}{\theta}\ln\left(1 + \frac{(e^{-\theta u_1}-1)(e^{-\theta u_2}-1)((e^{-\theta u_3}-1)}{(e^{-\theta}-1)}\right), -\infty < \theta + \infty \tag{8}$$

where, $\phi(t) = -\ln\left(\frac{e^{-\theta t}-1}{e^{-\theta}-1}\right)$ = generating function.

Similarly, the expression for 2-dimension Frank copula;

$$C_\theta^2(u_1, u_2) = \frac{-1}{\theta} \ln\left(1 + \frac{(e^{-\theta u_1}-1)(e^{-\theta u_2}-1)}{(e^{-\theta}-1)}\right), -\infty < \theta + \infty \tag{9}$$

where, $\phi(t) = -\ln\left(\frac{e^{-\theta t}-1}{e^{-\theta}-1}\right)$ = generating function; C_θ^2 & C_θ^3 = two-dimensional and three-dimensional copula with parameter θ; $u_1 = F_P(p), u_2 = F_V(v), u_3 = F_D(d)$ = marginal distribution of trivariate random characteristics.

For the 2-dimensional Gaussian copula;

$$C_\theta^2(u_1, u_2) = \Phi_\Sigma(\Phi^{-1}(u_1), \Phi^{-1}(u_2)), -1 < \theta + 1 \tag{10}$$

And, for 3-dimensional Gaussian copula;

$$C_\theta^3(u_1, u_2, u_3) = \Phi_\Sigma(\Phi^{-1}(u_1), \Phi^{-1}(u_2), \Phi^{-1}(u_3)), -1 < \theta + 1 \tag{11}$$

where Φ = cumulative distribution function of standard normal or gaussian distribution.

2.2. Estimation of copula dependence parameters

In this literature, the parameter of the 3-dimensional copula, also the 2-dimensional bivariate copulas are estimated using the ranked-based Maximum pseudo-likelihood estimations (MPL) estimation procedure [9,36,37]. The MPL estimators is the modified version of traditional maximum likelihood method where the rank based empirical distributions are used for estimating copula parameters and can be applied for both one or multi-parameter copula functions also, copula parameters are usually estimated independently from their univariate marginal distribution functions [9,38,39]. MPL estimation procedure required firstly, to transform the univariate flood marginal variables into uniformly distributed vectors using its empirical distribution function. After that, through the maximization of pseudo-loglikelihood function one can easily estimate copula dependence parameters.

Mathematically,

$$l(\theta) = \sum_{i=1}^{n} \log\left[c_\theta\{F_1(X_{i,1}), F_2(X_{i,2})\ldots\ldots\ldots, F_k(X_{i,k})\}\right] \tag{12}$$

where, θ = copula parameter; $l(\theta)$ = pseudo log-likelihood function; $F_1(X_{i,1}) = F_1(X_{i,2}) = \ldots\ldots = F_k(X_{i,k})$ = empirical CDFs. Eq 12 is estimated by putting the value of empirical cumulative density or CDFs into copula density function and taking the logarithm to the likelihood function of the copula. Also, the empirical CDF is used as a substitute for the unknown univariate marginals distribution. Finally, the copula parameter can be derived through maximizing Eq 12, as given below;

$$\frac{1}{n}\frac{\partial l(\theta)}{\partial \theta} = \frac{1}{n}\sum_{i=1}^{n} l_\theta\left[\theta, F_1(X_{i,1}), F_2(X_{i,2})\ldots\ldots\ldots, F_k(X_{i,k})\right] = 0 \tag{13}$$

After the estimation of copula dependence parameter "θ", it can be used for the representation of multivariate structure of flood characteristics and estimation of joint and conditional return periods that are needed for the hydrologic design.

2.3. Goodness-of-fit Statistics

In the estimation of multivariate copula joint distribution, the Cramer-von Mises test statistics is employed to evaluate the adequacy of hypothesized copulas fitted to trivariate (or bivariate) flood characteristics [40,41]. According to Genest et al., (i.e., [41]) and Reddy and Ganguli (i.e., [9]), this test makes the use of the Cramer-von Mises statistic "Sn" through a comparative assessment between empirical, and theoretical probability distribution, using the following mathematical algorithm as given below;

For testing the fitness level of 2-dimensional or bivariate copula function

$$S_n = n\int_{[0,1]^2} \{c_n(u_1, u_2) - C_\theta(u_1, u_2)\}^2 dC_n(u_1, u_2)$$

$$= \sum_{i=1}^n \{c_n(U_{1i,n}, U_{2i,n}) - C_\theta(U_{1i,n}, U_{2i,n})\}^2 \tag{14}$$

For testing the fitness consistency during 3-dimensional or trivariate copula construction

$$S_n = \int_{[0,1]^2} n \{c_n(u_1, u_2, u_3) - C_\theta(u_1, u_2, u_3)\}^2 dC_n(u_1, u_2, u_3)$$

$$= \sum_{i=1}^n \{c_n(U_{1i,n}, U_{2i,n}, U_{3i,n}) - C_\theta(U_{1i,n}, U_{2i,n}, U_{3i,n})\}^2 \tag{15}$$

where, $c_n(u_1, u_2, u_3)$ & $c_n(u_1, u_2)$ = trivariate and bivariate empirical copulas estimated using the "n" observational flood attribute pairs; C_θ = parametric copula derived under the null hypothesis; u_1, u_2, u_3 = univariate marginal distribution of flood characteristics say P,V and D; $U_{1i,n}, U_{2i,n}$ or $U_{1i,n}, U_{2i,n}, U_{3i,n}$ = pseudo-observations of C transformed from $(X_1, Y_1), (X_2, Y_2), \ldots\ldots (X_n, Y_n)$ or $(X_1, Y_1, Z_1), (X_2, Y_2, Z_2), \ldots\ldots (X_n, Y_n, Z_n)$. Numerically, the value of $U_{1i,n}, U_{2i,n}$ and $U_{3i,n}$ can be estimated by using following mathematical approach;

$$U_{1i,n} = \frac{1}{n+1}\sum_{j=1}^n 1(X_j \le X_i); \ U_{2i,n} = \frac{1}{n+1}\sum_{j=1}^n 1(Y_j \le Y_i); \ U_{3i,n} = \frac{1}{n+1}\sum_{j=1}^n 1(Z_j \le$$

$$Z_i), i \in \{1, \ldots, n\} \tag{16}$$

In this demonstration, the p-values for each fitted copulas are estimated using the parametric bootstrapping technique (i.e., [40]), during the simulation of bivariate copulas structure and using the faster multiplier approach (i.e., [42,43]) during the simulation of trivariate copulas function. Although the empirical processes involved in the multiplier and the parametric bootstrap-based test are asymptotically equivalent under the null, the finite-sample behaviour of the two tests might differ significantly. Mathematically, the parametric bootstrapping procedure can be formulated as given below;

$$p = \frac{1}{N}\sum_{i=1}^N 1(S_{n,t} \ge S_n) \tag{17}$$

where N = number of simulations.

This fitness statistics actually involve testing of null hypothesis H_0 against the against hypothesis H_a as given below;

Null hypothesis $(H_0) = C \in C_0$ {where, $C_0 = C_\theta$; $\theta \in O$).

Alternate hypothesis $(H_a) = C \notin C_0$.

where, O is the open subset of $\Re^q$ for some integer value q. On the other side, the test statistics "Rn" (i.e., [44]) is also incorporated for testing the adequacy of best-fitted trivariate copulas to flood characteristics. The "Rn" test is an information ratio statistic which is approximately equivalent to the "Tn" test, which is the PIOS (or Pseudo in-and-out-of-sample test). The acceptance or rejection of the considered copulas is based on estimated p-values. The null hypothesis must be accepted if the estimated p-value is larger than a significance level and which in result that copula must be considered as satisfactory performance otherwise will be liable for rejections. Overall, from the Eq 15, it must be conclude that minimum the value of "Sn" and "Rn" test value must indicates for minimum gap or distance between an empirical and derived parametric copulas word thus, most justifiable copula for establishing multivariate (trivariate and bivariate) joint relationship between flood variables.

2.4. Flood risks estimation

The study of the joint and conditional probability distribution for estimating the different notation of return periods (i.e., joint return periods, conditional joint return periods) is often considered as an essential concern for hydrologic design, that can be easily facilitated using the copulas function (i.e., [10,11,45]). Hydrologist and water practioner are mostly interested in the evaluation of the average inter-arrival duration between two design events and which usually defined in a year called the return period [10]. According to Yue and Rassumesen (i.e., [5]), the concurrence probability defines the chance that any hydrologic happening, which either characterizing through univariate or either multivariate exceeding certain a threshold level. Mathematically, the univariate return period that occurs once in a year can be defined from univariate cumulative distribution function or CDF of the variable (say "X") as given below;

$$T_{Univariate} = \frac{\mu}{total\ no.of\ flood\ per\ year} = \frac{1}{P(X \geq x)} = \frac{1}{(1-F(x))} = \frac{1}{1-CDF(x)} \qquad (18)$$

where, $T_{Univariate}$ is return period in years; F(x) is univariate CDF of random variable, X; $\mu = 1$, for annually basis or annual maxima-based flood analysis [5].

2.4.1. Derivation of joint return periods

According to Salvadori (i.e., [10]) and Zhang and Singh (i.e., [8]), the joint return periods of triplet flood characteristics can be estimated using the inclusive probability, also called "OR" and "AND" cases. The joint probability distributions for annual flood analysis can describe the following two situation such that in the first condition when all the flood variables (say, $P \geq p$, $V \geq v$, and $D \geq d$) simultaneously exceed certain threshold during a flood events and their associated return period called AND joint period and it can be written as;

A. For the trivariate joint distribution case;

$$T_{P,V,D}^{AND}(p,v,d) = \frac{1}{P\,(P\geq p \wedge V\geq v \wedge D\geq d)} = \frac{1}{(1-F(p)-F(v)-F(d)+H(p,v)+H(v,d)+H(p,d)-H(p,v,d))} =$$

$$\frac{1}{(1-F(p)-F(v)-F(d)+C(F(p),F(v))+C(F(v),F(d))+C(F(p),F(d))-C(F(p),F(v),F(d)))} \qquad (19)$$

B. For the bivariate distribution case (any flood combinations i.e., between P and V);

$$T_{P,V}^{AND}(p,v) = \frac{1}{P\,(P\geq p\ AND\ V\geq v)} = \frac{1}{(1-F(p)-F(v)+H(p,v))} = \frac{1}{(1-F(p)-F(v)+C(F(p),F(v)))} \qquad (20)$$

where $H(p,v,d)$ = trivariate joint CDF of random variable P, V and D; $H(p,v)$ = bivariate joint CDF of flood random variables; $C(F(p),F(v),F(d))$ = trivariate copulas CDFs for flood characteristics; $F(p) = F(v) = F(d)$ = univariate marginal distribution of flood variables.

In the second situation, probability either the first or second or third flood variable (say, $P \geq p$, $V \geq v$, and $D \geq d$) exceed given threshold and thus their associated return period called OR joint return period can be expressed as;

C. For trivariate case;

$$T_{P,V,D}^{OR}(p,v,d) = \frac{1}{P\,(P\geq p \vee V\geq v \vee D\geq d)} = \frac{1}{(1-H(p,v,d))} = \frac{1}{(1-C(F(p),F(v),F(d)))} \qquad (21)$$

D. For bivariate case (for any combination i.e., between P and V);

$$T_{P,V}^{OR} = \frac{1}{P\,(P\geq p \vee V\geq v)} = \frac{1}{(1-H(p,v))} = \frac{1}{(1-C(F(p),F(v)))} \qquad (22)$$

2.4.2. Derivation of return periods from conditional distribution

Besides the necessity of joint return periods, it could be an essential concern to investigate flood events in such a manner that one could highlights the priority of one design variables over another design variables therefore, from this prospects numerous demonstration focused towards defining the concept of the conditional distributional framework in order to derive the conditional return periods (i.e., [3,7,8,31,32]). For example, the conditional return period of flood peak series given various percentile value of flood volume or vice-versa or in another words, where the flood peak "P" exceeds a threshold "p" given that the volume "V" series exceeds a threshold "v". The conditional distributions based on the different conditions are firstly estimated thereafter the associated conditional return periods are derived.

A. For trivariate case,

The conditional distribution of peak (P), volume (V) given duration ($D \leq d$) in "OR" case is given by

$$F_{P,V,D}(p,v\backslash D \leq d) = P(P \leq p, V \leq v\backslash D \leq d) = \frac{H(p,v,d)}{F(d)} = \frac{C(p,v,d)}{F(d)} \qquad (23)$$

where, F(d) = univariate marginal CDF of flood variable, D. therefore under this condition, their corresponding return period can be estimated as,

$$T_{P,V\backslash D}(p, v\backslash D \leq d) = \frac{1}{1 - F_{P,V,D}(p,v\backslash D \leq d)} = \frac{1}{1 - \frac{C(p,v,d)}{F(d)}} \tag{24}$$

Similarly, the conditional return period of peak (P), duration (D) given volume (V ≤ v) in "OR" case is given by;

$$T_{P,D\backslash V}(p, d\backslash V \leq v) = \frac{1}{1 - F_{P,D,V}(p,d\backslash V \leq v)} = \frac{1}{1 - \frac{C(p,v,d)}{F(v)}} \tag{25}$$

Similarly, the conditional return period of Volume (V), duration (D) given peak (P ≤ p) in "OR" case is given by;

$$T_{V,D\backslash P}(v, d\backslash P \leq p) = \frac{1}{1 - F_{V,D,P}(v,d\backslash P \leq p)} = \frac{1}{1 - \frac{C(p,v,d)}{F(p)}} \tag{26}$$

Again, the conditional distribution of peak (P) given (volume(V ≤ v), duration(D ≤ d)) is given by,

$$F_{P\backslash V,D}(p\backslash V \leq v, D \leq d) = P(P \leq p \backslash V \leq v, D \leq d) = \frac{H(p,v,d)}{H(d,v)} = \frac{C(p,v,d)}{C(d,v)} \tag{27}$$

The corresponding return period can be estimated as;

$$T_{P\backslash DV}(p\backslash V \leq v, D \leq d) = \frac{1}{1 - F_{P,V,D}(p\backslash V \leq v, D \leq d)} = \frac{1}{1 - \frac{C(p,v,d)}{c(vd)}} \tag{28}$$

where, $C(d, v)$ = bivariate copula CDF of flood characteristics duration(D) and volume (V). Therefore, using Eq 27, it can be possible to estimate trivariate conditional return period for various possible combinations of flood characteristics.

B.   For bivariate distribution case;

The conditional return periods between flood peak (P) given volume (V ≤ v) (or vice-versa) can be obtained from the conditional probability distribution function is given by;

$$F(p\backslash V \leq v) = \frac{P(P \leq p, V \leq v)}{P(V \leq v)} = \frac{H_{P,V}(p,v)}{F(v)} = \frac{C(p,v)}{F(v)} \tag{29}$$

$$T_{(P\backslash V)}(p\backslash v) = T_{(p\backslash V \leq v)} = \frac{1}{1 - F(p\backslash V \leq v)} = \frac{F(v)}{F(v) - C(F(p), F(v))} \tag{30}$$

Overall, using Eq 29 we can easily estimate return periods of one variable conditioning to another variable for any possible combination of flood characteristics.

3. Case study

3.1. Trivariate flood characteristics of Kelantan River basin

To illustrate the trivariate distribution analysis of flood episodes, the 50 years (1961–2016) of daily streamflow discharge records of the Kelantan River basin at Gulliemard Bridge gauge station in Malaysia (which are collected from the Drainage and Irrigation Department, Malaysia) are employed. The Gulliemard bridge station is located at the downstream of Kelantan river near the Kuala Kari region. The geographical location of this river basin is Lat 4°30′ N to 6°15′ N and Long 101°E to 101°E to 102°45′ E. It is the longest river of Kelantan state, which originating from the Tahan mountain range to the South China Sea in the north-eastern part of Peninsular Malaysia. The river is about 248 km long with a drain area of 13100 km^2 and which occupying more than 85% of the state of Kelantan. The estimated runoff is about 500 m^{3}sec^{-1} and the variations of annual precipitations for this region in between 0 mm (dry period)–1750mm (wet or north-eastern monsoonal period) [27]. The major land use of this area is agriculture (i.e., paddy, rubber and oil palm) for midstream and downstream and forest for the upstream (i.e. near to Gua Musang).

In this study, we adopted the Annual (Maximum) series or AM approach, also called block (annual) maxima to delineate the triplet flood vectors i.e., flood peak discharge flow (P), volume (V) and duration (D) from the daily basis stream flow discharge records [9,37]. The characterizations of flood peak flow values are based on their maximum streamflow discharge records at an annual scale using Eq 31, such that at the targeted site there is only one flood episodes at each year (refer to Figure 1) [4,5,37]. Figure 1, illustrating the single-peaked flood hydrograph where both the flood duration (D), which are estimated by recognizing the time of rise and fall of the flood hydrograph (i.e., points at Qis and Qie in the Figure 1) and volume (V) series are obtained using algorithm which is reported in the literature (i.e., [4,5]) (see Eqs 32 and 33). Flood peak discharge often attains their maximum value but not mandatory for hydrograph volume and duration series [37].

Mathematically,

$$P_i = \max\{Q_{ij}, j = SD_i + SD_i + 1, \ldots \ldots, ED_i\} = \text{Annual flood peak series} \qquad (31)$$

$$\text{Volume} = V_i = V_i^{total} - V_i^{Baseflow} = \sum_{j=SD_i}^{ED} Q_{ij} - \frac{(1+D_i)(Q_{is}+Q_{ie})}{2} \qquad (32)$$

$$\text{Duration} = D_i = ED_i - SD_i \qquad (33)$$

where, Q_{ij} = jth days streamflow magnitude for the ith year; and Q_{is} & Q_{ie} = streamflow magnitude for the start date "SD_i" and end date "ED_i"of the flood runoff.
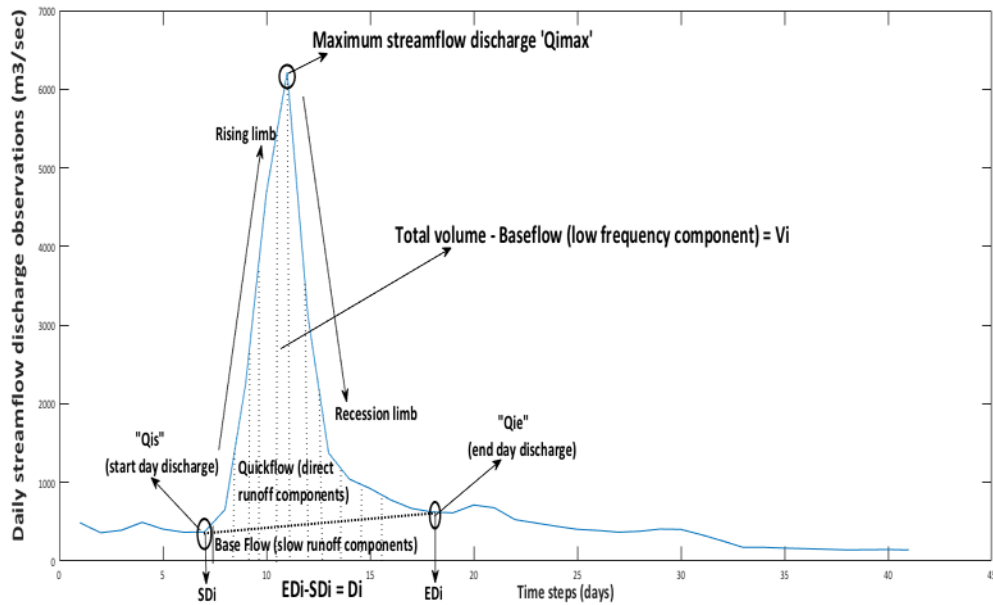
Figure 1. A typical hydrograph showing the flood characteristic.

3.2. *Descriptive behaviour and relationship between hydrological characteristics*

In this research methodology, the flood events are characterized based on annual maximum discharge series method which is also called partial data series based distribution modelling. Table 1 representing the descriptive behaviour of individual flood characteristics which indicating that each flood characteristics exhibiting positively skewed distribution. Figure 2a,b representing the histogram distributions plot and time-series visualization of the annual flood characteristics.

Table 1. Basic descriptive statistics of the annual flood characteristics.

Descriptive statistics	P (m^3/sec)	V (m^3)	D (days)
Sample Size	50	50	50
Range	19670	71558	57
Mean	6078	19122	19.04
Variance	21,520,084	213,845,800	117.75
Std. Deviation	4639	14623	10.851
Coef. of Variation	0.76324	0.76473	0.56993
Std. Error	656.05	2068.1	1.5346
Skewness (Pearson)	1.506	1.590	2.210
Kurtosis (Pearson)	1.883	2.864	6.252
Min	916.3	3182.3	7
50% Percentile (Median)	4961	15959	16
Max	20586	74740	64
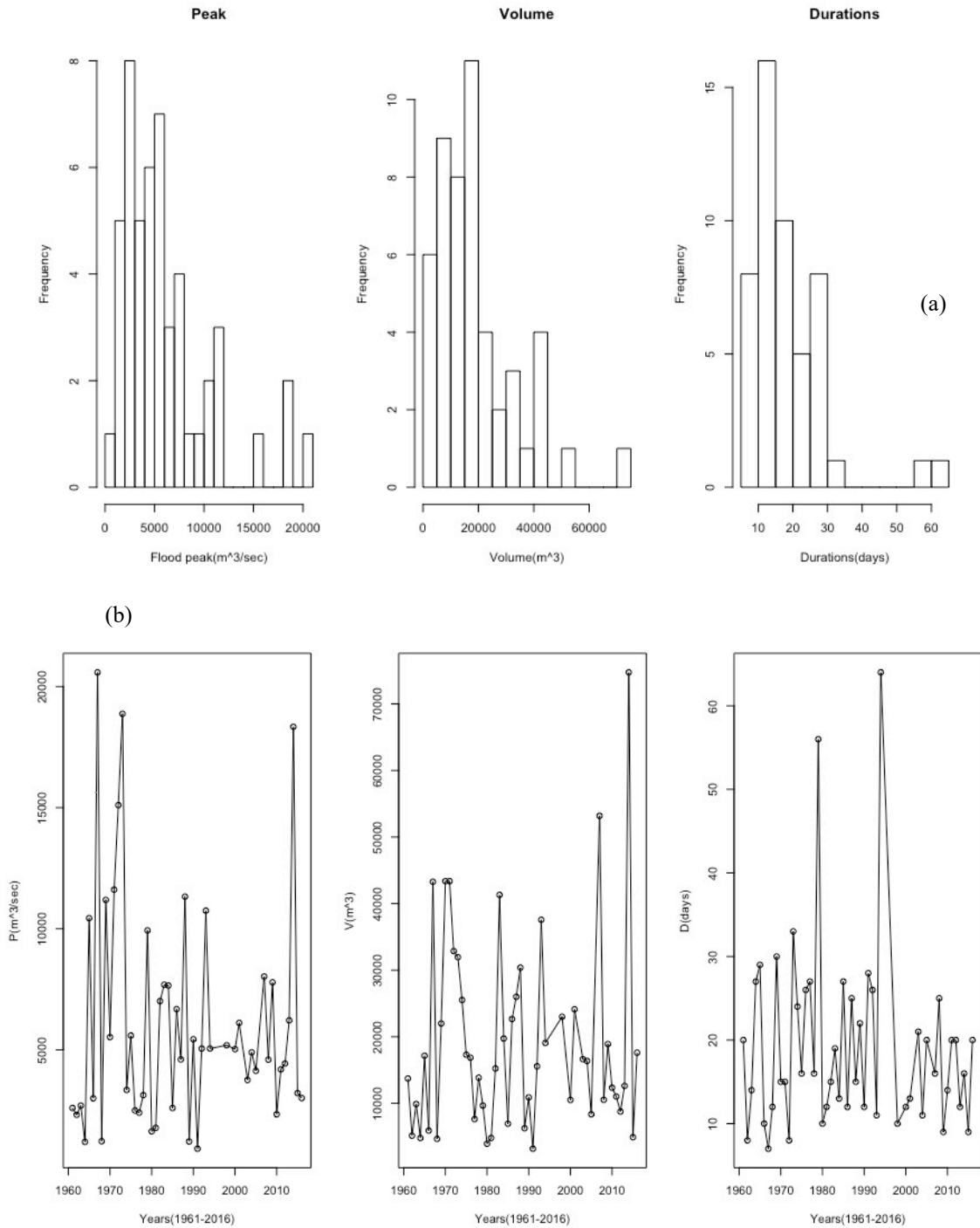
Figure 2. Visualizing the annual flood characteristics of the Kelantan River Basin at Gulliemard Bridge station between the year 1960–2016 in the context of (a) histogram distribution plot (b) Time series plot.

3.2.1. Dependency measures via analytical approach

The strength of dependency between the targeted flood vectors i.e., flood peak, volume and duration series are estimated using the Pearson's linear correlation (r), and the two non-parametric

dependence measures, also called the rank based correlations statistics such as Kendall's tau (t) and Spearman's rho (ρ) and their estimated values are listed in Table 2. The Pearson coefficient only capture the linear dependencies therefore might be incompatible for heavy-tailed distribution series. On other side, Kendall's tau (t) and Spearman's rho (ρ) are invariant under monotonic non-linear transformations without any assumption of underlying distribution structure which frequently used as effective dependence measures for the nonlinear modeling in multivariate statistics also, it possess high resistance to outliers [39].

Table 2. Correlation matrix of analysed flood attribute pairs.

Dependence measure	Peak-Volume (P-V)	Volume-Duration (V-D)	Peak-Duration (P-D)
Pearson's correlation (r)	0.7387784	−0.1079511	−0.0061526
Kendall's correlation(τ)	0.60759499	−0.0225141	−0.0741828
Spearman's correlation (ρ)	0.79425677	−0.0343127	−0.094851

3.2.2. Via graphical investigation

Graphical based dependency investigation among the flood characteristics are also undertaken using the scatter plots, chi plots (i.e., [46]) and Kendall's plots (i.e., [47]), as illustrated in the Figures 3–5. Chi-plot is actually a scatter pot of the pairs $(\lambda_i \chi_i)$, where it uses the data ranks and λ_i values is a measure of the distance of bivariate random observations (say $p_i v_i$) from the center of the data sets within the range of [−1(negatively correlated), 1(positively correlated)]. Also, the control limits χ_i are the another measuring factor in chi-plot that are placed at $\chi = \pm c_p/\sqrt{n}$ [46]. Thus, in case of stronger dependency the random pairs must be outside the control limit of chi-plot otherwise, inside the control limit region can be indicated for independence between random pairs. On other side, when the point data are largely on the upper side of control limits, which could indicate for positively correlated variables and for negatively correlated random variables, the data points are distributed in the lower side of the control limits. Similarly, the Kendall's plot are analogous to quantile-quantile (Q-Q) plot such that, deviation of random pairs from the main diagonal of K-plot is the indication of inter-dependence otherwise could be revealing for independence when the pot tends to be linear [21,47].
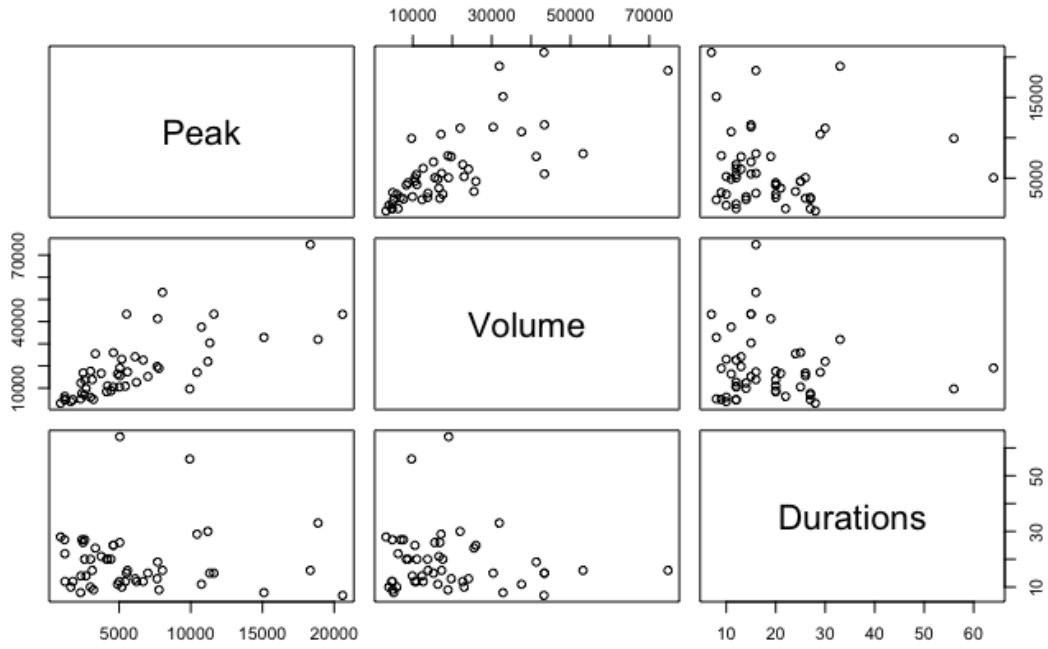
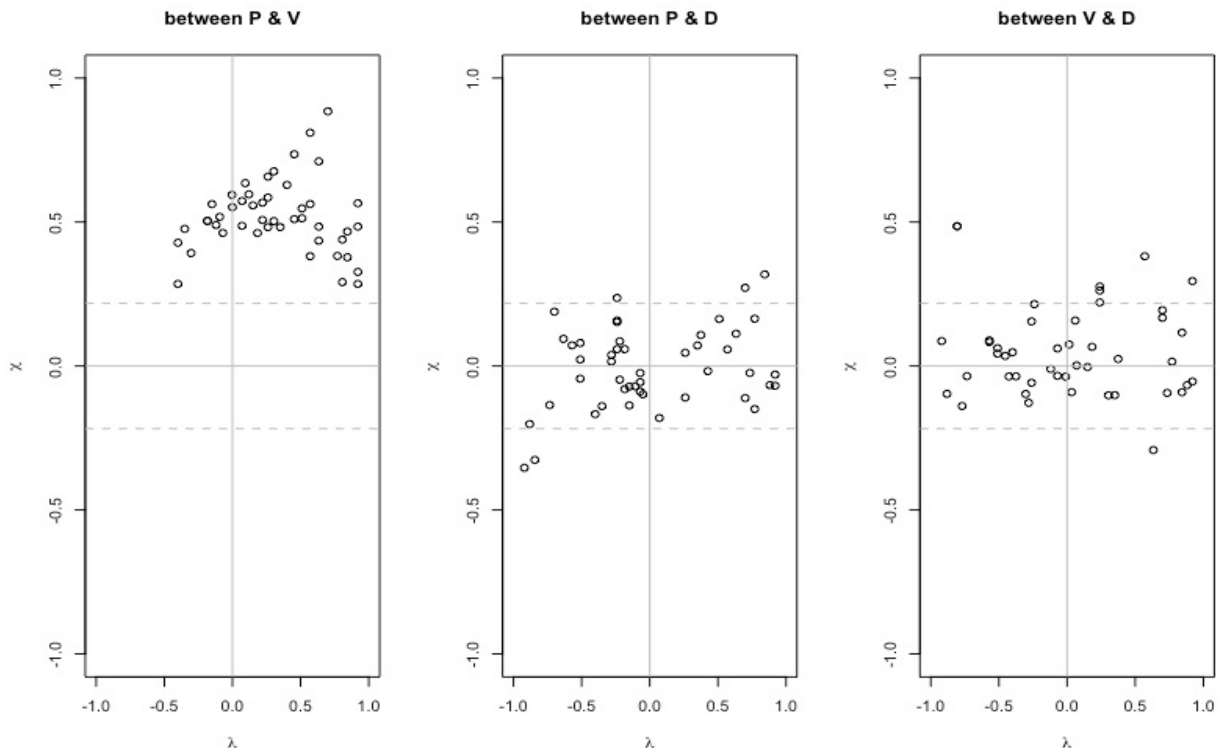**Figure 3.** Scatterplot of multidimensional flood characteristics.

Figure 4. Graphical interpretation to investigate strength of dependency among flood characteristics using Chi-plot between P-V, P-D and V-D.
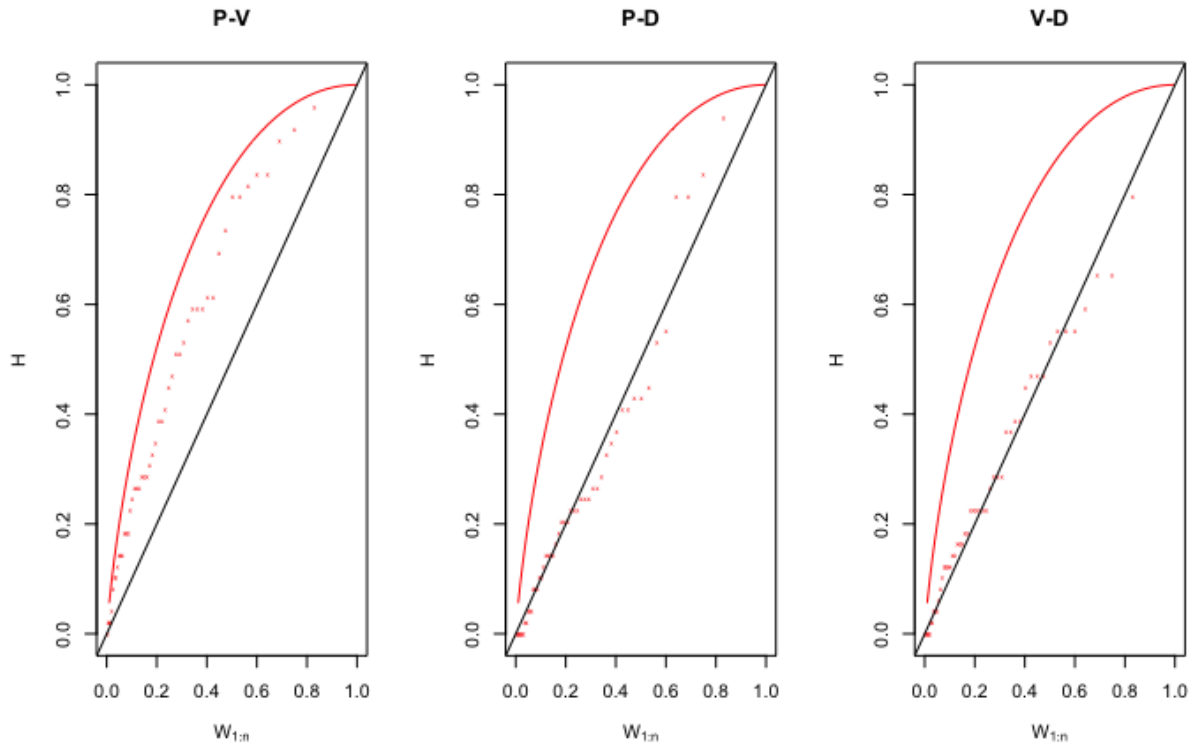
Figure 5. Kendall's plot (or K-plot) of flood characteristics i.e., between P-V (shows high and positive correlation structure), P-D (shows negatively correlated random pairs with weak dependency exhibited), V-D (negatively correlated random pairs).

3.3. *Estimating marginal distribution of flood characteristics*

3.3.1. Empirical probabilities

The empirical nonexceedance probabilities are estimated for the each individual flood characteristics by using commonly used the Gringorten based position-plotting formula [7,48], which usually compared with CDF of the fitted distributions for pointing the gaps and deviations between empirical and fitted samples, as given below,

$$\text{Empirical Cumulative frequency} = \text{P}(K \le k) = (k - 0.44)/(N + 0.12) \tag{34}$$

where N = length of the sample (i.e., the total number of flood observations); and $k = k^{th}$ smallest observations where the dataset is arranged in an ascending order.

3.3.2. Univariate flood marginal distributions

Selecting the most justifiable univariate probability distribution functions for defining flood marginal distribution is often a mandatory pre-requisite demands before the establishment of flood dependence structure. Several models often would fit the data equally well but, each would give different estimates of a given quantile especially, in the tails of the distribution and which is solely

based on the goodness-of-fit procedure to visualize the compatibility of the fitted distributions [49]. A distinguish varieties of univariate parametric family-based probability distribution are selected and introduced as a possible marginal distribution. The parameter of each distributions is first estimated using maximum likelihood estimation (MLE) (i.e., [50]), method of moments (MOM) (i.e., [1]), least square method (LS), and L statistics-based method of L-moments (i.e., [51]) and then the best fitted distributions are selected for each individual flood characteristics using different goodness-of-fit test statistics. All the univariate distribution fitting procedures are carried out using the Easyfit-distribution fitting software.

4. Results and discussions

4.1. Modeling of univariate marginal distribution

A distinct variety of univariate parametric families-based probability distributions (i.e., 1-parameter, 2-parameters, 3-parameters & 4-parameters) are introduced as a candidate models as listed in the Table 3 and their estimated parameters values are listed in the Table 4. Fitness level of each fitted distributions are examined through different analytical based goodness-of-fit measures such as based on distance criteria statistics Kolmogorov-Smirnov (or K-S) test and Anderson-Darling (or A-D) test (i.e., [37,52]), based on information criteria statistics such as Akaike Information criteria (or AIC) (i.e., [53]), Schwartz's Bayesian Information criteria (or BIC) (i.e., [54]) and Hannan-Quinn Information criteria (HQIC) (i.e., [55]), and also based on error indices statistics such as Mean square error (or MSE) and Root mean square error (or RMSE) (i.e., [56]). Table 5a–c listed the performance level of different univariate distributions for fitting the marginal distribution for the flood characteristics. Investigation results reveals that the Lognormal-2P distribution are much satisfactory for flood peak flow series, the Johnson SB (4P) for volume and the Gamma(3P) distribution for duration series because these distribution possess the minimum values of K-S, A-D, AIC, BIC HQC, MSE and RMSE test statistics in compare with their peer candidates function for each individual flood characteristics.

4.2. Bivariate modeling using the 2-dimensional copulas

Before initiating the fitting procedure of two-dimensional copulas function for establishing bivariate joint relationship among the flood characteristics, we investigated the level of dependency through both analytical and graphical procedure. The Pearson's linear correlation (r), Kendall's tau (t) and Spearman's rho (ρ) are used to measure the strength of dependency (see Table 2). Analytical investigation reveals that flood peak and volume pair exhibited strong positive correlation but the correlation structure between flood peak-duration, and flood volume-duration pair are very weak and negatively correlated. On the otherside, the graphical illustrations i.e., based on scatter plot (see, Figure 3), chi-plots (see, Figure 4) and Kendall's plots (see, Figure 5) are also in support of the analytical approach. Based on scatter plot, it clearly indicating the existence of positive and strong dependency between peak-volume pairs because the increased density of points are located near the diagonal region (i.e., close to 45° angle) but weak and negative dependencies are exhibited between

flood volume-duration and flood peak-duration pairs. Similarly, based on chi-plot, strong deviation from the control limit is observed for flood peak-volume pairs (indicates for high and positive correlation) but most of the data samples are within the region of control limit for peak-duration and volume-duration pairs. Similarly based on the Kendall's plot, peak-volume data pairs are much deviated from the main diagonal (high and positive correlation) but much closer to main diagonal for peak-duration and volume-duration pairs (low and negative correlation).

Referred to Table 6, the mono-parametric Archimedean copulas such as the Clayton copula, Gumbel copula, Frank copula, and Joe copula and also, one Elliptical copula such as the Gaussian or normal copula are introduced and tested for establishing bivariate joint distribution of flood characteristics. Actually, the Gumbel-Hougaard, Clayton and Joe copula cannot be used for negatively dependent flood characteristics (i.e., only applicable to model positively correlated random variables). The copulas dependence parameter are estimated using maximum pseudo log-likelihood (or MPL) estimation procedure, using Eqs 12 and 13 and their estimated values are listed in Table 7. Identification and selection of most parsimonious copulas for each flood attribute pairs are performed using the Cramer-von Mises distance statistics with parametric bootstrap procedure, using Eq 14. The test statistics "S_n" and its associated p-value have been computed from 1000 and 500 simulated random samples by the mean of parametric bootstrap procedure and their values are listed in the Table 6. Investigation reveals that the Gaussian copula exhibited minimum "S_n" statistics and highest p-value for flood peak-volume pair and thus identified as most appropriate for this pair. On other side, the Frank copula is identified as the most justifiable bivariate model for capturing the joint structure of both flood peak-duration and volume-duration pairs, referred to the same Table 7. Figures 6–8 represents the joint probability density function (JPDF) and joint cumulative distribution function (JCDF) (i.e., scatterplot and surface plot) derived from the best-fitted bivariate copulas for flood peak-volume, volume-duration and peak-duration pairs.

Table 3. The probability density functions (PDF) and vector of unknown statistical parameters of different univariate functions.

Parametric distribution functions	Probability density function (PDF)	Remarks
Frechet (2P)	$f(x) = \dfrac{\alpha}{\beta}\left(\dfrac{\beta}{x}\right)^{\alpha+1} e^{-\left(\frac{\beta}{x}\right)^{\alpha}}$	$\alpha > 0$ (shape), $\beta > 0$(scale), such that, $\gamma \equiv 0$ yield 2-parameter Frechet functions
Gamma (2P) & (3P)	$f(x) = \dfrac{(x-\gamma)^{\alpha-1}}{\beta^{\alpha}\Gamma(\alpha)} e^{\frac{-(x-\gamma)}{\beta}} \,\&\, f(x)$ $= \dfrac{x^{\alpha-1}}{\beta^{\alpha}\Gamma(\alpha)} e^{\frac{-x}{\beta}}$	$\alpha > 0, \beta > 0, \gamma > 0$ —shape, scale and locations parameter such that $\gamma \equiv 0$ yield 2-parameter gamma structure
GEV(3P)	$f(x) = \dfrac{1}{\sigma} e^{-(1+kz)^{-1/k}(1+kz)^{-1-1/k}}$ for $k \neq 0$ $\dfrac{1}{\sigma}e^{(-1-e^{(-z)})}$ for $k = 0$	k, σ, μ signifies for shape, scale & their location parameter, such that, $\sigma > 0$ & $z \equiv \dfrac{(x-\mu)}{\sigma}$ Domain: $1 + k\,(x-\mu)/\sigma$ for $k \neq 0$ & $-\infty < x < +\infty$ for $k = 0$
Gen. Gamma (3P)	$f(x) = \dfrac{k(x)^{k\alpha-1}}{\beta^{k\alpha}\Gamma(\alpha)} e^{-(x/\beta)^{k}}$	Domain:$y \leq x < +\infty; k > 0$ & $\alpha > 0$ (shape), $\beta > 0(scale), \gamma > 0(location)$
Inv. Gaussian (2P)	$f(x) = \sqrt{\dfrac{\lambda}{2\pi x^3}}\, e^{-\frac{\lambda(x-\mu)^2}{2\mu^2(x)}}$	$\lambda > 0,\ \mu > 0$ (continuous parameter, γ(location parameter) for $\gamma < x < +\infty$
Johnson SB(4P)	$f(x) = \dfrac{\delta}{\lambda\sqrt{2\pi}z(1-z)} e^{-0.5\left(\gamma+\delta\ln\frac{z}{1-z}\right)^2}$	Domain: $\xi \leq x \leq \xi + \lambda$ $\gamma,\ \delta > 0$ (shape); $\lambda > 0$ (scale); ξ location parameter)
Log-Gamma (2P)	$f(x) = \dfrac{(\ln x)^{\alpha-1}}{x\beta^{\alpha}\Gamma(\alpha)} e^{-\left(\frac{\ln x}{\beta}\right)}$	Domain: $0 < x < +\infty$ $\alpha > 0, \beta > 0$ (shape parameter)
Log-Logistic (2P)	$f(x) = \dfrac{\alpha}{\beta}\left(\dfrac{x}{\beta}\right)^{\alpha-1}\left(1+\left(\dfrac{x}{\beta}\right)^{\alpha}\right)^{-2}$	Domain: $\alpha > 0$ (shape); $\beta > 0$(scale)
Lognormal (3P) & (2P)	$f(x) = \dfrac{e^{-0.5\left(\frac{\ln(x-\gamma)-\mu}{\sigma}\right)^2}}{(x-\gamma)\sigma\sqrt{2\pi}} \,\&\, f(x)$ $= \dfrac{e^{-0.5\left(\frac{\ln(x)-\mu}{\sigma}\right)^2}}{(x)\sigma\sqrt{2\pi}}$	$\gamma < x < +\infty$; $\sigma > 0$ (shape parameter); γ (location parameter); μ (scale parameter)
Weibull (2P)	$f(x) = \dfrac{\alpha}{\beta}\left(\dfrac{x}{\beta}\right)^{\alpha-1} e^{-\left(\frac{x}{\beta}\right)^{\alpha}}$	Domain: $\alpha > 0$ (shape), $\beta > 0$ (scale)

Table 4. Estimated parameters of fitted univariate probability distributions.

Parametric Functions	Flood Peak (P)	Flood Volume (V)	Flood Durations (D)
Frechet (2P)	a = 1.576, b = 3207.5	a = 1.5703, b = 10017.0	a = 2.6001, b = 13.304
Gamma (2P)	a = 1.7166, b = 3540.6	a = 1.71, b = 11183.0	a = 3.0786, b = 6.1845
Gamma(3P)	a = 1.2106, b = 4290, g = 884.47	a = 1.0848, b = 14723.0, g = 3150.8	a = 1.4696, b = 8.3319, g = 6.7958
GEV(3P)	k = 0.22596, s = 2683.6, m = 3765.6	k = 0.20446, s = 8736.0, m = 11890.0	k = 0.20682, s = 6.0766, m = 13.987
Log-Gamma(2P)	a = 129.15, b = 0.06544	a = 164.32, b = 0.05839	a = 35.165, b = 0.08037
Log-Logistic (2P)	a = 2.2801, b = 4541.7	a = 2.2731, b = 14202.0	a = 3.6928, b = 16.426
Log-Normal (2P)	s = 0.7362, m = 8.4513	s = 0.74093, m = 9.5943	s = 0.47178, m = 2.826
Log-Normal (3P)	s = 0.75437, m = 8.4267, g = 85.951	s = 0.8237, m = 9.4858, g = 1115.2	s = 0.69194, m = 2.413, g = 4.8982
Weibull (2P)	a = 1.599, b = 6398.7	a = 1.5993, b = 20008.0	a = 2.5437, b = 20.375
Inverse. Gaussian (2P)	l = 10434.0, m = 6078.0	l = 32699.0, m = 19122.0	l = 58.617, m = 19.04
Johnson SB (4P)	g = 1.5161, d = 0.74495, l = 27319.0, x = 1304.2	g = 2.2027, d = 1.0357, l = 1.3052E+5, x = 961.8	g = 2.5314, d = 0.92215, l = 118.81, x = 8.2791
Gen. Gamma (3P)	k = 1.054, a = 1.8127, b = 3540.6	k = 1.0521, a = 1.8019, b = 11183.0	k = 1.0877, a = 3.4664, b = 6.1845

Table 5a. Fitness measures of univariate distributions Based on K-S and A-D test distance statistics.

(a)	Peak			Volume			Durations		
Functions	p-value	KSn (d-max)	ADn (d-max)	p-value	KSn (d-max)	ADn (d-max)	p-value	KSn (d-max)	ADn (d-max)
Frechet (2P)	0.32428	0.13147	1.0751	0.28744	0.1359	1.1173	0.36268	0.1272	0.58456
GEV(3P)	0.99655	0.05451	0.21667	0.99931	0.04897	0.24945	0.82259	0.086	0.35244
Log-Gamma (2P)	0.97557	0.06486	0.22646	0.95247	0.07004	0.26683	0.85726	0.08255	0.3451
Log-Logistic (2P)	0.96909	0.06655	0.24216	0.88242	0.07982	0.32827	0.73162	0.09416	0.49615
Gamma (2P)	0.81376	0.08684	0.44712	0.94562	0.07126	0.34627	0.54764	0.10968	1.1617
Gamma (3P) *	0.8802	0.08007	0.26953	0.98701	0.06089	0.21109	0.89254	0.07865	0.37708
Log-Normal (2P) *	0.9977	0.05293	0.19412	0.98539	0.06157	0.2338	0.60127	0.10511	0.4602
Log-Normal (3p)	0.99466	0.05638	0.20029	0.93057	0.07365	0.28195	0.79396	0.08867	0.33032
Weibull (2P)	0.81311	0.0869	0.73212	0.89172	0.07875	0.63575	0.23928	0.14235	1.5472
Inv. Gaussian (2P)	0.98175	0.06293	0.38095	0.81919	0.08633	0.48954	0.87056	0.08114	0.60496
Gen.Gamma (3P)	0.66896	0.09944	0.45939	0.89941	0.07782	0.36811	0.28097	0.13672	0.91168
Johnson SB (4P) *	0.84788	0.84788	14.822	0.99811	0.05222	0.17314	0.56249	0.1084	11.874

Notes. K-S test stands for Kolmogorov-Smirnov test; A-D test stands for Anderson-Darling test.

*, indicates that Lognormal (2P), Johnson SB (4P) and Gamma (3P) distribution exhibited minimum test statistics i.e., K-S and A-D values for describing flood peak, volume and duration series.

Table 5b. Fitness measures of univariate distributions based on Information criteria statistics such as AIC, BIC & HQC.

(b)	Peak			Volume			Duration		
Functions	AIC	BIC	HQIC	AIC	BIC	HQIC	AIC	BIC	HQIC
Frechet (2P)	−284.118	−280.294	−282.66	−274.569	−270.745	−273.11	−307.04	−303.22	−305.588
GEV(3P)	−374.335	−368.599	−372.15	−268.985	−263.249	−266.8	−336.32	−330.583	−334.135
Log-Gamma (2P)	−370.146	−366.322	−368.69	−359.914	−356.09	−358.46	−340.53	−336.709	−339.077
Log-Logistic (2P)	−360.392	−356.568	−358.94	−294.927	−291.103	−293.47	−321.32	−317.493	−319.861
Gamma (2P)	−335.861	−332.037	−334.4	−360.025	−356.201	−358.57	−260.55	−256.722	−259.089
Gamma (3P) *	−216.301	−210.565	−214.12	−210.107	−204.371	−207.92	−343.62	−337.88	−341.438
Log-Normal (2P) *	−379.344	−375.52	−377.89	−371.028	−367.204	−369.57	−327.46	−323.633	−326.001
Log-Normal (3p)	−285.412	−279.676	−283.23	−352.906	−347.17	−350.72	−340.76	−335.026	−338.578
Weibull (2P)	−329.681	−325.857	−328.23	−342.868	−339.044	−341.41	−292.91	−289.085	−291.453
Inv. Gaussian (2P)	−362.489	−358.665	−361.03	−344.722	−340.898	−343.27	−325.76	−321.938	−324.306
Gen.Gamma (3P)	−321.553	−315.817	−319.37	−338.918	−333.182	−336.73	−290.95	−285.21	−291.856
Johnson SB(4P) *	−340.899	−333.251	−337.99	−381.821	−374.173	−378.91	−223.65	−216.006	−220.742

Notes. AIC stands for Akaike information criteria; BIC stands for Bayesian information criteria; HQIC stands for Hannan-Quinn information criteria.

*, indicates that Lognormal (2P), Johnson SB (4P) and Gamma (3P) distribution exhibited minimum values of AIC, BIC and HQC test statistics for describing flood peak, volume and duration, thus could be further indicated for the better performance.

Table 5c. Fitness measures of univariate distributions based on error indices statistics such as MSE and RMSE.

(c)	Peak		Volume		Duration	
Functions	MSE	RMSE	MSE	RMSE	MSE	RMSE
Frechet (2P)	0.00314	0.05607	0.00380	0.06168	0.00199	0.04458
GEV(3P)	0.00049	0.02229	0.00409	0.06394	0.00106	0.03261
Log-Gamma (2P)	0.00056	0.02372	0.00069	0.02627	0.0010172	0.031894
Log-Logistic (2P)	0.00068	0.02615	0.00253	0.05032	0.00149	0.03865
Gamma (2P)	0.00111	0.03341	0.00068	0.02624	0.005037	0.070973
Gamma (3P)*	0.01173	0.10882	0.01327	0.11520	0.000918	0.030312
Log-Normal (2P)*	0.00046	0.02163	0.00055	0.02351	0.001321	0.03635
Log-Normal (3p)	0.00294	0.05425	0.00076	0.02762	0.000973	0.031191
Weibull (2P)	0.00126	0.03555	0.00097	0.03115	0.002637	0.05135
Inv. Gaussian (2P)	0.00066	0.02561	0.00094	0.03059	0.00137	0.03697
Gen.Gamma (3P)	0.00014	0.03780	0.00101	0.03177	0.00248	0.04977
Johnson SB* (4P)	0.00093	0.03053	0.00041	0.02028	0.00972	0.09861

Notes. MSE stands for Mean Square Error; RMSE stands for Root Mean Square Error.

*, indicates that Lognormal (2P), Johnson SB (4P) and Gamma (3P) distribution exhibited minimum values of MSE and RMSE test statistics for describing flood peak, volume and duration, thus could be further indicated for the better performance.

Table 6. Mathematical expressions for bivariate Archimedean copula families and their associated properties.

Copula family	Bivariate copula $C_\theta(u,v)$	Parameter range (θ)	Generating function (or generator) $\phi(t)$	Relation of Kendall's τ and θ (τ_θ)
Clayton	$\left[max\{u^{-\theta}+v^{-\theta}-1;0\}\right]^{-1/\theta}$	$0 \le \theta < \infty$	$\frac{1}{\theta}(t^{-\theta}-1)$	$\frac{\theta}{\theta+2}$
Frank	$\frac{-1}{\theta}\ln\left(1+\frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{(e^{-\theta}-1)}\right)$	$-\infty < \theta < \infty$	$-\ln\left(\frac{e^{-\theta t}-1}{e^{-\theta}-1}\right)$	$1+4\left(\frac{D_1(-\ln\theta)-1}{\ln\theta}\right)$ where $D_k(x)$ is the Debye function, for any positive integer k, $D_K(x)=\frac{k}{x^k}\int_0^x t^k/(e^t-1)\,dt$ (Zhang and Singh 2006 and Wang et al., 2009)
Gumbel-Hougaard	$exp\left\{-[(-\ln(u))^\theta+(-\ln(v))^\theta]^{\frac{1}{\theta}}\right\}$	$1 \le \theta < \infty$	$(-\ln t)^\theta$	$\frac{\theta-1}{\theta}$
Joe	$1-[(1-u)^\theta+(1-v)^\theta$ $\quad -(1-u)^\theta(1-v)^\theta]^{1/\theta}$	$1 \le \theta < \infty$	$-\ln(1-(1-t)^\theta)$	

Table 7. Estimating the value of parameter θ of 2-dimensional copulas function and their corresponding goodness-of-fit statistics for flood characteristics.

For (P-V) pair			N = 1000 (No. of bootstrap sampling)		N = 500 (No. of bootstrap sampling)		
Copula family	Parameter Estimates $\hat{\theta}$	Standard Error SE	Maximized log likelihood	*Sn*	(*p*-value) Sn	*Sn*	(*p*-value) Sn	Kendall's tau ($\tau *$) estimated from fitted copula
Gaussian*	0.8333772	0.052	26.98	0.013444	0.9356	0.013443	0.9411	0.6271915
Clayton	2.600312	0.716	26.57	0.035144	0.1923	0.035144	0.1806	0.5652469
Gumbel-Hougaard (GH)	2.311711	0.331	22.21	0.027751	0.2063	0.027751	0.2605	0.56742
Frank	7.878869	1.829	23.98	0.02383	0.464	0.02383	0.4361	0.5980901
Joe	2.553838	0.372	16.26	0.083346	0.0004995	0.083346	0.002498	0.4572527

Note: Bold letter indicated via * indicates that the Gaussian copula exhibiting minimum S_n value, which means performance for P-V is much consistence that the other copula functions also, ($\tau *$) in the last column of above table indicated the estimated kendall's tau value from derived copulas fitted to observed random series

For (P-D) pair								
Gaussian	−0.1276312	0.052	0.3041	0.032132	0.486	0.032132	0.48	−0.08147478
Clayton	NA	NA	NA	NA	NA	NA	NA	NA
Gumbel-Hougaard (GH)	NA	NA	NA	NA	NA	NA	NA	NA
Frank*	−0.6942	0.777	0.262	0.031215	0.4001	0.031215	0.3762	−0.07676464
Joe	NA	NA	NA	NA	NA	NA	NA	NA

Note: Bold letter indicated via * denotes that the performance of Frank copula is most satisfactory that other copulas. NA denotes that for Gumbel-Hougaard, Clayton and Joe copulas can't be used for negatively dependent data [i.e., only positively correlated random variables can be simulated (i.e., Kendall's tau > 0)].

For (V-D) pair				N = 1000 (No. of bootstrap sampling)		N = 500 (No. of bootstrap sampling)		
Copula family	Parameter Estimates $\hat{\theta}$	Standard Error SE	Maximized log likelihood	Sn	(p-value) Sn	Sn	(p-value) Sn	Kendall's tau ($\tau *$) estimated from fitted copula
Gaussian	−0.05098	0.163	0.0478	0.034466	0.3132	0.034466	0.3224	−0.03246895
Frank*	−0.225	0.86	0.03082	0.032761	0.2922	0.032761	0.3084	−0.02498735
Clayton	NA	NA	NA	NA	NA	NA	NA	NA
Gumbel-Hougaard (GH)	NA	NA	NA	NA	NA	NA	NA	NA
Joe	NA	NA	NA	NA	NA	NA	NA	NA

[Notes: NA denotes that for negatively dependent data the above following copulas can't be used, which is only applicable for positively correlated random variables. Bold letter with * indicates that the performance of Frank copula is much satisfactory than other functions.]
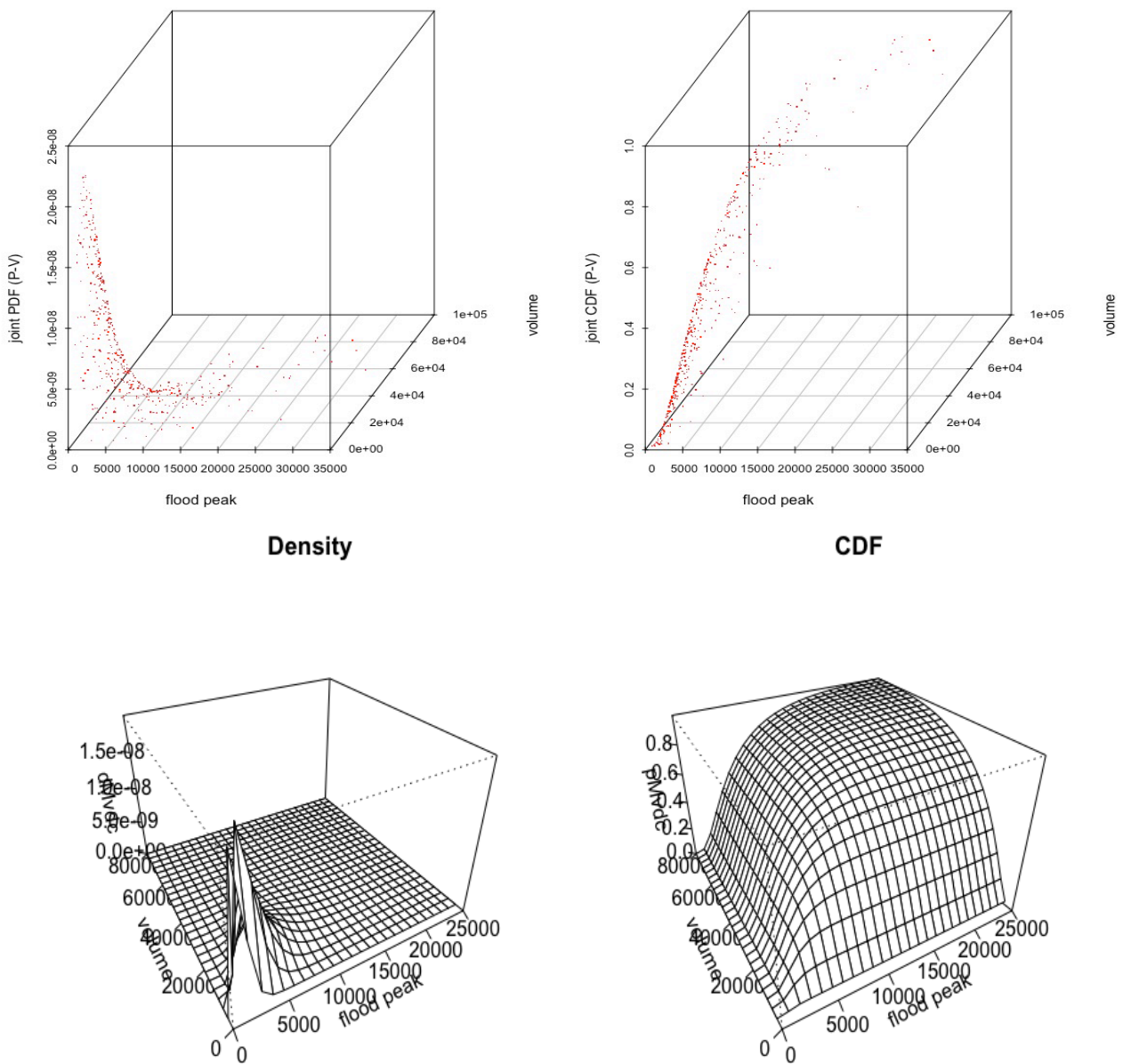
Figure 6. Joint probability density function (JPDF) and joint cumulative distribution function (JCDF) of flood peak-volume using best-fitted two-dimensional Gaussian copula.
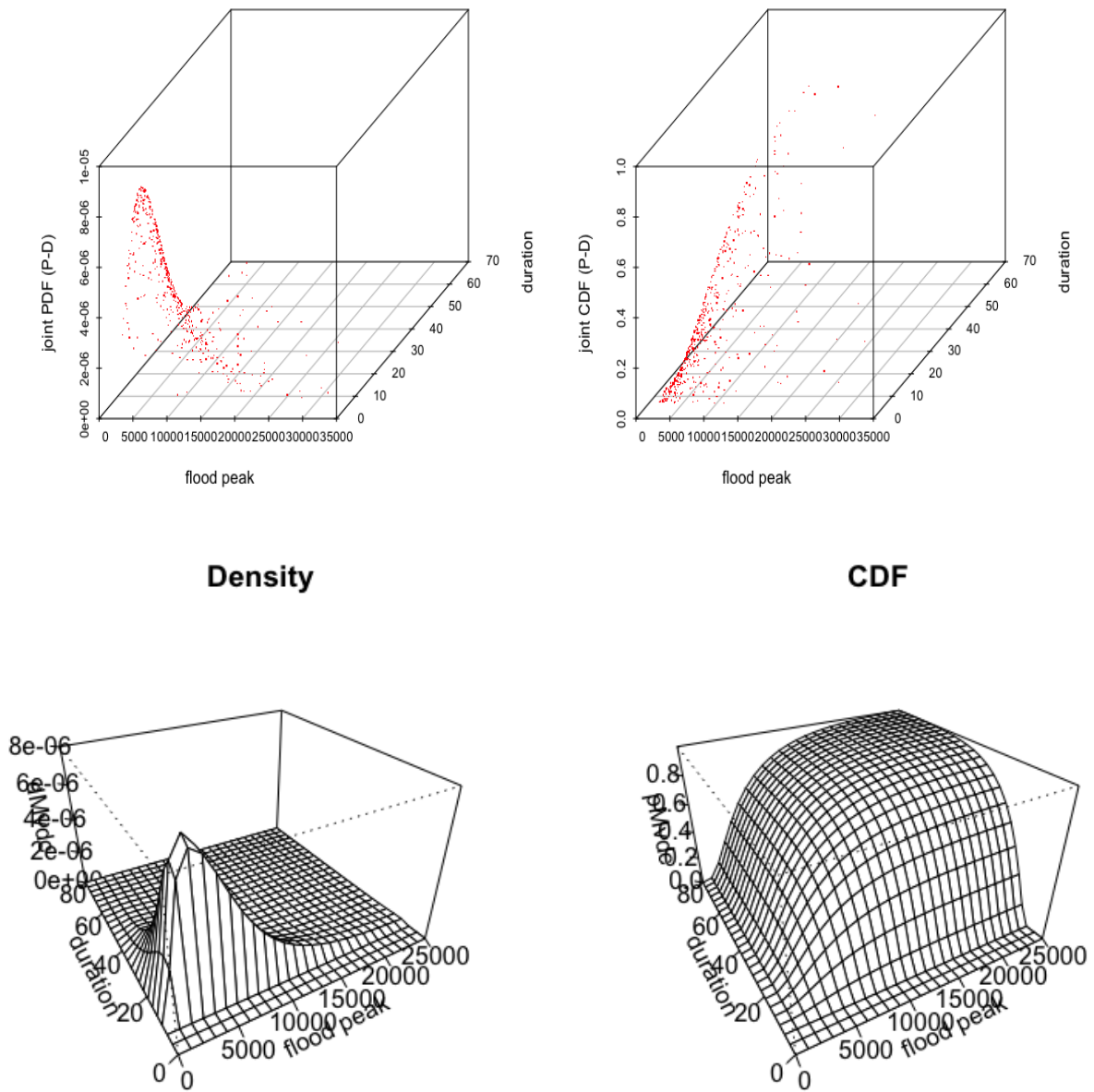
Figure 7. Joint probability density function (JPDF) and joint cumulative distribution function (JCDF) of flood peak-duration using Frank copula distribution.
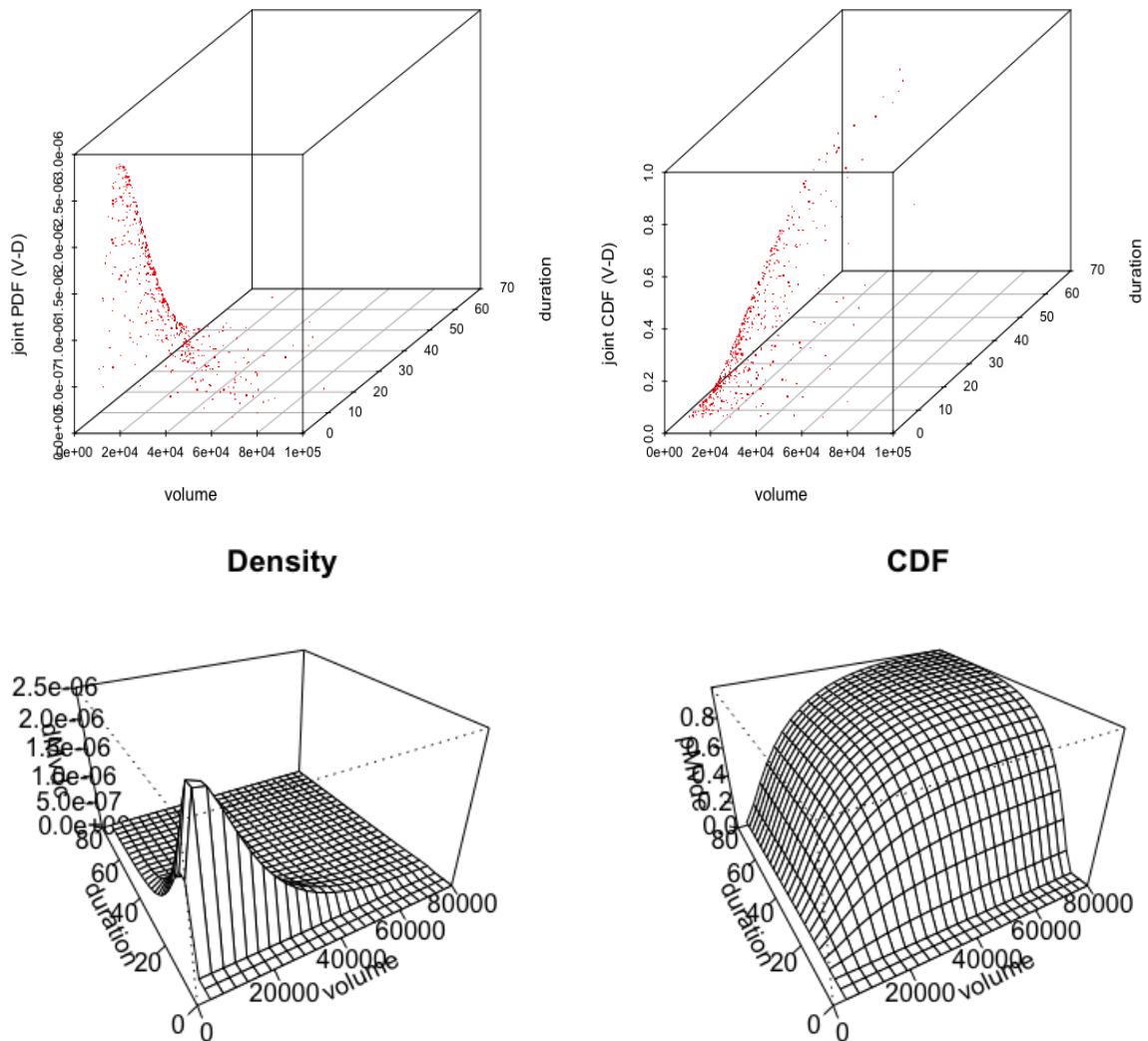
Figure 8. Joint probability density function (JPDF) and joint cumulative distribution function (JCDF) of flood volume and duration series using Frank copula distribution.

4.3. Trivariate modeling using the 3-dimensional copulas

The Archimedean class copula called the Frank copula and the elliptical copula called the Gaussian copula are incorporated (see Eqs 8 and 11) and their adequacy for establishing the trivariate joint distribution among flood characteristics flood peak flow, volume and duration series are investigated. The dependence parameter of trivariate copulas are estimated using the maximum pseudo log-likelihood (or MPL) estimation procedure of Eqs 12 and 13 and their estimated values are listed in the Table 8. To analytically validate and identify the best-fitted copula for describing trivariate joint distribution of flood characteristics, the Cramer-von Mises distance statistics are employed where the approximation of p-values for the test statistics are obtained by means of a faster multiplier bootstrapping approach (i.e., [42,43]) followed by Eq 15. For this purpose both the test statistics "Sn" and its associated p-values and also, the test statistics "Rn" (i.e., [57]) and its associated p-value has been computed from 1000 and 500 simulated random samples by the mean of

faster multiplier approach and their estimated values are listed in Table 8. Result reveals that the Gaussian copula is recognized as most consistence copula for establishing the trivariate joint distribution of flood characteristics, which exhibited minimum "Sn" test statistics (i.e., "Sn" = 0.082819) their p-value (i.e., p-value = 0.01748, for N = 1000 bootstrap samples and p-value = 0.01098, for N = 500 random bootstrap samples) than the Frank copula test statistics. Also, for the Gaussian copula the "Rn" test statistics is less than the Frank copula statistics (i.e., "Rn" = 1.2742) and their p-value (i.e., p-value = 0.1294, for N = 1000 bootstrap samples and p-value = 0.1307, for N = 500 random bootstrap samples) such that based on both test statistics it is concluded that the estimated p-values are exceeding the specified significance level (i.e., $\alpha = 0.05$), and which is higher than the Frank copulas, referred to same Table 8.

4.4. Probabilistic analysis of flood characteristics

4.4.1.   Via joint return periods

In order to analyse the critical hydrologic behaviour of flood episodes for tackling the several basin perspective water-related issues the multivariate frequency analysis is much comprehensive approach. The univariate return periods are derived from the best-fitted CDFs for each flood characteristics i.e., Lognormal-2P distribution for flood peak, Johnson SB-4P distribution for volume and Gamma-3P distribution for duration series using Eq 18 and their estimated values are listed in Table 9. It is already pointed in the section 1 that estimation of univariate return period as a design criterion would be problematic and might attributes for underestimations or overestimations of hydrologic risk. Therefore, the bivariate joint CDFs which are derived from the best-fitted copulas for each flood attribute pairs are employed to derived primary return period for both the "OR" and "AND" case using Eqs 20, 21 and their estimated values are listed in Table 10. It is revealed that the AND-joint cases produce higher return period than the OR-joint cases for different possible combination of flood characteristics i.e., $T^{AND}_{PV} > T^{OR}_{PV}$ or $T^{AND}_{VD} > T^{OR}_{VD}$ or $T^{AND}_{PD} > T^{OR}_{PD}$. In other words, the occurrence of bivariate flood characteristics simultaneously is less frequent in "AND" case in compare with "OR" case of joint return periods (i.e., more frequent). For example, a flood event with peak flow, P = 10463.8 m^3s^{-1}, volume, V = 17148 m^3, and duration, D = 29 days, then OR-joint return period between P-V is, T^{OR}_{PV} = 2.2037 years, between P-D, T^{OR}_{PD} = 3.7524779 years and between V-D pair is, T^{OR}_{VD} = 1.94310711 years. On otherside, joint return for "AND" case for same flood combination is, between P-V, T^{AND}_{PV} = 7.5013958 years, between P-D, T^{AND}_{PD} 66.4233248 years and between V-D pair, T^{AND}_{VD} = 17.1879396 years, (see Table 10). Also, the univariate return periods derived from flood peak, T(P) and volume, T(V) is higher than that derived from their joint distribution for "OR" cases but produces low return periods than "AND" joint cases i.e., $T(P) > T(V) > T^{OR}_{PV}$ and $T(P) < T(V) < T^{AND}_{PV}$. Similarly, univariate return periods derived from peak, T(P) and duration, T(D) as well as volume, T(V) and duration T(V) is higher than the joint return periods for "OR" cases for the same flood attributes but lower return periods for "AND" joint cases i.e., $T(V) > T(D) > T^{OR}_{VD}$ and $T(P) > T(D) > T^{OR}_{PD}$ also, $T(V) < T(D) < T^{AND}_{VD}$ and $T(P) < T(D) < T^{AND}_{PD}$.

Actually, the estimation of trivariate joint and conditional distribution and their associated return periods often required at first to determine bivariate joint copula distribution i.e., $C(p, v), C(v, d)$ or $C(p, d)$ of flood characteristics for various possible combinations (see Eqs 19, 27 & 28). At first, the trivariate return periods for the conditions (1) when all the flood characteristics (say, $P \geq p$, $V \geq v$, and $D \geq d$) simultaneously exceed certain threshold (also called "AND" primary joint return period) and (2) probability either any of the flood variable (say, $P \geq p$, $V \geq v$, and $D \geq d$) exceed given threshold (also called "OR" primary joint return period)) during a flood events are examined using Eqs 19 and 21 and their estimated values are listed in Table 11. For example, the flood event having peak, $P = 10463.8 \ m^3s^{-1}$, volume, $V = 17148 \ m^3$ and duration, $D = 29$ (days), in joint return period for "OR" and "AND" cases are $T^{AND}_{PVD} = 34.8401$ years and $T^{OR}_{PVD} = 1.87605$ years. Similarly, for $P = 18875.4 \ m^3s^{-1}$, $V = 31945.6 \ m^3$ and $D = 33$ (days), $T^{AND}_{PVD} = 547.92$ years and $T^{OR}_{PVD} = 4.12544$ years. It is also clearly revealed from Table 10 that for all the cases considering trivariate flood characteristics (i.e., P, V, D) the joint return periods in "AND" case is greater than "OR" case i.e., $T^{AND}_{PVD} > T^{OR}_{PVD}$. In other words, we can say that the occurrence of trivariate flood characteristics simultaneously is less frequent in "AND" case in compare with "OR" case of joint return periods (i.e., more frequent).

4.4.2. Via conditional joint return periods

The joint return period of two flood characteristics conditional on third flood characteristic, i.e., conditional distribution of peak (P), volume (V) given duration ($D \leq d$), $T(p, v \backslash D \leq d)$, $T(p, d \backslash V \leq v)$ and $T(v, d \backslash P \leq p)$ are estimated using Eqs 24–26 and their estimated values are listed in Table 11. For example, a flood episode characterized with peak flow, $P=10463.8 \ m^3s^{-1}$, volume, $V= 17148 \ m^3$ and duration, $D = 29$ (days), using Eq 24–26, then joint return period of, "P" and "V" conditional to "D" is $T(p, v \backslash D \leq d) = 2.19874$ years, $T(p, d \backslash V \leq v) =5.50286$ years and $T(p, d \backslash V \leq v) = 2.1822555$ years. Similarly, for $P = 20586.4 \ m^3s^{-1}$, $V = 43273.2 \ m^3$, $D = 7$ (days) the conditional return periods are $T(p, v \backslash D \leq d) = 83.650777$ years, $T(p, d \backslash V \leq v) = 1.0034823$ years and $T(v, d \backslash P \leq p) = 1.0032946$ years. On the other side, the joint return periods of one flood characteristics conditional on other two flood characteristics i.e., $T_{P \backslash DV}(p \backslash V \leq v, D \leq d)$, $T_{V \backslash PD}(v \backslash D \leq d, P \leq p)$, $T_{D \backslash PV}(d \backslash V \leq v, P \leq p)$ and are estimated using Eq 27. For example, a flood event with $P = 10463.8 \ m^3s^{-1}$, $V = 17148 \ m^3$ and $D =29$ (days), then the conditional return period of peak (P) given (volume($V \leq v$), duration($D \leq d$)) is $T_{P \backslash DV}(p \backslash V \leq v, D \leq d) = 26.386055$ years, $T_{V \backslash PD}(v \backslash D \leq d, P \leq p) = 2.7519296$ years and $T_{D \backslash PV}(d \backslash V \leq v, P \leq p) = 5.718863$ years. Similarly, for the flood events ($P = 18875.4 \ m^3s^{-1}$, $V = 31945.6 \ m^3$ and $D = 33$ (days)), $T_{P \backslash DV}(p \backslash V \leq v, D \leq d) = 404.30955$, years, $T_{V \backslash PD}(v \backslash D \leq d, P \leq p) = 7.25924$ years and $T_{D \backslash PV}(d \backslash V \leq v, P \leq p) = 10.4372045$ years. Similarly, for the flood episode ($P = 4603 \ m^3s^{-1}$, $V=25999 \ m^3$ and $D = 25$ (days)), $T_{P \backslash DV}(p \backslash V \leq v, D \leq d) = 2.44463$, years, $T_{V \backslash PD}(v \backslash D \leq d, P \leq p) = 19.3565$ years and $T_{D \backslash PV}(d \backslash V \leq v, P \leq p) = 3.7111744$ years.

Table 8. Values of parameter "θ" of 3-dimensional copulas function and their corresponding goodness-of-fit statistics.

Copula family	Parameter Estimates $\hat{\theta}$	Standard Error SE	N = 1000 (No. of bootstrap sampling)		N = 500 (No. of bootstrap sampling)		N = 1000 (No. of bootstrap sampling)		N = 500 (No. of bootstrap sampling)	
			Rn	p-value	Rn	p-value	Sn	p-value	Sn	p-value
Gaussian	0.2595	0.067	1.2742	0.1294	1.2743	0.1307	0.082819	0.01748	0.082831	0.01098
Frank	1.347	0.464	2.3196	0.1384	2.3196	0.1427	0.10173	0.003497	0.10173	0.01098

Table 9. Univariate return period derived from best-fitted marginal distribution function of flood characteristics.

P (m^3s^{-1})	V (m^3)	D (days)	T(P)	T(V)	T(D)
2597	13729.8	20	1.26865844	1.85501224	2.8085943
10436.8	17148	29	7.24346213	2.32921063	6.96912677
20586.4	43273.2	7	45.2067321	13.394053	1.0032606
11192.4	21994.2	30	8.46033619	3.21967868	7.73634535
18875.4	31945.6	33	34.3443917	6.24648635	10.6134579
15103.7	32864.7	8	17.9245588	6.64098818	1.04288336
11324.5	30381.1	15	8.69006457	5.62841222	1.76270469
8028.4	53185.7	16	4.31295471	26.8384326	1.92808252
5435.5	10887.75	12	2.38328056	1.53997782	1.36798906
7786	18911.1	9	4.0857416	2.6204764	1.10282765

Table 10. Bivariate joint return periods of flood characteristics derived from the joint cumulative function of best-fitted 2-dimensional copula function.

P (m³s⁻¹)	V (m³)	D (days)	T_{PV}^{AND}(years)	T_{PV}^{OR}(years)	T_{VD}^{AND} (years)	T_{VD}^{OR}(years)	T_{PD}^{AND}(years)	T_{PD}^{OR}(years)
2597	13729.8	20	1.8956685	1.2503190	5.39125825	1.40915619	3.74255457	1.14013769
10436.8	17148	29	7.5013958	2.3037387	17.1879396	1.94310711	66.4233248	3.7524779
20586.4	43273.2	7	55.800594	12.680757	13.4420881	1.00299213	45.4055525	1.00316311
11192.4	21994.2	30	9.1821312	3.1261582	26.7236536	2.48490441	87.0822263	4.23773322
2495.4	16867.15	26	2.3047989	1.2388910	12.3189158	1.8119749	6.74531088	1.17517145
18875.4	31945.6	33	36.917886	6.1682820	72.5134819	4.15766772	505.997238	8.2399106
11324.5	30381.1	15	11.323725	4.8915594	10.3391302	1.54259078	17.6820031	1.59787989
10746.3	37576	11	13.632454	6.0245175	11.7799366	1.22726507	10.4717863	1.21380131
11612.5	43375.9	15	19.311904	7.636906	24.91585169	1.663022131	18.76059677	1.606189886

Table 11. Trivariate joint and conditional return periods.

P (m³s⁻¹)	V (m³)	D (days)	T_{PVD}^{OR}(years)	T_{PVD}^{AND}(years)	$T(p,v \backslash D \leq d)$ (years)	$T(p,d \backslash V \leq v)$(years)	$T(v,d \backslash P \leq p)$ (years)	$T_{P \backslash DV}(p \backslash V \leq v, D \leq d)$ (years)	$T_{V \backslash PD}(v \backslash D \leq d, P \leq p)$ (years)	$T_{D \backslash PV}(d \backslash V \leq v, P \leq p)$ (years)
2597	13729.8	20	1.116254	5.189694921	1.1929357	1.29191472	1.96774417	1.5593051	6.5498644	2.0842240
10436.8	17148	29	1.876052	34.84014218	2.19874894	5.50286046	2.18225548	26.386055	2.7519296	5.7188639
11192.4	21994.2	30	2.328124	34.13868707	2.89985406	5.79627005	2.83235363	22.050297	3.9473251	6.2026968
5052.6	19073.8	64	1.603469	9.263537116	1.60648426	2.52890799	3.28159236	2.5417539	3.3087039	4.4579073
2495.4	16867.15	26	1.145157	10.01044791	1.18705747	1.29087479	2.81989688	1.3944448	6.6836202	2.9185839
18875.4	31945.6	33	4.125447	547.9258783	6.11278641	10.2044853	4.55212964	404.30955	7.2592403	10.437204
3755	16635.4	21	1.2552974	6.916888717	1.42940401	1.57722517	2.13667663	2.2197248	6.431746	2.36817306
3007.3	17604.1	20	1.181136401	7.111431617	1.31259713	1.35648649	2.27223994	1.7095796	22.76157	2.36767201
9929.3	9667.4	56	1.372616776	40.52366099	1.37654485	11.1918456	1.47211228	12.247682	1.4777196	11.3114697

Table 12. Bivariate conditional return periods of flood characteristics derived from conditional joint distribution of best-fitted copulas function.

P (m³s⁻¹)	V (m³)	D (days)	T(P/V ≤ v) (years)	T(V/P ≤ p) (years)	T(V/D ≤ d) (years)	T(D/V ≤ v) (years)	T(P/D ≤ d) (years)	T(D/P ≤ p) (years)
2597	13729.8	20	1.76790171	18.3162609	1.82115367	2.70232014	1.23590059	2.38333189
10436.8	17148	29	120.216827	2.9117633	2.3077213	6.68939406	6.96346327	6.71113133
20586.4	43273.2	7	220.337776	17.2346824	12.1815919	1.00323433	33.5532361	1.0032349
11192.4	21994.2	30	74.1984419	4.3722186	3.18753928	7.50664468	8.15947394	7.48706776
2495.4	16867.15	26	1.52155663	54.8218787	2.25774928	4.91627523	1.22745507	4.148129
18875.4	31945.6	33	413.811146	7.29971652	6.19127366	10.4428152	33.3736912	10.525197
11324.5	30381.1	15	30.725477	9.90295038	5.34514704	1.74744446	7.3939753	1.73258314
10746.3	37576	11	15.856349	23.8523655	8.41910424	1.26267802	6.18093928	1.25367329
11612.5	43375.9	15	16.2994709	39.8692511	12.72433967	1.756267201	7.82704575	1.73424829

The bivariate conditional return periods for different possible combination of flood characteristics are also estimated using Eq 30 and their values are listed in the same Table 12. For example, a flood episode characterized with flood peak, P = 10436.8 m^3s^{-1} and volume, V = 17148 m^3 then, the conditional return periods $T(P/V \leq v)$ = 120.216827 years and $T(V/P \leq p)$ = 2.9117633 years. Similarly, for the flood events (P = 20586.4 m^3s^{-1} and D = 7 (days), then the conditional return periods $T(P/D \leq d)$ = 33.5532361 years and $T(D/P \leq p)$ = 1.0032349 years. Similarly, the flood episode which has volume of V = 31945.6 m^3 and duration D = 33 days, the return period of volume given duration or vice versa is $T(V/D \leq d)$ = 6.19127366 years and $T(D/V \leq v)$ = 10.4428152 years. Again, the flood episodes characterized based on peak, P = 18875.4 m^3s^{-1} and duration, D = 33days then conditional return periods, $T(P/D \leq d)$ = 33.3736912 years and $T(D/P \leq p)$ = 10.525197 years.

5. Research conclusions

This literature incorporated the copula-based methodology to establishing the trivariate distribution modelling of the flood episodes for the Kelantan River basin in Malaysia. Firstly, a distinguish varieties of parametric families-based probability functions are tested for defining the univariate marginal structure of each flood characteristics. Results reveals that the Lognormal (2P), Johnson SB (4P) and Gamma (3P) distribution are recognized as most justifiable for describing marginal distribution of flood peak, volume and duration series. Based on the correlation measuring statistics, via the analytical approach such as the Pearson, Kendall's tau and Spearman rho correlation coefficient as well as the graphical visual inspection (i.e., based on ranked based scatter plot, K-plot and Chi-plot). It is found that flood peak flow and volume pair exhibited higher and positive dependence structure but both flood volume and duration pairs as well as peak flow and duration pairs are found to be negatively correlated random pairs with very weak correlation and thus considered for flood frequency analysis. The adequacy of one elliptical copula, the Gaussian copula and one Archimedean copula, the Frank copula are introduced to model the trivariate joint distribution of flood characteristics. The copula dependence parameter of fitted trivariate copulas are estimated using maximum pseudo log-likelihood (or MPL) estimation procedure. The best fitted trivariate copulas are selected using the Cramer-von Mises distance statistics where the approximation of p-values for the test statistics are obtained using the faster multiplier bootstrapping approach. The test statistics "Sn" and "Rn" and their associated p-values are computed from 1000 and 500 simulated random samples by the mean of the faster multiplier approach. Result reveals that the Gaussian copula is recognized as most justifiable copula function for establishing the trivariate flood dependence structure as it exhibited the minimum values for "Sn" and "Rn" test statistics. The estimation of trivariate joint probability distribution often required at first to determine bivariate joint copula distribution. Therefore, the mono-parametric Archimedean copulas such as the Clayton copula, Gumbel copula, Frank copula, Joe copula and one Elliptical copula such as the Gaussian or normal copula are tested for establishing the bivariate joint distribution of the flood characteristics. Investigation reveals that the Gaussian copula is identified as most appropriate for flood peak flow and volume pair and the Frank copula is for volume and duration and peak flow and duration pairs. Finally, the cumulative distribution function or CDF of the best fitted trivariate copula is further

employed to derive trivariate joint and conditional return periods. The bivariate and univariate return periods are also estimated and compared with trivariate return periods. It reveals that the trivariate joint return period for "OR" case is less than return periods for "AND" case for triplet flood characteristics. In other words, the occurrence of trivariate flood characteristics simultaneously is less frequent in "AND" case in compare with "OR" case of joint return periods (i.e., more frequent). Overall, it is concluded that copula function effectively preserving the flood dependence structure and thus found as very flexible and dynamic tools for the assessments of multidimensional extreme episodes i.e., flood. From the estimated trivariate return periods it could be revealed that for an effective flood risk assessments, it could be an essential concern to take the accountability of trivariate return periods, by considering all the inter-associated random vector simultaneously, instead of just pair-wise joint association or bivariate return periods.

Acknowledgements

Conflict of interest

All authors declare no conflicts of interest in this manuscript.

References

1. Rao AR, Hameed KH (2000) *Flood frequency analysis*. CRC Press, Boca Raton, Fla.
2. Zhang L (2005) Multivariate hydrological frequency analysis and risk mapping. Doctoral dissertation, Beijing Normal University.
3. Ganguli P, Reddy MJ (2013) Probabilistic assessments of flood risks using trivariate copulas. *Theor Appl Climatol* 111: 341–360.
4. Yue S (2000) The bivariate lognormal distribution to model a multivariate flood episode. *Hydrol Processes* 14: 2575–2588.
5. Yue S, Rasmussen P (2002) Bivariate frequency analysis: discussion of some useful concepts in hydrological applications. *Hydrol Processes* 16: 2881–2898.
6. Yue S, Wang CY (2004) A comparison of two bivariate extreme value distribution. *Stoch Environ Res Risk Assess* 18: 61–66.
7. Zhang L, Singh VP (2006) Bivariate flood frequency analysis using copula method. *J Hydrol Eng* 11: 150–164.
8. Zhang L, Singh VP (2007) Trivariate flood frequency analysis using the Gumbel-Hougaard copula. *J Hydrol Eng* 12: 431–439.
9. Reddy MJ, Ganguli P (2012) Bivariate Flood Frequency Analysis of Upper Godavari River Flows Using Archimedean Copulas. *Water Resour Manage* 26: 3995–4018.
10. Salvadori G (2004) Bivariate return periods via-2 copulas. *Stat Methodol* 1: 129–144.

11. Graler B, van den Berg M, Vandenberg S, et al. (2013) Multivariate return periods in hydrology: a critical and practical review focusing on synthetic design hydrograph estimation. *Hydrol Earth Syst Sci* 17: 1281–1296.

12. Krstanovic PF, Singh VP (1987) A multivariate stochastic flood analysis using entropy. In: Singh VP (Ed.), *Hydrologic Frequency Modelling*, Baton Rouge, U.S.A., 515–539.

13. Escalante-Sanboval CA, Raynal-Villasenor JA (1998) Multivariate estimation of floods: the trivariate gumble distribution. *J Stat Comput Simul 61*: 313–340.

14. Sandoval CE, Raynal-Villasenor J (2008) Trivariate generalized extreme value distribution in flood frequency analysis. *Hydrol Sci J* 53: 550–567.

15. Song S, Singh VP (2010) Meta-elliptical copulas for drought frequency analysis of periodic hydrologic data. *Environ Res Hazard Assess* 24: 425–444.

16. De Michele C, Salvadori G (2003) A generalized Pareto intensity-duration model of storm rainfall exploiting 2-copulas. *J Geophys Res Atmos* 108: 4067.

17. Grimaldi S, Serinaldi F (2006) Asymmetric copula in multivariate flood frequency analysis. *Adv Water Resour* 29: 1155–1167.

18. Salvadori G, De Michele C (2006) Statistical characterization of temporal structure of storms. *Adv Water Resour* 29: 827–842.

19. Saklar A (1959) Functions de repartition n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris* 8: 229–231.

20. Nelsen RB (2006) *An introduction to copulas*, Springer, New York.

21. Genest C, Favre AC (2007) Everything you always wanted to know about copula modelling but were afraid to ask. *J Hydrol Eng* 12: 347–368.

22. Favre AC, El Adlouni S, Perreault L, et al. (2004) Multivariate hydrological frequency analysis using copulas. *Water Resour Res* 40.

23. Renard B, Lang M (2007) Use of a Gaussian copula for multivariate extreme value analysis: Some case studies in hydrology. *Adv Water Resour* 30: 897–912.

24. Serinaldi F, Grimaldi S (2007) Fully nested 3-copula procedure and application on hydrological data. *J Hydrol Eng* 12: 420–430.

25. Genest C, Favre AC, Beliveau J, et al. (2007) Metaelliptical copulas and their use in frequency analysis of multivariate hydrological data. *Water Resour Res* 43: W09401.

26. Li F, Zheng Q (2016) Probabilistic modelling of flood events using the entropy copula. *Adv Water Resour* 97: 233–240.

27. Drainage and Irrigation Department Malaysia (2004) *Annual flood report of DID for Peninsular Malaysia. Unpublished report*. DID: Kuala Lumpur.

28. Malaysian Meteorological Department (2007) *Report on Heavy Rainfall that Caused Floods in Kelantan and Terengganu. Unpublished report.* MMD: Kuala Lumpur.

29. Adnan NA, Atkinson PM (2011) Exploring the impact of climate and land use changes on streamflow trends in a monsoon catchment. *Int J Climatol* 31:815–831.

30. Madadgar S, Moradkhani H (2013) Drought Analysis under Climate Change Using Copula. *J Hydrol Eng* 18: 746–759.

31. Salvadori G, De Michele C (2010) Multivariate multiparameters extreme value models and return periods: A Copula approach. *Water Resour Res* 46.

32. Shiau JT (2006) Fitting drought duration and severity with two dimensional copulas. *Water Resour Manage* 20: 795–815.

33. Zhang R, Chen X, Cheng Q, et al. (2016) Joint probability of precipitation and reservoir storage for drought estimation in the headwater basin of the Huaihe River, China. *Stoch Environ Res Risk Assess* 30: 1641–1657.

34. Kamarunzaman IF, Zin WZW, Ariff NM (2018) A Generalized Bivariate Copula for Flood Analysis in Peninsular Malaysia. *Preprints*, 2018080118.

35. Couasnon A, Sebastian A, Morales-Napoles O (2018) A Copula-Based Bayesian Network for Modeling Compound Flood Hazard from Riverine and Coastal Interactions at the Catchment Scale: An Application to the Houston Ship Channel, Texas. *Water* 10: 1190.

36. Genest C, Ghoudi K, Rivest LP (1995) A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika* 82: 543–552.

37. Xu Y, Huang G, Fan Y (2015) Multivariate flood risk analysis for Wei River. *Stoch Environ Res Risk Assess* 31: 225–242.

38. De Michele C, Salvadori G, Canossi M, et al. (2005) Bivariate statistical approach to check the adequacy of dam spillway. *J Hydrol Eng* 10: 50–57.

39. Klein B, Pahlow M, Hundecha Y, et al. (2010) Probability analysis of hydrological loads for the design of flood control system using copulas. *J Hydrol Eng* 15: 360–369.

40. Genest C, Rémillard B (2008) Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. *Annales de l'Institut Henri Poincare: Probabilites et Statistiques* 44: 1096–1127.

41. Genest C, Rémillard B, Beaudoin D (2009) Goodness-of-fit tests for copulas: A review and a power study. *Insur Math Econ* 44: 199–214.

42. Kojadinovic I, Yan J, Holmes M (2011) Fast large-sample goodness-of-fit tests for copulas. *Stat Sin* 21: 841–871.

43. Kojadinovic I, Yan J (2011) A goodness-of-fit test for multivariate multiparameter copulas based on multiplier central limit theorems. *Stat Comput* 21: 17–30.

44. Zhang S, Okhrin O, Zhou QM, et al. (2016) Goodness-of-fit Test for Specification of Semiparametric Copula Dependence Models. *J Econometrics* 193: 215–233.

45. Salvadori G, De Michele C (2004) Frequency analysis via copulas: theoretical aspects and applications to hydrological events. *Water Resour Res* 40: W12511.

46. Fisher NI, Switzer P (2001) Graphical assessments of dependence: is a picture worth 100 tests? *Am Stat* 55: 233–239.

47. Genest C, Boies JC (2003) Detecting dependence with Kendall plots. *Am Stat* 57: 275–284.

48. Gringorten II (1963) A plotting rule of extreme probability paper. *J Geophys Res* 68: 813–814.

49. Karmakar S, Simonovic SP (2008) Bivariate flood frequency analysis. Part-1: Determination of marginal by parametric and non-parametric techniques. *J Flood Risk Manage* 1: 190–200.

50. Cohn TA, Lane WL, Baier WG (1997) An algorithm for computing moments-based flood quantile estimates when historical flood information is available. *Water Resour Res* 33: 2089–2096.

51. Hosking JRM, Walis JR (1987) Parameter and quantile estimations for the generalized Pareto distributions. *Technometrics* 29: 339–349.

52. Anderson TW, Darling DA (1954) A test of goodness of fit. *J Am Stat Assoc* 49: 765–769.

53. Akaike H (1974) A new look at the statistical model identification. *IEEE Trans Autom Control* 19: 716–723.

54. Schwarz GE (1978) Estimating the dimension of a model. *Ann Stat* 6: 461–464.

55. Hannan EJ, Quinn BG (1979) The Determination of the Order of an Autoregression. *J R Stat Soc Series B Stat Methodol* 41: 190–195.

56. Moriasi DN, Arnold JG, Van Liew MW, et al. (2007) Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Trans ASABE* 50: 885–900.

57. Genest C, Huang W, Dufour JM (2013) A regularized goodness-of-fit test for copulas. *J Soc Fr Stat* 154: 64–77.