

Probabilistic assessment of flood risks using trivariate copulas

Poulomi Ganguli · M. Janga Reddy

Received: 4 October 2011 / Accepted: 19 April 2012 / Published online: 25 May 2012
© Springer-Verlag 2012

Abstract In this paper, a copula-based methodology is presented for probabilistic assessment of flood risks and investigated the performance of trivariate copulas in modeling dependence structure of flood properties. The flood is a multi-attribute natural hazard and is characterized by mutually correlated flood properties peak flow, volume, and duration of flood hydrograph. For assessing flood risk, many studies have used bivariate analysis, but a more effective assessment can be possible considering all three mutually correlated flood properties simultaneously. This study adopts trivariate copulas for multivariate analysis of flood risks, and applied to a case study of flood flows of Delaware River basin at Port Jervis, NY, USA. On evaluation of various probability distributions for representation of flood variables, it is found that the flood peak flow and volumes can be best represented by Fréchet distribution, whereas flood duration by log-normal distribution. The joint distribution is modeled using four trivariate copulas, namely, three fully nested form of Archimedean copulas: Clayton, Gumbel–Hougaard, Frank copulas; and one elliptical copula: Student's t copula. Based on distance-based performance measures, graphical tests, and tail-dependence measures, it is found that the Student's t copula best representing the trivariate dependence structure of flood properties as compared to the other copulas. Similar results are found for bivariate copula modeling of flood variables pairs, where Student's t copula performed better than the other copulas. The obtained copula-based joint distributions are used for multivariate analysis of flood risks, in terms of primary and

secondary return periods. The resultant trivariate return periods are compared with univariate and bivariate return periods, and addressed the necessity of multivariate flood risk analysis. The study concludes that the trivariate copula-based methodology is a viable choice for effective risk assessment of floods.

1 Introduction

Floods have become most common natural hazards, increasingly posing a significant risk to human life and environment. At the drainage basin scale, consideration of flood risk plays an important role in planning of water infrastructure projects. In design of hydraulic structures (e.g., dam spillways, diversion canals, dikes, and river channels), urban drainage systems, cross-drainage structures (e.g., culverts and bridges), reservoir management, and flood hazard mapping, it requires risk analysis of floods as a design criteria. However, because of limited hydrological data and higher uncertainty associated with hydrological variables, it is also very important to use more effective method for assessment of flood risks. Traditionally for probabilistic assessment of flood risks (also referred as flood frequency analysis), various methods were developed. Basically, flood frequency analysis consists of obtaining relationship between flood quantiles and their non-exceedance probability (also referred as return periods). Conventional method of frequency analysis involves use of observed annual peak discharges as a function of recurrence interval or exceedance probability (Stedinger et al. 1993). In general, such applications are incapable of giving adequate information of floods since flood is a multivariate phenomenon characterizing flood peak, volume, and duration of the flood hydrograph.

P. Ganguli · M. J. Reddy (✉)
Department of Civil Engineering,
Indian Institute of Technology Bombay,
Powai, Mumbai 400076, India
e-mail: mjreddy@civil.iitb.ac.in

In most of the hydrologic literature, the application of multivariate modeling is mainly contained to bivariate cases. Krstanovic and Singh (1987) used bivariate normal and exponential distributions for modeling joint distribution of flood peak and volumes as well as for obtaining conditional distribution of flood volume given peak discharge. Similarly, there are several studies that have used different conventional bivariate distributions for flood frequency analysis, bivariate generalized extreme value distribution (Adamson et al. 1999; Raynal and Salas 1987; Yue et al. 1999; Yue 2001a; Nadarajah and Shiau 2005; Escalante 2007), bivariate gamma distribution (Yue 2001b; Yue et al. 2001), bivariate normal distribution (Goel et al. 1998; Yue 1999), bivariate lognormal distribution (Yue 2000), bivariate exponential distribution (Choulakian et al. 1998) etc. Very few studies dealt with trivariate distributions for flood frequency analysis. For example, trivariate gumbel distribution (Escalante and Raynal 1998), trivariate generalized extreme value distribution (Escalante and Raynal 2008) are used for trivariate flood frequency analysis.

Many of the conventional models have major limitation in terms of assuming fixed probability distribution for all flood properties (i.e., assuming all flood properties are well represented by a single probability distribution). However, in practice, this may not be the case, as the flood variables may follow different distributions. Moreover, conventional estimates of flood exceedance are heavily dependent on right tail of the underlying frequency distribution, which is most difficult to estimate from observed flood data. Hence, it is desired to have evaluation of marginal distribution of flood variables separately and at the same time a function is needed that can join these marginal distributions by preserving the dependence structure of flood variables. In this context, the theory of copulas provides some advantages. Copulas are multivariate distribution functions, used for capturing the association or dependence between two or more random variables (Joe 1997). Also, the other advantage of copulas for multivariate analysis is that, it is invariant to monotonic transformations of marginal variables (i.e., data transformations, such as taking natural logarithms or applying Box–Cox transformations do not influence copula).

Recently, copulas have been applied in the field of hydrology (Favre et al. 2004; Salvadori and De Michele 2004; Grimaldi and Serinaldi 2006; Zhang and Singh 2007; Chowdhary et al. 2011; Salvadori et al. 2011). Although many studies in the past have focused on bivariate analysis of peak flow and volume for analyzing flood risks, a more complete analysis can be possible by considering mutually correlated flood properties—peak flow, volume and duration of flood events, and modeling by copulas. Very few studies used copulas for trivariate analysis of flood flows. Grimaldi and Serinaldi (2006) compared trivariate

symmetric Frank copula, asymmetric or fully nested form of Frank copula and Gumbel logistic distribution for Kanawha watershed, West Virginia and found better results using asymmetric scheme of copula. Zhang and Singh (2007) used symmetric Gumbel–Hougaard copula for trivariate modeling of peak flow, volume and duration; and noted that copula fits empirical joint distribution better than trivariate normal distributions. Serinaldi and Grimaldi (2007) applied fully nested Archimedean copulas to two different hydrologic data set, viz., trivariate flood frequency analysis and multivariate sea wave frequency analysis. Genest et al. (2007) applied meta-elliptical copulas for multivariate frequency analysis of annual spring flood at Romaine River in Québec, Canada and concluded that meta-elliptical copulas can help in better modeling the dependence structure of random variables when observed differences between the bivariate margins restricts the use of exchangeable copula families (i.e., the Archimedean copulas). Salvadori et al. (2011) illustrated application of copula in multivariate flood quantiles estimation by employing joint distribution of annual maximum flood peak, volume, and the corresponding initial water levels of the dam.

In most of the studies, the researchers applied either bivariate copulas or symmetric trivariate copulas for flood analysis. They consider simplifications in terms of deriving return periods, like assuming the dependence between only two flood properties and/or independence of the third variable. Then, copulas used to model pair-wise dependency in bivariate way. But in this study, asymmetric copulas are applied for modeling trivariate flood properties peak flow, volume and duration simultaneously (instead of considering pair-wise variable separately). The main aims of present study are: (1) to develop trivariate models based on Archimedean and elliptical class of copulas for modeling flood characteristics, (2) to evaluate efficacy of copulas in modeling hydrological extremes and demonstrate potential of copula models for flood risk analysis through a case study, and (3) to evaluate the importance of multivariate return periods for estimation of flood risks through comparative analysis of the trivariate return periods with univariate and bivariate return periods.

In the following, first the definition of copula function and its properties are presented. Then, the details of copulas adopted in the study (i.e., three fully nested Archimedean families of copulas and one elliptical Student's t copula) are presented. The subsequent section describes various performance measures including distance based statistics, graphical tests as well as tail dependence measures to select suitable copula model for analyzing multivariate dependence of flood properties. Then details of flood properties and procedures for estimation of joint return periods using copulas are explained. Later, application of the methodology

to a case study is presented and the results are discussed. Finally, brief summary and conclusions of the study are presented.

2 Multivariate modeling using copulas

2.1 Copula function

In general, copula function is a multivariate distribution function, used for capturing the dependence between two or more random variables. Let $\mathbf{X}=(X_1, \dots, X_d)$ be a random vector with continuous marginal cumulative distribution functions (CDF) F_1, \dots, F_d . By following Sklar theorem (Sklar 1959), the relation between the joint CDF $H(\mathbf{X})$ and copula C can be represented as,

$$H(\mathbf{X}) = C\{F_1(x_1), \dots, F_d(x_d); \theta\} \mathbf{X} \in R^d \tag{1}$$

where the function $C: [0,1]^d \rightarrow [0,1]$ is called a d -dimensional *copula*, with association parameter θ . More details on theoretical background and properties of various copula families can be found in Nelsen (2006). In the following brief details of copulas used in the present study are presented.

2.2 Multivariate 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 several desirable properties, such as, symmetry and associativity. The d -dimensional Archimedean copula (Nelsen 2006) is expressed as,

$$C(u_1, \dots, u_d) = \phi^{-1}(\phi(u_1) + \phi(u_2) + \dots + \phi(u_d)) \tag{2}$$

where $u_i=F_i(X_i)$ is marginal CDF of variable $X_i(i=1, \dots, d)$; $\phi(\bullet)$ is known as generator of the copula and $\phi^{-1}(\bullet)$ is the pseudo inverse of $\phi(\bullet)$. If $\phi(0)=\infty$ the pseudo-inverse describes an ordinary inverse function (i.e., $\phi^{-1}=\phi^{-1}$) and in this case ϕ is known as strict generator. Equation 2 represents a d -dimensional Archimedean copula if and only if the generator $\phi(\bullet)$ is a completely monotonic function. This symmetric form of copula is developed under the assumption of homogeneous dependence across the variables and has the limitation that in this case all dependencies are averaged to a same value. But for hydrologic variables such assumption is not feasible in practice. In order to increase flexibility and to allow for heterogeneous dependence—fully nested Archimedean (FNA) copulas were suggested in recent past (Whelan 2004; Savu and Trede 2010). The structure of FNA copula joins two or more ordinary bivariate or higher-dimensional Archimedean copulas by another Archimedean

copula. The dependence structure of FNA in trivariate case (Savu and Trede 2010) is expressed as,

$$\begin{aligned} C(u_1, u_2, u_3) &= \phi_2(\phi_2^{-1} \circ \phi_1[\phi_1^{-1}(u_1) + \phi_1^{-1}(u_2)] + \phi_2^{-1}(u_3)) \\ &= C_2[C_1(u_1, u_2), u_3] \end{aligned} \tag{3}$$

with condition that the first derivative of $\phi_2^{-1} \circ \phi_1$ is completely monotonic. In above expression, ϕ_2 and ϕ_1 are Laplace transforms. In Eq. 3, the pair (u_1, u_2) has bivariate margin of the form of Eq. 2 with Laplace transform ϕ_1 , whereas (u_1, u_3) and (u_2, u_3) has the bivariate margins of the form Eq. 2 with laplace transform ϕ_2 . Thus, the three variable asymmetric copula is composed of two bivariate copulas C_1 and C_2 , in which C_1 is the copula describing the dependence between variables u_1 and u_2 ; and the outer copula C_2 is a function of the inner copula C_1 and u_3 . It follows that this asymmetric scheme can be only applied in cases when the correlation values among two variables (the most nested) is stronger than the correlation between these variables and the third variable.

The properties of trivariate FNA structures for Clayton, Gumbel–Hougaard, and Frank copulas are presented in Table 1. To apply FNA three-dimensional copula, it has to satisfy the condition that the rank correlation coefficients (i.e., Kendall’s τ or Spearman’s ρ) between inner pair (u_1, u_2) is higher than that of between the other pairs (u_1, u_3) and (u_2, u_3) .

2.3 Student’s t copula

The Student’s t copula belongs to the class of elliptical copulas. The Student’s t copula is specified by multivariate Student’s t distribution. Let $\Sigma \in R^{d \times d}$ for $x \in R^d$ denotes a symmetric shape parameter matrix (i.e., Σ is correlation matrix of multiple variables in d -dimension), then multivariate Student’s t copula for $u=(u_1, \dots, u_d) [0,1]^d$ with ϑ degrees of freedom is defined as (Mashal and Zeevi 2002),

$$C(u_1, \dots, u_d; \vartheta, \Sigma) = t_{\vartheta, \Sigma}^d(t_{\vartheta}^{-1}(u_1), \dots, t_{\vartheta}^{-1}(u_d)) \tag{4}$$

where,

$$t_{\vartheta, \Sigma}^d(x) = \int_{-\infty}^x \frac{\Gamma((\vartheta + d)/2)}{\Gamma(\vartheta/2)(\vartheta\pi)^{d/2} \sqrt{|\Sigma|}} (1 + y^T \Sigma^{-1} y / \vartheta)^{-(\vartheta+d)/2} dy$$

Here $y = \{y_1, \dots, y_d\}$ and $y_i = t_{\vartheta}^{-1}(u_i)$.

Equation 4 represents t copula with parameters (ϑ, Σ) . For $\vartheta > 2$, the shape parameter matrix Σ is nothing but as

Table 1 Mathematical expressions for trivariate Archimedean family of copulas and their associated properties

Copula family	$C(u_1, u_2, u_3)$	$\phi_2^{-1} \circ \phi_1(\bullet)$	Conditions
Clayton	$\left[(u_1^{-\theta_2} + u_2^{-\theta_2} - 1)^{\theta_1/\theta_2} + u_3^{-\theta_1} - 1 \right]^{\frac{1}{\theta_1}}$	$\frac{1}{\theta_1} \left\{ (\theta_2 t + 1)^{\theta_1/\theta_2} - 1 \right\}$	$\theta_1 < \theta_2, \theta \in (0, \infty)$
Gumbel–Hougaard	$\exp \left\{ - \left(\left[(-\log u_1)^{\theta_2} + (-\log u_2)^{\theta_2} \right]^{\theta_1/\theta_2} + (-\log u_3)^{\theta_1} \right)^{\frac{1}{\theta_1}} \right\}$	t^{θ_1/θ_2}	$\theta_1 < \theta_2, \theta \in [1, \infty)$
Frank	$-\theta_1^{-1} \log \left\{ 1 - c_1^{-1} \left(1 - \left[1 - c_2^{-1} (1 - e^{-\theta_2 u_1}) \cdot (1 - e^{-\theta_2 u_2}) \right]^{\theta_1/\theta_2} \right) \cdot (1 - e^{-\theta_1 u_3}) \right\}$ where, $c_1 = 1 - e^{-\theta_1}$ and $c_2 = 1 - e^{-\theta_2}$	$-\ln \frac{1 + e^{-t} (e^{-\theta_2 - 1})^{\theta_1/\theta_2} - 1}{e^{-\theta_1} - 1}$	$\theta_1 < \theta_2, \theta \in (0, \infty)$

θ_2 and θ_1 represents dependence parameters; $u_1, u_2,$ and u_3 denotes marginal CDFs

positive-definite correlation matrix. The density of multivariate t copula can be expressed as,

$$c(u; \vartheta, \Sigma) = \frac{\Gamma((\vartheta + d)/2) [\Gamma(\vartheta/2)]^{d-1}}{[\Gamma((\vartheta + 1)/2)]^d |\Sigma|^{1/2}} \left[\prod_{i=1}^d (1 + y_i^2/\vartheta)^{-(\vartheta+1)/2} \right] (1 + y^T \Sigma^{-1} y/\vartheta)^{-(\vartheta+d)/2} \tag{5}$$

2.4 Copula parameter estimation

The parameter estimation is performed using maximum pseudo-likelihood (MPL) method. The multivariate copula model can have $d-1$ parameters if the dependency is modeled by FNA copula, and $d(d-1)/2$ parameters for elliptical t copula. The MPL estimation method does not require any prior assumptions regarding marginal distributions of the dependent variables. The procedure consists of transforming the marginal variables into uniformly distributed vectors using its empirical distribution function. Then, the copula parameters are estimated using maximization of pseudo log-likelihood function.

Let $X_1 = (X_{1,1}, \dots, X_{1,d}), \dots, X_n = (X_{n,1}, \dots, X_{n,d})$ be n sample of observations in d -dimensional case. The empirical CDF of variable X_k can be computed by,

$$F_k(X_{i,k}) = \frac{R_{i,k}}{n+1}, i \in \{1, \dots, n\}, k \in \{1, \dots, d\} \tag{6}$$

where $R_{i,k}$ is rank, which is given by

$$R_{i,k} = \sum_{j=1}^n I(X_{j,k} \leq X_{i,k}); \tag{7}$$

where $I(A)$ is a logical indicator function results in either 1 (if A is true) or 0 (if A is false). The rescaling $(n+1)$ at the denominator is used instead of n to avoid numerical problems at the boundaries of $[0, 1]^2$ (a standard convention in the probabilistic modeling). The empirical distribution function is used as a surrogate for the unknown marginals. Substituting empirical CDFs into copula density and

applying logarithm to the likelihood function of the copula yields the following form,

$$\ell(\theta) = \sum_{i=1}^n \log [c_\theta \{F_1(X_{i,1}), \dots, F_d(X_{i,d})\}] \tag{8}$$

Then, the copula parameter $\hat{\theta}$ can be obtained by maximizing this pseudo log-likelihood function $\mathcal{L}(\theta)$.

For obtaining Σ in three-dimensional case, it involves computing σ_{ij} for $i, j \in \{1, 2, 3\}$ with the values of Kendall’s tau, $\tau_{i,j}$, for three random variables. For elliptical family of copulas, $\sigma_{ij} = \sin(\pi \tau_{i,j}/2)$ and the Σ is given as

$$\Sigma = \begin{bmatrix} 1 & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & 1 & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & 1 \end{bmatrix} \tag{9}$$

However, sometimes it may be possible that the estimate of Σ is not positive definite. In that case, a procedure based on eigenvalue decomposition is used to transform correlation matrix into positive definite (McNeil et al. 2005). Using the relationship in Eq. 9, the correlation matrix of Student’s t copula is estimated and then a numerical search technique is employed for estimating $\hat{\vartheta}$ (Mashal and Zeevi 2002). In this study, a real-coded genetic algorithm (GA) is applied to find the optimal parameters of the FNA and Student’s t copulas.

3 Selection of suitable copula model

3.1 Performance measures

The appropriate dependence structure is selected by minimizing the distance between fitted parametric copula and the

empirical copula. Anderson–Darling (AD) and Integrated Anderson–Darling (IAD) statistics are used as a distance measures between empirical copulas and fitted parametric copulas.

The empirical copula C_n from the pseudo-observations $(U_{1,n}, V_{1,n}, W_{1,n}), \dots, (U_{n,n}, V_{n,n}, W_{n,n})$, is given by (Genest et al. 2009):

$$U_{i,n} = \frac{1}{n+1} \sum_{j=1}^n 1(X_j \leq X_i), \quad V_{i,n} = \frac{1}{n+1} \sum_{j=1}^n 1(Y_j \leq Y_i), \quad W_{i,n} = \frac{1}{n+1} \sum_{j=1}^n 1(W_j \leq W_i), i \in \{1, \dots, n\} \tag{11}$$

The expressions for AD and IAD distance measures are given as (Ané and Kharoubi 2003):

$$AD = \max_{1 \leq i \leq n, 1 \leq j \leq n, 1 \leq k \leq n} \frac{\left| \widehat{C}_n\left(\frac{i}{n}, \frac{j}{n}, \frac{k}{n}\right) - C_{p\theta}\left(\frac{i}{n}, \frac{j}{n}, \frac{k}{n}\right) \right|}{\sqrt{C_{p\theta}\left(\frac{i}{n}, \frac{j}{n}, \frac{k}{n}\right) \left[1 - C_{p\theta}\left(\frac{i}{n}, \frac{j}{n}, \frac{k}{n}\right)\right]}} \tag{12}$$

$$IAD = \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \frac{\left[\widehat{C}_n\left(\frac{i}{n}, \frac{j}{n}, \frac{k}{n}\right) - C_{p\theta}\left(\frac{i}{n}, \frac{j}{n}, \frac{k}{n}\right) \right]^2}{C_{p\theta}\left(\frac{i}{n}, \frac{j}{n}, \frac{k}{n}\right) \left[1 - C_{p\theta}\left(\frac{i}{n}, \frac{j}{n}, \frac{k}{n}\right)\right]} \tag{13}$$

$$E[f_{X_1, X_2, X_3}(x_1, x_2, x_3)] = - \int \int \int f_{X_1, X_2, X_3}(x_1, x_2, x_3) \ln[f_{X_1, X_2, X_3}(x_1, x_2, x_3)] dx_1 dx_2 dx_3 \tag{14}$$

The entropy offers a distance measure based on copula density, whereas the AD and IAD statistics are computed using CDFs. For copula representation of trivariate density, the entropy of $f_{X_1, X_2, X_3}(x_1, x_2, x_3)$ is equal to the sum of the entropies of individual marginal distribution plus the entropy of the copula distribution function. The entropy of the univariate distribution $f_X(x)$ is defined as,

$$E[f_X(x)] = - \int f_X(x) \ln[f_X(x)] dx \tag{15}$$

Therefore, the discrete entropy of the copula model can be expressed as,

$$E[f_{X_1, X_2, X_3}(x_1, x_2, x_3)] = E[f_{X_1}(x_1)] + E[f_{X_2}(x_2)] + E[f_{X_3}(x_3)] + E\{c[F_{X_1}(x_1), F_{X_2}(x_2), F_{X_3}(x_3)]\} \tag{16}$$

where,

$$E\{c[F_{X_1}(x_1), F_{X_2}(x_2), F_{X_3}(x_3)]\} = - \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n c_{p\theta}\left(\frac{i}{n}, \frac{j}{n}, \frac{k}{n}\right) \ln \left[c_{p\theta}\left(\frac{i}{n}, \frac{j}{n}, \frac{k}{n}\right) \right] \tag{17}$$

$$C_n(u, v, w) = \frac{1}{n} \sum_{i=1}^n I(U_{i,n} \leq u, V_{i,n} \leq v, W_{i,n} \leq w), (u, v, w) \in [0, 1] \tag{10}$$

where $(U_{i,n}, V_{i,n}, W_{i,n})$ are pseudo-observations computed from the collected observational data $(X_1, Y_1, W_1), \dots, (X_n, Y_n, W_n)$,

where i, j , and k represents order statistics of the random variable u_1, u_2 , and u_3 . Since empirical copula is defined on a lattice \mathcal{L} , the distance statistics are defined with discrete norms. The copula family with minimum AD and IAD distance is chosen as the best fitted copula. Apart from these distance-based measures, the relevance of each copula model is measured using the concept of entropy, which measures uncertainty of the distribution $f_{X_1, X_2, X_3}(x_1, x_2, x_3)$:

For visual inspection, a graphical comparison between observed data and simulated samples from each copula family is performed. This also offers qualitative assessment in finding a suitable copula model.

3.2 Tail dependence of flood characteristics

The tail dependence coefficient (TDC) captures the concordance between extreme values in the lower left quadrant tail and upper right quadrant tails of the variables. In order to study occurrence of extreme events, the pair-wise analysis of upper tail dependence of flood variables is performed for the fitted copula models. If u be a threshold value then upper tail dependence between two variables X and Y , denoted as λ_U is given as

$$\lambda_U = \lim_{u \rightarrow 1^-} \{F_X(x) > u | F_Y(y) > u\} \tag{18}$$

in terms of copula the above equations can also be expressed as (Nelsen et al. 2008),

$$\lambda_U = \lim_{u \rightarrow 1^-} \frac{1 - 2u + C(u, u)}{1 - u} = 2 - \lim_{u \rightarrow 1^-} \frac{1 - C(u, u)}{1 - u} = 2 - \delta'_C(1^-) \tag{19}$$

where, the function $\delta_C(\bullet)$ is the *diagonal section* of copula C and given by $\delta_C(u) = C(u, u)$ for every $u \in [0, 1]$. The estimate λ_U measures the concordance between extremely high values of random variables. If $\lambda_U \in (0, 1]$, then $F_X(x)$ and $F_Y(y)$ are said to show upper tail dependence or extremal dependence in upper tail.

To study nonparametric TDC, several methods have been suggested in literature such as λ_U^{LOG} estimator (Coles et al. 1999), λ_U^{SEC} estimator (Joe et al. 1992), Capéraá-Frahm-Genest estimator λ_U^{CFG} (Capéraá et al. 1997; Frahm et al. 2005), and Schmidt–Stadtmüller estimator λ_U^{SS} (Schmidt and

Stadtmüller 2006). The estimator λ_U^{LOG} can be interpreted as logarithm of the copula diagonal (Coles et al. 1999). The estimate λ_U^{SEC} can be interpreted as the slope of the secant along the copula diagonal (i.e., close to the 45° line) and hence can specify wrongly the value of TDC when the data are not accumulated along the diagonal. Except for λ_U^{CFG} , all other estimators require specification of a threshold. This study adopts λ_U^{CFG} for TDC. If $\{(u_1, v_1), \dots, (u_n, v_n)\}$ be random sample obtained from Copula $C(\bullet)$, the bivariate upper TDC using λ_U^{CFG} is given by (Capéraá et al. 1997),

$$\widehat{\lambda}_U^{\text{CFG}} = 2 - 2 \exp \left[\frac{1}{n} \sum_{i=1}^n \log \left\{ \frac{\sqrt{\log \left(\frac{1}{u_i} \right) \log \left(\frac{1}{v_i} \right)}}{\log \left(\frac{1}{\max(u_i, v_i)^2} \right)} \right\} \right] \tag{20}$$

Though λ_U^{CFG} estimator assumes that the underlying copula can be approximated by an extreme-value copula, the estimator performs well even if the copula does not belong to extreme value classes as discussed by Frahm et al. (2005). For a comparison purpose, the estimates of λ_U^{LOG} and λ_U^{SS} are also presented, although these estimators can show higher variance as compared to $\widehat{\lambda}_U^{\text{CFG}}$. If the diagonal section of copula $C(u, u)$ is differentiable for $u \in (1-\varepsilon, 1)$ for any $\varepsilon > 0$, then,

$$\begin{aligned} \lambda_U &= 2 - \lim_{u \rightarrow 1^-} \frac{1 - C(u, u)}{1 - u} = 2 - \lim_{u \rightarrow 1^-} \frac{dC(u, u)}{du} \\ &= 2 - \lim_{u \rightarrow 1^-} \frac{\log C(u, u)}{\log(u)} \end{aligned} \tag{21}$$

The log estimator is based on Eq. 21 and can be expressed as

$$\widehat{\lambda}_U^{\text{LOG}} = 2 - \frac{\log \widehat{C}_n \left(\left(1 - \frac{k}{n} \right), \left(1 - \frac{k}{n} \right) \right)}{\log \left(1 - \frac{k}{n} \right)} \tag{22}$$

where $k \in \{1, \dots, n-1\}$ represents the threshold to be selected. An optimal threshold is selected by applying a plateau-finding algorithm as described in Frahm et al. (2005). In first step, the curve of $\widehat{\lambda}_k$ is smoothed by nonparametric kernel function. A kernel smoother defines a set of weights $\{W_i(x), i=1, \dots, n\}$ for each x and can be expressed as,

$$\widehat{f}(x) = \sum_{i=1}^n W_i(x) y_i \tag{23}$$

where y_i is the observations to be smoothed. For a given kernel band width b , the weight sequence is defined as,

$$W_i(x) = \frac{K\left(\frac{x-x_i}{b}\right)}{\sum_{i=1}^n K\left(\frac{x-x_i}{b}\right)}, \sum_{i=1}^n W_i(x_i) = 1 \text{ and } \int K(u) du = 1 \tag{24}$$

In this study, the box kernel with specified bandwidth (say, $b = \lfloor 0.005n \rfloor, b \in N$) is chosen as suggested by Frahm et al. (2005). The analytical expression of $K(\bullet)$ for box-kernel estimator can be given using following *window function*

$$K(x) = \begin{cases} 1, & |x| \leq 1/2, \\ 0, & \text{otherwise} \end{cases} \tag{25}$$

where $K(x)$ defines a unit interval centered at the origin.

Thus the kernel smoothed map of $k \mapsto \widehat{\lambda}_k$ leads to the means of $2b + 1$ successive points of $\widehat{\lambda}_1, \dots, \widehat{\lambda}_n$ to a new smoothed map of $\widetilde{\lambda}_1, \dots, \widetilde{\lambda}_{n-2b}$. In next step, a vector $p_k = (\widetilde{\lambda}_k, \dots, \widetilde{\lambda}_{k+m-1})$, $k = 1, \dots, n - 2b - m + 1$ is defined with a plateau length $m = \sqrt{n - 2b}$. This procedure continues and the algorithm stops at the first plateau p_k , whose elements fulfill the following condition,

$$\sum_{i=k+1}^{k+m-1} |\widetilde{\lambda}_i - \widetilde{\lambda}_k| \leq 2\sigma \tag{26}$$

where σ represents the standard deviation of $\widetilde{\lambda}_1, \dots, \widetilde{\lambda}_{n-2b}$. Then the TDC λ_U^{LOG} is estimated as arithmetic mean of the vector corresponding to the plateau,

$$\widehat{\lambda}_U^{\text{LOG}}(k) = \frac{1}{m} \sum_{i=1}^m \widetilde{\lambda}_{k+i-1} \tag{27}$$

If no plateau fulfills the stopping condition, the TDC is estimated as zero and the procedure is repeated with a different set of parameters.

The expression for λ_U^{SS} is given as (Schmidt and Stadtmüller 2006),

$$\widehat{\lambda}_U^{\text{SS}}(k) = \frac{n}{k} \widehat{C}_n \left(\frac{k}{n}, \frac{k}{n} \right) \approx \frac{1}{k} \sum_{i=1}^n I \{ (R_i > n - k) \wedge (S_i > n - k) \} \tag{28}$$

where $\bar{C}_n(u, v) = P(U > 1 - u, V > 1 - v) = u + v - 1 + C(1 - u, 1 - v)$ denotes empirical survival copula (Nelsen 2006).

4 Trivariate flood characteristics

The annual maximum peak discharge, volume, and duration values of flood events are obtained from daily stream flow data. Annual peak flow magnitude is selected from the portion of daily stream flow hydrographs having the highest peak flow from each year’s stream flow record. The single-peaked flood hydrograph is shown in Fig. 1. Flood duration (D) can be determined by identifying the time of rise (point “s” in Fig. 1) and fall (point “e” in Fig. 1) of the flood hydrograph. The start of the surface runoff is marked by the sharp rise of the hydrograph and end of the flood runoff is identified by the inflection point on the receding limb of the hydrograph. Between these two points, the total flood volume is estimated. If time of rise of the flood hydrograph is denoted by SD (day) and fall by ED (day), the flood volume (V) of each flood event is determined using following expression (Yue 2001a, b)

$$V_i = (V_i^{\text{total}} - V_i^{\text{baseflow}}) = \sum_{j=SD}^{ED} q_{ij} - \frac{D_i}{2}(q_{is} + q_{ie}), \quad \forall i = 1, 2, \dots, n \quad (29)$$

where, q_{ij} is the j th day observed stream flow value for i th year, q_{is} and q_{ie} are the observed daily stream flow values on start and end date of the flood hydrograph for

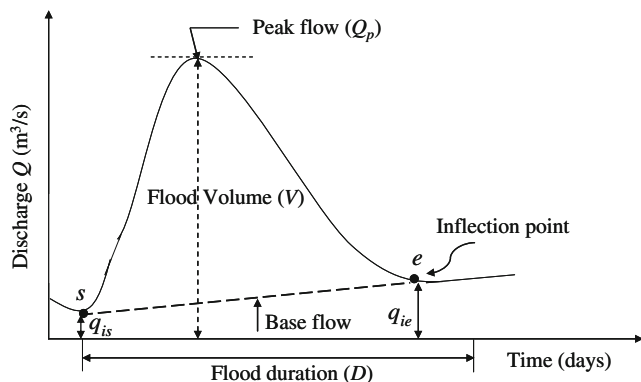


Fig. 1 Typical flood hydrograph showing flood flow characteristics

the i th year. The annual flood peak series is constructed by,

$$Q_i = \max\{q_{ij}, j = SD_i, SD_i + 1, \dots, ED_i\}, \quad \forall i = 1, 2, \dots, n \quad (30)$$

The duration series is given by the difference between starting and ending day of the flood event, and can be expressed as,

$$D_i = ED_i - SD_i, \quad \forall i = 1, 2, \dots, n \quad (31)$$

5 Estimation of flood risks

The copula models can form the basis for the estimation of various quantities, which can be very useful for risk analysis of floods, such as estimation of conditional probability distributions as well as conditional and joint return periods. The return period of a prescribed event is generally adopted as a criterion for design purposes in hydrologic projects, which provides a simple means for risk analysis. Usually, the return period is defined as “the average time elapsing between two successive realizations of the given event” (Salvadori 2004). The basic concepts of return periods are thoroughly discussed in Yue and Rasmussen (2002) and Salvadori and De Michele (2004).

5.1 Primary return period

The objective of frequency analysis of hydrologic data is to relate the magnitude of extreme events to their frequency of occurrence through the use of probability distributions (Chow et al. 1988). For multivariate case, in which X_1, X_2, \dots, X_d exceeds their respective thresholds ($X_1 > x_1, \dots, X_d > x_d$), the joint return period is computed using inclusive probability (“OR” and “AND” cases) of all three events, known as primary return periods (Salvadori 2004). For trivariate case, the joint primary return period in “OR” case $T_{X_1, X_2, X_3}^{\text{OR}}$ (for annual flood analysis) is given by,

$$T_{X_1, X_2, X_3}^{\text{OR}}(x_1, x_2, x_3) = \frac{1}{P(X_1 \geq x_1, \vee X_2 \geq x_2, \vee X_3 \geq x_3)} = \frac{1}{1 - P(X_1 \leq x_1, \vee X_2 \leq x_2, \vee X_3 \leq x_3)} = \frac{1}{1 - F_{X_1, X_2, X_3}(x_1, x_2, x_3)} = \frac{1}{1 - C(u_1, u_2, u_3)} \quad (32)$$

The joint primary return period in “AND” case T_{X_1, X_2, X_3}^{AND} (for annual flood analysis) can be expressed as,

$$T_{X_1, X_2, X_3}^{AND}(x_1, x_2, x_3) = \frac{1}{P(X_1 \geq x_1 \wedge X_2 \geq x_2 \wedge X_3 \geq x_3)} = \frac{1}{1 - F_{X_1}(x_1) - F_{X_2}(x_2) - F_{X_3}(x_3) + F_{X_1, X_2}(x_1, x_2) + F_{X_2, X_3}(x_2, x_3) + F_{X_1, X_3}(x_1, x_3) - F_{X_1, X_2, X_3}(x_1, x_2, x_3)}$$

$$= \frac{1}{1 - F_{X_1}(x_1) - F_{X_2}(x_2) - F_{X_3}(x_3) + C(u_1, u_2) + C(u_2, u_3) + C(u_1, u_3) - C(u_1, u_2, u_3)} \tag{33}$$

where $C(u_1, u_2)$, $C(u_2, u_3)$, and $C(u_1, u_3)$ are bivariate copula CDFs for flood characteristics.

The conditional distribution function of X_1, X_2 given $(X_3 \leq x_3)$ in “OR” case is given by,

$$F_{X_1, X_2, X_3}(x_1, x_2 | X_3 \leq x_3) = P[X_1 \leq x_1, X_2 \leq x_2 | X_3 \leq x_3]$$

$$= \frac{F_{X_1, X_2, X_3}(x_1, x_2, x_3)}{F_{X_3}(x_3)} = \frac{C(u_1, u_2, u_3)}{u_3} \tag{34}$$

where $F_{X_3}(x_3)$ is the marginal CDF of random variable X_3 . The corresponding conditional return period under this condition can be expressed as,

$$T_{X_1, X_2 | X_3}(x_1, x_2 | X_3 \leq x_3) = \frac{1}{1 - F_{X_1, X_2 | X_3}(x_1, x_2 | X_3 \leq x_3)} \tag{35}$$

Similarly, the conditional distribution of X_1 given $(X_2 \leq x_2, X_3 \leq x_3)$ is given by,

$$F_{X_1 | X_2, X_3}(x_1 | X_2 \leq x_2, X_3 \leq x_3) = P[X_1 \leq x_1 | X_2 \leq x_2, X_3 \leq x_3]$$

$$= \frac{F_{X_1, X_2, X_3}(x_1, x_2, x_3)}{F_{X_2, X_3}(x_2, x_3)} = \frac{C(u_1, u_2, u_3)}{C(u_2, u_3)} \tag{36}$$

where $C(u_2, u_3)$ is a bivariate copula CDF. The corresponding conditional return period can be expressed as,

$$T_{X_1 | X_2, X_3}(x_1 | X_2 \leq x_2, X_3 \leq x_3) = \frac{1}{1 - F_{X_1 | X_2, X_3}(x_1 | X_2 \leq x_2, X_3 \leq x_3)} \tag{37}$$

5.2 Secondary return period

Flood events can be subcritical, critical, and supercritical. In this context, the concept of secondary return period is

practically useful for design of hydraulic structures (Salvadori 2004; Vandenberghe et al. 2011). Salvadori and De Michele (2004) used Kendall distribution function to define the “secondary return period”. The primary return period predicts that a critical event is expected to appear once in a given time interval (i.e., it gives an average forecast), where as the secondary return period provides the average time between the occurrence of two supercritical events. The probability of supercritical event for any realization can be computed using Kendall distribution function in place of $C(\mathbf{u})$ in computation of joint primary return period in OR case.

For a d -dimensional distribution $F = C_{\hat{\theta}}(\mathbf{U})$ and $t \in (0, 1]$ the critical probability level \bar{p} is defined as (Salvadori et al. 2011)

$$\bar{p} = \{F(\mathbf{X}) = t, \mathbf{X} \in \mathbf{R}^d\} \tag{38}$$

Thus the isosurface \bar{p} partitions \mathbf{R}^d into three non-overlapping regions, viz., the subcritical region consisting the points which are less than \bar{p} ; the critical layer \bar{p} , where $F(\mathbf{X}) = t$; the supercritical region consisting the points which are greater than \bar{p} . The return period of a supercritical or potentially dangerous event associated with critical probability level \bar{p} can be computed as

$$\bar{T}_{X_1, X_2, X_3}(x_1, x_2, x_3) = \frac{1}{\bar{K}_{C_{\hat{\theta}}}(\bar{p})} = \frac{1}{1 - K_{C_{\hat{\theta}}}(\bar{p})} \tag{39}$$

where $K_{C_{\hat{\theta}}}(\cdot)$ is Kendall’s distribution function associated with trivariate copula $C_{\hat{\theta}}(\mathbf{U})$ at critical probability level \bar{p} . $\bar{K}_{C_{\hat{\theta}}}$ denotes survival probability and Kendall’s distribution of the region of the supercritical events; a multivariate form of univariate event $\{X > x\}$. Thus, any

Table 2 Statistic properties of flood variables

Variables	Maximum	Minimum	Mean	Standard deviation	Coefficient of skewness	Coefficient of kurtosis
Peak flow (Mm ³ /day)	398.8	36.70	126.20	73.60	2.060	8.02
Volume (Mm ³)	1,242	19.5	283.30	202.6	2.065	9.66
Duration (days)	27	5	11.69	4.63	1.10	4.52

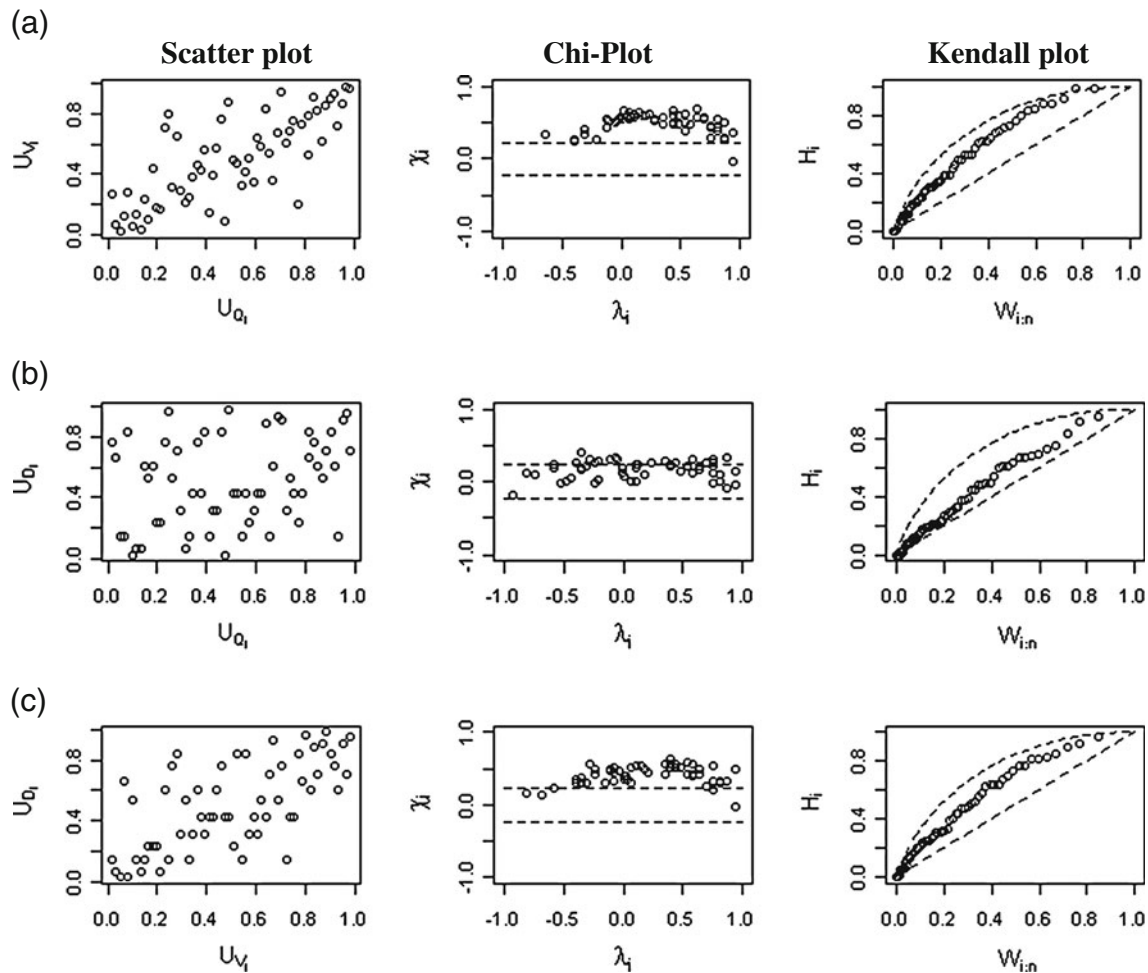


Fig. 2 Graphical representation of strength of dependence of flood variables using ranked scatter plot, chi-plot, and Kendall plot. First row (a) peak flow–volume combination; second row (b) peak flow–

duration combination; and third row (c) volume–duration combination of flood variable. U_Q , U_V , and U_D denotes pseudo-samples of peak flow, volume and duration respectively

point in the subcritical region will have a smaller joint CDF function value than any point in the supercritical region. The function $K_{C_\theta}(\cdot)$ helps to project multivariate information into a single dimension. For bivariate Archimedean copulas, an explicit expression for K_{C_θ} is available; though for FNA structure, the form of Kendall distribution function is complex as shown by Okhrin et al. (2009). For elliptical class of copulas, the function K_{C_θ} can be constructed numerically using Monte Carlo simulation.

6 Application

6.1 Study area

The case study of Delaware River basin is chosen for illustration of the methodology, the basin drains an area of 34,447 km² in the states of New York, Pennsylvania, New Jersey, and Delaware. Major flood events in the basin took place in the years 1955, 1999, 2001, 2003, 2004, 2005, 2006,

Table 3 Correlations to measure dependencies among flood variables

Dependence measure	Peak flow–volume	Volume–duration	Peak flow–duration
Pearson’s r	0.785 ^a (6.905e–014)	0.613 ^a (1.487e–007)	0.283 (0.027)
Spearman’s ρ	0.756 ^a (1.834e–012)	0.665 ^a (4.839e–009)	0.277 (0.030)
Kendall’s τ	0.584 ^a (3.007e–011)	0.508 ^a (1.980e–008)	0.209 (0.021)

Bracketed values represent p value of the estimate

^a Correlation is significant at 5 % significance level

Table 4 The mathematical expressions for probability density functions and parameters of different probability distributions

Marginal distribution function	Density function	Remarks
Log-normal	$f_X(x) = \frac{1}{x\sqrt{2\pi\sigma_y^2}} \exp\left(-\frac{(\ln(x)-\mu_y)^2}{2\sigma_y^2}\right)$ $x > 0, \sigma_y > 0, -\infty < \mu_y < \infty$ $Y = \ln(X)$	μ_y and σ_y are mean and standard deviations of Y $\mu_y = \frac{\sum y_i}{n}, \sigma_y^2 = \frac{\sum y_i^2 - n\bar{y}^2}{n-1}$
Gamma	$f_X(x) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}$; $x \geq 0, \alpha, \beta > 0$; $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$ for $\alpha > 0$	Positively skewed distribution α =shape parameter β =scale parameter $A = \ln(\bar{x}) - \frac{\sum \ln(x)}{n}$ $\alpha = \frac{1}{4A} \left[1 + \sqrt{1 + \frac{4A}{3}} \right], \beta = \frac{\alpha}{\bar{x}}$
Gumbel	$f_X(x) = \frac{1}{\beta} e^{-(x-\gamma)/\beta} \exp[-e^{-(x-\gamma)/\beta}]$ $-\infty < x < \infty$	Negatively skewed distribution $-\infty < \gamma < \infty, \beta > 0$ γ =location parameter
Fréchet	$f(x) = \frac{\alpha}{\beta} \left(\frac{\beta}{x-\gamma}\right)^{\alpha+1} \exp\left(-\left(\frac{\beta}{x-\gamma}\right)^\alpha\right)$	$x > 0, \alpha > 0, \beta > 0, \gamma > 0$

and 2010 (DRBC 2011). US Geological Survey stream gauge 01438500 (41° 18' 33" latitude and 74° 47' 43" longitude) at Port Jervis, New York is used in the study, as part of Delaware

River flows. The location has a drainage area of 9,013 km² and datum gauge is 112.7 m above sea level. About 61 years daily stream flow data from the years 1949 to 2009 are

Table 5 Performance of various probability models for fitting marginal distributions for flood variables

Flood variables	Distributions	Estimated parameter ^a	RMSE	AIC	AD _n	KS _n	<i>p</i> value ^b
Peak flow	Gamma	$\hat{\alpha} = 3.93$ (2.79, 5.52) $\hat{\beta} = 32.13$ (22.33, 46.22)	0.0408	-386.35	0.89	0.10	0.26
	Log normal	$\hat{\mu}_y = 4.70$ (4.57, 4.83) $\hat{\sigma}_y = 0.51$ (0.43, 0.62)	0.0330	-412.02	0.55	0.07	0.60
	Gumbel	$\hat{\beta} = 104.5$ (89.16, 122.40) $\hat{\gamma} = 168.7$ (140.74, 196.72)	0.1234	-251.31	27.1	0.24	0.001
	Fréchet	$\hat{\alpha} = 0.16$ (-0.03, 0.35) $\hat{\beta} = 43.42$ (34.93, 53.97) $\hat{\gamma} = 93.06$ (80.79, 105.33)	<i>0.0299</i>	<i>-422.02</i>	<i>0.47</i>	<i>0.06</i>	<i>0.77</i>
Volume	Gamma	$\hat{\alpha} = 2.24$ (1.61, 3.12) $\hat{\beta} = 126.5$ (87.25, 183.54)	0.0269	-437.30	0.44	0.07	0.75
	Log normal	$\hat{\mu}_y = 5.41$ (5.21, 5.60) $\hat{\sigma}_y = 0.75$ (0.63, 0.91)	0.0330	-412.23	0.86	0.10	0.27
	Gumbel	$\hat{\beta} = 400.0$ (317.85, 482.17) $\hat{\gamma} = 307.1$ (264.44, 356.64)	0.1077	-267.88	27.5	0.24	0.0007
	Fréchet	$\hat{\alpha} = 0.14$ (-0.06, 0.33) $\hat{\beta} = 126.0$ (101.4, 156.65) $\hat{\gamma} = 191.1$ (155.31, 226.9)	<i>0.0216</i>	<i>-461.67</i>	<i>0.30</i>	<i>0.06</i>	<i>0.77</i>
Duration	Gamma	$\hat{\alpha} = 7.12$ (5.04, 10.08) $\hat{\beta} = 1.64$ (1.14, 2.35)	0.0275	-434.57	0.44	0.08	0.52
	Log normal	$\hat{\mu}_y = 2.39$ (2.29, 2.48) $\hat{\sigma}_y = 0.38$ (0.32, 0.46)	<i>0.0254</i>	<i>-443.88</i>	<i>0.34</i>	<i>0.067</i>	<i>0.81</i>
	Gumbel	$\hat{\beta} = 14.20$ (12.67, 15.72) $\hat{\gamma} = 5.72$ (4.85, 6.74)	0.0807	-303.14	22.83	0.18	0.03
	Fréchet	$\hat{\alpha} = 0.06$ (-0.16, 0.28) $\hat{\beta} = 3.39$ (2.72, 4.21) $\hat{\gamma} = 9.52$ (8.54, 10.5)	0.0274	-432.75	0.38	0.074	0.81

Best estimate is shown in italics

^a Bracketed value denotes 95 % confidence interval for the estimate of parameters

^b *p* value for KS test statistics

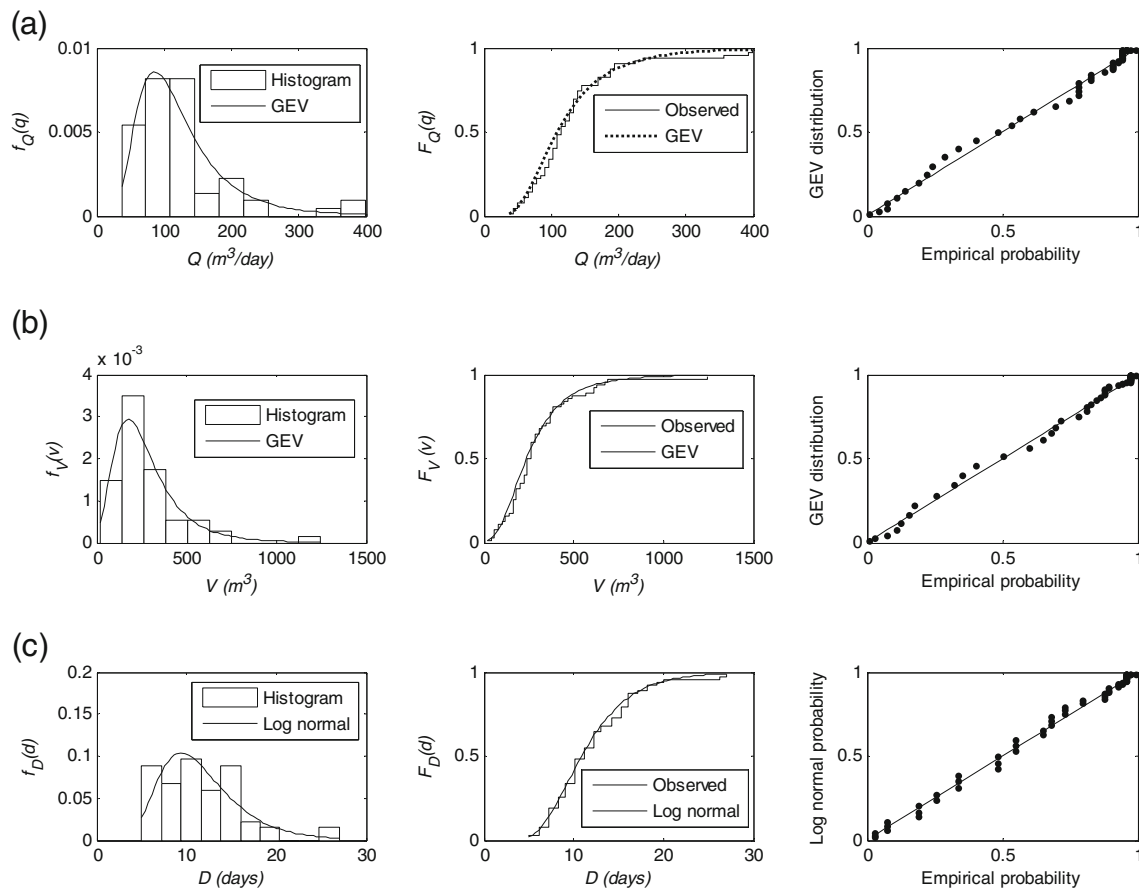


Fig. 3 Fitted marginal distributions for flood variables **a** peak flow, **b** volume, and **c** duration. The first, second, and third column shows probability density functions, CDFs, and P–P plots, respectively, for three flood characteristics

analyzed. Table 2 presents statistical properties of flood variables during the study period. The high kurtosis and positively skewed in nature of flood variables suggests that they can be best modeled by heavy tailed distributions.

6.2 Dependence of flood variables

The pair-wise association among flood variables as well as the strength of dependency is studied by employing graphical tools such as ranked scatter plots, chi-plots, and Kendall plots, since it is often difficult to judge random nature and nonlinear behavior of data from a simple scatter plot. In Fig. 2, ranked scatter plots, chi plots, and Kendall plots of pair-wise flood variables are presented. For peak flow–volume pair, the increased density of points close to 45° line in the ranked scatterplot indicates that dependence between this variable pair is strongest whereas a weak dependence is observed between peak flow and duration pair. The shape of the clusters of ranked observations for peak flow–volume and volume–duration pairs in regions around upper right corner of the unit square suggests presence of stronger upper tail dependence between these flood variables. The control limits for chi plot are set to enclose a p value of 0.95. A

strong deviation from the control limit is observed for flood variable peak flow–volume and volume–duration pair. For peak flow–duration pair, deviation is observed from the center of the main diagonal in Kendall plot, indicating presence of positive association between the flood variables. For quantitative assessment, the sample estimates of Pearson’s linear correlation r and two nonparametric dependence measures viz., Spearman’s ρ , Kendall’s τ with associated p values of the estimate are listed in Table 3. The dependencies are found to be significant as tested by standard two-tailed t test.

6.3 Modeling marginal distributions

For fitting marginal distribution two-parameter log normal, two-parameter gamma, extreme value type I (Gumbel distribution) and extreme value type II (Fréchet distribution) have been evaluated. The Gumbel and Fréchet distribution are special case of generalized extreme value (GEV) distribution. For Gumbel distribution, the shape parameter (α) of GEV takes the form of $\alpha \rightarrow 0$ and corresponds to an unbounded and thin upper tails. If $\alpha > 0$, then the GEV distribution is termed as Fréchet; the distribution is unbounded above and has polynomially decreasing tail function, which

Table 6 Performance of various bivariate copula models for representing dependence structure (joint distributions) of flood variables

Copula	Parameters	LL	AIC
Flood peak–volume			
Clayton	$\theta=1.801$	21.284	-40.568
Gumbel–Hougaard	$\theta=7.507$	26.428	-50.856
Frank	$\theta=2.402$	28.969	-55.937
Student’s t^a	$\vartheta=3.81, \sigma_{12}=0.794$	<i>29.155</i>	<i>-56.309</i>
Flood volume–duration			
Clayton	$\theta=1.425$	16.097	-30.194
Gumbel–Hougaard	$\theta=5.356$	17.071	-32.141
Frank	$\theta=1.823$	16.203	-30.405
Student’s t^a	$\vartheta=86.41, \sigma_{12}=0.716$	<i>18.889</i>	<i>-35.778</i>
Flood volume–duration			
Clayton	$\theta=0.383$	1.868	-1.737
Gumbel–Hougaard	$\theta=1.719$	2.260	-2.520
Frank	$\theta=1.229$	2.419	-2.838
Student’s t^a	$\vartheta=83.79, \sigma_{12}=0.323$	<i>2.458</i>	<i>-2.916</i>

Best estimate is shown in italics

LL log-likelihood value, AIC Akaike information criteria ($AIC=-2 \times \log(LL)+2 \times m$), where m =number of parameters of copula model

^a For t -copula, since $\sigma_{12}=\sigma_{21}$, the correlation matrix $\Sigma = \begin{bmatrix} 1 & \sigma_{12} \\ \sigma_{21} & 1 \end{bmatrix}$

corresponds to a long- or heavy-tailed distribution. The density form and corresponding parameters of the distributions are presented in Table 4. The parameters of the distribution functions are estimated using method of maximum likelihood. Table 5 lists various performance measures such as Akaike information criteria (AIC), Anderson–Darling (AD_n), and Kolmogorov–Smirnov (KS_n) statistics for the flood data fitted with parametric distributions. The table shows heavy tailed Fréchet distribution fits well for peak flow and volume data and log-normal distribution performed better for modeling flood duration data. The probability density functions (PDFs), cumulative distribution functions CDFs and corresponding probability–probability (P-P) plot for each marginal flood variables fitted with Fréchet distributions are shown in Fig. 3, which shows good correspondence between theoretical distributions with the observed data.

Table 7 Estimated parameters of trivariate models for various copula families

Copula classes	Parameters	Log-likelihood value	
Archimedean	Clayton	$\hat{\theta}_1 = 0.89, \hat{\theta}_2 = 1.71$	29.82
	Gumbel–Hougaard	$\hat{\theta}_1 = 1.44, \hat{\theta}_2 = 2.35$	35.88
	Frank	$\hat{\theta}_1 = 3.27, \hat{\theta}_2 = 7.21$	34.42
Elliptical	Student’s t	$\hat{\vartheta} = 5.08, \hat{\sigma}_{1,2} = 0.79, \hat{\sigma}_{2,3} = 0.72, \hat{\sigma}_{1,3} = 0.32$	59.77

Best estimate is shown in bold followed by next best estimator in italics

6.4 Modeling joint dependence structure using copulas

6.4.1 Bivariate copula models

First, bivariate copula models are developed for representing the joint dependence of flood variable pairs such as flood peak–volume, volume–duration, and flood peak–duration. Four copula models namely Clayton, Gumbel–Hougaard and Frank copulas, and Student’s t copula are fitted and their corresponding results are presented in Table 6. The parameters of copula functions are estimated using GA-based maximum pseudo-likelihood approach. The following GA parameters adopted: population size of 20, generations of 200, single point cross-over with cross-over rate of 0.8, Gaussian mutation function with mutation rate of 0.01, and selection strategy as stochastic uniform. From Table 6, it can be seen that the Student’s t copula resulted in higher log-likelihood and minimum AIC values for all three combinations of flood–variable pairs. These bivariate models are used for computing the return periods of bivariate and trivariate flood characteristics.

6.4.2 Trivariate copula models

The fully nested form of Clayton, Gumbel–Hougaard and Frank copulas, and one elliptical Student’s t copula are chosen to model trivariate flood characteristics. The copula functions are fitted with maximum pseudo-likelihood estimator, where parameters are estimated using genetic algorithm. The GA parameters used involves, for nested Archimedean class of copula: population size of 20, generations of 500, single point cross-over with cross-over rate of 0.8, Gaussian mutation function with mutation rate of 0.01, and selection strategy as stochastic uniform; whereas for Student’s t copula: a population size of 50, generations of 200, and the remaining parameters are the same as above. The estimated copula parameters and corresponding log-likelihood function values are presented in Table 7, which shows highest log-likelihood function value for Student’s t copula followed by Gumbel–Hougaard copula. The performance of copula families are compared using distance-based statistics AD, IAD and entropy tests, which are presented in Table 8. From this table, it can be seen that overall Student’s

Table 8 Comparison of copula models performance in representing the trivariate flood properties

Performance statistics	Clayton	Gumbel–Hougaard	Frank	Student's <i>t</i>
AD	0.22	0.29	1.67	0.11
IAD	0.65	0.85	7.67	0.21
Entropy	-190.68	-485.03	-283.23	-743.97

Best estimate is shown in bold followed by next best estimator in italics

t copula performed best for trivariate flood data as compared to other copula families. The low performance of Frank copula can be attributed to its radially symmetric structure whereas flood events are often upper tail dependent.

For further visual illustration, a set of random sample of size=500 are generated from both Gumbel–Hougaard and Student's *t* copulas and transformed back into their original units using corresponding marginal distribution functions and compared with observed data as shown in Fig. 4 (a–c for Student's *t* and d–f for Gumbel–Hougaard copula). From

these plots, it can be observed that the Student's *t* copula is performing satisfactorily, as the random pairs generated from this copula (gray dots) are adequately overlapped with the dependence pattern of sample data (black dots). Corresponding Kendall's τ value from the simulated data is also presented in Fig. 4. There is restriction in the application of fully nested Gumbel–Hougaard copula as the correlation between two pairs should be identical and lower than the third pair (such as rank-based correlation of $\tau_{P,V} \geq \tau_{V,D} = \tau_{P,D}$) as observed from the pair-wise simulated data.

The maximum distance, d_{\max} computed between non-exceedance probability and Student's *t* copula-based joint distribution is found to be 0.062. The critical values of the Kolmogorov–Smirnov test at 5 % significance level is $D_{n=61}^{\alpha=0.05} = 0.17$. Hence, Student's *t* copula is verified as an acceptable model at 5 % significance level. A scatter plot between the empirical joint CDF obtained using Grigorton plotting position formula and theoretical joint CDF obtained using Student's *t* copula is presented in Fig. 5. A good agreement between empirical and theoretical probabilities at the tails of the distribution is observed. Hence, Student's

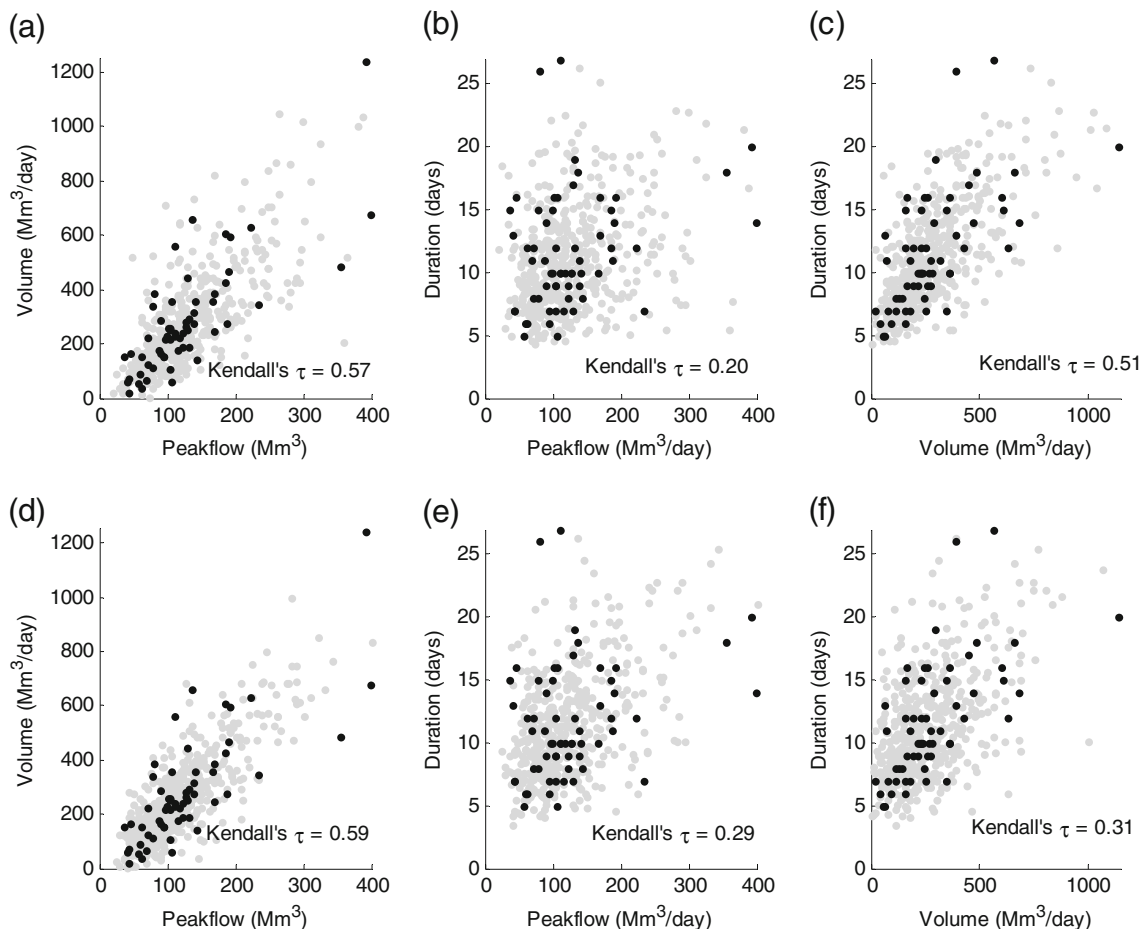


Fig. 4 Scatter plots of observed versus 500 simulated samples of flood variable pairs from Student's *t* copula (a–c) and fully nested Gumbel–Hougaard copula model (d–f). Solid black dots observed samples, light gray dots simulated samples

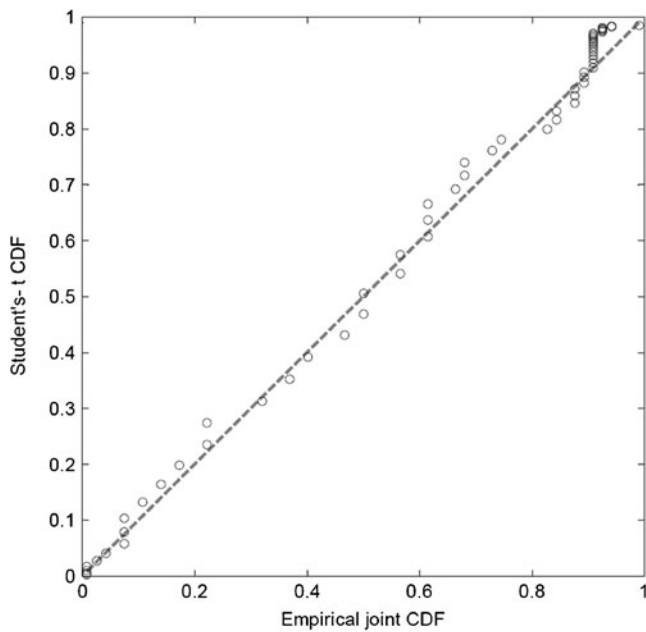


Fig. 5 Comparison of empirical joint probability distribution (obtained using plotting position formula) and theoretical joint probability distribution obtained from Student’s *t* copula

t copula is suitable for representing trivariate joint distribution of flood properties.

6.5 Analyzing tail dependence

In order to test the tail properties of simulated sample from the copula families, a pair-wise tail-dependence test is performed. The upper tail dependence coefficient for Clayton and Frank copulas are equal to zero. Hence, these two copulas are excluded from tail-dependence analysis. The TDC of empirically transformed (i.e., from ECDF) flood variable pairs is computed using $\hat{\lambda}_U^{cfg}$, $\hat{\lambda}_U^{LOG}$ and $\hat{\lambda}_U^{SS}$ estimators. For $\hat{\lambda}_U^{SS}$ and $\hat{\lambda}_U^{LOG}$ estimate of observed data (of sample length=61), the optimal plateau is selected using box-kernel of bandwidth ($b = [0.05n] = 3$). The empirical TDC of observed flood variable pairs is listed in Table 9.

For comparing TDC estimate of Gumbel–Hougaard and Student’s *t* copula, bivariate random numbers are generated for sample size 500. The optimal plateau for $\hat{\lambda}_U^{SS}$ and $\hat{\lambda}_U^{LOG}$ is

Table 9 Empirical tail dependence coefficient for three flood variable pairs

Flood variable pairs	$\hat{\lambda}_U^{cfg}$	$\hat{\lambda}_U^{LOG}$	$\hat{\lambda}_U^{SS}$
Peak flow–volume	0.68	0.63	0.67
Volume–duration	0.55	0.36	0.25
Peak flow–duration	0.29	0.12	0.29

selected using box-kernel of bandwidth ($b = [0.005n] \approx 3$). The computation is repeated for ten different runs (i.e., $\hat{\lambda}_{n,i}$, $i = 1, \dots, 10$). Then each TDC estimate is compared using sample mean $\hat{\mu}(\hat{\lambda}_U)$ and standard deviation $\hat{\sigma}(\hat{\lambda}_U)$ of ten different runs. The results are summarized in Table 10. The resulting box-plot for simulated tail-dependence coefficient of three different estimators is presented in Fig. 6. The asymmetrical nature of box-plots show that sample variance of $\hat{\lambda}_U^{LOG}$ and $\hat{\lambda}_U^{SS}$ is higher as compared to $\hat{\lambda}_U^{CFG}$. From Table 10 and Fig. 6, it can be observed that Student’s *t* copula is able to simulate observed upper TDC estimate well as compared to Gumbel–Hougaard copula, especially for $\hat{\lambda}_U^{CFG}$ estimator. The sample variance of $\hat{\lambda}_U^{CFG}$ estimator for Student’s *t* copula is less than that of Gumbel–Hougaard copula. The parametric upper TDC of Student’s *t* copula involving dependence parameter $\vartheta=5.08$ and σ_{ij} is (Frahm et al. 2005) $\lambda_{U_{ij}}^{param} = 2 - 2$

$t_{\vartheta+1} \left(\sqrt{(\vartheta + 1)} \sqrt{\frac{(1-\sigma_{ij})}{(1+\sigma_{ij})}} \right)$, $i, j \in \{1, 2, \dots, n\}$ where $t_{\vartheta+1}$ is a CDF of Student’s *t* random variable with $\vartheta+1$ degrees of freedom. The pair-wise parametric TDC estimate using Student’s *t* copula are: peak flow and volume $\lambda_{U_{pv}}^{param} = 0.43$, volume and duration $\lambda_{U_{vd}}^{param} = 0.35$ and peak flow and duration $\lambda_{U_{pd}}^{param} = 0.13$ respectively.

7 Probabilistic analysis of flood variables

7.1 Primary return periods of flood peak flow conditional to volume and duration

The frequency analysis of multivariate extreme events are helpful in understanding critical hydrologic behavior of flood in a River basin scale through use of various combinations of flood characteristics. It can also be helpful in taking nonstructural safety measures, and to delineate flood plains and developing flood mitigation strategies. From hydrological perspective, these scenarios are important, as extreme flood events with high peak flow and long duration and hence high volume may be devastating at watershed scale; on the other hand, short-duration events with high peak discharge and moderate volume may cause flash floods.

Table 11 presents return period obtained using univariate marginal distributions of peak flow, volume, and duration; and joint return periods for “AND” and “OR” cases for bivariate as well as trivariate distributions. The joint return period in “AND” case is longer than the joint return period in “OR” case when same univariate return period is assumed. The effect of including third variable (duration in this case) in computing joint return period can also be

Table 10 Bivariate upper TDC estimate for Student’s t and Gumbel–Hougaard copulas

Variable pair	Sample estimate	Student’s t			Gumbel–Hougaard		
		$\hat{\lambda}_U^{cfig}$	$\hat{\lambda}_U^{log}$	$\hat{\lambda}_U^{SS}$	$\hat{\lambda}_U^{cfig}$	$\hat{\lambda}_U^{log}$	$\hat{\lambda}_U^{SS}$
Peak flow—volume	$\hat{\mu}(\hat{\lambda}_U)$	0.653	0.628	0.656	0.662	0.651	0.727
	$\hat{\sigma}(\hat{\lambda}_U)$	0.008	0.021	0.016	0.014	0.043	0.050
Volume—duration	$\hat{\mu}(\hat{\lambda}_U)$	0.568	0.487	0.496	0.364	0.390	0.445
	$\hat{\sigma}(\hat{\lambda}_U)$	0.017	0.040	0.036	0.042	0.048	0.054
Peak flow—duration	$\hat{\mu}(\hat{\lambda}_U)$	0.264	0.204	0.256	0.362	0.376	0.426
	$\hat{\sigma}(\hat{\lambda}_U)$	0.022	0.043	0.059	0.032	0.055	0.107

$\hat{\mu}(\hat{\lambda}_U)$ and $\hat{\sigma}(\hat{\lambda}_U)$ denote sample mean and standard deviations, respectively

observed from Table 11. For all the cases considering trivariate flood properties, the joint return period in “AND” case $T_{QVD}^{AND}(q, v, d)$ is greater than the other cases of return periods; and the joint return period in “OR” case $T_{QVD}^{OR}(q, v, d)$ is lower than the pair-wise bivariate joint return periods. Hence, it also infers that the occurrence of trivariate flood characteristics simultaneously is less frequent in “AND” case and more frequent in “OR” case.

The joint return periods of two flood variables conditional on third flood variable, viz., $T_{Q,V|D}(q, v|D \leq d)$, $T_{V,D|Q}(v, d|Q \leq q)$ and $T_{Q,D|V}(q, d|V \leq v)$ are computed using Eq. 35. For example, let us consider a flood event with the following flood characteristics: annual maximum peak discharge, $q=398.8$ Mm³/day; flood volume, $v=676.6$ Mm³; and flood duration, $d=14$ days. By using Eq. 35, the corresponding conditional return periods are obtained as $T_{Q,V|D}(q, v|D \leq d)=106$ years, $T_{V,D|Q}(v, d|Q \leq q)=4$ years,

$T_{Q,D|V}(q, d|V \leq v)=4.4$ years. The contour plots of specific joint return periods for all the three combinations are plotted in Fig. 7. As seen from the scatterplots, most of the flood events have shorter primary return periods. However, there exists only one event both in $T_{Q,V|D}(q, v|D \leq d)$ and $T_{V,D|Q}(v, d|Q \leq q)$ cases, which have primary return period more than 100 years. One can obtain desired information for a particular use from contours of joint return periods (viz., flood volume and duration at given peak discharge, flood peak and duration at given flood volume, and flood peak and volume at given duration). This information will be helpful for design of flood control structures, reservoirs, spillways etc., where it seek design flood hydrographs.

The conditional return periods of peak flow at given volume and constant duration, and conditional return period of peak flow at given duration and constant volume are presented in Fig. 8a and b, respectively. It should be noted that in Fig. 8, the volume and duration values correspond to various percentile

Fig. 6 Box–whisker plots of upper TDC for simulated sample from copula. First row (a–c) shows Student’s t copula, second row (d–f) depicts Gumbel–Hogaard copula family. $\lambda_{U_{PQ}}$, $\lambda_{U_{PD}}$ and $\lambda_{U_{VD}}$ denotes upper TDC (λ_U) estimate of peak flow–volume, peak flow–duration, and volume–duration combination

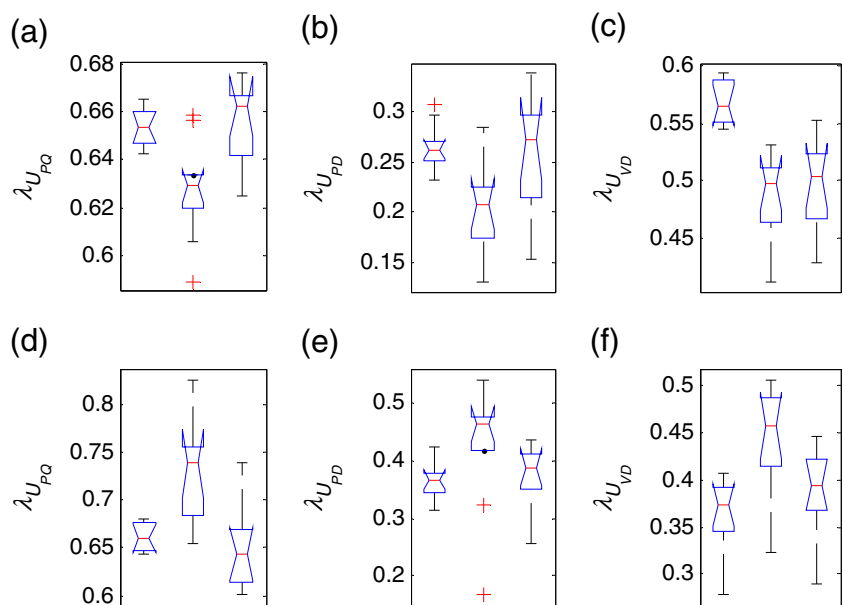


Table 11 Comparison of univariate, bivariate, and trivariate return periods for flood characteristics

T	Peak flow (Mm ³ /day)	Volume (Mm ³)	Duration (days)	T_{QV}^{AND}	T_{VD}^{AND}	T_{QD}^{AND}	T_{QV}^{OR}	T_{VD}^{OR}	T_{QD}^{OR}	T_{QVD}^{AND}	T_{QVD}^{OR}	\bar{T}_{QVD}
5	166.56	400.89	15	7.6	8.6	13.7	3.7	3.5	3.1	14.8	2.9	9.2
10	210.41	523.14	17.7	16.7	20.3	36	7.1	6.6	5.8	40.9	5.3	23.8
20	257.64	652.8	20.3	35.5	46.8	87.8	13.9	12.6	11.2	109	10.1	54.2
50	327.34	840.74	23.8	93.6	140.6	269.3	34.1	30.5	27.7	398.2	24.5	186
100	386.71	998.13	26.4	192.4	316.2	597.7	67.7	59.6	54.8	1,055.5	48	458.7

T return period associated with univariate marginal distributions, \bar{T}_{QVD} secondary return period of flood variables

values of the respective data. The skewness of the conditional return period curves for flood peak shows increasing trend for the first case (i.e., at given conditional volume for various percentile levels and constant duration), and decreasing trend for the second case (i.e., at given conditional duration and decreasing trends for the second case and constant volume). In a similar way, conditional return period of flood volume given peak discharge and duration, and/or conditional return period of flood duration given volume and peak discharge can also be obtained. These scenarios can be helpful in assessing flood risk for hydrologic

design purposes such as design of spillways and construction of various flood protection structures (for example levees, flood walls, diversion works).

7.2 Analysis of secondary return period

The secondary return period can be useful for analyzing risk of supercritical flood events. The relationship between joint primary return period $T_{QVD}(q,v,d)$ and the secondary return period $\bar{T}_{QVD}(q,v,d)$ is also investigated. The relationships

Fig. 7 Contour plots for conditional return periods (in years) of flood characteristics: **a** $T_{QV|D}(q,v|D \leq d)$, **b** $T_{VD|Q}(v,d|Q \leq q)$, and **c** $T_{QD|V}(q,d|V \leq v)$. Historical events are shown as solid dots on the graph

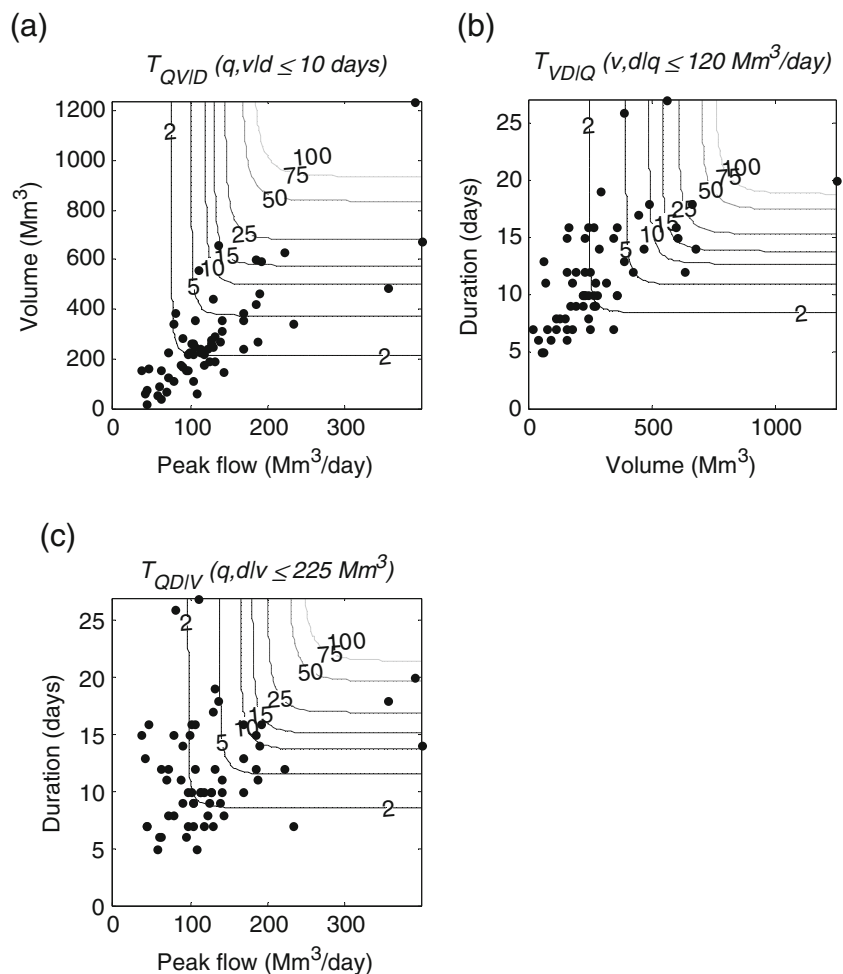
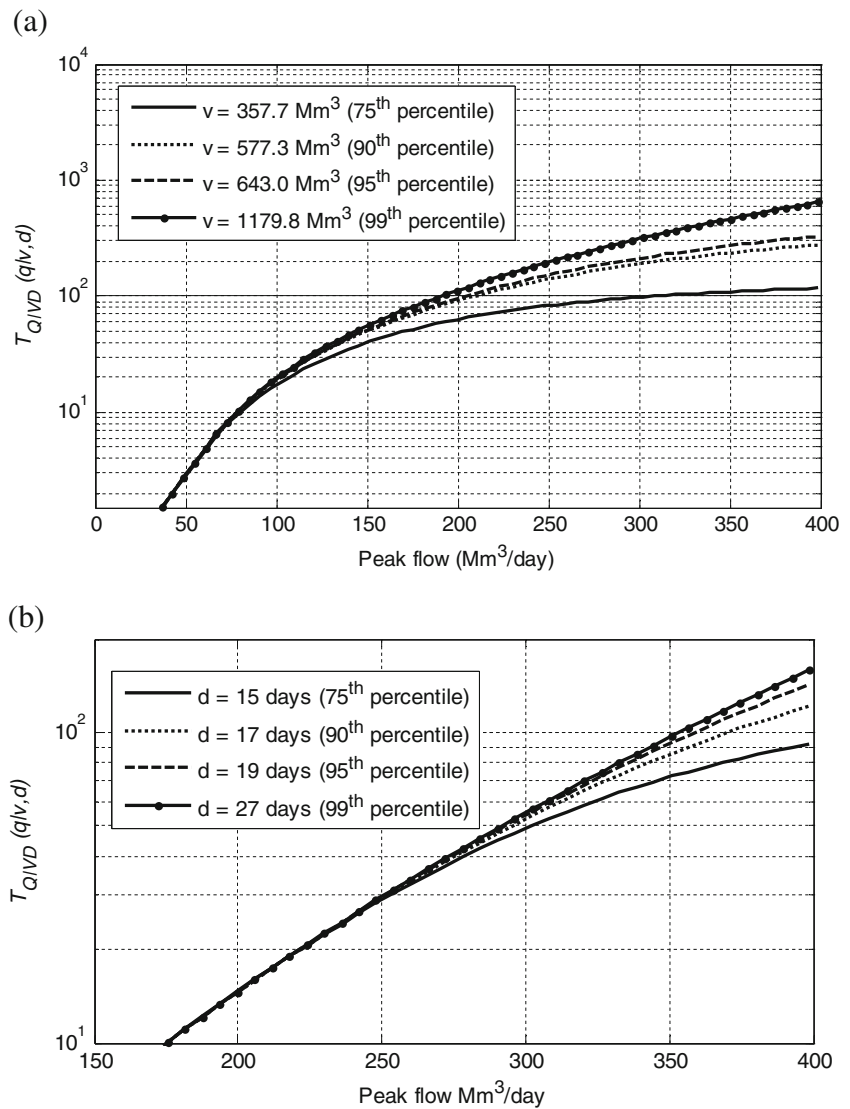


Fig. 8 Conditional return periods (in years) of flood characteristics: **a** $T_{Q|VD}(q|V \leq v, D \leq d)$ of peak flow at given volume and constant duration ($d=5$ days); **b** $T_{Q|VD}(q|V \leq v, D \leq d)$ of peak flow at given duration and constant volume ($v=120 \text{ Mm}^3$)



of $\bar{K}_{C_{\Sigma, \theta}}$ and the secondary return period $\bar{T}_{QVD}(q, v, d)$ against primary return period in ‘OR’-case $T_{QVD}^{OR}(q, v, d)$ are shown in Fig. 9a and b, respectively. For higher value of critical probability level \bar{p} , higher primary return period is observed, and the survival probability function $\bar{K}_{C_{\Sigma, \theta}}(\bar{p})$ tend to decrease, which corresponds to decrease in probability of supercritical events. As the probability of supercritical events decreases, the secondary return periods or mean occurrence time between two supercritical events increases and vice versa. It is also observed that the secondary return period is always greater than that of primary return period. A similar behavior is observed between $\bar{K}_{C_{\Sigma, \theta}}$ and the primary return period in ‘AND’ case $T_{QVD}^{AND}(q, v, d)$. It is also observed that the secondary return period is always greater than that of primary return period in OR case $T_{QVD}^{OR}(q, v, d)$, but lower than that of primary return period in AND case $T_{QVD}^{AND}(q, v, d)$.

An illustration is the 20-day flood event that occurred during 2006, which has peak flow of $391.45 \text{ Mm}^3/\text{day}$ and volume of $1,242 \text{ Mm}^3$ (which corresponds to 99th percentile value of flood volume). The secondary return period associated with this extreme event using Eq. 39 is more than 100 years ($\bar{T}_{QVD}(q, v, d) \approx 108 \text{ years}$) whereas primary return period associated with the event for ‘OR’ and ‘AND’ cases are $T_{QVD}^{OR}(q, v, d) = 16.5 \text{ years}$ and $T_{QVD}^{AND}(q, v, d) = 507.1 \text{ years}$. Similarly, for a 14-day flood event that occurred during 1955, with peak flow $q = 398.79 \text{ Mm}^3/\text{day}$ (which corresponds to 99th percentile value of annual maximum flood peak) and volume $v = 676.58 \text{ Mm}^3$, has secondary return period $\bar{T}_{QVD}(q, v, d) = 14.5 \text{ years}$, while primary return periods $T_{QVD}^{OR}(q, v, d)$ is 3.8 years and $T_{QVD}^{AND}(q, v, d)$ is 175 years, respectively. The secondary return period of flood variables presented in Table 11 also support this inference. This means that the structure could be underdimensioned if it is designed

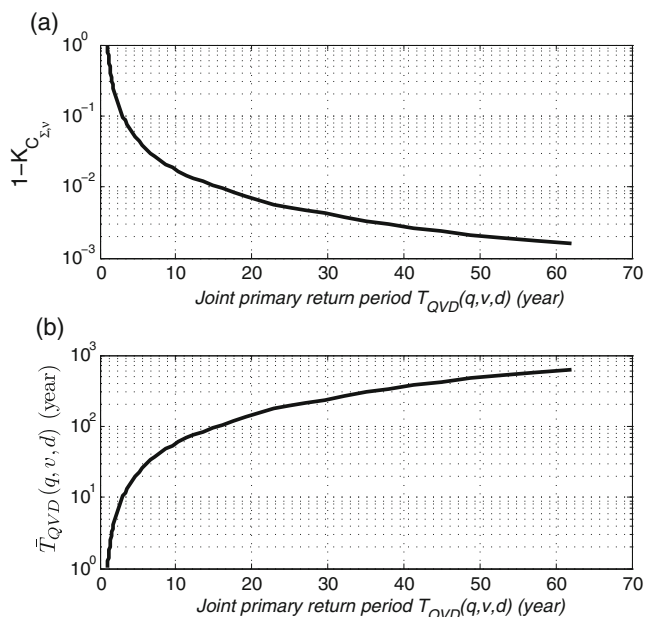


Fig. 9 The relation of primary return period $T_{QVD}^{OR}(q, v, d)$ versus **a** survival probability function $\bar{K}_{C_{\Sigma, \theta}}(\bar{p}) = 1 - K_{C_{\Sigma, \theta}}(\bar{p})$ and **b** the secondary return period $\bar{T}_{QVD}(q, v, d)$

considering only the primary return period in “OR” case and overdimensioned if it is designed with the primary return period in “AND” case. Thus, the presented methodology can be very useful for risk based hydrological design of hydraulic and water resources projects.

However, in recent times many concerns are raised over climate change and its influence on hydrological extremes. It would be interesting to see how the climate change may affect future floods, and quantifying the associated risks. These tasks can be considered for future studies to have comprehensive evaluation of flood risks.

8 Summary and conclusions

In this study, a trivariate copula-based approach is presented for risk analysis of flood flows of Delaware River basin at Port Jervis, New York. The association among the three mutually correlated flood properties (viz., annual flood peak flow, volume, and duration) is explored and used for dependence structure modeling using copulas. From graphical tests and rank correlations of flood variable pairs, it is found that the dependence among the flood variables is statistically significant and therefore copula based methodology is adopted for flood risk analysis. Three fully nested forms of Archimedean copulas (viz., Clayton, Gumbel–Hougaard, Frank copulas); and one elliptical copula (i.e., Student’s *t*

copula) are used to model the joint distribution of flood variables. To test the performance of copulas in modeling flood extreme events, apart from standard performance measures, the nonparametric tail-dependence coefficients are evaluated and used to verify the efficacy of copulas in modeling hydrological extremes. The return periods for univariate, bivariate, and trivariate cases are estimated and comparative analysis is performed. Also, the importance of primary and secondary return periods is analyzed.

The specific conclusions of the study are as follows:

- The copula method is found to be very effective tool for multivariate modeling of flood risks, as copulas are effectively preserving the dependence structure of multiple flood characteristics.
- In general, the upper tail-dependent copula families performed better in modeling extreme flood events and also capable of capturing tail behavior of data well. In relative comparison, the Student’s *t* copula is found to be better than the other copulas in describing bivariate and trivariate dependence structure of flood variables. This may be because of the weak dependences are averaged in other copulas. Also, in case of Student’s *t* copula, the nonparametric tail probability measures showed a close similarity with observed TDC, which suggests that the Student’s *t* copula is well capturing the extremes of the observed data.
- The comparative analysis of different return periods showed that it is very important to compute trivariate return periods of flood characteristics to know the expected flood risks and their magnitude of influence if they occur simultaneously. It also noted that the hydraulic structures could be underdimensioned if it is designed considering only the return period in “OR” case; and overdimensioned if it is designed with the return period in “AND” case. The primary and secondary return periods of flood characteristics can be very useful for effective risk-based design of water resources projects.

Appendix

Chi plot

Chi-plot is a scatter plot of the pairs (λ_i, χ_i) , where it uses the data ranks (Fisher and Switzer 2001). The values are defined as,

$$\chi_i = \frac{H_i - F_i G_i}{\sqrt{F_i(1 - F_i)G_i(1 - G_i)}} \tag{A1}$$

where

$$\begin{aligned}
 H_i &= \frac{\sum_{j \neq i} I(x_j \leq x_i, y_j \leq y_i)}{n-1}; \\
 F_i &= \frac{\sum_{j \neq i} I(x_j \leq x_i)}{n-1}, \text{ and} \\
 G_i &= \frac{\sum_{j \neq i} I(y_j \leq y_i)}{n-1} \text{ and} \\
 \lambda_i &= 4 \operatorname{sgn}_i \max\{(F_i - 0.5)^2, (G_i - 0.5)^2\}
 \end{aligned}
 \tag{A2}$$

where $\operatorname{sgn}_i = \operatorname{sgn}\{(F_i - 0.5)(G_i - 0.5)\}$. The values of λ_i is a measure of the distance of the bivariate observation (x_i, y_i) from the center of the data set and $\lambda_i \in [-1, 1]$. If X and Y are positively associated, λ_i will tend to be positive and vice versa for negatively correlated variable. The chi-plot also include control limits that are placed at $\chi = \pm c_p / \sqrt{n}$, where c_p values are provided by Fisher and Switzer (2001). Using simulation studies, it is found that C_p values 1.54, 1.78, and 2.18 correspond to p values of 0.90, 0.95, and 0.99, respectively. In case of independence, the points lie within the control limits of the plot. When the scatter of data points are largely on the upper side of the control limits, it indicates positive dependence; whereas in case of negative dependence the scatter of data points are largely in the lower side of the control limits.

Kendall plot

The Kendall plots are analogous to $Q-Q$ plots (Genest and Boies 2003). The Kendall plot, consists of plotting random variable $W_{i:n}$ in ordinate against order statistics of the sample $H_{(i)}$ in abscissa for $i \in \{1, \dots, n\}$. $W_{i:n}$ is the expected value of the i th order statistics of a random variable $W = H(X, Y) = C(U, V)$ of sample size of n from distribution K_n and is defined as (Genest and Favre 2007)

$$\begin{aligned}
 W_{i:n} &= n \binom{n-1}{i-1} \\
 &\times \int_0^1 wk(w) \{K(w)\}^{i-1} \{1 - K(w)\}^{n-1} dw
 \end{aligned}
 \tag{A3}$$

where $k = \frac{dK(w)}{dw}$; $K(w) = P(UV \leq w) = \int_0^1 P(U \leq w/v) dv = w - w \log(w)$ and $\binom{n-1}{i-1}$ is a binomial coefficient. The deviation from the main diagonal is the indication of dependence in Kendall plots. The plot tends to be linear in case of

independence whereas the curve tends to be $K(t) = w - w \log(w)$ in case of perfect positive dependence. In case of perfect negative dependence, all points will lie on x axis and a flat line will result with height zero.

References

Adamson PT, Metcalfe AV, Parmentier B (1999) Bivariate extreme value distributions: an application of the Gibbs sampler to the analysis of floods. *Water Resour Res* 35(9):2825–2832

Ané T, Kharoubi C (2003) Dependence structure and risk measure. *J Bus* 76(3):411–438

Capéraá P, Fougères A-L, Genest C (1997) A non-parametric estimation procedure for bivariate extreme value copulas. *Biometrika* 84(3):567–577

Choulakian V, El-Jabi N, Issa M (1998) On the distribution of flood volume in partial duration series analyses of flood phenomenon. *Stoch Hydrol Hydraul* 4(3):217–226

Chow VT, Maidment DR, Mays LW (1988) *Applied hydrology*. McGraw Hill, New York, p 572

Chowdhary H, Escobar LA, Singh VP (2011) Identification of suitable copulas for bivariate frequency analysis of flood peak and flood volume data. *Hydrol Res* 42.2(3):193–216

Coles SG, Heffernan J, Tawn JA (1999) Dependence measures for extreme value analyses. *Extremes* 2:339–365

DRBC (2011) Delaware River Basin Commission—information. Available from: www.drbc.net. Accessed on 10 Apr 2011

Escalante C (2007) Application of bivariate extreme value distribution to flood frequency analysis: a case study of Northwestern Mexico. *Nat Hazard* 42(1):37–46

Escalante C, Raynal JA (1998) Multivariate estimation of floods: the trivariate Gumbel distribution. *J Stat Comp Simul* 61:313–340

Escalante C, Raynal JA (2008) Trivariate generalized extreme value distribution in flood frequency analysis. *Hydrol Sci* 53(3):550–567

Favre AC, Adlouni SE, Perreault L, Thiemonge N, Bobée B (2004) Multivariate hydrological frequency analysis using copulas. *Water Resour Res*. doi:10.1029/2003WR002456

Fisher NI, Switzer P (2001) Graphical assessment of dependence: is a picture worth 100 tests? *Am Stat* 55(3):233–239

Frahm G, Junker M, Schimdt R (2005) Estimating the tail-dependence coefficient: properties and pitfalls. *Insur Math Econ* 37:80–100

Genest C, Boies JC (2003) Detecting dependence with Kendall plots. *Am Stat* 57(4):275–284

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

Genest C, Favre AC, Beliveau J, Jacques C (2007) Metaelliptical copulas and their use in frequency analysis of multivariate hydrological data. *Water Resour Res*. doi:10.1029/2006WR005275

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

Goel NK, Seth SM, Chandra S (1998) Multivariate modeling of flood flows. *J Hydraul Eng* 124(2):146–155

Grimaldi S, Serinaldi F (2006) Asymmetric copula in multivariate flood frequency analysis. *Adv Water Resour* 29(8):1155–1167

Joe H (1997) *Multivariate models and dependence concepts*. Vol. 73. Monographs on statistics and applied probability. Chapman and Hall, New York, p 399

- Joe H, Smith RL, Weissman I (1992) Bivariate threshold models for extremes. *J R Stat Soc B* 54:171–183
- Krstanovic PF, Singh VP (1987) A multivariate stochastic flood analysis using entropy. In: Singh VP (ed) *Hydrologic frequency modeling*. Reidel, Dordrecht, pp 515–539
- Mashal R, Zeevi A (2002) Beyond correlation: extreme co-movements between financial assets. Technical report, Columbia University
- McNeil AJ, Frey R, Embrechts P (2005) *Quantitative risk management: concepts, techniques, and tools*. Princeton University Press, Princeton, p 538
- Nadarajah S, Shiao JT (2005) Analysis of extreme flood events for the Pachang River, Taiwan. *Water Resour Manag* 19:363–374
- Nelsen RB, Quesada-Molina JJ, Rodríguez-Lallena JA, Úbeda-Flores M (2008) On the construction of copulas and quasi-copulas with given diagonal sections. *Insur Math Econ* 42:473–483
- Nelson RB (2006) *An introduction to copulas*. Springer, New York
- Okhrin O, Okhrin Y, Schmid W (2009) Properties of hierarchical Archimedean copulas. SFB 649 discussion paper no. 2009-014. Humboldt University, Berlin
- Raynal JA, Salas JD (1987) A probabilistic model for flooding downstream of the junction of two rivers. In: Singh VP (ed) *Hydrologic Frequency Modeling*. Reidel, Dordrecht, pp 595–602
- Salvadori G (2004) Bivariate return periods via 2-copulas. *Stat Methodol* 1:129–144
- Salvadori G, De Michele C (2004) Frequency analysis via copulas: theoretical aspects and applications to hydrological events. *Water Resour Res*. doi:10.1029/2004WR003133
- Salvadori G, De Michele C, Durante F (2011) Multivariate design via copulas. *Hydrol Earth Syst Sci Discuss* 8(3):5523–5558
- Savu C, Tiede M (2010) Hierarchies of Archimedean copulas. *Quant Finance* 10(3):295–304
- Schmidt R, Stadtmüller U (2006) Non-parametric estimation of tail dependence. *Scand J Stat* 33:307–335
- Serinaldi F, Grimaldi S (2007) Fully nested 3-copula: procedure and application on hydrologic data. *J Hydrol Eng* 12(4):420–430
- Sklar A (1959) Fonctions de repartition a n dimensions et leurs marges. *Publ Inst Stat Univ Paris* 8:229–231
- Stedinger JR, Vogel RM, Foufoula-Georgiou E (1993) Frequency analysis of extreme events. In: Meidment DR (ed) *Handbook of hydrology*. McGraw-Hill, New York, pp 18.1–18.66
- Vandenberghe S, Verhoest NEC, Onof C, De Baets B (2011) A comparative copula-based bivariate frequency analysis of observed and simulated storm events: a case study on Bartlett–Lewis modeled rainfall. *Water Resour Res*. doi:10.1029/2009WR008388
- Whelan N (2004) Sampling from Archimedean copulas. *Quant Finance* 4(3):339–352
- Yue S (1999) Applying the bivariate normal distribution to flood frequency analysis. *Water Int* 24(3):248–252
- Yue S (2000) The bivariate lognormal distribution to model a multivariate flood episode. *Hydrol Processes* 14:2575–2588
- Yue S (2001a) A bivariate extreme value distribution applied to flood frequency analysis. *Nordic Hydrol* 32:49–64
- Yue S (2001b) A bivariate gamma distribution for use in multivariate flood frequency analysis. *Hydrol Processes* 15(6):1033–1045
- Yue S, Ouarda TBMJ, Bobée B, Legendre P, Bruneau P (1999) The Gumbel mixed model for flood frequency analysis. *J Hydrol* 226(1–2):88–100
- Yue S, Ouarda TBMJ, Bobée B (2001) A review of bivariate gamma distributions for hydrological application. *J Hydrol* 246(1–4):1–18
- Yue S, Rasmussen P (2002) Bivariate frequency analysis: discussion of some useful concepts in hydrological application. *Hydrol Process* 16:2881–2898
- Zhang L, Singh VP (2007) Trivariate flood frequency analysis using the Gumbel–Hougaard copula. *J Hydrol Eng* 12(4):431–439