



Bivariate flood distribution analysis under parametric copula framework: a case study for Kelantan River basin in Malaysia

Shahid Latif¹ · Firuza Mustafa¹

Received: 30 October 2019 / Accepted: 25 April 2020 / Published online: 14 May 2020
© Institute of Geophysics, Polish Academy of Sciences & Polish Academy of Sciences 2020

Abstract

Flood is becoming an intensive hydro-climatic issue at the Kelantan River basin in Malaysia. Univariate frequency analysis would be unreliable due to multidimensional behaviour of flood, which often demands multivariate flow exceedance probabilities. The joint distribution analysis of multiple interacting flood characteristics, i.e. flood peak, volume and duration, is very useful for understanding critical hydrologic behaviour at a river basin scale. In this paper, a copula-based methodology is incorporated for multivariate flood frequency analysis for the 50-year annual basis flood characteristics of Kelantan River basin at Guillemard bridge station in Malaysia. Investigation reveals that the Lognormal (2P), Johnson SB-4P and Gamma-3P are selected as marginal distributions for the flood peak flow, volume and duration series. Several bivariate families such as mono-parametric, bi-parametric (i.e. mixed version) and rotated version of Archimedean copulas and also the elliptical copula are introduced to cover a large dependence pattern of flood characteristics. The dependence parameter of bivariate copulas is estimated by the method of moments (MOM) based on the inversion of Kendall's tau and maximum pseudo-likelihood estimator. To analytically validate and recognize most parsimonious copulas, GOF test and Cramer–von Mises distance statistics (S_n) with the parametric bootstrap method are employed. The Gaussian copula is identified as the most justifiable model for joint modelling of the flood peak–volume and peak–duration combination for MOM-based parameter estimation procedure. Similarly, the Frank copula is selected as the best-fitted structure for modelling peak–duration combination based on MPL estimators, but the MOM estimator recognized Gaussian copula as most suitable for peak–volume pair. Furthermore, the best-fitted copulas are used for obtaining the joint and conditional return periods of the flood characteristics.

Keywords Flood · Kelantan River basin · Copula function · Joint distribution · Conditional distribution · Return period

Introduction

Probabilistic assessments of flood episodes for estimating the flow exceedance probability or design variable quantiles in the light of univariate or multivariate joint distribution framework often facilitate an insightful concern or design criteria in the planning and designing of hydraulic or flood defence infrastructure projects (i.e. dam spillway, river channel, dikes, etc.), urban drainage system or flood hazard mapping (Cunane 1987; Bobee and Rasmussen 1994; Goel et al. 1998; Sen

1999; Rao and Hameed 2000; Katz et al. 2002; Poulin et al. 2007 and references therein). According to Rakhecha and Sigh (2009), flood signifies an inundation due to the river overflowing their banks in respect of heavy or intense rainfall structure or either melting of a large amount of snow. Flood is a multidimensional stochastic consequence usually characterized completely through its trivariate intercorrelated vectors such as flood peak discharge flow, volume and duration of flood hydrograph (Zhang 2005; Zhang and Singh 2006). Flood frequency analysis (FFA) statistically defines through an inter-association between flood quantiles with nonexceedance probabilities or return period by fitting the probability distribution functions or PDFs (Goel et al. 1998; Yue 1999; Yue 2001; Yue and Rasmussen 2002). In other words, it is an approach to relate the strength of extreme event quantiles to their frequency of occurrence in the light of probability distribution framework.

Hydro-meteorological simulations based on the extension of historical rainfall samples to recognize catchments

✉ Shahid Latif
macet.shahid@gmail.com

Firuza Mustafa
firuza@um.edu.my

¹ Department of Geography, Faculty of Arts and Social Sciences, University of Malaya, 50603 Kuala Lumpur, Malaysia

profile or based on joint simulations in conjunction with univariate or multivariate statistical framework over the variables of interest are the two distinct ways to address the risk assessments for the extreme flood scenario. Numerous attempts such as Calver and Lamb (1995), Blazkova and Beven (2004) and Lawrence et al. (2014) retrieved flood frequency curve through integrating the hydrologic models in conjunction with probabilistic rainfall models for demonstrating catchment's rainfall–runoff profile. Such incorporation usually adapted the conventional-based lumped and distributed models or either via the continuous or event-based hydro-climatic simulations. But, all such incorporations often required longer computational analysis to justify the demands of high spatial and temporal resolutions for revealing a satisfactory demonstration of flood stimulations procedure and thus would attribute for an ineffective characterization of catchments behaviour (Requena et al. 2016).

Earlier efforts frequently incorporated the univariate frequency analysis of flood series for estimating univariate return periods, i.e. either flood peak or volume as a function with their nonexceedance probabilities (Cunnane 1987, Bobee and Rasmussen 1994). But many studies highlighted the unreliability of univariate return period, which would be incapable to providing the full screen of flood or inflow hydrograph and also underestimation (i.e. low design value might increase the risk of failure) or overestimations (i.e. increasing the cost of hydraulic construction) of associated hydrologic risk of correlated flood characteristics (Yue et al. 1999; Yue and Wang 2004). For example, flood episodes with peak discharge value of 100-year return period would be less damaging than the same flood events describing through the joint concurrency of flood peak–volume/or peak–durations/or volume–durations. Actually, the potential damage could likely be a function of several associated flood variables and ignorance of spatial dependency among its multiple vectors might be attributed for underestimation of uncertainty distributed over the estimated design quantiles and thus often demands more flood variables (Graler et al. 2013), in particular, from the prospects of hydraulic designing procedures where accountability of multivariate design parameters could be a feasible desire based on their multivariate exceedance probabilities (Salvadori 2004; Reddy and Ganguli 2012a).

The necessity of estimating design hydrograph or multivariate return periods instead of just estimating the design quantiles derived from the single variable flood episodes motivated numerous literature towards the incorporation of joint distribution analysis among multiple flood characteristics, i.e. between peak–volume, or peak–duration, etc., through introducing a distinguished variety of traditional bivariate or few trivariate probability distribution functions, i.e. Krstanovic and Singh (1987), Escalante and Raynal (1998), Yue (1999), Yue (2000, 2001), Yue and Wang (2004), Nadarajah and Shiau (2005), Escalante

and Raynal (2008) and references therein. But such distribution-based flood modelling approach or FFA often surrounded with several statistical constraints and limitations such as (1) each flood vector must assume to have Gaussian or normal distributions or either transformed to have normal distributions; (2) if the number of variables is increased, then mathematical formulation becomes more complex and complicated; (3) statistical parameter of univariate marginal structure is often employed to model their joint dependence structure; (4) limited space is available to justify joint dependence structure, etc., (Zhang 2005; Zhang and Singh 2006; Schmidt et al. 2007; Song and Singh 2010). Besides this, conventional multivariate models often attribute for the heavy dependency of flood exceedance on the right tail which might result for complexity during the demonstrations of observed samples; thus, it could be demanding for the separate modelling of margins from their joint dependence structure for securing their joint association significantly (Zhang and Singh 2006; Reddy and Ganguli 2012a).

Therefore, after encountering the above limitations it motivated firstly, the De Michele and Salvadori (2003) firstly incorporated copulas function, for establishing the joint dependence structure of storm intensity and duration series. The copula function is recognized as a highly flexible multivariate tool which segregated the modelling of individual univariate flood vectors and their joint dependence structure separately into two distinct stages and thus attributes the higher flexibility in the selection of best-fitted marginal distributions and their joint structure to capture a wider extent of mutual concurrency and preservation in their joint association (Saklar 1959; De Michele and Salvadori 2003; Salvadori 2004; Salvadori and De Michele 2004; Nelsen 2006). Genest and Favre (2007) provided an extended review of the application of copulas in the field of engineering and sciences. A series of literature incorporated copulas function, for example, for flood samples (i.e. Salvadori and De Michele 2004; De Michele et al. 2005; Grimaldi and Serinaldi 2006; Poulin et al. 2007; Genest et al. 2007; Fan and Zheng 2016; and references therein), for rainfall characteristics (i.e. Salvadori and De Michele 2006), for drought modelling (i.e. Shiau 2006). Besides their extended applicability in extreme event modelling, copula-based joint distributions are significantly applied in the field of groundwater modelling (i.e. Reddy and Ganguli, 2012b) and also modelling of hydro-climatic samples (i.e. Cong and Brady 2011), etc. For the essential mathematical terminologies and theorems associated with copula function, readers are advised to follow the literature such as Saklar (1959) and Nelsen (2006), also 'International Association of Hydrological Sciences (or IAHS)' for extended details and lists of their applicability in the field of hydro-climatological observations.

The Kelantan River basin is one of the largest basins of Malaysia, which are known to be flood-prone. From the past few decades, it is often subjected to the most severe monsoonal flooding and perceived for increasing in terms of their frequency and magnitude (DID 2000, 2003, 2004; MMD 2007; Adnan and Atkinson 2011). Few studies, such as Chan (1997) and DID (2000, 2003, 2004), report the expectation of the occurrence of catastrophic flooding has increased from once in every 50–15 years from 2004 in Kelantan. For example, intense and prolonged precipitation in the year 2002 caused flooding of a total area of 1640 km² and affected the population of 714,287 (Adnan and Atkinson 2011). Similarly, in the early month of December 2014, much heavy precipitation occurred for many of days triggered the flood event in several parts of the east coast of the Kelantan river basin and it was the worst flood ever recorded in history and affected more than 200,000 people (Adnan and Atkinson 2011; Rahman et al. 2018). The maximum length and breadth of this catchment area are 150 km and 140 km, respectively, and it is about 248 km long and drains an area of 13,100 km², occupying more than 85% of the state of Kelantan. The basin has an annual rainfall of about 2500 mm much of which occurs during the north-east monsoon (or wet season) between mid-October and mid-January. Few studies are performed over this river basin; for example, according to Hussain and Ismail (2013) investigation, Guillemard Bridge, Lebir and Galas stations have highest in flood frequency rather than Nenggiri station and also the value of damaged property got increased according to the frequency of flood happening. Also, Abdulkareem and Sulaiman (2015) investigated the variability of precipitation in flood source area of Kelantan river basin through the trend analysis using the annual maximum series of 24-h precipitation data and annual maximum flood data and revealed that no statistically significant trend was detected in the annual maximum series of 24-h precipitation between for the period 1984–2014 while the AMF series were significant at 5% level at the targeted locations. Similarly, Nashwan et al. (2018) revealed that the downstream area of the Kelantan River basin is the highest risk of devastating flood events. Besides this, a lot of attention has been also pointed out in this region in the context of the impact of land-use changes which may affect the catchments response; for example, Hassan (2004) report revealed the substantial land-use changes in this region which might influence the rate of evapotranspiration and infiltrations (Wooldridge et al. 2001). Wan (1996) and Jamaliah (2007) literature also pointed the existence of rapid land-use changes from the year 1970 to 2000s, mainly due to deforestation and conversion from natural land into agriculture for oil palm and rubber.

The cause of frequent failure of hydrologic or flood defence infrastructure in Malaysia due to the impact of moderately severe of flood episodes might be attributed

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. For example, according to Gaál et al. (2015) in the designing of retention basins and spillways of reservoirs or any other flood defence hydraulic structures where the storage is involved, in such circumstances the estimation of hydrograph volume must be required along with peak discharge, in order to calculate the impact of inflow on the storage. Therefore, multivariate designs and their associated return periods could be a comprehensive way of tackling such extreme issues through a defensive risk-based decision-making in this river basin.

In this study, the multivariate flood frequency analysis is performed for the Kelantan River basin at Guillemard bridge station in Malaysia, in order to overcome the limitations associated with univariate analysis, by introducing the bivariate copula distribution framework for partial series of flood characteristics, i.e. flood peak discharge flow (P), volume (V) and duration (D) series. At-site event-based or block (annual) maxima-based methodology is adopted for the 50-year (1961–2016) continuously distributed stream flow characteristics of this river basin. Evaluating the performance of parametric family functions for constructing univariate marginal distribution of flood characteristics, comparing two types of parameter estimation procedure, i.e. method of moment (MOM) and maximum pseudo-likelihood (MPL) estimations procedure, for fitting a distinguished variety of bivariate copula families such as the Archimedean copulas, elliptical copulas also, rotated version of Archimedean and mixed version of Archimedean copulas family which are introduced and tested to analyse and statistically evaluate the bivariate joint probability of the flood variables occurrence, i.e. between flood peak–volume and peak–duration series and also their associated joint and conditional return periods estimations are the main research concern which is addressed in this literature.

As a prerequisite, brief details on the theoretical aspect of copula functions, procedure for the estimations of copula dependence parameters, performance measure of copulas via the goodness-of-fit test statistics to selecting most parsimonious copulas for establishing joint relationship among flood characteristics as well as mathematical approach towards the return periods of flood characteristics are discussed in the next sections. In the third section, the copula-based methodology is applied to a case study to establish a joint distribution, conditional distributional distribution and their associated return periods of the flood attribute pairs (i.e. P–V and P–D). The fourth section provides the result and discussions, and the fifth section provides the research conclusion. Figure 1 illustrates the flow diagram indicating the steps of the analysis.

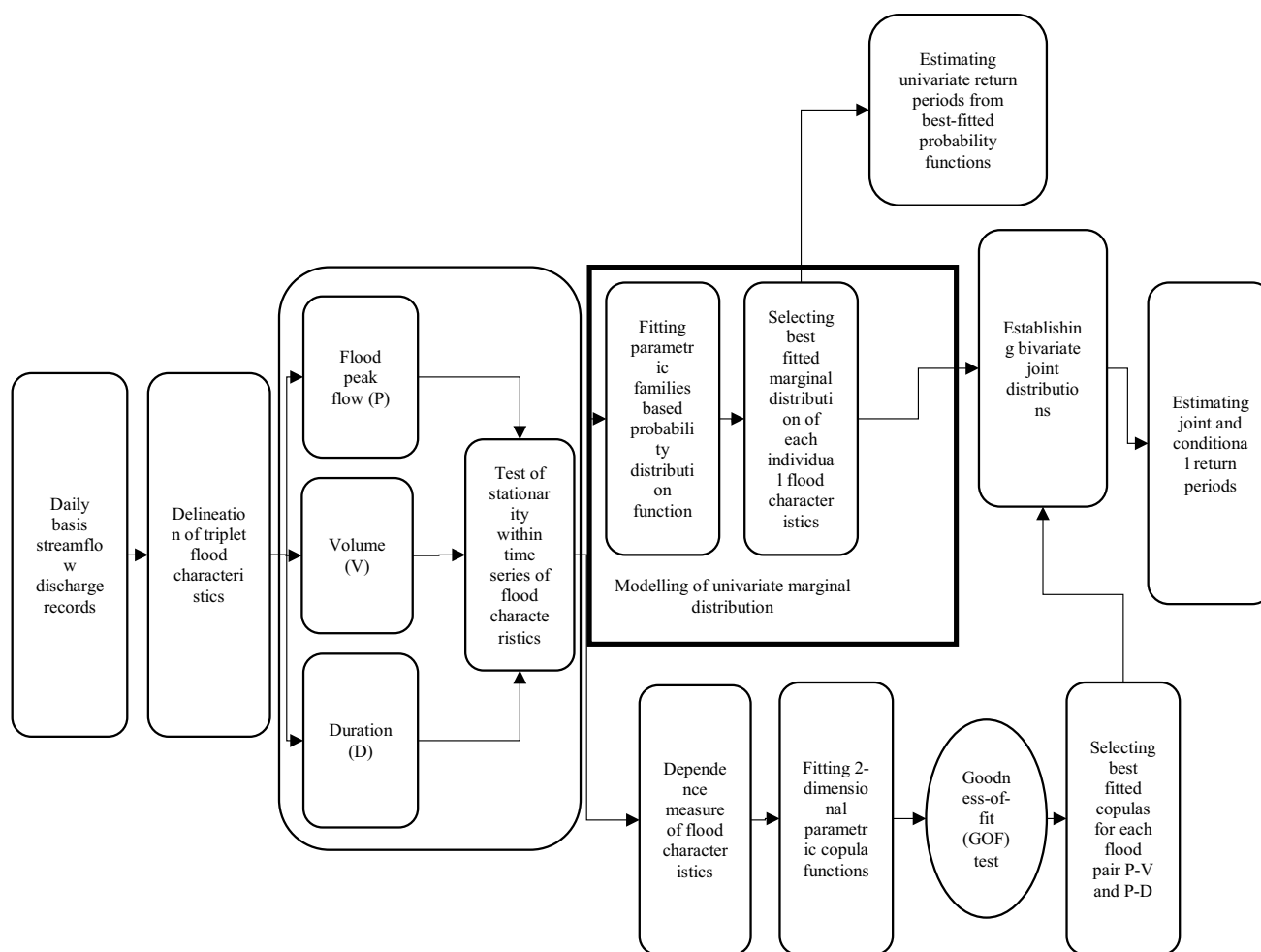


Fig. 1 Flow diagram of the copula-based bivariate joint analysis of flood characteristics

Theoretical research framework

Concept of copula function

The ideas of the copula method have been developed by Saklar (1959). According to Nelsen (2006), copula function connects multivariate probability distributions to their univariate marginal functions. According to Sklar's theorem (Nelsen 2006), if (X, Y) be the bivariate random variables with continuous marginal distributions $u = F_X(x) = P(X \leq x)$ and $v = F_Y(y) = P(Y \leq y)$, then it can be characterized uniquely by its associated dependence function called Copula or C which can be defined on the unit square and can be expressed as:

$$H_{X,Y}(x, y) = C[F_X(x), F_Y(y)] = C(u, v) \quad (1)$$

where C = any type of bivariate copulas under consideration; $F_X(x) = F_Y(y)$ = cumulative distribution functions of random variables 'X' and 'Y'; $H_{X,Y}(x, y)$ = bivariate joint

distribution which can be expressed using bivariate copula function $C[F_X(x), F_Y(y)]$, as revealed from Eq. (1). According to Shiau (2006) and Zhang and Singh (2006), the copula C must be unique if $F_X(x)$ and $F_Y(y)$ are continuous, which can capture the wider extent of dependencies among random variables. Conversely, if $F_X(x)$, $F_Y(y)$ and the copula functions $C[x, y]$ are given, then Eq. (1) must define the bivariate joint distribution functions with its marginal distributions $F_X(x)$ and $F_Y(y)$. Similarly, if $f_X(x)$ and $f_Y(y)$ are the probability density function of variables X and Y , then the joint probability density of the two random variables can be expressed as:

$$f_{X,Y}(x, y) = c(F_X(x), F_Y(y))f_X(x)f_Y(y) \quad (2)$$

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

$$c(u, v) = \frac{\partial^2 c(u, v)}{\partial u \partial v} \quad (3)$$

in which $u = F_X(x)$ and $v = F_Y(y)$.

For the extended mathematical details about copula functions, readers are advised to follow Genest and Rivest (1993) and Nelsen (2006).

In this study, we introduced the Archimedean copula families (i.e. Clayton, Gumbel, Frank, Joe), mixed version of two-parameter Archimedean copulas such as BB1 (i.e. Clayton–Gumbel copula) (i.e. Constantino et al. 2008), BB6 (i.e. Joe–Gumbel copula) (i.e. Manner 2010), BB7 (i.e. Joe–Clayton copula) (i.e. Li et al. 2016) and BB8 (i.e. Joe–Frank copula) (i.e. Tang et al. 2015), one elliptical copula family (i.e. Gaussian copula) (i.e. Zhang et al. 2016), one rotated version of Archimedean copula family (rotated Clayton by 90° of rotation for capturing negative dependence as well) (i.e. Manner 2007). The rotated Clayton copula is introduced to cover negative dependence structure and can't be used for modelling positively correlated random pair. Actually, the bivariate Clayton copula is only capable of capturing positive dependency (i.e. $0 \leq \theta < \infty$); this means by considering the rotation (i.e. by 90°), it becomes possible to capture negative dependency for the same random variables (i.e.

$-\infty < \theta \leq 0$). The consideration of the Joe copula would be a good decision in case of higher positive correlation exhibited between random variables (McNeil et al. 2015) which often plays an important role, especially for highlighting the upper tail dependence structure in extreme value analysis (Alina 2018). For the Clayton and Gumbel family, the Kendall's tau ≥ 0 and is only significant to capture the positive dependency while the Frank family exhibited higher versatility due to its capability in accommodating the entire range of dependencies (i.e. $\tau\theta \in [1, -1]$) and is the only member that justified radial symmetry as well (i.e. symmetric to $u + v = 1$) (De Michele and Salvadori 2003; Favre et al. 2004; Nelsen 2006). Table 1 represents the summary description of the bivariate Archimedean copula functions incorporated in this study.

Mathematically, the copula function (i.e. $[C: [0,1]^2 \rightarrow [0,1]]$) approximates the bivariate Archimedean class copula, if it justifies the representation as given below:

$$C(u, v) = \phi^{-1}(\phi(u) + \phi(v)) \text{ for } u, v \in, [0, 1] \tag{4}$$

Table 1 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 $\theta(\tau_\theta)$
Clayton	$[\max\{u^{-\theta} + v^{-\theta} - 1; 0\}]^{-1/\theta}$	$0 \leq \theta < \infty$	$\frac{1}{\theta}(t^{-\theta} - 1)$	$\frac{\theta}{\theta+2}$
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; Wang et al. 2009)
Gumbel–Hougaard	$\exp \left\{ - \left[(-\ln(u))^\theta + (-\ln(v))^\theta \right]^{\frac{1}{\theta}} \right\}$	$1 \leq \theta < \infty$	$(-\ln t)^\theta$	$\frac{\theta-1}{\theta}$
Joe	$1 - \left[(1-u)^\theta + (1-v)^\theta - (1-u)^\theta(1-v)^\theta \right]^{1/\theta}$	$1 \leq \theta < \infty$	$-\ln(1 - (1-t)^\theta)$	
BB1	$\left(1 + \left[(u^{-\theta} - 1)^\delta + (v^{-\theta} - 1)^\delta \right]^{1/\delta} \right)^{-1/\theta}$	$0 < \theta < \infty;$ $1 \leq \delta < \infty$		
BB6	$1 - \left(1 - \exp - \left[((-\ln(1-u)^\theta)^\delta + ((-\ln(1-(1-v)^\theta))^\delta) \right]^{1/\delta} \right)^{1/\theta}$	$1 \leq \theta < \infty;$ $1 \leq \delta < \infty$		
BB7	$1 - \left[1 - \left((1-u^\theta)^{-\delta} + (1-v^\theta)^{-\delta} - 1 \right)^{-1/\delta} \right]^{1/\theta}$	$1 \leq \theta < \infty;$ $0 \leq \delta < \infty$	$(1 - (1-t)^\theta)^{-\delta} - 1$	
BB8	$\frac{1}{\theta} \left(1 - \left[1 - \frac{1}{1-(1-\delta)^\theta} (1 - (1-\delta u)^\theta(1-\delta v)^\theta) \right]^{\frac{1}{\delta}} \right)$	$1 \leq \theta < \infty;$ $0 \leq \delta \leq 1$	$-\ln \left[\frac{1-(1-\delta t)^\theta}{1-(1-\delta)^\theta} \right]$	

θ is the copula dependence parameter of mono-parametric (or 1 parameter) copulas; θ and δ jointly represents the dependence parameter for bi-parametric (or 2 parameter) Archimedean copulas like BB1, BB6, BB7 & BB8

where $(\mathcal{O}(\cdot))$ and $\mathcal{O} - 1$ signify the generator function of the specified Archimedean copulas and their inverse such that the generator $(\varphi: I \rightarrow R^+)$ signifies for the positive, convex and decreasing function and could be approximated for $(\mathcal{O}(1) = 0$ and $\mathcal{O}(1) = \infty)$ (Nelsen 2006). The elliptical family-based Gaussian copula is also introduced for testing their adequacy in the establishment of bivariate distribution of flood characteristics and can capture both the positive and negative dependencies (Favre et al. 2004). The Gaussian copula is an implicit copula which can be expressed as an integral over the density of X and can be expressed mathematically as given below (Zhang et al. 2016):

$$C_{\theta}(u, v) = \int_{-\infty}^{-\phi^{-1}(u)} \int_{-\infty}^{-\phi^{-1}(v)} \frac{1}{2\pi(1-\theta^2)^{1/2}} \exp\left[-\frac{s^2 - 2\theta st + t^2}{2(1-\theta^2)}\right] ds dt \quad (5)$$

The Gaussian copula shows almost no dependence in the tails of distribution which is mostly distributed around centre (Zhang et al. 2016; Alina 2018).

Estimating copula dependence parameter

Several methods are often motivated such as the method of moment based rank-based nonparametric measures, i.e. Kendall's tau (τ) or Spearman's rho (ρ) (i.e. Genest and Rivest 1993; Reddy and Ganguli 2012a), exact maximum likelihood or EML (i.e. Zhang et al. 2016), inference functions for marginal or IFM (i.e. Joe 1997), canonical maximum likelihood or CML (Genest et al. 1995), maximum pseudo-likelihood estimations (MPL) (Genest et al. 1995). In this experiment, both the MOM estimator based on the inversion of Kendall's tau (τ) and MPL estimators are applied separately for estimating copula dependence parameter.

MOM estimators based on inversion of Kendall's tau (τ) is quite popular and highly flexible for the Archimedean families of copula which often follows the definition of empirical copula. This method is based on the relationship between the copula parameter ' θ ' and sample rank correlation coefficient and if the corresponded between them exhibited then the copula parameter can be estimated by putting the empirical values of the rank correlation into the relation as given below (Nelsen 2006):

$$\hat{\theta} = f(\hat{\tau}) \quad (6)$$

According to Nelsen (2006) and Veronika and Halmova (2014), the mathematical relationship between Kendall's tau (τ), copula function and generator function can be expressed as;

$$\tau = 4 \int_{[0,1]^2} C(u, v) dC(u, v) - 1 \quad (7)$$

and

$$\tau = 1 + 4 \int_0^1 \frac{\phi(t)}{\phi'(t)} dt \quad (8)$$

where $\phi(\cdot)$ is the generator function of Archimedean copula family and $\phi'(t)$ is the first derivative of the generator function. Functional relationship of Kendall's tau (τ) with copula parameter for various Archimedean families of copula, which are incorporated in this demonstration, is presented in Table 1.

Schweizer and Wolff (1981) explored that the two standard nonparametric correlation measuring statistics, i.e. Kendall's correlation and Spearman's correlation, can be expressed in terms of copula function. Mathematically, the Kendall's tau (τ) statistics can be estimated by using the following equation:

$$\left(\frac{n}{2}\right)^{-1} \sum_{i < j} \text{sign}[(x_i - x_j)(y_i - y_j)] \quad (9)$$

where $\text{sign} = 1$; if $[(x_i - x_j)(y_i - y_j) > 0]$ and $\text{sign} = -1$; if $[(x_i - x_j)(y_i - y_j) < 0]$, for $i = j = 1, 2, \dots, n$.

The mathematical range of Kendall's tau (τ) is $[-1, 1]$; for example, $\tau = 1$ represents for concordant and $\tau = -1$ for discordant and zero represents no concordance will be exhibited between random pairs. Therefore, determination of Kendall's tau (τ) from the given bivariate random observations is firstly required to estimate unknown copula dependence parameters. MOM estimators can only be used for mono-parametric copulas functions, i.e. having single dependence parameter can't be used for deriving more than one copula parameters.

The MPL estimators are a modified version of traditional maximum likelihood method where the rank-based empirical distributions are used for estimating copula parameters independently from their univariate marginal distribution functions and can be applied for both one- and multi-parameter copula functions (i.e. De Michele et al. 2005; Klein et al. 2010; Kojadinovic and Yan 2010; Reddy and Ganguli 2012a). Based on maximizing the pseudo-log-likelihood function, one can easily estimate the copula parameter as follows:

$$l(\theta) = \sum_{i=1}^n \log [c_{\theta}\{F_1(X_{i,1}), F_1(X_{i,2})\}] \quad (10)$$

where θ is the copula parameter; $l(\theta)$ is the pseudo-log-likelihood function; $F_1(X_{i,1}) = F_1(X_{i,2})$ is the empirical CDFs.

Goodness-of-fit test for bivariate copulas

In this demonstration, Cramer–von Mises test statistics is employed to evaluate adequacy of hypothesized copula fitted to bivariate flood characteristics which is often considered as one of the most powerful models for compatibility testing (Genest and Rémillard 2008; Genest et al. 2009). According to Genest et al. (2009), this test is based on parametric bootstrapping procedure and makes use of the Cramer–von Mises statistic S_n , which can be computed as follows:

$$\begin{aligned} S_n &= n \int_{[0,1]^2} \{c_n(u, v) - C_\theta(u, v)\}^2 dC_n(u, v) \\ &= \sum_{i=1}^n \{c_n(U_{i,n}, V_{i,n}) - C_\theta(U_{i,n}, V_{i,n})\}^2 \end{aligned} \quad (11)$$

where C_n = empirical copula estimated using n observational flood attribute pairs and C_θ = parametric copula derived under the null hypothesis. In this experiment, the p -values for each copula are estimated using the parametric bootstrapping procedure (i.e. followed by Genest and Rémillard (2008)) which can be mathematically formulated as given below:

$$p = \frac{1}{N} \sum_{i=1}^N 1(S_{n,t} \geq S_n) \quad (12)$$

where N = number of simulations.

This fitness statistics actually involve testing of null hypothesis H_0 against the alternate 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 \mathfrak{R}^q for some integer value q . The acceptance or rejection of the considered copulas is based on estimated p values. If the estimated p value is larger than a significance level (α), then the null hypothesis must be accepted, resulting that copula must be considered as satisfactory performance; otherwise, it will be liable for rejections. Therefore, from Eq. (11), it must be concluded that minimum value of S_n test value indicates minimum gap or distance between an empirical and derived parametric copula; in other words, it indicates the most justifiable copula for establishing bivariate joint relationship between flood variables.

Return periods of the flood characteristics and conditional distribution

Hydrology and hydraulic applications mostly interested in the evaluation of the mean inter-arrival period between two design events which usually are defined in a year called the return period (Shiau 2003; Salvadori 2004). In particular, the design quantiles define a higher return period that often seems a feasible practice in the hydraulic structure designs

(Requena et al. 2016). Concurrence probability usually defines the probability of any extreme happening (i.e. flood episodes), which either characterizes through univariate (say flood peak discharge or ‘ P ’) or multivariate variables (say flood peak and volume series ‘ P ’, ‘ V ...’) exceeding certain a threshold level say ‘ p ’ (or ‘ p ’, ‘ v ...’ for the multivariate structure) (Yue and Rasmussen 2002; Shiau 2003; Salvadori 2004). Yue and Rasmussen (2002) and Salvadori and De Michele (2004) thoroughly discussed the basic concept of return periods. Mathematically, univariate return period of the targeted flood characteristics that occurs once in a year can be defined from the univariate cumulative distribution function or CDF of the variable (say ‘ X ’) as given below:

$$\begin{aligned} T_{\text{Univariate}} &= \frac{\mu}{\text{total no. of flood per year}} \\ &= \frac{1}{P(X \geq x)} = \frac{1}{(1 - F(x))} = \frac{1}{1 - \text{CDF}(y)} \end{aligned} \quad (13)$$

where μ is the mean inter-arrival time between two consecutive episodes and that could be equal to unity (i.e. $\mu = 1$) for annual maxima-based flood modelling (Yue and Rasmussen 2002).

Derivation of return periods from bivariate joint distribution of flood attribute pairs

The notation of return period under univariate concept (Eq. 13) might be useful only if the concentration of single hydrologic or flood attribute will justify the requirements of the design process, and in other words, it will also indicate the existence of no significant inter-association exhibited between multiple relevant vectors (Veronika and Halmova 2014). But for justifying much practical and effective risk analysis, accountability of multiple potential vectors for characterizing flood events based on joint probability density function or JPFD and joint cumulative distribution functions or JCDF could be the essential demands for hydrologist and water practitioner. Estimating multivariate design variable quantiles under different notations of return such as based on joint probability distribution functions, conditional distribution analysis or either based on the Kendall’s distribution (or survival functions) is often an essential concern in the hydrologic risk assessments (Salvadori 2004; Graler et al. 2013). In multidimensional risk framework, return periods can be derived from the exceedance probabilities of flood attributes pair, such as joint return period derived from the joint exceedance probabilities. Actually, each separate approach of return periods has their own significance, and that will be solely based on the nature of the undertaking problem, which cannot be interchanged and is also impossible to decide for the most consistent ways (Sernaldi 2015). Selection of return periods is depending upon

the importance of undertaken structure as well as its consequences of failure where their appropriate selection often attributed an impact over the strength of design variables quantiles. Shiau (2003), Salvadori (2004), Salvadori and De Michele (2004, Salvadori and De Michele 2007), and Serinaldi (2015) pointed an extended mathematical framework towards the deriving of different notations of return periods under copula-based methodology.

The joint probability distributions for annual flood analysis can describe the following two situation such that in the first condition both the flood variables (say, $P \geq p$ AND $V \geq v$) simultaneously exceed certain threshold during a flood event and their associated return period called AND joint period and it can be written in the form of:

$$T_{p,v}^{\text{AND}} = \frac{1}{P(P \geq p \text{ AND } V \geq v)} = \frac{1}{(1 - F(p) - F(v) + H(p, v))} = \frac{1}{(1 - F(p) - F(v) + C(F(p), F(v)))} \quad (14)$$

Equation represents the joint return period of flood variable $P \geq p$ and $V \geq v$ where both variables simultaneously exceed threshold value. $H(p, v)$ is the joint cumulative distribution functions between flood variables (i.e. P and V) and that can be expressed using bivariate copula function $C(F(p), F(v))$.

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

$$T_{p,v}^{\text{OR}} = \frac{1}{P(P \geq p \text{ OR } V \geq v)} = \frac{1}{(1 - H(p, v))} = \frac{1}{(1 - C(F(p), F(v)))} \quad (15)$$

Equation represents the OR joint return period of flood variable, i.e. $P \geq p$ OR $V \geq v$ for the bivariate events where either of the variables exceeds threshold value. Using the above two equations, it could be possible to derive the OR and AND joint return periods between flood peak flow–volume and peak flow–duration series.

Derivation of return periods from conditional distribution of flood attribute pairs

In most of the hydrologic design requirements, it would be demanding to define events through highlighting the significance or priority of one design variables over another design vectors, and thus, from this prospect several studies pointed out the necessity of conditional distributional framework for defining the concept of conditional return periods such as Salvadori and De Michele (2004), Shiau (2006), Zhang and Singh (2006), Zhang and Singh

(2007a), Salvadori and De Michele (2010), Reddy and Ganguli (2012a, 2013), Veronika and Halmova (2014), Zhang et al. (2016) and Tosunglou and Kisi (2016). The conditional return period relies on a conditional probability relationship between flood characteristics given that some condition is fulfilled such as the conditional return period of flood peak series given various percentile values of flood volume or vice versa or in other words where the flood peak ‘ P ’ exceeds a threshold ‘ p ’ given that the volume ‘ V ’ series exceeds a threshold ‘ v ’. For example, probability of flood peak conditional to volume (or duration) or either the flood volume conditional to peak (or duration) or either flood duration conditional to flood peak (or volume) information would be benefited from the hydraulic design prospects. Therefore, using the bivariate copula functions, the conditional return periods between flood characteristics (say, P and V) can be obtained from the conditional probability distribution function as:

$$H'_{P \setminus V}(p \setminus v) = H(p \setminus V \leq v) = \frac{P(P \leq p, V \leq v)}{P(V \leq v)} = \frac{H_{P,V}(p, v)}{F(v)} = \frac{C(F(p), F(v))}{F(v)} \quad (16)$$

and

$$T_{(P \setminus V)(p \setminus v)} = T_{(p \setminus V \leq v)} = \frac{1}{1 - H'_{P \setminus V}(p \setminus v)} = \frac{F(v)}{F(v) - C(F(p), F(v))} \quad (17)$$

where $H_{P,V}(p, v)$ is the joint CDF of flood vector P and V which can be estimated using the bivariate copula functions $C(F(p), F(v))$ with $F(p)$ and $F(v)$ being the marginal distribution of flood peak and volume series. Actually, Eq. (16) represents the conditional distribution function of flood peak ‘ P ’ given volume ‘ $V \leq v$ ’. Thus, using Eq. (17), it could be possible to obtain return periods for various possible occurrences of flood characteristics (i.e. return period of volume condition to peak or either between volume–duration and peak–duration series).

Similarly, the conditional distribution of flood peak ‘ P ’ given ‘ $V \geq v$ ’ or vice versa can be obtained using the following equations:

$$H'_{P \setminus V}(p \setminus v) = H(p \setminus V \geq v) = \frac{P(P \leq p, V \geq v)}{P(V \geq v)} = \frac{F(p) - H_{P,V}(p, v)}{1 - F(v)} = \frac{F(p) - C(F(p), F(v))}{1 - F(v)} \quad (18)$$

and

$$T_{(P \setminus V)(p \setminus v)} = T_{(p \setminus V \geq v)} = \frac{1 - F(v)}{1 - F(v) - F(p) + C(F(p), F(v))} \quad (19)$$

Therefore, by using the copula-based methodologies for estimating the joint and conditional distribution of the event-based or block (annual) maxima-based mutually correlated flood characteristics, the joint and conditional return periods are computed which can be an essential requirement for the Kelantan River basin for tackling several water-related queries.

Application

Details of study area

Monsoonal flood happening seems to be increased in the Kelantan River basin in Malaysia from the last few decades in terms of frequency as well as magnitude, according to the report of Drainage and Irrigation Department, Malaysia (DID 2000, 2003, 2004) and Malaysia Meteorological Department (MMD 2007). The geographical location of this river basin is Lat $4^{\circ} 30' N$ to $6^{\circ} 15' N$ and Long $101^{\circ} E$ to $102^{\circ} 45' E$, and it is the longest river of Kelantan state, which originates 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 $13,100 \text{ km}^2$ occupying more than 85% of the state of Kelantan. The estimated runoff is about $500 \text{ m}^3 \text{ s}^{-1}$, and the variations in annual precipitations for this region are in between 0 mm (dry period) and 1750 mm (wet or north-eastern monsoonal period) (DID 2000). The major land use of this area is agriculture (i.e. paddy, rubber and oil palm) for midstream and downstream and forest for upstream (i.e. near to Gua Musang). Few studies over this region such

as Chan (1997), Jamaliah (2007) and Adnan and Atkinson (2011) pointed that such extreme hydrologic consequences are mainly due to rapid human intervention from natural to land-use activities in the form of deforestations or land clearance either for promoting the agricultural activities through palm oil and rubber plantations or due to logging activities. In this literature, copula-based methodology is adopted for the 50 years (1961–2016) of daily basis streamflow discharge records for the Kelantan River basin which are collected and provided by the Drainage and Irrigation Department (DID), Malaysia, at the Guillemard Bridge gauge stations, at the downstream of Kelantan River near the Kuala Kari region.

Delineation of trivariate flood characteristics

Flood probability construction via the partial data series only focuses the extreme hydrograph portion, i.e. either high flow (for flood episodes) or low flow (for drought events), instead of visualizing the entire hydrograph (Correia 1987; Hosking et al. 1985; Rao and Hameed 2000). Annual (maximum) series (AM) also called block (annual) maxima and peak over threshold (or POT) are the two frequently modelling techniques widely accepted in the extreme probability simulations (Hosking et al. 1985; Madsen et al. 1997). The annual maximum flood peak discharge, volume and duration value of the flood episodes are retrieved from daily basis stream flow data. The characterizations of flood peak flow values are based on their maximum streamflow discharge records at an annual scale which means for each year there will be only one flood episode at the targeted site (Fig. 2) (Yue 2000; Yue and Rasmussen 2002; Xu et al. 2015). Similarly, corresponding

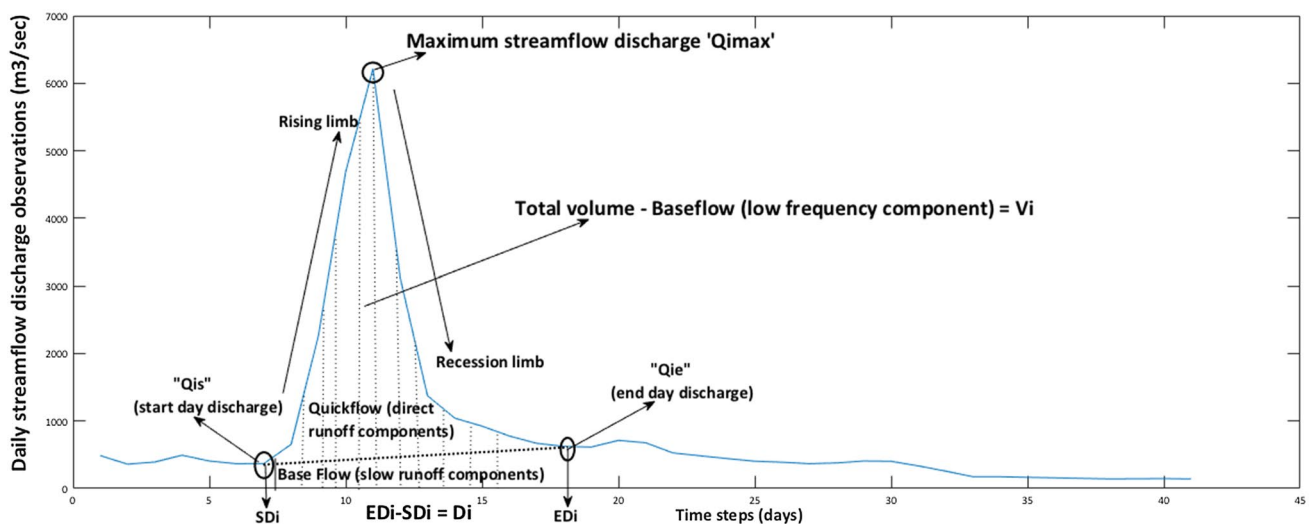


Fig. 2 A typical hydrograph showing flood characteristic for the i th flood episodes

to each flood peak value of the annual basis flood events, volume and duration series are derived from the streamflow hydrograph and using the algorithm reported by the literature such as Yue et al. (1999), Yue (2000) and Yue and Rasmussen (2002). Actually, flood duration (D) can be estimated by recognizing the time of rise and fall of the flood hydrograph as illustrated in Fig. 2 (i.e. points at Q_{is} and Q_{ie}) using Eq. (22). Flood peak discharge often attains their maximum value but not mandatory for hydrograph volume and duration series (Sraj et al. 2014; Xu et al. 2015). Mathematically, we are required to derive triplet flood characteristics for each of the i th years using the following equations:

$$P_i = \max\{Q_{ij}, j = SD_i + SD_i + 1, \dots, ED_i\} \quad (20)$$

= Annual flood peak series

$$\text{Volume} = V_i = V_i^{\text{total}} - V_i^{\text{Baseflow}} = \sum_{j=SD_i}^{ED} Q_{ij} - \frac{(1 + D_i)(Q_{is} + Q_{ie})}{2} \quad (21)$$

$$\text{Duration} = D_i = ED_i - SD_i \quad (22)$$

where Q_{ij} = j th days streamflow magnitude for the i th year and Q_{is} & Q_{ie} = streamflow magnitude for the start date 'SD_{*i*}' and end date 'ED_{*i*}' of the flood runoff. Once the flood characteristics are obtained from daily stream flow data using equations, it can be used for copula-based flood frequency

analysis. Table 2 represents the descriptive statistics of triplet flood characteristics extracted from the daily streamflow discharge value at an annual scale (i.e. partial data series). Each univariate flood vector exhibited positively skewed distributions in which the duration series exhibited quite higher degree of unsymmetrical behaviour; Fig. 3a shows histogram distribution plot of flood characteristics. Figure 3b represents time series visualization of flood peak, volume and duration series.

Estimating marginal probability distribution of flood characteristics

Empirical univariate nonexceedance probabilities

The empirical nonexceedance probabilities or cumulative distribution, i.e. $P(K \leq k)$, for observed values of flood characteristics is derived from the commonly used Gringorten-based position-plotting formula (Gringorten 1963; Cunnane 1978; Zhang and Singh 2006; Karmakar and Simonovic 2008), which is usually compared with CDF of the fitted distributions for pointing the gaps and deviations between empirical and fitted samples and which can be mathematically formulated as:

$$\begin{aligned} \text{Empirical Cumulative frequency} \\ = P(K \leq k) = (k - 0.44)/(N + 0.12) \end{aligned} \quad (23)$$

where N = length of the sample (i.e. the total number of flood observations) and k = k th smallest observations in the data set arranged in ascending order.

Table 2 Summary statistics of the flood event characteristics

Statistics	Peak	Volume	Durations
nbr.val	5.00E+01	5.00E+01	5.00E+01
nbr.null	0.00E+00	0.00E+00	0.00E+00
nbr.na	0.00E+00	0.00E+00	0.00E+00
Min	9.16E+02	3.18E+03	7.00E+00
Max	2.06E+04	7.47E+04	6.40E+01
Range	1.97E+04	7.16E+04	5.70E+01
Sum	3.04E+05	9.56E+05	9.52E+02
Median	4.96E+03	1.60E+04	1.60E+01
Mean	6.08E+03	1.91E+04	1.90E+01
SE.mean	6.56E+02	2.07E+03	1.53E+00
CI.mean.0.95	1.32E+03	4.16E+03	3.08E+00
Var	2.15E+07	2.14E+08	1.18E+02
Std.dev	4.64E+03	1.46E+04	1.09E+01
Coef.var	7.63E−01	7.65E−01	5.70E−01
Skewness	1.46E+00	1.54E+00	2.14E+00
Skew.2SE	2.17E+00	2.29E+00	3.19E+00
Kurtosis	1.69E+00	2.63E+00	5.89E+00
Kurt.2SE	1.28E+00	1.99E+00	4.45E+00
Normtest.W	8.40E−01	8.50E−01	7.84E−01
Normtest.p	8.42E−06	1.54E−05	3.78E−07

Fitting univariate marginal distribution of flood characteristics

Selecting the most justifiable and parsimonious univariate probability distribution functions for defining flood marginal density is often a mandatory prerequisite desire before introducing random vectors into multivariate copula framework (Reddy and Ganguli 2013; Tosunoglu and Kisi 2016). In this study, the flood characteristics that are targeted are flood peak discharge flow, volume and its durations for the partial series of data called block (annual) maxima-based flood distribution analysis. Parametric distribution-based modelling is often based on the assumption that the undertaken samples must be following some specific distributions or having their predefined PDF. Actually, in hydrologic data modelling, no universally accepted distributions are assigned from any literature or in favour of any probability distribution functions to model any extreme series (Adamowaski 1985, 1989; Silverman 1986; Dooge 1986; Yue et al. 1999; Santhosh and

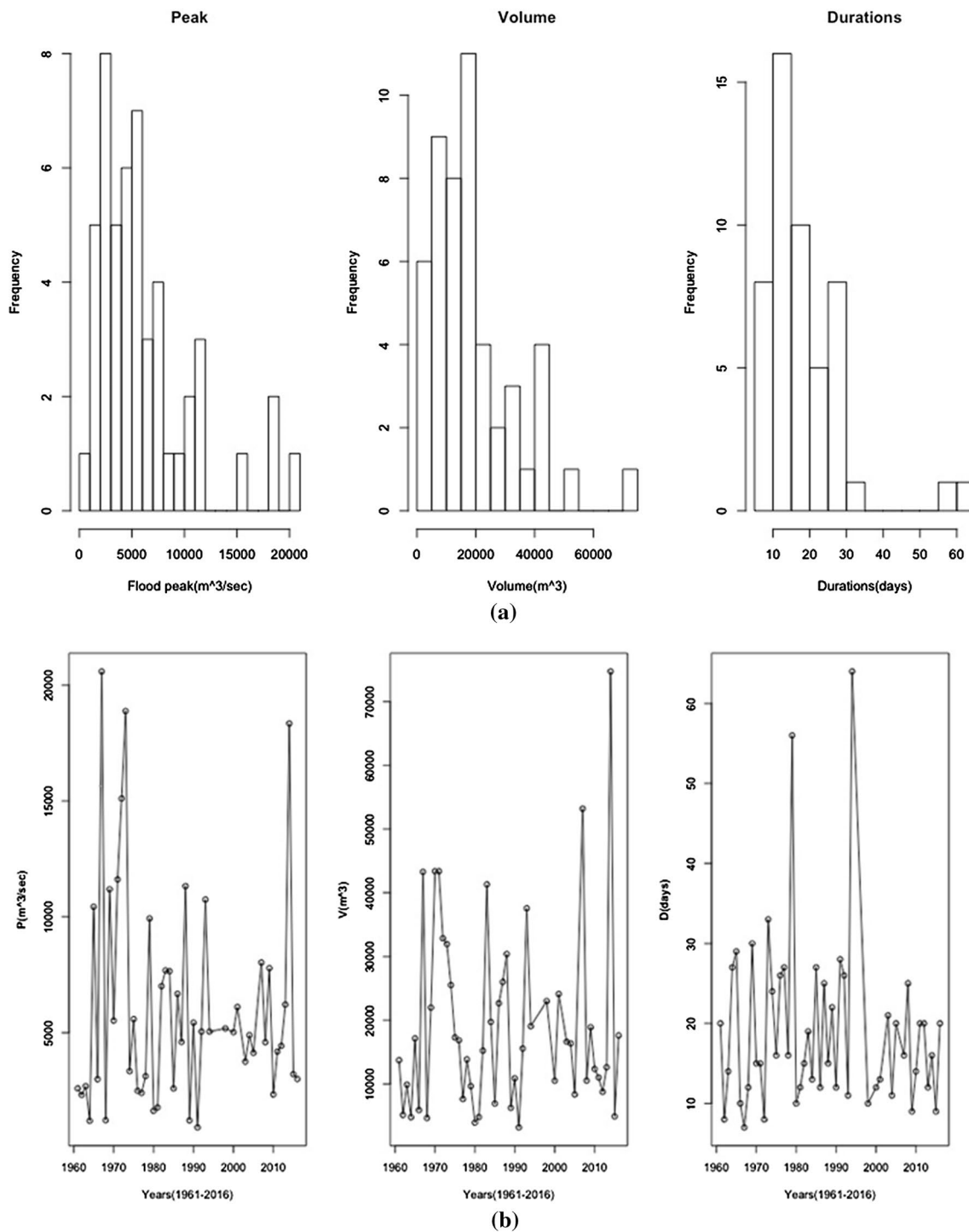


Fig. 3 Visualizing the block (annual) maxima-based flood characteristics of the Kelantan River Basin at Guillemard Bridge station between the years 1960 and 2016 in the context of **a** histogram plot and **b** time series plot of peak, volume and duration series

Srinivas 2013). 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, which is solely based on the goodness-of-fit procedure

to visualize the compatibility of the fitted distributions (Karmakar and Simonovic 2008). An interactive set of 1-dimensional parametric functions with varying numbers of the unknown statistical parameter (i.e. 1 parameter, 2

Table 3 Mathematical expressions for density functions and vector of unknown statistical parameters of different univariate probability distribution functions

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

parameters, 3 parameters and 4 parameters) are introduced as a candidate function in modelling univariate marginals of the flood characteristics. Table 3 lists the mathematical expressions (i.e. PDF) of distinct varieties of univariate parametric models and their associated vector of unknown statistical parameters or model parameters which are targeted for modelling marginal distribution of flood characteristics. The vector of the unknown statistical parameter of the fitted distributions for each flood characteristic is estimated based on the maximum likelihood estimation (MLE) (i.e. Owen 2008), method of moments (MOM) (i.e. Bain and Engelhardt 1991; Rao and Hameed 2000), least square method (LS) and L statistics-based method of L-moments (i.e. Hosking and Walis 1987). All the univariate distribution fitting procedures and their parameter estimation are carried out using Easyfit software (Mathwave Technologies 2004, 2017).

Result and discussions

Test for stationarity within time series of flood characteristics

The individual flood characteristics need to be stationary or time independency behaviour before introducing into the univariate or multivariate copula distribution framework. Before modelling the joint dependency between floods using bivariate copula functions, it is mandatory to investigate whether the individual time series associated with each flood characteristic exhibits no serial correlations or stationary. For this, Ljung and box (1978)-based hypothesis testing also called Q-statistics is undertaken for investigating whether the individual series are time independent and no serial correlation (Cong and Brady 2011). Statistically, the Q-statistics

Table 4 The Q-statistics and their corresponding p value for peak flow, volume and duration series

Flood vectors	Box–Ljung test	Lag size (30)	Lag size(20)	Lag size(15)	Lag size(10)	Lag size(5)
Peak flow	X-squared (Q-statistics)	18.869	14.88	13.338	9.7188	4.89
	df	30	20	15	10	5
	p value	0.9427	0.7828	0.5762	0.4655	0.4295
Volume	X-squared (Q-statistics)	14.724	11.981	10.683	7.4968	0.49837
	df	30	20	15	10	5
	p value	0.9912	0.9167	0.7747	0.6779	0.9922
Duration	X-squared (Q-statistics)	23.343	15.702	13.958	6.1707	3.035
	df	30	20	15	10	5
	p value	0.8009	0.7349	0.5287	0.8007	0.6946
Note	Critical value (at 5% level of significance)	43.77	31.41	25	18.31	11.07

usually follows a Chi-square distribution with ‘ h ’ degree of freedom under the null hypothesis H_0 (Ljung and Box 1978; Daneshkhan et al. 2016). Thus, Q-statistics for the sample size ‘ n ’ with ‘ t ’ is the total no. of lags being tested with sample autocorrelations at the specific lag, i.e. $\hat{\rho}_t$, as given below:

$$Q = n(n+2) \sum_{t=1}^h \hat{\rho}_t^2 / n - t \quad (24)$$

where null hypothesis (H_0) = zero autocorrelation or independent distributions and alternative hypothesis (H_a) = existence of serial correlation (or autocorrelation). The estimated Q-statistics and their associated p value for different lag sizes (i.e. 30, 20, 15, 10 and 5) are listed in Table 4 which points for almost negligible or zero first-order autocorrelations as their estimated statistics are below their critical value for each of the univariate series by accepting the null hypothesis (H_0) at 5% or 0.05 significance level against their alternative hypothesis (H_a) (Table 4). Figure 4a, b illustrates the graphical visualizations for investigating the existence of serial correlation within flood characteristics in the context of autocorrelation plot and sample autocorrelation plots. The nonparametric rank-based Mann–Kendall or M–K test is also incorporated for visualizing the monotonic trend within the historical series of flood characteristics under the null hypothesis H_0 against their alternative hypothesis H_a (Mann 1945; Kendall 1975), and their estimated values are listed in Table 5. The test is incorporated under two-tailed hypothesis attempts of the null hypothesis H_0 (i.e. zero trend with flood characteristics) against alternative hypothesis H_a which can be mathematically expressed as:

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sgn}(T_j - T_i) \quad (25)$$

$$\text{sgn}(T_j - T_i) = \begin{cases} 1, & \text{if } T_j > T_i \\ 0, & \text{if } T_j = T_i \\ -1, & \text{if } T_j < T_i \end{cases} \quad (26)$$

where T_j and T_i represent the annual value in year ‘ j ’ and ‘ i ’, respectively. Test statistics reveals the acceptance of null hypothesis (H_0) which points the existence of zero monotonic trend at the 5% or 0.05 level of significance level within flood series. In conclusion, no significant trends are detected for the flood characteristics; therefore, detrend or prewhitening procedure is not adopted (i.e. Razawi and Vogel 2018) before introducing the flood samples into univariate probability distribution framework.

Dependence measures of flood characteristics

The Pearson’s linear correlation (r) and two nonparametric dependence measures also called the rank-based correlations statistics such as Kendall’s tau (t) and Spearman’s rho (ρ) are used to measure the strength of dependency between flood characteristics and their estimated values are listed in Table 6. Actually, the Pearson coefficient only captures the linear dependencies and therefore might be incompatible for heavy-tailed distribution series (Tosunoglu and Kisi 2016). According to Favre et al. (2004), the Pearson correlation coefficient is not invariant to monotonic transformations to Kendall’s and Spearman’s correlation measures. Also, it can be strongly affected by outliers. On the other hand, the nonparametric measures of dependence called the Kendall’s tau and Spearman’s correlation measures are invariant under monotonic nonlinear transformations, without any assumption of underlying distribution structure, and thus frequently used as effective dependence measures for

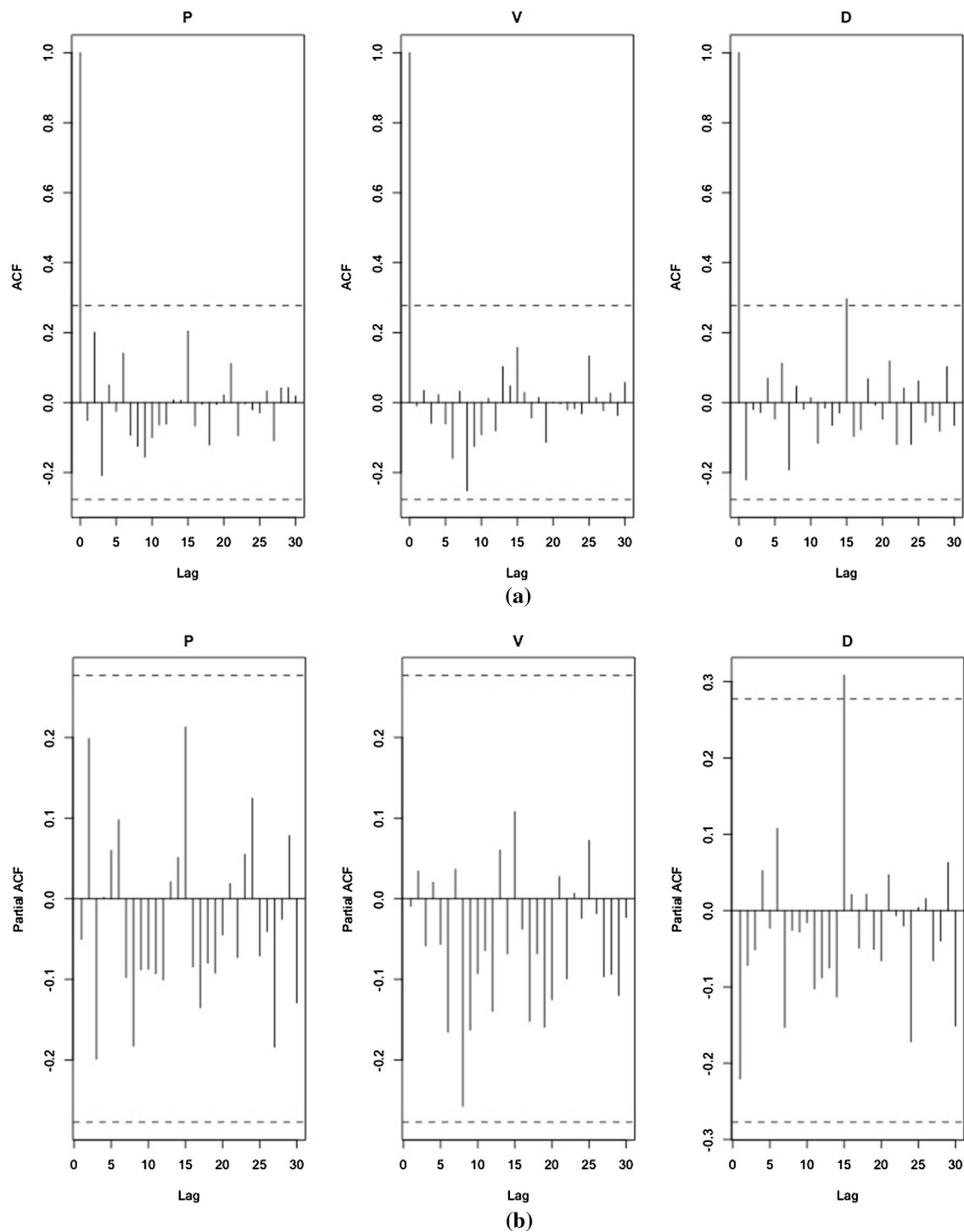


Fig. 4 **a** Autocorrelation plot and **b** sample autocorrelation plot of flood characteristics (P , V and D)

the nonlinear modelling in multivariate statistics. Both the Kendall's tau and Spearman's are calculated using ranking of random series; also both exhibited high resistance to outliers (Klein et al. 2011). From Table 6, it can be observed that flood peak and volume pair exhibited strong positive correlation. But the correlation structure between flood

peak–duration pair and flood volume–duration pair is very weak where both are having negative dependence structure or negatively correlated.

Several graphical tools such as scatter plot, chi plots (i.e. Fisher and Switzer 2001) and Kendall's plots (i.e. Genest and Boies 2003) of the pair-wise flood characteristics are

Table 5 Mann–Kendall (or M–K) test for identifying the existence of monotonic trend within time series of flood variables

M–K Test	Peak	Volume	Duration
z	0.066919	0.058556	−0.40259
p value	0.9466	0.9533	0.6872
S	9	8	−4.9000e+01
varS	1.429167e+04	1.429067e+04	1.42150e+04
τ	7.346939e−03	6.533279e−03	−4.08422e−02

also presented for analysing the degree of inter-association and strength of dependency between flood attribute pairs, as illustrated in Figs. 5, 6 and 7. Chi-plot is actually a scatter plot of the pairs (λ_i, χ_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 centre of the data sets within the range of $[-1, 1]$. Thus, for positively or negatively correlated random pairs, their values will tend to positive or negative. Also, the control limits χ_i are another measuring

Table 6 Correlation coefficient between flood peak flow–volume, volume–duration and peak flow–duration series

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

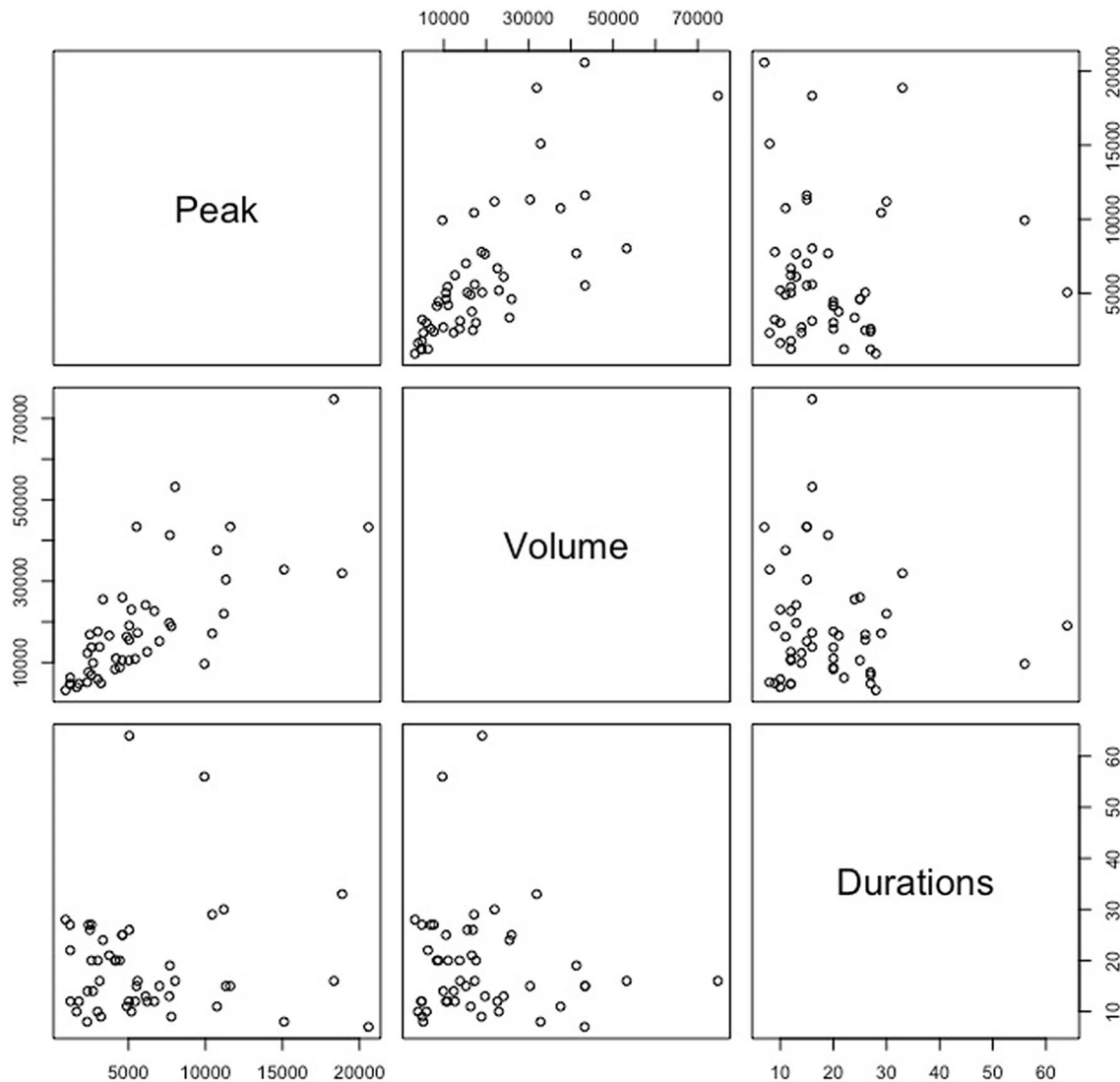


Fig. 5 Scatter plot of the multivariate flood characteristics

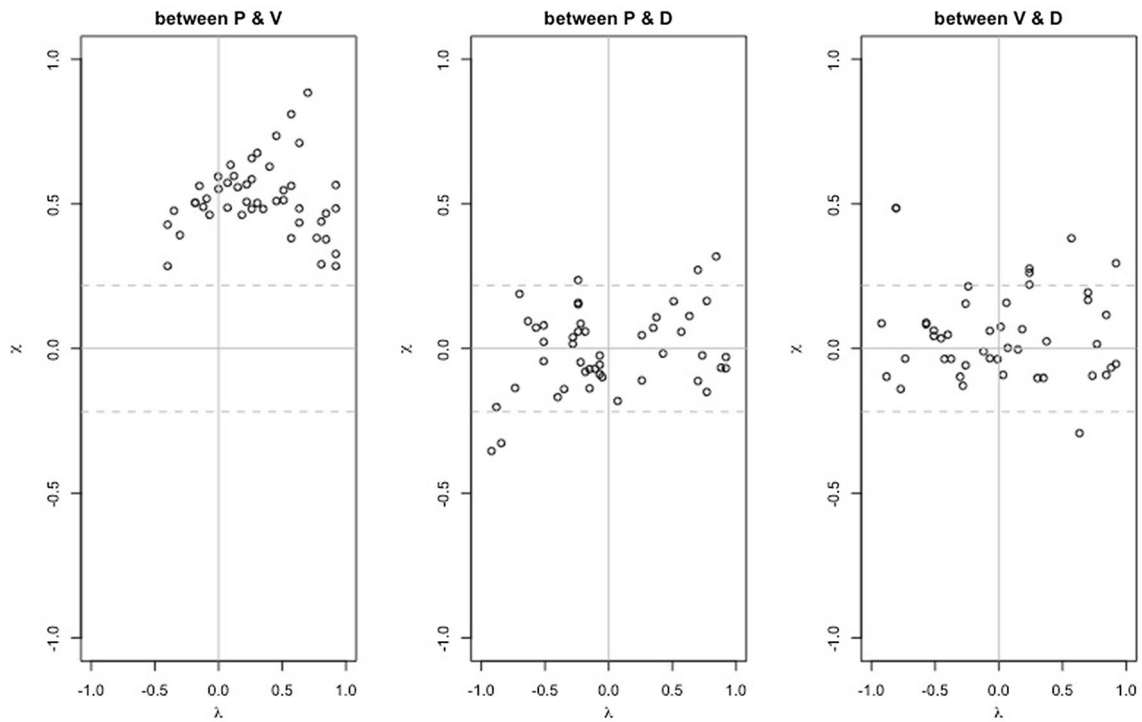


Fig. 6 Graphical interpretation of strength of dependence of pair-wise flood characteristics using Chi-plot between P–V, P–D and V–D

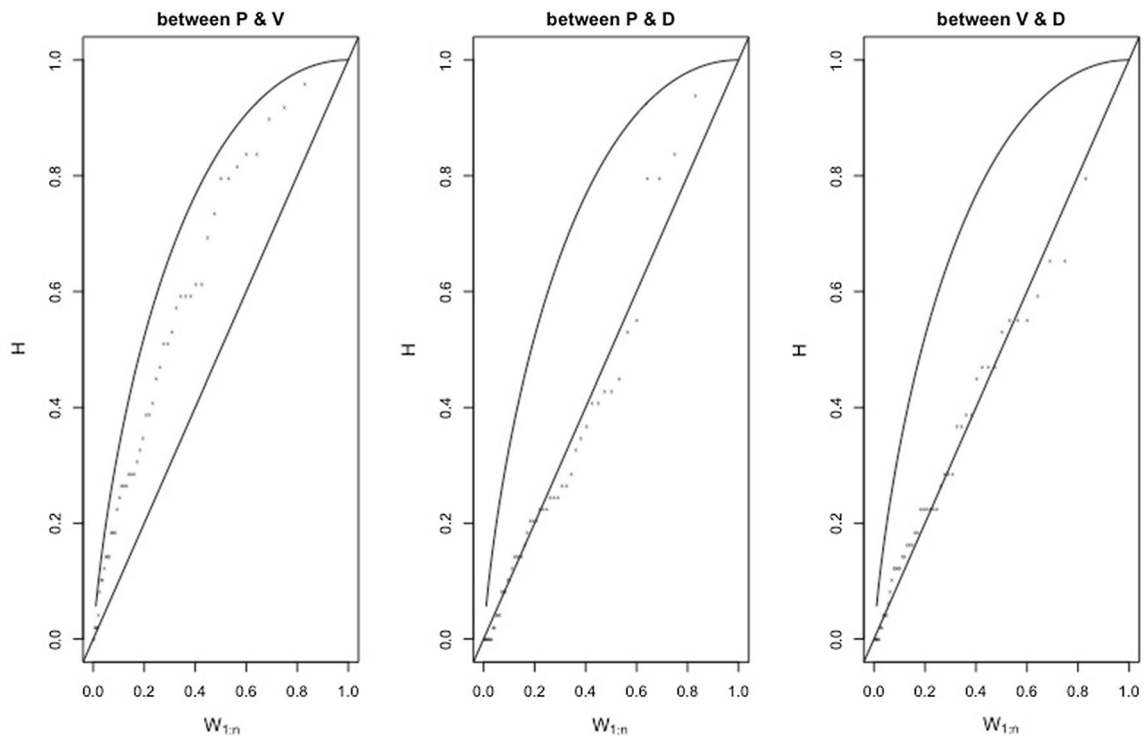


Fig. 7 Kendall's plot (or K-plot) of the pair-wise flood characteristics such as between P–V (shows high and positive correlation structure), P–D (shows negatively correlated random pairs with weak dependency exhibited), V–D (negatively correlated random pairs)

factor in chi-plot that are placed at $\chi = \pm c_p / \sqrt{n}$ (Fisher and Switzer 2001). Therefore, in case of stronger dependency the random pairs must be outside the control limit of chi-plot otherwise inside the control limit region which can be indicated for independence between random pairs. On the other hand, the Kendall's plot is 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 it could be revealed for independence when the plot tends to be linear (Genest and Boies 2003; Genest and Favre 2007). The scatter plot between flood peak–volume pair of Fig. 5 clearly indicates the existence of positive and strong dependency because the increased density of points is located near the diagonal region (i.e. close to 45° angle). From Fig. 5, it reveals the existence of weak dependency between flood volume–duration and flood peak–duration pairs. Based on ranked chi-plot for flood peak–volume series, it indicated the evidence of strong dependency between these random pairs because a strong deviation from the control limit is observed for these random pairs. Between flood peak–duration and volume–duration pairs, very weak dependency is observed because most of the data samples are within the region of control limit. Similarly, based on Kendall's plot, it also indicates strong positive dependency between flood peak–volume pair because the data pairs are much deviated from the main diagonal, but the flood peak–volume and volume–duration random pairs are much

closer to main diagonal, thus indicating negative and weak dependencies between these flood attribute pairs.

Modelling of univariate flood marginal distributions

From Tables 4 and 5, it is already revealed that no significant trends are detected for the flood characteristics; therefore, the prewhitening procedure is not adopted (i.e. Razawi and Vogel 2018) before introducing the flood samples into univariate probability distribution framework.

A distinct variety of parametric probability distribution families (i.e. 1 parameter, 2 parameters, 3 parameters & 4 parameters) are introduced as a candidate model for demonstrating the univariate marginal distribution of each individual flood characteristic, and their PDFs are listed in Table 3. The vector of the unknown statistical parameters is estimated using maximum likelihood estimation (MLE), method of moments (MOM), least square method (LS) and method of L-moments, and the estimated parameter values of different univariate functions are listed in Table 7. Different analytical-based goodness-of-fit measures such as Kolmogorov–Smirnov (or K–S) (i.e. Conover 1999; Xu et al. 2015; Kong et al. 2015) and Anderson–Darling (or A–D) (i.e. Anderson and Darling 1954; Scholz and Stephens 1987; Farrel and Stewart 2006) distance criteria statistics, information criteria statistics such as Akaike Information criteria (or AIC) (i.e. Akaike 1974), Schwartz's Bayesian Information criteria (or BIC) (i.e.

Table 7 Estimated parameters of marginal distribution of flood characteristics

Parametric Functions	Flood peak (<i>P</i>)	Flood volume (<i>V</i>)	Flood durations (<i>D</i>)
Exponential (1P)	$l=1.6453E-4$	$l=5.2295E-5$	$l=0.05252$
Exponential (2P)	$l=1.9373E-4, g=916.3$	$l=6.2735E-5, g=3182.3$	$l=0.08306, g=7.0$
Frechet (2P)	$a=1.576, b=3207.5$	$a=1.5703, b=10017.0$	$a=2.6001, b=13.304$
Frechet (3P)	$a=3.1238, b=7764.6, g=-4076.2$	$a=2.8923, b=22571.0, g=-11,129.0$	$a=3.6283, b=20.616, g=-6.7647$
Gamma (2P)	$a=1.7166, b=3540.6$	$a=1.71, b=11,183.0$	$a=3.0786, b=6.1845$
Gamma (3P)	$a=1.2106, b=4290, g=884.47$	$a=1.0848, b=14,723.0, g=3150.8$	$a=1.4696, b=8.3319, g=6.7958$
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)	$a=2.2801, b=4541.7$	$a=2.2731, b=14,202.0$	$a=3.6928, b=16.426$
Log-Logistic (3P)	$a=2.0775, b=4217.5, g=423.18$	$a=1.8662, b=12,305.0, g=2091.6$	$a=2.3027, b=10.393, g=5.654$
Log-Pearson (3P)	$a=663.54, b=-0.02887, g=27.608$	$a=1787.0, b=-0.01771, g=41.234$	$a=14.523, b=0.12506, g=1.0099$
Gumbel max (2P)	$s=3617.0, m=3990.2$	$s=11402.0, m=12,541.0$	$s=8.4608, m=14.156$
Lognormal (2P)	$s=0.7362, m=8.4513$	$s=0.74093, m=9.5943$	$s=0.47178, m=2.826$
Lognormal (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)	$a=1.599, b=6398.7$	$a=1.5993, b=20008.0$	$a=2.5437, b=20.375$
Weibull (3P)	$a=1.1175, b=5389.8, g=899.42$	$a=1.0689, b=16,369.0, g=3155.6$	$a=1.1951, b=12.878, g=6.9279$
Inv. Gaussian (2P)	$l=10,434.0, m=6078.0$	$l=32,699.0, m=19,122.0$	$l=58.617, m=19.04$
Johnson SB (4P)	$g=1.5161, d=0.74495$ $l=27,319.0, x=1304.2$	$g=2.2027, d=1.0357,$ $l=1.3052E+5, x=961.8$	$g=2.5314, d=0.92215,$ $l=118.81, x=8.2791$
Gen. gamma (3P)	$k=1.054, a=1.8127, b=3540.6$	$k=1.0521, a=1.8019, b=11,183.0$	$k=1.0877, a=3.4664, b=6.1845$

Table 8 Performance of different univariate models for fitting marginal distribution for flood characteristics (a) based on K–S and A–D test distance statistics, (b) based on information criteria statistics such as AIC, BIC and HQC, (c) based on error indices statistics such as MSE and RMSE

Functions	Peak			Volume			Durations		
	<i>p</i> value	KS _{<i>n</i>} (d-max)	AD _{<i>n</i>} (d-max)	<i>p</i> value	KS _{<i>n</i>} (d-max)	AD _{<i>n</i>} (d-max)	<i>p</i> value	KS _{<i>n</i>} (d-max)	AD _{<i>n</i>} (d-max)
<i>(a)</i>									
Fretch (2P)	0.32428	0.13147	1.0751	0.28744	0.1359	1.1173	0.36268	0.1272	0.58456
Fretch (3P)	0.99732	0.05351	0.21153	0.96141	0.06828	0.3033	0.68038	0.09849	0.36878
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
Log-logistic (3P)	0.9968	0.05421	0.23129	0.9471	0.07101	0.36216	0.6921	0.09751	0.38531
Gumbel (2P)	0.4966	0.11417	0.90135	0.62555	0.10307	0.74771	0.51472	0.11255	1.0798
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
Exponential (1P)	0.03558	0.19698	2.3258	0.0784	0.17643	2.1603	0.0000381	0.32306	6.9597
Exponential (2P)	0.45829	0.11768	2.3535	0.9265	0.07425	2.094	0.25721	0.13985	1.661
Lognormal (2P)	0.9977	0.05293	0.19412	0.98539	0.06157	0.2338	0.60127	0.10511	0.4602
Lognormal (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
Weibull (3P)	0.86868	0.08134	0.28905	0.99653	0.05454	0.194	0.88156	0.07992	0.45987
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	0.84788	0.84788	14.822	0.99811	0.05222	0.17314	0.56249	0.1084	11.874
Functions	Peak			Volume			Duration		
	AIC	BIC	HQIC	AIC	BIC	HQIC	AIC	BIC	HQIC
<i>(b)</i>									
Fretch (2P)	-284.118	-280.294	-282.66	-274.569	-270.745	-273.11	-307.04	-303.22	-305.588
Fretch (3P)	-371.057	-365.32	-368.87	-353.796	-348.06	-351.61	-331.1	-325.361	-328.912
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
Log-logistic (3P)	-371.549	-365.813	-369.36	-350.302	-344.566	-348.12	-330.41	-324.673	-328.225
Gumbel (2P)	-294.924	-291.1	-291.43	-308.477	-304.653	-307.02	-293.6	-289.775	-292.143
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
Exponential (1P)	-242.501	-240.589	-164.07	-248.425	-246.513	-247.7	-179.33	-177.416	-178.6
Exponential (2P)	-249.916	-246.092	-248.46	-345.901	-342.076	-344.44	-280.58	-276.758	-279.126
Lognormal (2P)	-379.344	-375.52	-377.89	-371.028	-367.204	-369.57	-327.46	-323.633	-326.001
Lognormal (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
Weibull (3P)	-200.361	-194.625	-200.9	-376.477	-370.741	-374.29	-200.15	-194.418	-197.97
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
Functions	Peak		Volume		Duration				
	MSE	RMSE	MSE	RMSE	MSE	RMSE			
<i>(c)</i>									
Fretch (2P)		0.00314	0.05607	0.00380	0.06168	0.00199	0.04458		

Table 8 (continued)

Functions	Peak		Volume		Duration	
	MSE	RMSE	MSE	RMSE	MSE	RMSE
Fretch (3P)	0.00053	0.02304	0.00075	0.02737	0.00118	0.03436
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
Log-logistic (3P)	0.00053	0.02293	0.00081	0.02835	0.001197	0.034594
Gumbel (2P)	0.00263	0.05136	0.00193	0.04394	0.0026007	0.05099
Gamma (2P)	0.00111	0.03341	0.00068	0.02624	0.005037	0.070973
Gamma (3P)* (for D)	0.01173	0.10882	0.01327	0.11520	0.000918	0.030312
Exponential (1P)	0.00075	0.08672	0.00668	0.08174	0.02661	0.16312
Exponential (2P)	0.00623	0.07890	0.00091	0.03022	0.003374	0.058087
Lognormal (2P)* (for P)	0.00046	0.02163	0.00055	0.02351	0.001321	0.03635
Lognormal (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
Weibull (3P)	0.01612	0.12699	0.00048	0.02182	0.01619	0.12726
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)* (for V)	0.00093	0.03053	0.00041	0.02028	0.00972	0.09861

Bold letter such as Lognormal (2P), Johnson SB-4P and Gamma-3P is recognized as best-fitted probability functions for defining marginals of flood characteristics P, V and D because it exhibited minimum test statistics values

K-S test Kolmogorov–Smirnov test, *A–D test* Anderson–Darling test, *AIC* Akaike Information Criteria, *BIC* Bayesian information criteria, *HQIC* Hannan–Quinn information criteria, *RMSE* root mean square error, *MSE* mean square error

Schwarz 1978) and Hannan–Quinn Information criteria (HQIC) (i.e. Hannan and Quinn 1979) and error indices statistics such as mean square error (or MSE) and root mean square error (or RMSE) (i.e. Singh and Demissie 2004; Moriasi et al. 2007) are incorporated for selecting the possible marginal structure of peak flow, volume and duration series based on the comparative assessments between their empirical cumulative and theoretical probabilities. The empirical nonexceedance probabilities are estimated from the Gringorten-based position-plotting formula using Eq. (23). Performance of different univariate models for fitting marginal distribution for flood characteristics is listed in Table 8a–c. It reveals that the performance of Lognormal (2P) distribution is much satisfactory for flood peak flow samples in comparison with other candidate functions such as the K–S value and A–D test statistics (KS_n (d-max) = 0.05293 with p value 0.9977) and (AD_n (d-max) = 0.19412) where the D-critical value for K–S and A–D test for sample size 50 is 0.1884 and 2.5018 at 5% significance level. Similarly, the performance of Johnson SB-4P and Gamma-3P is much consistent for representing flood volume and duration series such as K–S value and A–D test value for Johnson SB-4P distribution (KS_n (d-max) = 0.0522 with p value 0.99811) and (AD_n (d-max) = 0.17314), and for Gamma-3P distribution it is (KS_n (d-max) = 0.07865 with p value 0.89254) and (AD_n

(d-max) = 0.37708). Statistically, if the estimated K–S and A–D values are below their critical level at the significance level ‘ α ’, then it often indicated better model performance with the observed flood characteristics. Information criteria-based statistics are also incorporated to find out the acceptability of the distribution functions which are pointed on the basis of K–S and A–D test and which indicates that the AIC, BIC and HQIC values are at minimum for Lognormal (2P) (i.e. AIC = – 379.344, BIC = – 375.52, HQC = – 377.89) for representing flood peak flow value, Johnson SB-4P distribution (i.e. AIC = – 381.821, BIC = – 374.173, HQC = – 378.91) for representing flood volume samples and Gamma-3P distribution (i.e. AIC = – 343.62, BIC = – 337.886, HQIC = – 259.089) for representing flood duration series. Minimum the AIC, BIC and HQIC value always indicates for the most suitable model. Error indices statistics such as based on RMSE and MSE measures are also in support of the distribution function selected from distance criteria statistics and information criteria statistics such as Lognormal (2P) distribution (MSE = 0.00046, RMSE = 0.02163) for flood peak, Johnson SB-4P (MSE = 0.0004112, RMSE = 0.02163) for volume and Gamma-3P distribution (MSE = 0.000918804, 0.030312) for representing duration samples. Overall, after summarizing all the analytical testing measures, it is pointing towards the Lognormal (2P) distributions for

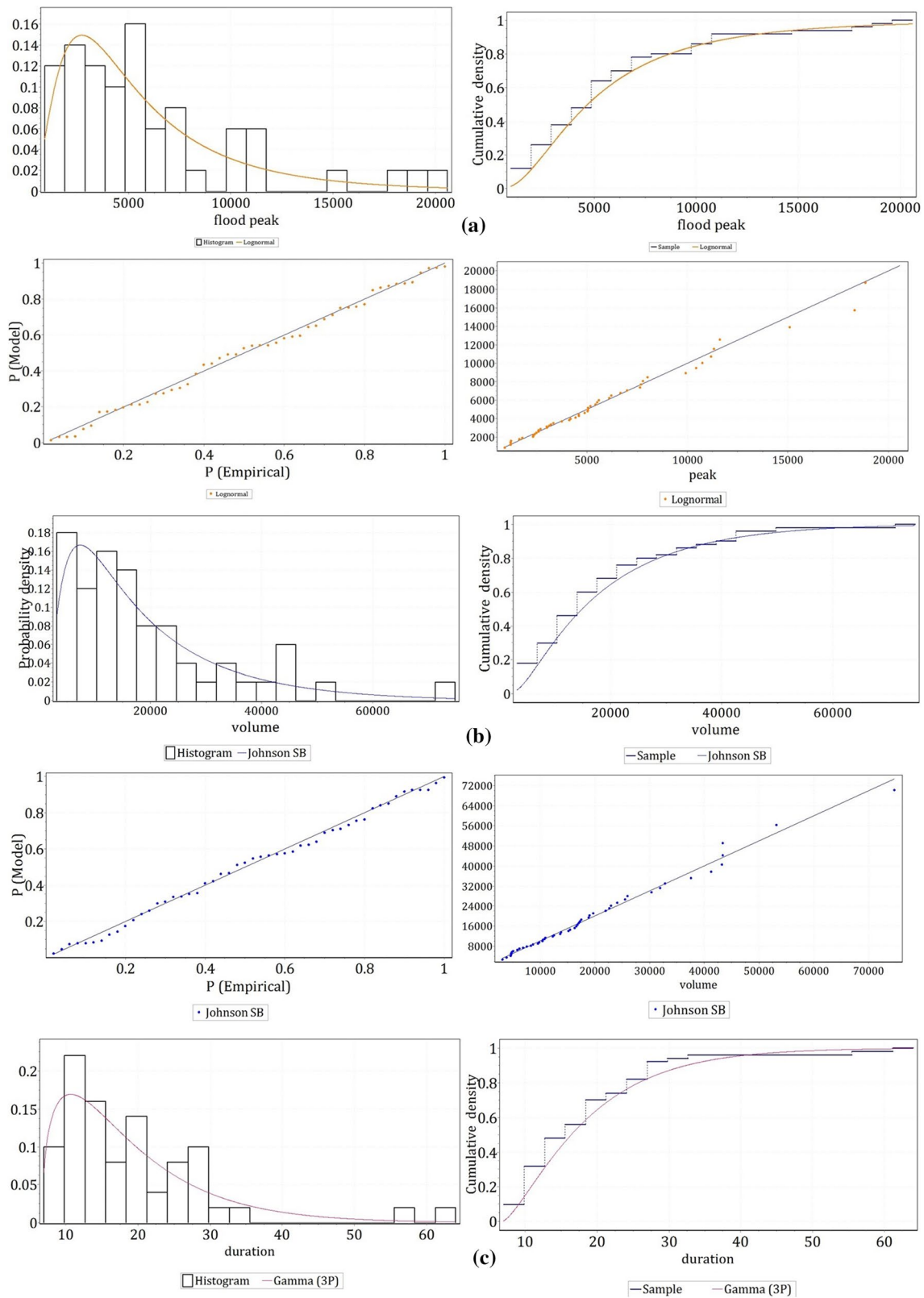


Fig. 8 Fitted marginal distribution and their probability density functions (PDFs), cumulative distribution functions (CDFs), plot for flood characteristics **a** flood peak flow, **b** volume and **c** duration series

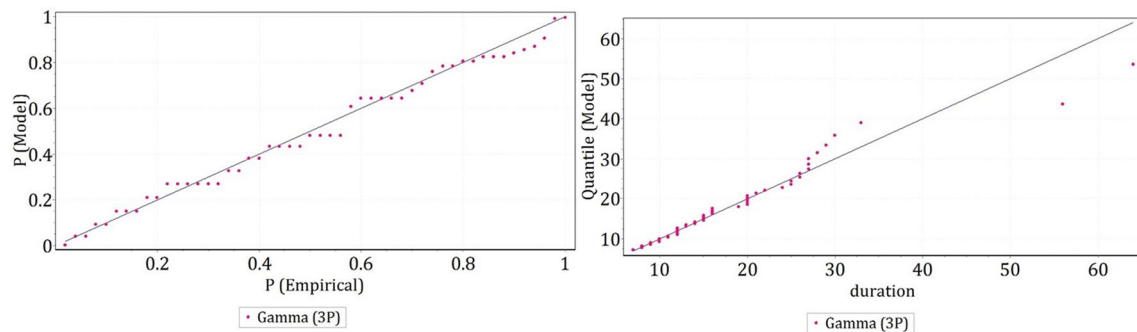


Fig. 8 (continued)

flood peak discharge flow, Johnson SB-4P for volume and Gamma-3P for modelling flood duration series. The probability density or PDFs plot and cumulative distribution or CDFs plot of the best-fitted marginal distribution of the flood characteristics are illustrated in Fig. 8a–c.

Modelling of joint dependence structure using bivariate copulas

In this demonstration, several bivariate copula families such as one elliptical family, i.e. Gaussian copula, mono-parametric and mixed version of bi-parametric Archimedean families, i.e. Clayton, Gumbel, Frank, Joe, BB1, BB6, BB7 and BB8 copulas, also rotated versions Archimedean copula families, i.e. rotated Gumbel by 90 degrees, are incorporated, and then, most appropriate copulas are selected for bivariate joint distribution analysis of flood characteristics. The rotated Archimedean copulas are included to cover negative dependence as well. In the modelling of flood peak–volume series, the rotated Clayton copula can't be used because of positively dependent random variables between these random pairs which is only applicable to capture negative flood correlation structure (i.e. Kendall's tau < 0). Similarly, the Gumbel–Hougaard, Clayton, Joe, BB1, BB6, BB7, BB8 copulas can't be used for negatively dependent data (i.e. Kendall's tau < 0) such as between P–D, which is only applicable for positively correlated random variables such as between flood pair, P–V.

The vector of unknown statistical or dependence parameters of copula families fitted to observed flood attribute pairs is estimated using method of moment-like (MOM) estimators based on the inversion of Kendall's tau and maximum pseudo-likelihood (MPL) estimators using Eqs. (6)–(9) for MOM estimator and based on Eq. (10) for MPL-based copula parameter estimation. The MOM estimators can't be used for the estimation of bi-parametric copulas (i.e. having more than one dependence parameter) such as BB1, BB6, BB7 and BB8 copula, but MPL estimator can be used

for estimating one parameter or more than one parameter copula. The estimated copula parameters fitted to P–V and P–D random pairs using both MPL and MOM estimators are listed in Tables 9a, b and 10a, b.

To analytically validate and identify the most justifiable copula for describing joint distribution for both the estimators (i.e. MOM and MPL), it is investigated by means of Cramer–von Mises distance statistics with parametric bootstrap method which is employed using Eqs. (11) and (12). For this purpose, the test statistics ' S_n ' and its associated p value have been computed from 1000 and 500 simulated random samples by means of parametric bootstrap procedure as given in Tables 9a, b (for MPL-based estimation procedure) and 10a, b (for MOM-based copulas estimation procedure). Table 9a, b shows the result based on MPL-based copula simulations; it reveals that the Gaussian copula exhibited minimum ' S_n ' statistics (i.e. ' S_n ' = 0.013444) and the highest p value (i.e. p value = 0.9356 for N = 1000 bootstrap samples and p value = 0.9411 for N = 500 random bootstrap samples) for flood peak–volume pair, as compared to another copula function. From the same table, it reveals that the frank copula is identified as the most suitable bivariate model for capturing joint structure of flood peak–duration (i.e. ' S_n ' = 0.031215, p value = 0.4001 for N = 1000 bootstrap samples and p value = 0.3762 for N = 500 bootstrap samples). Similarly, Table 10a, b shows the estimations based on MOM estimators, the Gaussian copula recognized as most suitable for constructing joint distribution of flood peak–volume and peak–duration samples (i.e. ' S_n ' = 0.01579, p value = 0.9156 for N = 1000 bootstrap samples and p value = 0.9012 for N = 500 bootstrap samples), for P–V random pairs, (i.e. ' S_n ' = 0.030559, p value = 0.4101 for N = 1000 bootstrap samples and p value = 0.3982 for N = 500 bootstrap samples), for P–D random pairs. Also, the Kendall's tau τ values estimated from the derived copulas fitted to flood attribute pairs are also listed in Tables 9 and 10 for each copula function. Besides the analytical fitness measures, graphical visual inspections are also carried

Table 9 Estimating copula dependence parameters based on the maximum pseudo-likelihood (MPL) estimator and their corresponding goodness-of-fit statistics via parametric bootstrap technique for (a) flood peak flow and volume and (b) peak flow and duration

(a) For (P–V) samples via MPL estimator				N= 1000 (No. of bootstrap sampling)		N= 500 (No. of bootstrap sampling)		Kendall’s tau (τ *) estimated from fitted copula
Copula family	Parameter estimates $\hat{\theta}$	Standard error SE	Maximized log-likelihood	S_n	(p value) S_n	S_n	(p value) S_n	
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
Rotated Clayton (90°)	NA	NA	NA	NA	NA	NA	NA	NA
BB1	PAR1 = THETA = 1.44, PAR2 = DELTA = 1.49	NA	27.99	0.016513	0.6988	0.016513	0.6768	0.6098018
BB6	PAR1 = THETA = 1.001, PAR2 = DELTA = 2.310108	NA	22.21	0.02777	0.1513	0.02777	0.1707	0.5673704
BB7	PAR1 = THETA = 1.756445, PAR2 = DELTA = 2.4499998	NA	27.77	0.0189	0.49	0.0189	0.4721	0.5955592
BB8	PAR1 = THETA = 6, PAR2 = DELTA = 0.734783	NA	21.98	0.036858	0.08541	0.036858	0.09281	0.5510257

(b) For (P–D) samples				N= 1000 (No. of bootstrap sampling)		N= 500 (No. of bootstrap sampling)		Kendall’s tau (τ *) estimated from fitted copula
Copula family	Parameter estimates $\hat{\theta}$	Standard error SE	Maximized log-likelihood	S_n	(p value) S_n	S_n	(p value)	
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
Rotated Clayton (90°)	– 0.3116	NA	1.134	0.049363	0.2672	0.049363	0.2864	– 0.1347984
Joe	NA	NA	NA	NA	NA	NA	NA	NA
BB1	NA	NA	NA	NA	NA	NA	NA	NA
BB6	NA	NA	NA	NA	NA	NA	NA	NA
BB7	NA	NA	NA	NA	NA	NA	NA	NA
BB8	NA	NA	NA	NA	NA	NA	NA	NA

NA denotes that SE values are not available for this copula function. Also, rotated Clayton copula can’t be used for positively dependent random variables, which is only applicable to capture negative flood correlation structure. Bold letter indicated via * denotes that the Gaussian copula exhibiting minimum S_n value, which means the performance of Gaussian copula for P–V is much consistent than the other copula functions. (τ *) in the last column of above table indicated the estimated kendall’s tau value from derived copulas fitted to observed random series]

Bold letter indicated via * denotes that the performance of Frank copula is most satisfactory that other copulas for the flood pair P–D. NA denotes that for Gumbel–Hougaard, Clayton, Joe, BB1, BB6, BB7, BB8 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)]

out based on scatter plot (Fig. 9a, b) of the 1000 sample observations which are simulated from the joint distribution using copula functions which are recognized as most consistent (i.e. based on analytically goodness-of-fit measure) for both MPL- and MOM-based copula dependence parameter estimation procedures. The Kendall’s τ values of the simulated samples are also shown in Fig. 9a, b. Similarly,

Fig. 10a, b illustrates the chi-plot and K-plot of the set of 1000 random samples drawn from the best-fitted bivariate copulas derived from MOM and MPL estimators for P–V and P–D combinations. From the comparison of scatter plots of simulated data (green colour) with overlapped observed flood characteristics (red colour), it can be revealed that the Gaussian copula performed satisfactorily for capturing the

Table 10 Dependence parameter estimates of copulas function based on the Inverse of Kendall’s tau [method of moment (or MOM)] and their corresponding goodness-of-fit statistics based on the parametric bootstrap technique for (a) flood peak flow and volume and (b) peak flow and duration pairs

(a) For (P–V) samples			N= 1000 (No. of bootstrap sampling)		N= 500 (No. of bootstrap sampling)		Kendall’s tau (τ^*) estimated from fitted copula	Verification via iTau (calculated value), calibration of fitted copula parameter
Copula family	Parameter estimates $\hat{\theta}$	Standard error SE	S_n	(p value) S_n	S_n	(p value) S_n		
Gaussian*	0.8159716	0.054	0.01579	0.9156	0.015797	0.9012	0.6075949	0.8159716
Clayton	3.097	0.774	0.027508	0.2073	0.027508	0.1727	0.6067123	3.096775
Gumbel–Hougaard (GH)	2.548387	0.387	0.019973	0.6199	0.019973	0.5818	0.6075949	2.848388
Frank	8.135964	1.009	0.022772	0.4461	0.022772	0.508	0.607595	8.135964
Joe	3.921185	NA	0.036051	0.03447	0.036051	0.02495	0.607595	3.921185
Rotated Clayton (90°)	NA	NA	NA	NA	NA	NA	NA	NA
BB1	NA	NA	NA	NA	NA	NA	NA	NA
BB6	NA	NA	NA	NA	NA	NA	NA	NA
BB7	NA	NA	NA	NA	NA	NA	NA	NA
BB8	NA	NA	NA	NA	NA	NA	NA	NA

(b) For (P–D) samples			N= 1000 (No. of bootstrap sampling)		N= 500 (No. of bootstrap sampling)		Kendall’s tau (τ^*) estimated from fitted copula	Verification via iTau (calculated value), calibration of fitted copula parameter
Copula family	Parameter estimates $\hat{\theta}$	Standard error SE	S_n	(p value)	S_n	(p value)		
Gaussian*	–0.1163	0.178	–0.0742068	0.030559	0.4101	0.030559	0.3982	–0.1162625
Frank	–0.6706	0.003	–0.0741785	0.030644	0.3881	0.030644	0.4401	–0.6706383
Rotated Clayton	–0.1602536, par2=0	NA	–0.0741827	0.031419	0.537	0.031419	0.488	–0.1602536
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
BB1	NA	NA	NA	NA	NA	NA	NA	NA
BB6	NA	NA	NA	NA	NA	NA	NA	NA
BB7	NA	NA	NA	NA	NA	NA	NA	NA
BB8	NA	NA	NA	NA	NA	NA	NA	NA

The bold letter indicated via * represents that Gaussian copula poses the minimum value of S_n statistics and thus most consistent bivariate functions than other copulas for the flood pair P-V. NA denotes that the parameters of the two parameters BB1, BB6, BB7 and BB8 copula can’t be estimated via MOM estimators hence, standard error values and S_n statistics are not available. Rotated Clayton copula can’t be used for positively dependence random series. Also, (τ^*) represented the value of Kendall’s tau estimated from the fitted copulas with the flood characteristics

The bold letter indicated via * represents that Gaussian copula poses the minimum value of S_n statistics for the flood pair P-D as compared with the other bivariate copula functions; NA denotes that all such copulas cannot be used for negative dependence modelling between the random variables, only applicable for positively correlated flood characteristics

observed dependence of flood variables for P–V and P–D series in case of MOM-based estimators since the simulated data are adequately overlapped with the natural dependence of observed flood characteristics. Similarly, for MPL estimators, the performance of Gaussian copula (for P–V samples) and Frank copula (for P–D samples) is much satisfactorily and in support of analytical fitness test result.

Joint and conditional distributions and their associated return periods

The univariate return period of the flood characteristics is derived from best-fitted univariate probability distribution functions (i.e. Lognormal (2P) for flood peak, Johnson

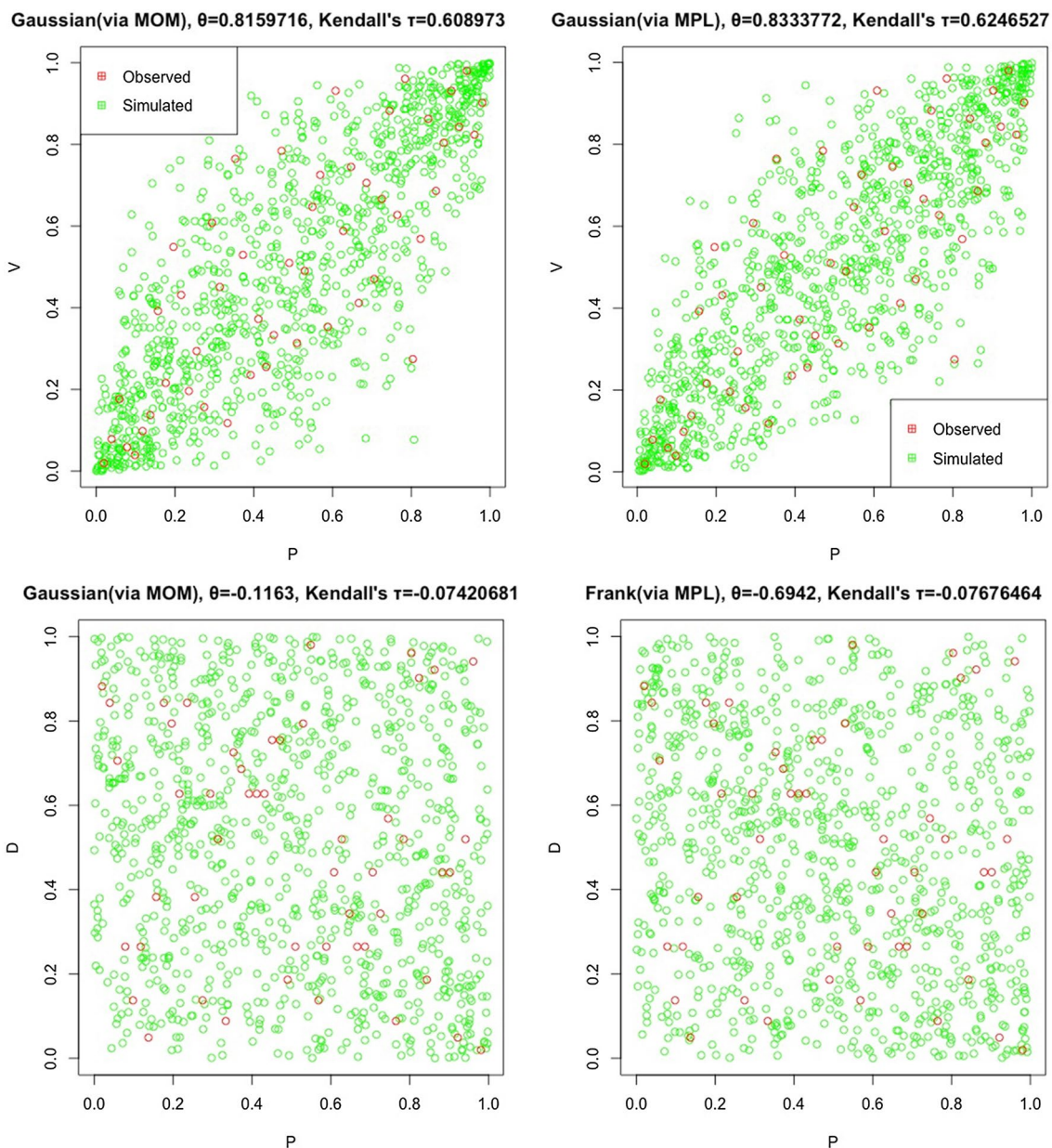


Fig. 9 Scatter plot for comparing observed (red colour) flood peak–volume and peak–duration series with the set of 1000 sample observations simulated (green colour) from best-fitted bivariate copula

functions for both MPL- and MOM-based dependence parameter estimation procedures

SB-4P for flood volume and Gamma-3P for flood duration series) using Eq. (13) and represented in Fig. 11.

Univariate return periods would attribute for underestimations or overestimations of hydrologic risk. This approach would be useful when only one flood characteristic is significant in design criteria, but in practical the considerations of multiple design variables are required for revealing effective risk practices. The best-fitted copulas which are identified in the previous section for both MOM and MPL estimators are employed for establishing

joint distribution relationship, and their associated bivariate joint return between flood peak–volume and peak–duration pairs for both the ‘OR’ and ‘AND’ joint cases is estimated using Eqs. (14) and (15). Figures 12 and 13 represent the joint probability density function (JPDF) and joint cumulative distribution function (JCDF) plots (i.e. three-dimensional scatter plot, surface plot and contour plot) of flood peak–volume pairs using best-fitted Gaussian copula whose parameter is estimated using MOM and MPL estimator. Similarly, Figs. 14 and 15

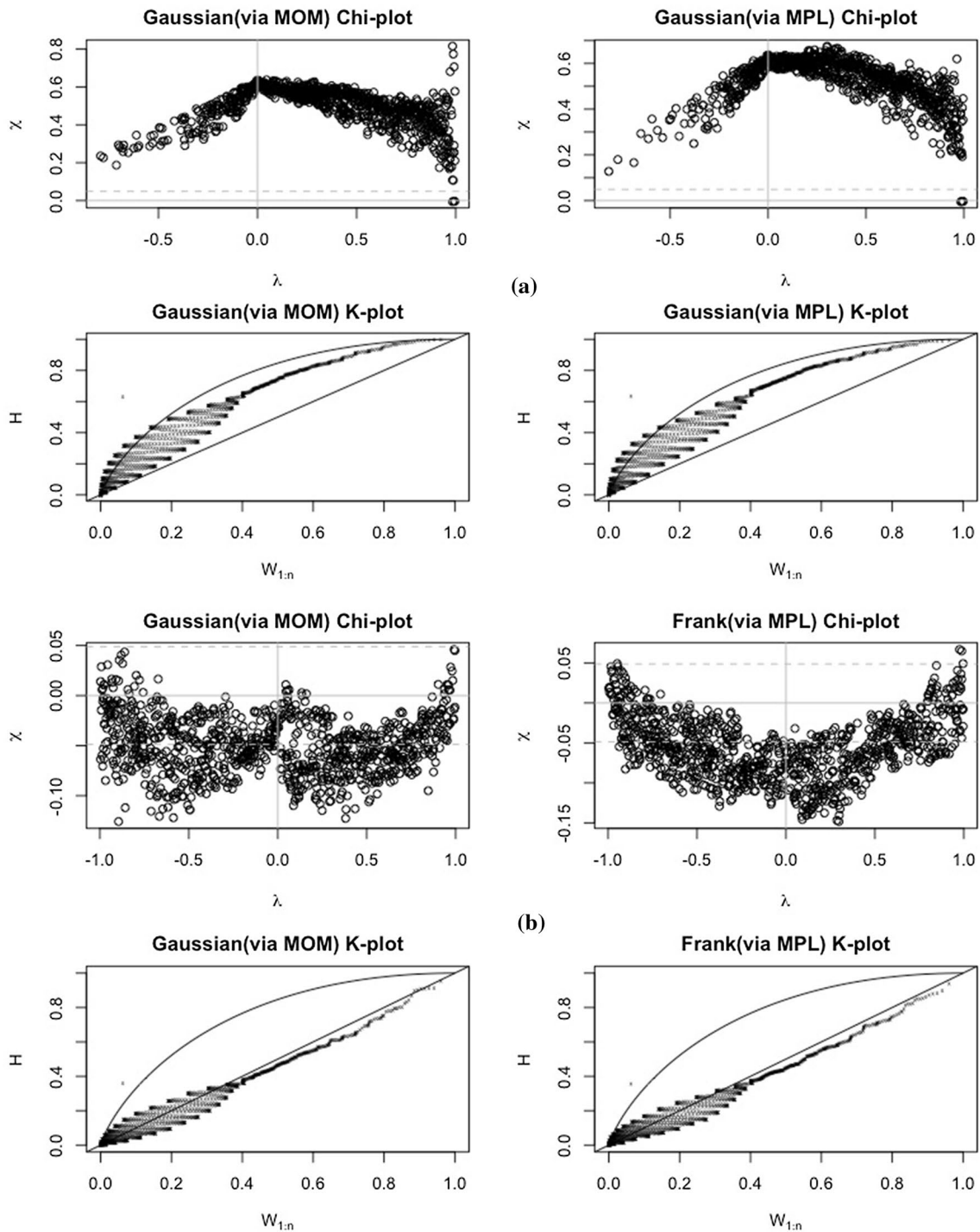


Fig. 10 Chi-plots and K-plots of the set of 1000 random sample drawn from best-fitted copulas using MOM and MPL-based dependence parameter estimation procedure fitted to **a** flood peak and vol-

ume (based on Gaussian copula) and **b** flood peak flow and duration series (Gaussian copula via MOM estimator and Frank copula via MPL estimator)

represent the JPDFs and JCDFs of peak–duration pair using the Gaussian copula (i.e. best-fitted based on MOM estimator) and the Frank copula (best-fitted copula based on MPL estimator). The joint return periods of the flood characteristics, such as between flood peak–volume pair,

and peak–duration pair for both OR and AND joint return cases, are estimated using Eqs. (14) and (15). It should be noted that for a given return period (or joint probability distribution), there may exhibit various possible combinations among flood characteristic combinations. Figure 16a,

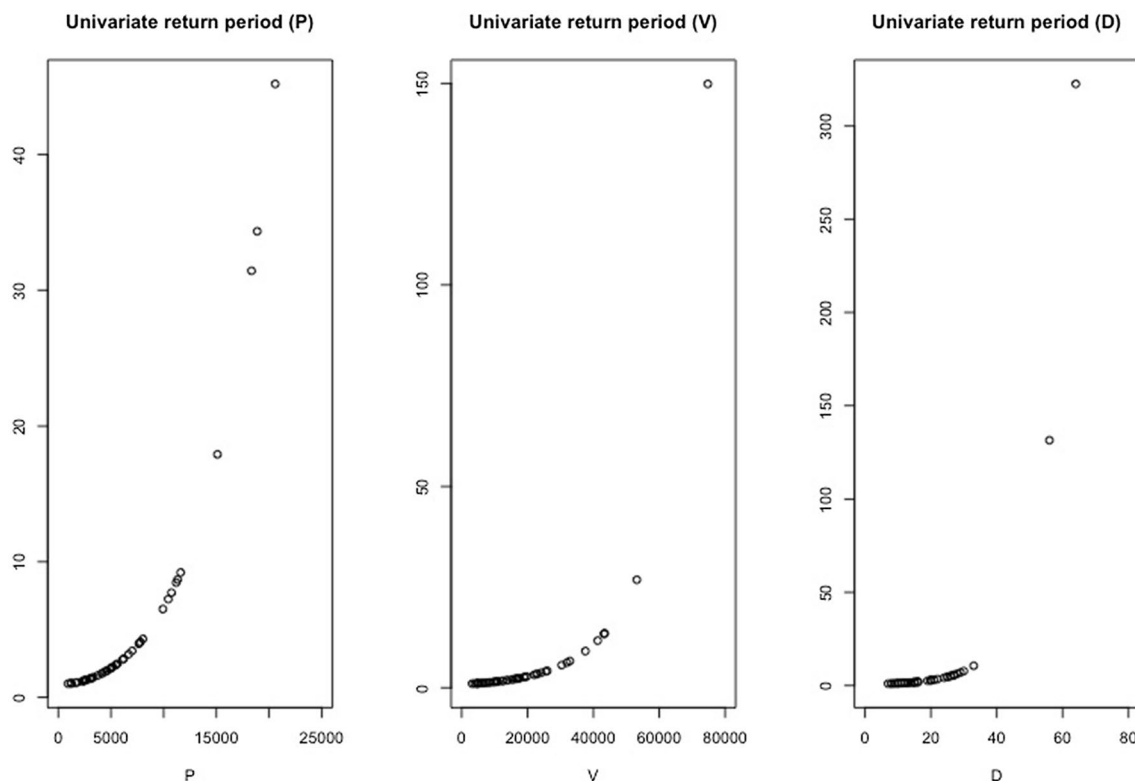


Fig. 11 Univariate return period of the flood characteristics during the period of 1961–2016 (Kelantan river basin at Guillemard Bridge gauge station in Malaysia) derived from the Lognormal (2P) distri-

bution for flood peak flow series, Johnson SB-4P for volume and Gamma-3P for duration series. *Note:* Brown circle in the Joint CDF contour plot represents observed flood peak and volume samples

b represents the scatter plot for comparing observed (in red colour) flood peak–volume and peak–duration combinations with the set of 1000 sample observations simulated (in green colour) from the joint distribution of the best-fitted bivariate copula functions and their marginal distributions for both MPL- and MOM-based parametric estimators. Figure 17 illustrates the OR and AND joint return periods derived from the best-fitted Gaussian copula for flood peak–volume series whose parameter is estimated via MOM and MPL estimators. Similarly, Fig. 18 illustrates the OR and AND joint return periods for flood peak–duration combination using the Gaussian (based on MOM estimator) and Frank copula (based on MPL estimator). Actually, in most of the engineering-based water related issues, a flood episode may be considered as dangerous if its peak discharge flow value is larger either its volume or durations is also large. In other words, performing the flood risk by considering only the single variable return periods would be problematic which often demands the joint occurrence of multiple flood characteristics. Therefore, Tables 11 and 12 show the estimated values of joint return period for both the OR and AND joint cases for few combinations between flood peak–volume and

peak–duration. From the estimated return periods, it is revealed that the OR joint return period is smaller than AND joint return periods for every possible combination of flood characteristic, i.e. $T^{OR} < T^{AND}$. For example, if a flood episode characterized by peak flow, $P = 8089.2353 \text{ m}^3 \text{ s}^{-1}$, volume, $V = 13185.141 \text{ m}^3$, the OR joint return period is $T_{PV}^{OR} = 1.765498249$ years and the AND return period is $T_{PV}^{AND} = 4.521252343$ years. Similarly, for $P = 17,019.5015 \text{ m}^3 \text{ s}^{-1}$ and $V = 53,420.634$, the OR- joint return period is $T_{PV}^{OR} = 17.614561$ years and the AND joint return period is $T_{PV}^{AND} = 50.919642$ years (Table 11b). Similarly, if a flood episode with peak, $P = 12639.4182 \text{ m}^3 \text{ s}^{-1}$, duration, $D = 14.262899$, the $T_{PD}^{OR} = 1.54$ years and $T_{PD}^{AND} = 21.60$ years and so on (Table 12a). Similar results are observed for joint return periods between flood peak–duration combinations which also reveals that the AND joint return is higher than OR joint return periods (i.e. $T_{PD}^{OR} < T_{PD}^{AND}$). It is also revealed that univariate return periods are derived from either the flood peak series, T_P or volume series, $T_V > T_{PV}^{OR}$ but produces low return periods than T_{PV}^{AND} , i.e. both $(T_P$ and $T_V) > T_{(PV)}^{AND}$ (Table 11a, b). Similarly, Table 12a, b, shows between the flood peak–duration pairs, both $(T_P$ and

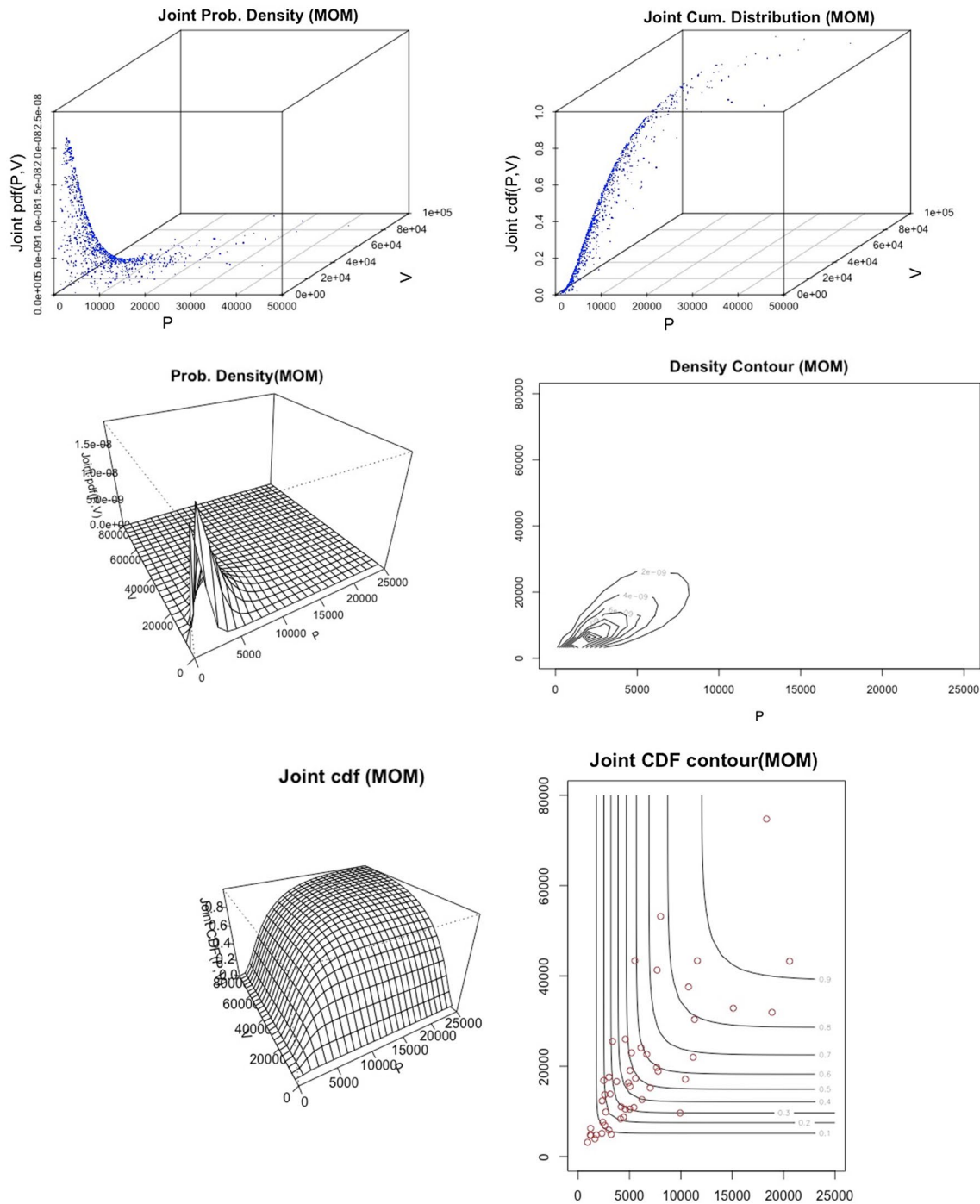


Fig. 12 Joint probability density function (JPDF) and joint cumulative distribution function (JCDF) of flood peak and volume pair using Gaussian copula derived from MOM estimator with Lognormal (2P)

and Johnson SB-4P as a best-fitted marginal distribution to P and V . *Note:* Brown circle in the Joint CDF contour plot represents observed flood peak and volume samples

$T_D) > T_{PD}^{OR}$ and $(T_P$ and $T_D) < T_{PD}^{AND}$. Overall, based on the above results and the mathematical inequalities exhibited between univariate and bivariate joint return periods or between OR and AND joint return periods, it is concluded that for an effective risk-based flood risk analysis, it could

be an essential concern and more practical to the hydrologist and water practitioner to take the accountability of both the cases of joint return periods, i.e. OR and AND joint cases, including the univariate return periods. In other words, considering just only the OR joint or AND

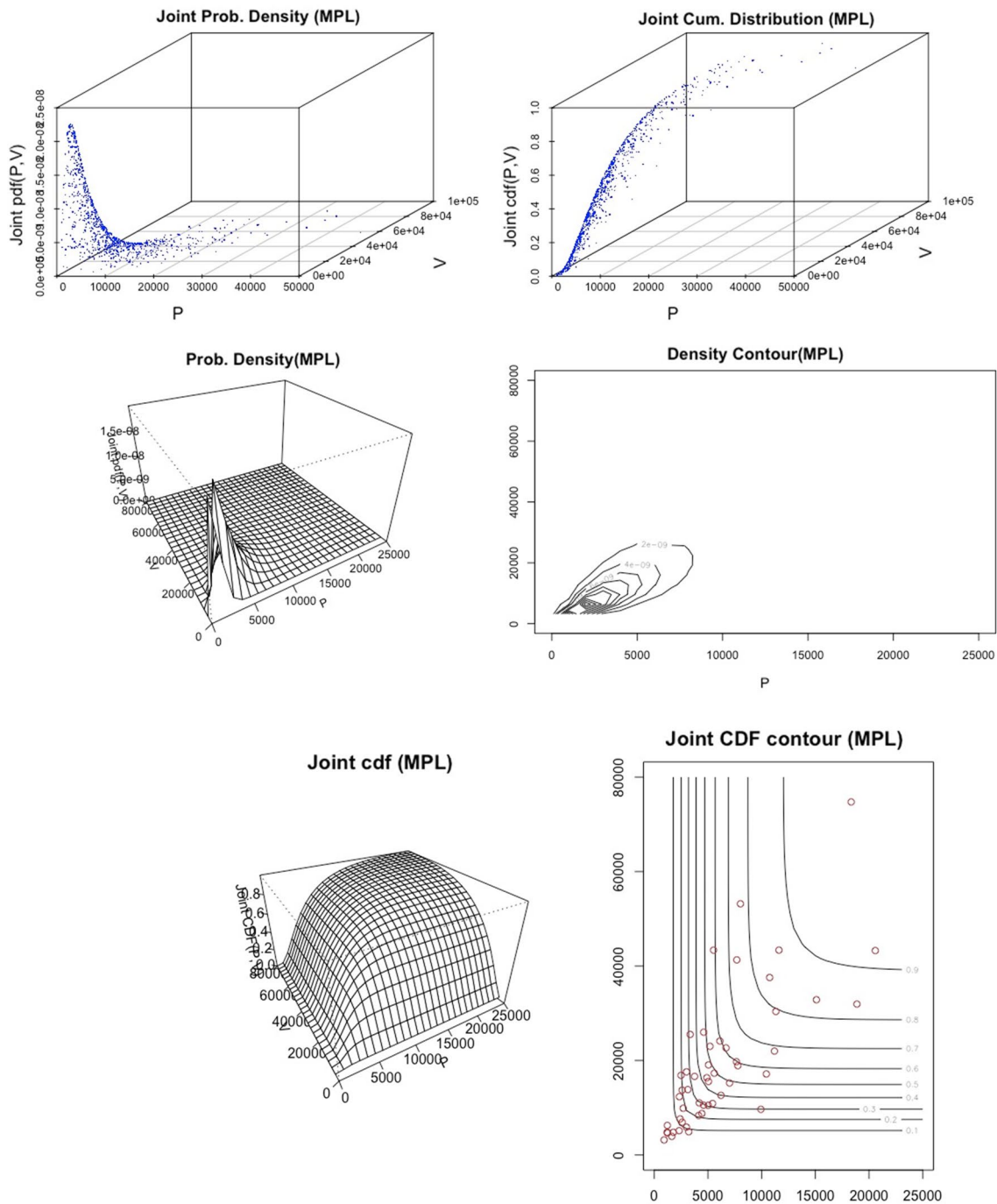


Fig. 13 Joint probability density function (JPDF) and joint cumulative distribution function (JCDF) of flood peak and volume series using best-fitted Gaussian copula derived from MPL estimators with

Lognormal (2P) and Johnson SB-4P as a best-fitted marginal to P and V. *Note:* Brown circle in Joint CDF contour plot represents the observed flood peak and duration samples

joint return periods in the flood risk analysis might be problematic and might result for the underestimation or overestimation of hydrologic risk.

Similarly, the bivariate conditional return periods of the flood peak given various percentile values of flood volume or

duration, i.e. $(T_{(P|V \leq v)}$ and $T_{(V|P \leq p)}$) or $(T_{(P|D \leq d)}$ and $T_{(D|P \leq p)}$) are estimated by using Eqs. (17) and (19), and their estimated values are listed in Tables 11 and 12. For example, if a flood episode is characterized with peak value,

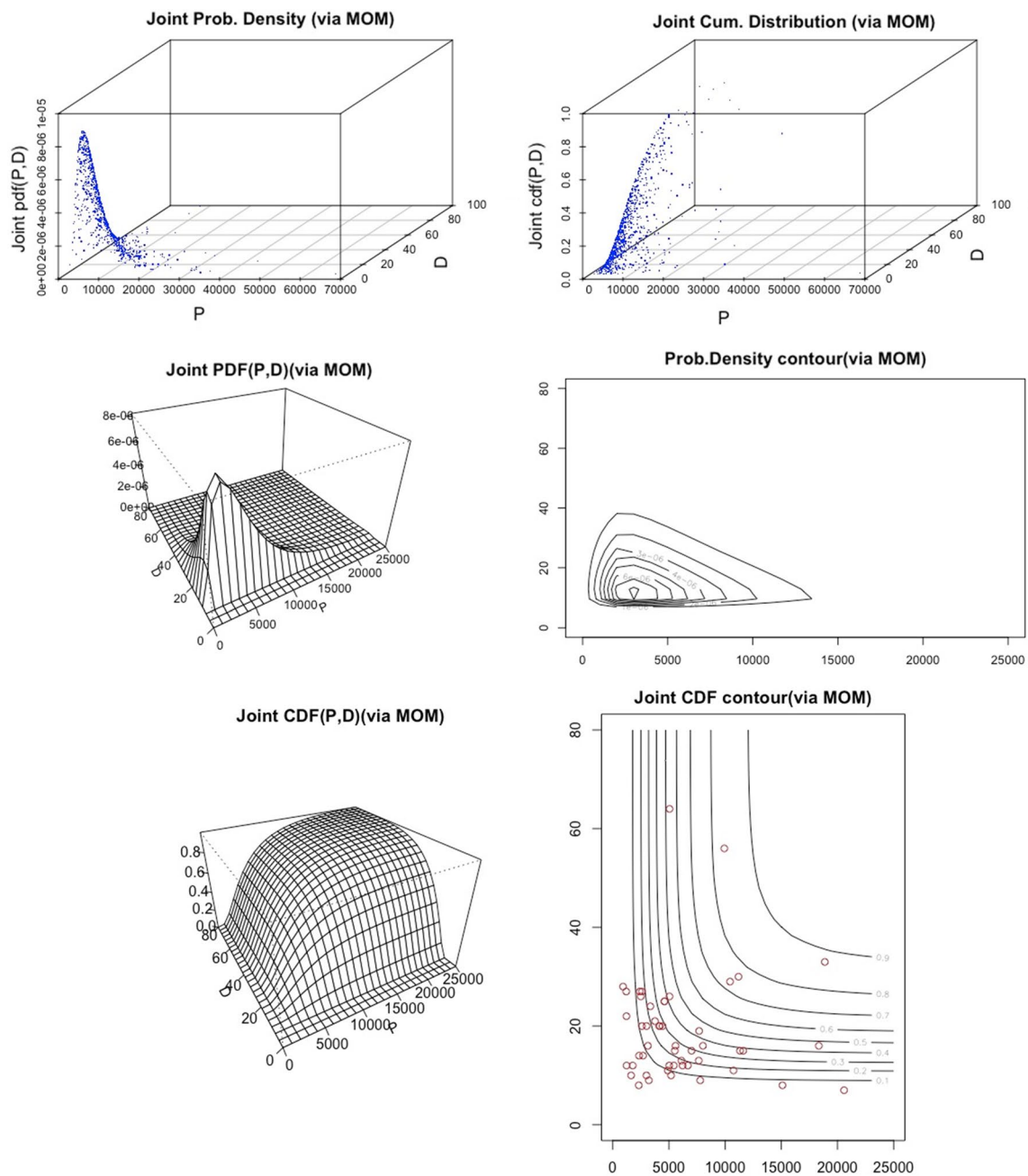


Fig. 14 MOM-based Joint probability density function (JPDF) and joint cumulative distribution function (JCDF) of flood peak and duration series using best-fitted Gaussian copula with Lognormal (2P) and

Gamma-3P as a best-fitted marginal to P and D series. *Note:* Brown circle in Joint CDF contour plot represents the observed flood peak and duration samples

$P=8089.2353 \text{ m}^3 \text{ s}^{-1}$, and volume, $V=13,185.141 \text{ m}^3$, then the conditional return periods $T(P/V \leq v)=58.33509$ years and $T(V/P \leq p)=2.28404$ years. From Table 11a, b it is also revealed that both the $T(P/V \leq v)$ and $T(V/P \leq p)$ conditional return periods are higher than the T_{PV}^{OR} , i.e. $((T(P/V \leq v)$ and $T(V/P \leq p)) > T_{PV}^{OR}$). The same mathematical inequality existed of return periods exhibited for flood pairs P–D, i.e. $T(P/D \leq d) > T_{PD}^{OR}$ and $T(D/P \leq p) > T_{PD}^{OR}$. But for the same

flood pair, $T(P/D \leq d) < T_{PD}^{AND}$ and $T(D/P \leq p) < T_{PD}^{AND}$. From Table 11, it is also revealed that $T(P/V \leq v)$ produced higher return periods than the return period derived from flood peak, T_p , and also, $T(V/P \leq p)$ is higher than volume, T_v , series. Therefore, accounting both the case of joint return periods and the conditional return periods is an essential concern for tackling several water-related queries.

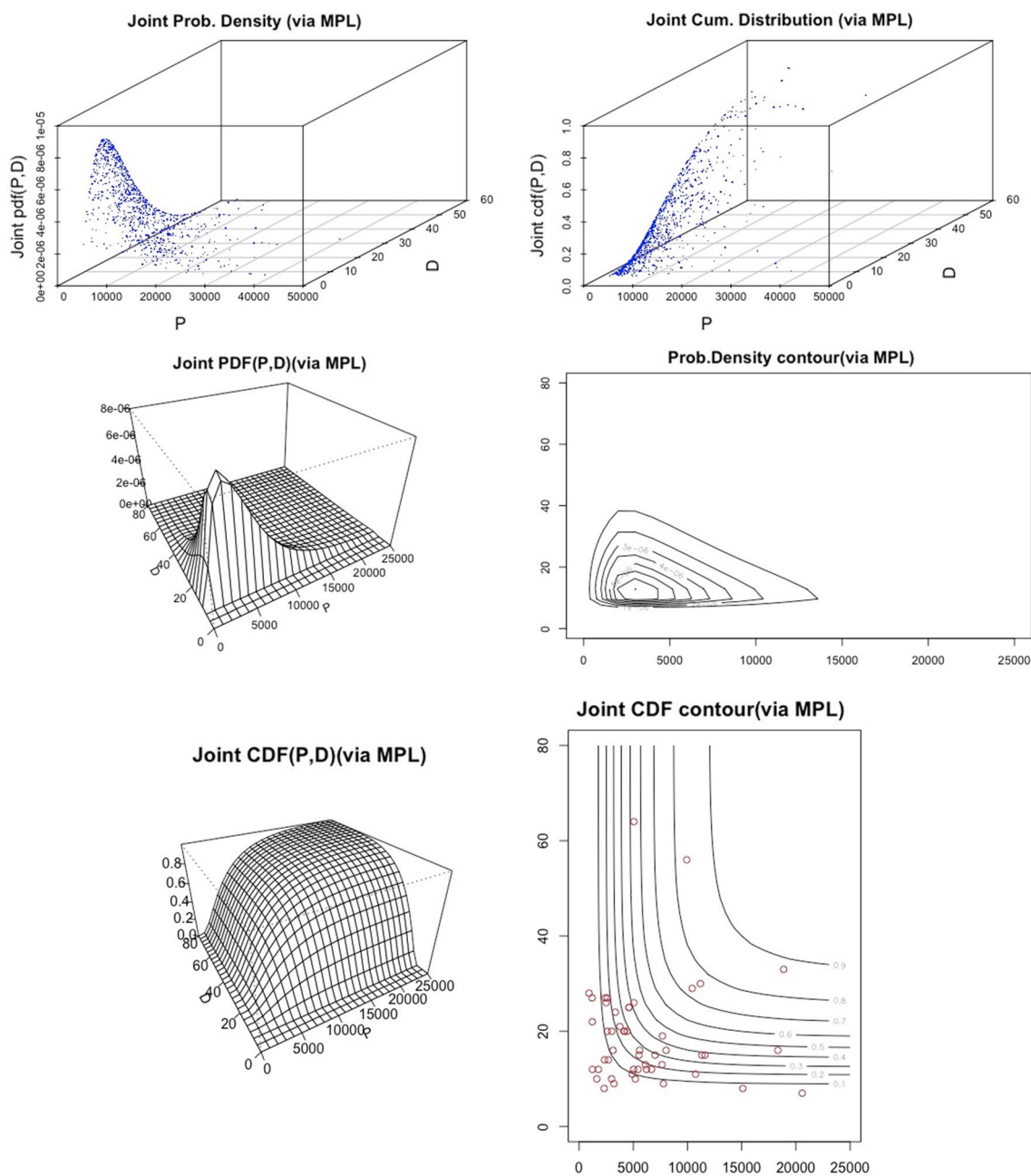


Fig. 15 MPL-based Joint probability density function (JPDF) and joint cumulative distribution function (JCDF) of flood peak and duration series using best-fitted Frank copula with Lognormal (2P) and

Gamma-3P as a best-fitted marginal to P (peak) and D (duration) series. *Note.* Brown circle in Joint CDF contour plot represents the observed flood peak and duration samples

Research conclusions

Joint probability distribution analysis of multiple interacting flood is very useful for understanding critical hydrologic behaviour of flood episodes at a river basin scale, for tackling several water resources’ planning, managements or either flood defence infrastructure designing projects. In actual, univariate flood frequency analysis or return periods would be problematic due to multidimensional

characteristics of flood episodes and might be attributed for underestimation or overestimation of hydrologic risk. Therefore, estimating multivariate design quantiles under different notations of return periods, i.e. based on conditional distribution and joint distributions, could be an effective attempt for assessing the hydrologic risk. The Kelantan River basin in Malaysia is often subjected to the most severe monsoonal flooding and perceiving for increasing in terms of their frequency and magnitude.

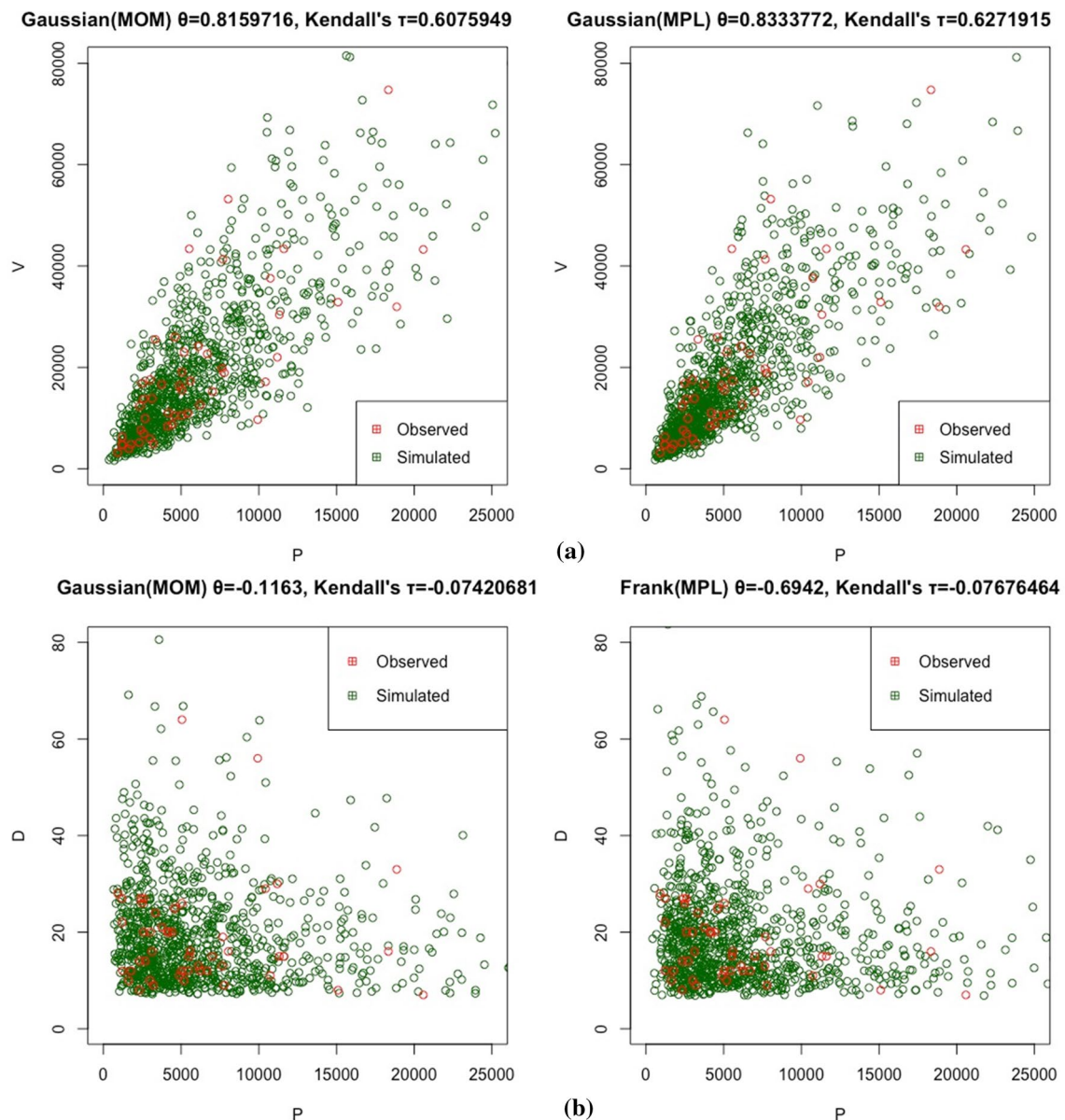


Fig. 16 Scatter plot for comparing observed flood peak–volume, peak–duration and volume–duration series with the set of 1000 sample observations simulated from joint distribution using best-fitted

copula functions and their marginal distributions for both MPL- and MOM-based dependence parameter estimation procedures **a** peak–volume and, **b** peak–duration pairs

In this study, copula-based methodology is incorporated for establishing bivariate frequency analysis of flood characteristics, i.e. flood peak discharge flow (P), volume (V) and duration (D) of flood hydrograph, and applied for a case study of the Kelantan River basin at Guillemard bridge gauge station in Malaysia. At-site event-based or block (annual) maxima-based methodology is adopted for the 50-year (1961–2016) continuously distributed streamflow characteristics of this river basin. Based on the correlation measuring statistics, from both the analytical approach (i.e. Pearson, Kendall's tau and Spearman rho correlation coefficient) and

graphical visual inspection (i.e. based on rank-based scatter plot, K-plot and Chi-plot), it is found that flood peak–flow flood volume pair exhibited higher and positive dependence structure, but both flood volume–duration and peak–duration pairs are found to be negatively correlated random pairs with very weak correlation and thus considered for flood frequency analysis. A distinguished variety of probability distribution functions (i.e. 1 parameter, 2 parameters, 3 parameters and 4 parameters) are tested under parametric approach, for defining the univariate marginal distribution of each individual flood characteristic. The vector of the

Fig. 17 OR and AND joint return period derived from both MPL- and MOM-based bivariate joint distributions of the best-fitted copula functions, i.e. Gaussian copula for flood peak (variable = a) and volume (variable = b) series

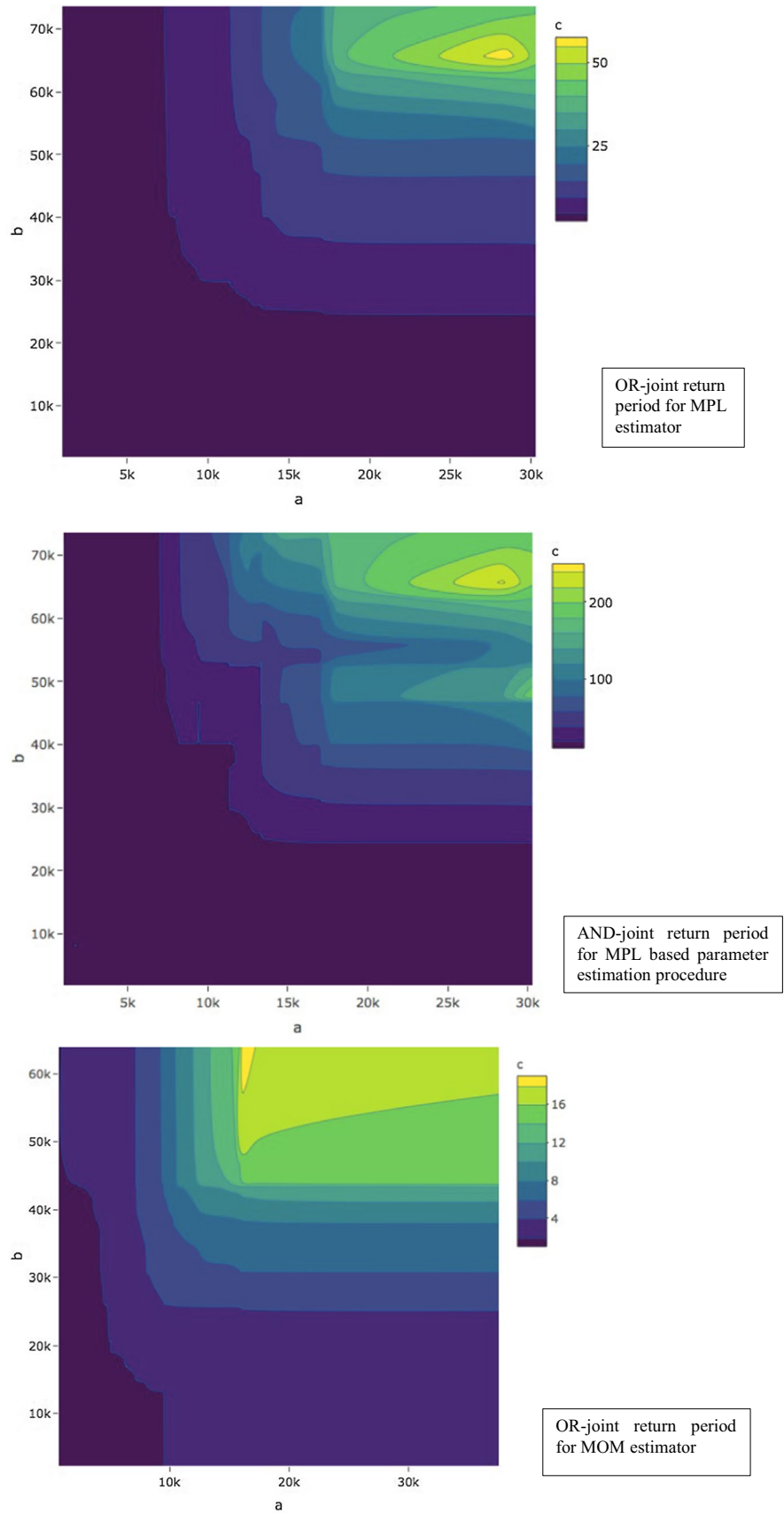
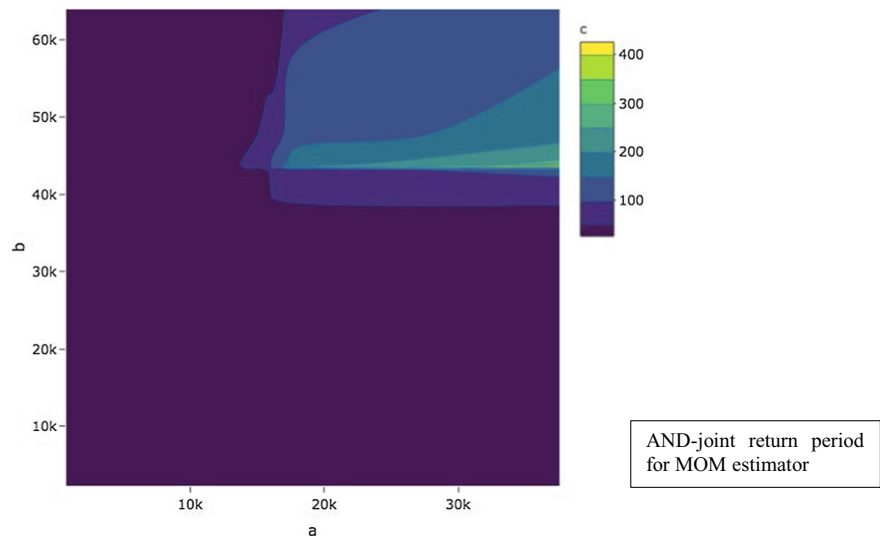


Fig. 17 (continued)



unknown statistical parameter of the fitted distributions is estimated based on maximum likelihood estimation (MLE), method of moments (MOM), least square method (LS) and L statistics-based method of L-moments. Based on fitness test statistics, best-fitted marginal distributions for each flood characteristic are introduced for copula dependence modelling. A distinguished variety of bivariate copula families such as the mono-parametric (or 1 parameter), Archimedean copulas (i.e. Clayton, Frank, Gumbel, Joe), mixed form of bi-parametric (i.e. 2 parameters) Archimedean copulas (i.e. BBI based on mixing between Clayton and Gumbel copula, BB6 mixing between Joe and Gumbel copula, BB7 mixing between Joe and Clayton copula, and BB8 mixing between Joe and Frank copula), elliptical family copula (i.e. Gaussian), and also rotated version of Archimedean copula (i.e. rotated Clayton by 90° of rotation) have been introduced and tested to model the bivariate joint probability distribution of flood variables occurrence, i.e. between flood peak–volume, and peak–duration pairs. The parameter of copulas function is estimated using the method of moment (MOM) based on inversion of Kendall’s tau (τ) and maximum pseudo-likelihood (MPL) estimators. The performance of each fitted copula is tested using both graphical and analytical approaches (i.e. based on Cramer–von Mises test statistics), and thus, best-fitted copulas are used for defining the joint and conditional return periods of flood characteristics.

From the present demonstrations, the following specific research conclusions are drawn, as given below:

- Based on the Q-statistics and their estimated p value for different lag sizes, it is concluded that the time series of each individual flood characteristic exhibited almost negligible or zero first-order autocorrelations. Also, based on nonparametric rank-based Mann–Kendall or M–K test, it reveals the acceptance of null hypothesis (H_0) which further points the existence of zero monotonic trends at the 5% or 0.05 level of significance within each flood series. In conclusion, no significant trends are detected for the flood characteristics.
- Based on parametric distribution fitting procedure for defining univariate flood marginals, it is pointing towards the Lognormal (2P) distributions for flood peak discharge flow, Johnson SB-4P for volume and Gamma-3P for modelling flood duration series.
- On performing the standard GOF test for each derived copula based on MOM and MPL estimators for each flood attribute pair, it is found that the Gaussian copula is the most parsimonious or best-fitted copula for flood peak–volume and peak–duration series using MOM-based parametric copula estimations. On the other hand, the Gaussian copula is recognized as best-fitted for defining flood peak–volume dependence structure and the frank copula is the most justifiable for both flood peak–duration pairs. The best-fitted copulas are employed for estimating the joint and conditional probability distributions, and their associated return periods for various possible combinations of flood characteristics (i.e. P–V and P–D combinations) are retrieved.
- Overall, copula function is found to be an effective way for establishing joint dependency simulation as well as preserving dependence structure of multiple flood characteristics, also revealing an effectiveness in retrieving both the conditional and joint return periods of the flood characteristics and thus could provide a better understanding of critical hydrologic behaviour at a river basin scale for assessing multivariate hydrologic risk.
- The present study investigated the adequacy of copula-based methodologies under the parametric setting for establishing the bivariate joint dependence structure of the flood characteristics, i.e. between flood peak–vol-

Fig. 18 OR and AND joint return period derived from both MPL- and MOM-based bivariate joint distributions of the best-fitted copula functions, i.e. Gaussian copula via MOM estimator and Frank copula via MPL estimator for flood peak (axis variable = a) and duration (axis variable = b) series

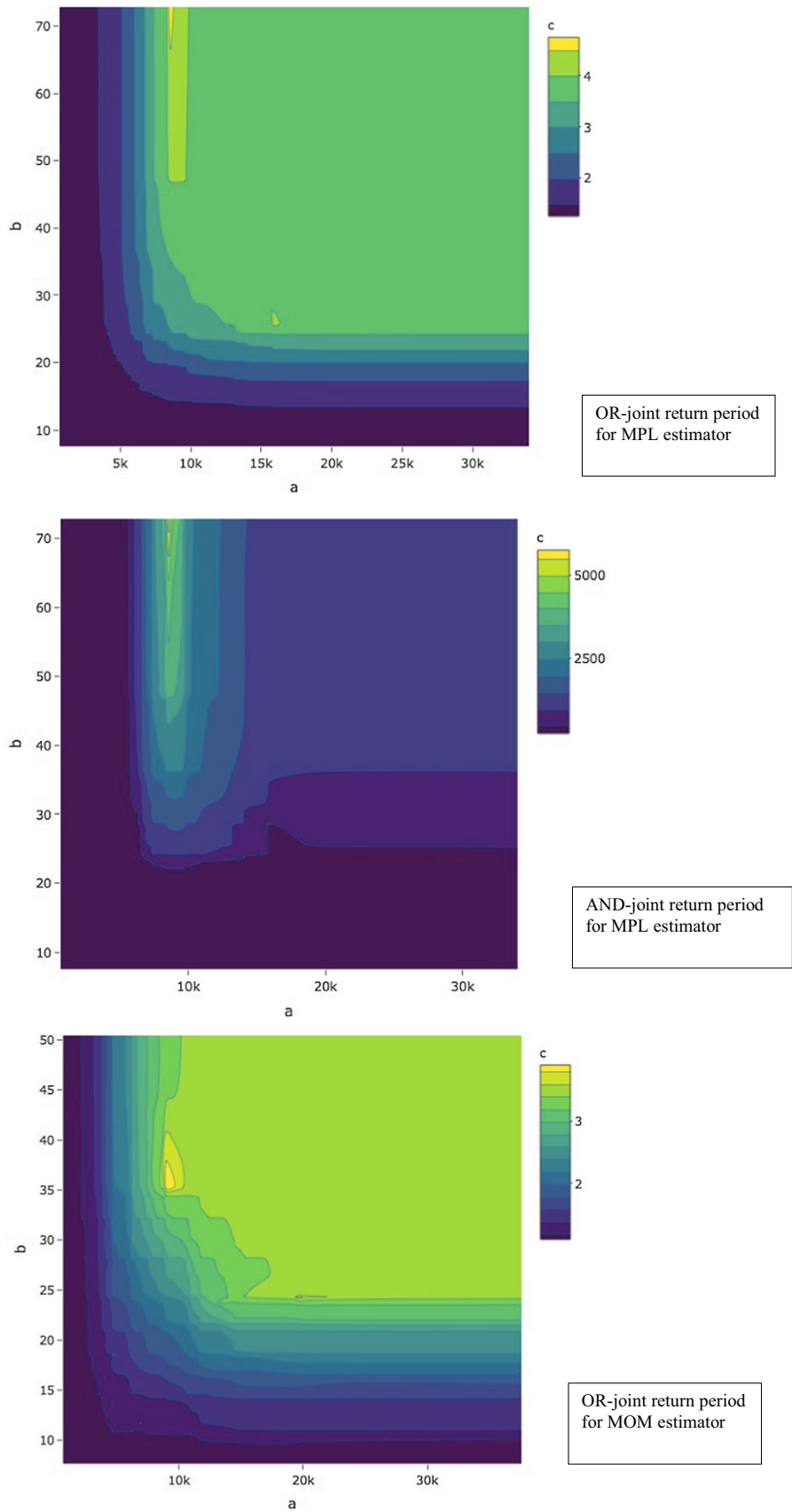


Fig. 18 (continued)

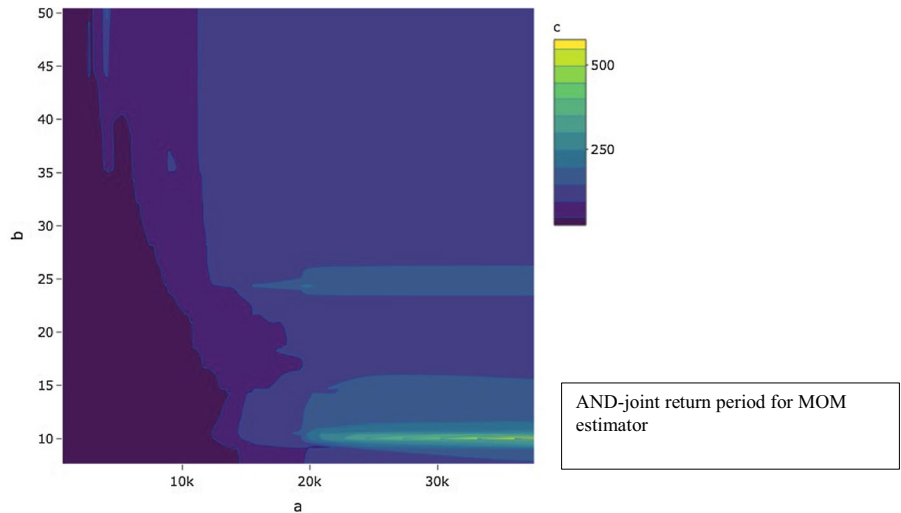


Table 11 Joint and conditional return period for the flood pairs P–V (a) using Gaussian copula via the MOM estimator (b) using Gaussian copula via the MPL estimator

<i>P</i>	<i>V</i>	T_P	T_V	OR return period	AND return period	$T_{(P/V \leq v)}$	$T_{(P/V \geq v)}$	$T_{(V/P \leq p)}$	$T_{(V/P \geq p)}$
<i>(a)</i>									
8089.2353	13,185.141	4.37177	1.789	1.765498249	4.521252343	58.33509	2.5267	2.28404	1.0341913
13,002.924	34,279.336	12.1050	7.298	6.272564178	16.61251954	38.5015	2.27608	11.9429	1.3723
7086.1589	44,015.568	3.488	14.092	3.436903244	15.00366134	4.22291	1.06465	165.52573	4.3007995
15,011.3722	43,029.274	17.627	13.171	10.23909076	28.58361499	42.4963	2.170068	23.043161	1.6215485
37,562.7721	43,439.063	427.350	13.546	13.52882137	445.6862253	9620.75019	32.900055	13.938419	1.0429058
10,195.7231	47,777.959	6.888	18.284	6.449169084	22.32079831	9.4183711	1.220724	86.444828	3.24031
12,131.9342	30,708.24	10.213	5.752	5.097759141	13.22832215	37.02280787	2.299743	9.1811456	1.295185
17,745.1725	43,338.874	28.457	13.453	11.80852775	40.34627106	89.3973	2.998938	19.474	1.417768
6316.0124	39,637.633	2.923	10.460	2.880260163	11.05593695	3.594725	1.05694	127.73792	3.7815
<i>(b)</i>									
7392.7301	26,155.411	3.739	4.248	3.06361	5.670598	8.39867	1.334688727	12.41253947	1.516261286
11,354.2573	40,038.679	8.742	10.748	6.866647	16.184238	17.2453769	1.505781479	28.33849245	1.851153112
9240.6956	25,793.39	5.624	4.147	3.564997	7.225181	19.2635315	1.742063419	8.005439649	1.284637202
18,014.5378	55,731.996	29.779	32.299	20.673153	61.850248	55.6538407	1.914883668	65.33433532	2.076931317
17,019.5015	53,420.634	25.144	27.292	17.614561	50.919642	47.8540464	1.865695677	56.48008763	2.025074156
12,479.5852	71,122.448	10.936	109.051	10.866282	116.520212	11.9581986	1.068490348	1545.701074	10.65460822
28,439.1291	65,676.413	140.252	69.300	57.362111	242.313219	328.183117	3.496579749	96.3660487	1.727693251
14,672.097	73,626.882	16.5672	135.685	16.358702	151.504661	18.4642567	1.116589351	1221.036381	9.144821336
13,465.7503	58,472.781	13.225	39.603	12.373302	49.898128	17.541202	1.259927732	177.4544419	3.772797457

ume and flood peak–duration series. But no one could deny from the facts that the more comprehensive flood risk analysis can be achieved through accounting all the trivariate random vectors simultaneously instead of just the pair-wise joint analysis (i.e. Reddy and Ganguli 2013; Fan and Zheng 2016; Daneshkhan et al. 2016). Also, according to Graler et al. (2013) and Renard and Lang (2007), multiple relevant vectors of the specified hydrologic episodes could likely depend upon the

potential damage where the ignorance of spatial dependency among these uncertain flood characteristics may responsible for the underestimation of uncertainty. On the other hand, the multivariate FFA either with or without copulas has been applied frequently with parametric distributions where the parametric functions are often employed to modelled their marginals and the parametric copula function for the joint dependence structure. But, the parametric functions always imposed an assumption

Table 12 Joint and conditional return period for pairs P–D (a) using Gaussian copula via the MOM estimator (b) using Frank copula via the MPL estimator

P	D	T_P	T_D	OR return period	AND return period	$T_{(P/D \leq d)}$	$T_{(P/D \geq d)}$	$T_{(D/P \leq p)}$	$T_{(D/P \geq p)}$
<i>(a)</i>									
2545.9308	21.69847	1.256	3.316	1.15	4.38	1.23026661	1.32126917	2.78291225	3.49
3684.82	43.42169	1.593	32.679	1.56	62.30	1.58559794	1.90632719	25.6093696	39.09
8982.7458	35.364881	5.3194	13.655	3.98	102.83	5.19883884	7.5299273	12.7865912	19.33
3061.2817	32.251128	1.392	9.803	1.33	15.17	1.37719753	1.5468327	7.82083647	10.89
12,639.4182	14.262899	11.282	1.6523	1.54	21.60	9.32455987	13.0740416	1.6306209	1.91
3806.5193	34.754081	1.637	12.792	1.54	24.35	1.61853103	1.90367478	10.4936186	14.87
15,632.8656	17.176828	19.716	2.1480	2.01	53.11	16.7590427	24.7246732	2.12508829	2.69
4131.1398	50.474034	1.762	71.073	1.74	158.37	1.75718281	2.22824568	55.7789385	89.86
20358.4714	9.163743	43.610	1.1141	1.11	51.89	27.9895339	46.5776097	1.11245641	1.19
<i>(b)</i>									
5123.0281	34.730555	2.2161	12.759	1.998604	34.187126	2.18	2.68	11.172204	15.4262569
10,734.5316	11.871255	7.7041	1.3540	1.28158	11.361914	6.26	8.39	1.33776764	1.4747764
33,994.7685	12.432854	282.485	1.4164	1.413803	450.856831	222.40	318.30	1.41588469	1.59603318
4735.5027	40.545997	2.0253	23.877	1.929045	58.075908	2.01	2.43	20.5290458	28.6738181
4288.911	17.304232	1.8271	2.1736	1.287246	4.338436	1.70	2.00	1.9721007	2.37442586
15,795.1923	25.729495	20.296	4.972	4.117002	133.961793	19.11	26.94	4.9101567	6.60029753
8562.9629	72.837062	4.853	877.192	4.831214	5791.09214	4.85	6.60	820.803022	1193.0808
5177.77	31.060031	2.244	8.646	1.929462	23.315032	2.20	2.70	7.62075281	10.387313
3144.0293	36.324186	1.4171	15.133	1.370462	23.798763	1.41	1.57	12.2347316	16.7928834

that the random samples are coming from the known populations whose PDF are predefined, but, in actual, no universal rules and studies are imposed to model any hydrologic vectors through any fixed or predefined distribution functions, which would follow different distributions and desire to model separately (i.e. Adamowski 1989; Silverman 1986; Kim and Heo 2002; Karmakar and Simonovic 2008). Therefore, in the past few decades, an attempt via kernel density estimators (KDE) is recognized as a much flexible and stable nonparametric data smoothing procedure and thus implemented in conjunction with parametric copula function under the semiparametric settings (i.e. Karmakar and Simonovic 2009; Reddy and Ganguli 2012a). Also, if the targeted copulas and univariate marginal distributions belong to some specific parametric families, it might be problematic, if the underlying assumption is violated. Therefore, in such circumstances, the nonparametric copula distribution framework can ameliorate these modelling issues and can be able to produce a significant outcome without assuming a particular form for the univariate marginal or multivariate copula distributions. All the above raised issues can be considered as a future research purpose.

Compliance with ethical standards

Conflict of interest This manuscript exhibited no conflict of interest.

References

- Abdulkareem JH, Sulaiman WNA (2015) Trend analysis of precipitation data in flood source areas of Kelantan River basin, Malaysia. In: The 3rd international conference in water resources, ICWR-2015, pp 174–188
- Adamowaski K (1985) Nonparametric kernels estimation of flood frequencies. *Water Resour Res* 21(11):1585–1590
- Adamowski K (1989) A monte Carlo comparison of parametric and nonparametric estimations of flood frequencies. *J Hydrol* 108:295–308
- 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
- Akaike H (1974) A new look at the statistical model identification. *IEEE Trans Autom Control* AC 19(6):716–722
- Alina B (2018) Copula Modeling for world's biggest competitors. Master Thesis, Amsterdam School of Economics, University of Amsterdam
- Anderson TW, Darling DA (1954) A test of goodness of fit. *J Am Stat Assoc* 49(268):765–769
- Bain LJ, Engelhardt M (1991) *Statistical Analysis of Reliability and Life-Testing Models*. Dekker, New York

- Blazkova S, Beven K (2004) Flood frequency estimation by continuous simulation of subcatchments rainfalls and discharges with the aim of improving dam safety assessments in a large basin in the Czech Republic. *J Hydrol* 292:153–172
- Bobee B, Rasmussen PF (1994) Statistical analysis of annual flood series. In: Menon J (ed) *Trend in hydrology*, vol 1. Council of Scientific Research Integration, New Delhi, pp 117–135
- Calver A, Lamb R (1995) Flood frequency estimation using continuous rainfall-runoff modelling. *Phys Chem Earth* 20(479):483
- Chan NW (1997) Institutional arrangement of flood hazard management in Malaysia: an evaluation using criteria approach. *Disasters* 21:206–222
- Cong RG, Brady M (2011) The interdependence between rainfall and temperature: copula analyses. *Sci World J* 2011, 405675
- Conover WJ (1999) *Practical nonparametric statistics*. Wiley, New York
- Constantino M, Larran M, Brebbia CA (2008) *Computational finance and its applications III*, Volume 41 of WIT transactions on information and communication technologies, WIT Press, 2008
- Correia FN (1987) Multivariate partial duration series in flood risk analysis. In: Singh VP (ed) *Hydrologic frequency modelling*. Reidel, Dordrecht, pp 541–554
- Cunnane C (1978) Unbiased plotting positioning—a review. *J Hydrol* 37:205–222
- Cunnane C (1987) Review of statistical models for flood frequency estimation. In: Singh VP (ed) *Hydrologic frequency modeling*. Reidel, Dordrecht, pp 49–95
- Daneshkhan A, Remesan R, Omid C, Holman IP (2016) Probabilistic modelling of flood characteristics with parametric and minimum information pair-copula model. *J Hydrol* 540:469–487
- De Michele C, Salvadori G (2003) A generalized Pareto intensity-duration model of storm rainfall exploiting 2-copulas. *J Geophys Res* 108(D2):4067. <https://doi.org/10.1029/2002JD002534>
- De Michele C, Salvadori G, Canossi M, Petaccia A, Rosso R (2005) Bivariate statistical approach to check the adequacy of dam spillway. *J Hydrol Eng* 10(1):50–57
- DID (Drainage and Irrigation Department) (2000) Annual flood report of DID for Peninsular Malaysia. Unpublished report. DID, Kuala Lumpur
- DID (Drainage and Irrigation Department Malaysia) (2003) Annual flood report of DID for Peninsular Malaysia. Unpublished report. DID, Kuala Lumpur
- DID (Drainage and Irrigation Department Malaysia) (2004) Annual flood report of DID for Peninsular Malaysia. Unpublished report. DID, Kuala Lumpur
- Dooge JCE (1986) Looking for hydrologic laws. *Water Resour Res* 22(9):465–485
- Escalante C, Raynal JA (1998) Multivariate estimation of floods: the trivariate gumble distribution. *J Stat Comput Simul* 61:313–340
- Escalante C, Raynal JA (2008) Trivariate generalized extreme value distribution in flood frequency analysis. *Hydrol Sci* 53(3):550–567
- Fan L, Zheng Q (2016) Probabilistic modelling of flood events using the entropy copula. *Adv Water Resour* 97:233–240
- Farrel PJ, Stewart KR (2006) Comprehensive study of tests for normality and symmetry: extending the Spiegelhalter test. *J Stat Comput Simul* 2006(76):803–816. <https://doi.org/10.1080/10629360500109023>
- Favre A-C, Adlouni SE, Perreault L, Thiémonge N, Bobee B (2004) Multivariate hydrological frequency analysis using copulas. *Water Resour Res*. <https://doi.org/10.1029/2003WR002456>
- Fisher NI, Switzer P (2001) Graphical assessments of dependence: is a picture worth 100 tests? *Am Stat* 55(3):233–239
- Gaál L, Szolgay J, Kohnová S, Hlavčová K, Parajka J, Viglione A, Blöschl G (2015) Dependence between flood peaks and volumes: a case study on climate and hydrological controls. *Hydrol Sci J* 60(6):968–984
- 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 modelling but were afraid to ask. *J Hydrol Eng* 12:347–368
- Genest C, Rémillard B (2008) Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. *Ann l'Inst Henri Poincaré Prob Stat* 44:1096–1127
- Genest C, Rivest LP (1993) Statistical inference procedures for bivariate Archimedean copulas. *J Am Stat Assoc* 88(423):1034–1043
- Genest C, Ghoudi K, Rivest LP (1995) A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika* 82(3):543–552
- Genest C, Favre AC, Beliveau J, Jacques C (2007) Metaelliptical copulas and their use in frequency analysis of multivariate hydrological data. *Water Resour Res* 43:W09401. <https://doi.org/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* 44:199–214
- Goel NK, Seth SM, Chandra S (1998) Multivariate modelling of flood flows. *J Hydraul Eng* 124(2):146–155
- Graler B, Berg MJV, Vandenberg S, Petroselli A, Grimaldi S, Baets BD, Verhost NEC (2013) Multivariate return periods in hydrology: a critical and practical review focusing on synthetic design hydrograph estimation. *Hydrol Earth Syst Sci* 17:1281–1296
- Grimaldi S, Serinaldi F (2006) Asymmetric copula in multivariate flood frequency analysis. *Adv Water Resour* 29:1155–1167
- Gringorten II (1963) A plotting rule of extreme probability paper. *J Geophys Res* 68(3):813–814
- Hannan EJ, Quinn BG (1979) The determination of the order of an autoregression. *J R Stat Soc Ser B Stat Methodol* 41:190–195
- Hassan AAG (2004) *Growth, Structural change and regional inequality in Malaysia*. Ashgate Publishing Ltd, Malaysia
- Hosking JRM, Wallis JR (1987) Parameter and quantile estimations for the generalized Pareto distributions. *Technometrics* 29(3):339–349
- Hosking JRM, Wallis JR, Wood EF (1985) Estimation of the general extreme value distribution by the method of probability weighted moments. *Technometrics* 27(3):251–261
- Hussain STPR, Ismail H (2013) Flood frequency analysis of Kelantan River Basin, Malaysia. *World Appl Sci J* 28(12):1989–1995. <https://doi.org/10.5829/idosi.wasj.2013.28.12.1559>
- Jamaliah J (2007) Emerging Trends of Urbanization in Malaysia [online]. <http://www.statistics.gov.my/eng/images/stories/files/journalDOSM/V104ArticleJamaliah.pdf>. Accessed 9 Dec 2018
- Joe H (1997) *Multivariate models and dependence concept*. CRC Press, Boca Raton
- Karmakar S, Simonovic SP (2008) Bivariate flood frequency analysis. Part-1: determination of marginal by parametric and non-parametric techniques. *J Flood Risk Manag* 1:190–200
- Karmakar S, Simonovic SP (2009) Bivariate flood frequency analysis. Part-2: a copula-based approach with mixed marginal distributions. *J Flood Risk Manag* 2(1):1–13
- Katz RW, Parlange MB, Naveau P (2002) Statistics of extremes in hydrology. *Adv Water Resour* 25(8):1287–1304
- Kendall MG (1975) *Rank correlation methods*, 4th edn. Charles Griffin, London, p 1975
- Kim KD, Heo JH (2002) Comparative study of flood quantiles estimation by nonparametric models. *J Hydrol* 260:176–193
- Klein B, Schumann AH, Pahlow M (2011) Copulas-New risk assessment methodology for dam safety, flood risk assessment and management. Springer, pp. 149–185.

- Klein B, Pahlow M, Hundedcha Y, Schumann A (2010) Probability analysis of hydrological loads for the design of flood control system using copulas. *J Hydrol Eng ASCE* 15:360–369
- Kojadinovic I, Yan J (2010) Modeling multivariate distributions with continuous margins using the copula R package. *J Stat Softw* 34(9):1–20
- Kojadinovic I, Yan J, Holmes M (2011) Fast large-sample goodness-of-fit tests for copulas. *Stat Sin* 21:841–871
- Kong XM, Huang GH, Fan YR, Li YP (2015) Maximum entropy-Gumbel-Hougaard copula method for simulation of monthly streamflow in Xiangxi river, China. *Stoch Environ Res Risk A* 29:833–846
- Krstanovic PF, Singh VP (1987) A multivariate stochastic flood analysis using entropy. In: Singh VP (ed) *Hydrologic frequency modelling*. Reidel, Dordrecht, pp 515–539
- Lawrence D, Paquet E, Gailhard J, Fleig AK (2014) Stochastic semi-continuous simulations for extreme flood estimations in catchments with combined rainfall-snowmelt flood regimes. *Nat Hazard Earth Syst* 14:1283–1298
- Li F (2016) Modeling covariate-contingent correlation and tail-dependence with copulas. <https://arxiv.org/pdf/1401.0100.pdf>
- Ljung GM, Box GEP (1978) On a measure of lack of fit in time series models. *Biometrika* 65:297–303
- Madsen H, Rasmussen PF, Rosbjerg D (1997) Comparison of annual maximum series and partial duration series methods for modelling extreme hydrologic events. 1. At-site modelling. *Water Resour Res* 33(4):747–757
- Mann HB (1945) Nonparametric test against trend. *Econometrics* 13:245–259
- Manner H (2007) Estimation and model selection of copulas with an application to exchange rates. Department of Quantitative Economics, Maastricht University, Maastricht
- Manner H (2010) Modelling asymmetric and time-varying dependence. <https://cris.maastrichtuniversity.nl/portal/files/667227/guid-ae8195ad-cf0b%20-4744-8bb1-6a44fbc10fe7-ASSET1.0>. Accessed 15 May 2019
- Mathwave Technologies (2004–2017) <http://mathwave.com/en/home.html>
- McNeil AJ, Frey R, Embrechts P (2015) *Quantitative risk management*. Princeton University Press, Princeton. ISBN: 978-0-691-03407-16627-8
- MMD (2007) Malaysian Meteorological Department (MMD). Report on heavy rainfall that caused floods in Kelantan and Terengganu. Unpublished report. MMD, Kuala Lumpur
- Moriassi DN, Arnold JG, Van Liew MW, Bingner RL, Harmel RD, Veith TL (2007) Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Trans ASABE* 50(3):885–900
- Nadarajah S, Shiao J (2005) Analysis of extreme flood events for the Pachang River, Taiwan. *Water Resour Manag* 19:363–375
- Nashwan MS, Ismail T, Ahmed Kamal (2018) Flood susceptibility assessment in Kelantan river basin using copula. *Int J Eng Technol* 7(2):584–590
- Nelsen RB (2006) *An introduction to copulas*. Springer, New York
- Owen CEB (2008) Parameter estimation for the beta distribution. All Thesis and Dissertation, p 1614. <https://scholarsarchive.byu.edu/etd/1614>
- Poulin A, Huard D, Favre AC, Pugin S (2007) Importance of tail dependence in bivariate frequency analysis. *J Hydrol Eng* 12(4):394–403
- Rahman HAA, Ibrahim NA, Abdul Hamid AA, Abdul Hamid THT (2018) Effect of 2014 massive flood on well water qualities: a case study on Kelantan River basin, Malaysia. *J Water Land Dev* 38:127–136. <https://doi.org/10.2478/jwld-2018-0049>
- Rakhecha PR, Singh VP (2009) *Applied hydrometeorology*. Capital Publishing Company, New Delhi
- Rao AR, Hameed KH (2000) *Flood frequency analysis*. CRC Press, Boca Raton
- Razawi S, Vogel R (2018) Pre-whitening of hydroclimatic time series? Implications for inferred change and variability across time scales. *J Hydrol* 557(2018):109–115
- Reddy MJ, Ganguli P (2012a) Bivariate flood frequency analysis of upper Godavari River flows using archimedean copulas. *Water Resour Manag*. <https://doi.org/10.1007/s11269-012-0124-z>
- Reddy MJ, Ganguli P (2012b) Risk assessments of hydro-climatic variability on ground water levels in the Manjra basin aquifer in India using Archimedean copulas. *J Hydrol Eng*. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000564](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000564)
- Reddy MJ, Ganguli P (2013) Probabilistic assessments of flood risks using trivariate copulas. *Theor Appl Climatol* 111:341–360
- 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
- Requena AI, Flores I, Mediero L, Garrote L (2016) Extension of observed flood series by combining a distributed hydro-meteorological model and a copula-based model. *Stoch Environ Res Risk Assess* 30:1363–1378. <https://doi.org/10.1007/s00477-015-1138-x>
- Saklar A (1959) *Functions de repartition n dimensions et leurs marges*. Publ Inst Stat Univ Paris 8:229–231
- Salvadori G (2004) Bivariate return periods via-2 copulas. *J R Stat Soc Ser B* 1:129–144
- Salvadori G, De Michele C (2004) Frequency analysis via copulas: theoretical aspects and applications to hydrological events. *Water Resour Res* 40:W12511. <https://doi.org/10.1029/2004WR003133>
- Salvadori G, De Michele C (2006) Statistical characterization of temporal structure of storms. *Adv Water Resour* 29:827–842
- Salvadori G, De Michele C (2007) On the use of copulas in hydrology: theory and practices. *J Hydrol Eng* 12:369–380
- Salvadori G, De Michele C (2010) Multivariate multiparameters extreme value models and return periods: a Copula approach. *Water Resour Res*. <https://doi.org/10.1029/2009WR009040>
- Santhosh D, Srinivas VV (2013) Bivariate frequency analysis of flood using a diffusion kernel density estimators. *Water Resour Res* 49:8328–8343. <https://doi.org/10.1002/2011WR010077>
- Schmidt T (2007) Coping with Copulas. In: Rank J (ed) *Copulas: from theory to application to finance*, chap. 1. Risk Books, London, pp 3–34
- Scholz FW, Stephens MA (1987) K-sample Anderson–Darling tests. *J Am Stat Assoc* 82(399):918–924
- Schwarz GE (1978) Estimating the dimension of a model. *Ann Stat* 6(2):461e464
- Schweizer B, Wolff EF (1981) On parametric measures of dependence for random variables. *Ann Stat* 9:879–885
- Sen Z (1999) Simple risk calculations in dependent hydrological series. *Hydrol Sci J* 44(6):871–878
- Serinaldi F (2015) Dismissing return periods! *Stoch Environ Res Risk Assess* 29(4):1179–1189
- Shiao JT (2003) Return period of bivariate distributed hydrological events. *Stoch. Environ. Res. Risk Assess.* 17(1–2):42–57
- Shiao JT (2006) Fitting drought duration and severity with two dimensional copulas. *Water Resour Manag* 20(5):795–815
- Silverman BW (1986) *Density estimation for statistics and data analysis*, 1st edn. Chapman and Hall, London
- Singh J, Knapp HV, Demissie M (2004) Hydrologic modeling of the Iroquois River watershed using HSPF and SWAT. ISWS CR 2004-08. Illinois State Water Survey, Champaign, Ill. www.sws.uiuc.edu/pubdoc/CR/ISWSCR2004-08.pdf. Accessed 10 Apr 2019

- Song S, Singh VP (2010) Metaelliptical copulas for drought frequency analysis of periodic hydrologic data. *Environ Res Hazard Assess* 24(3):425–444
- Sraj M, Bezak N, Brilly M (2014) Bivariate flood frequency analysis using the copula function: a case study of the Litija station on the Sava River. *Hydrol Process*. <https://doi.org/10.1002/hyp.10145>
- Tang Y, Huynh VN, Lawry J (2015) Integrated uncertainty in knowledge modelling and decision making- 4th International Symposium, Proceedings, LNAI 9376, Springer-Verlag, 2015.
- Tosunoglu F, Kisi O (2016) Joint modelling of annual maximum drought severity and corresponding duration. *J Hydrol*. <https://doi.org/10.1016/j.jhydrol.2016.10.018>
- Veronika BM, Halmova D (2014) Joint modelling of flood peak discharges, volume and duration: a case study of the Danube River in Bratislava. *J Hydrol Hydromech* 62(3):186–196
- Wan I (1996) Urban growth determinants for the state of Kelantano of the state's policy makers. *Penerbitan Akademik Fakulti Kejuruteraan dan Sains Geoinformasi Buletin Ukur* 7:176–189
- Wang C, Ni-Bin Chang N-B, Yeh G-T (2009) Copula-based flood frequency (COFF) analysis at the confluences of river systems. *Hydrol Process* 23:1471–1486
- Wooldridge S, Kalma J, Kuczera G (2001) Parameterisation of a simple semi-distributed model for assessing the impact of land-use on hydrologic response. *J Hydrol* 254:16–32
- Xu Y, Huang G, Fan Y (2015) Multivariate flood risk analysis for Wei River. *Stoch Environ Res Risk Assess*. <https://doi.org/10.1007/s00477-015-1196-0>
- 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 Process* 14:2575–2588
- Yue S (2001) A bivariate gamma distribution for use in multivariate flood frequency analysis. *Hydrol Process* 15:1033–1045
- Yue S, Rasmussen P (2002) Bivariate frequency analysis: discussion of some useful concepts in hydrological applications. *Hydrol Process* 16:2881–2898
- Yue S, Wang CY (2004) A comparison of two bivariate extreme value distribution. *Stoch Environ Res Risk A* 18:61–66
- Yue S, Ouarda TMBJ, Bobee B, Legendre P, Bruneau P (1999) The Gumbel mixed model for flood frequency analysis. *J Hydrol* 226(1–2):88–100
- Zhang L (2005) Multivariate hydrological frequency analysis and risk mapping. Doctoral dissertation, Beijing Normal University
- Zhang L, Singh VP (2006) Bivariate flood frequency analysis using copula method. *J Hydrol Eng* 11(2):150
- Zhang L, Singh VP (2007) Trivariate flood frequency analysis using the Gumbel-Hougaard copula. *J Hydrol Eng* 12(4):431–439
- Zhang R, Xi Chen, Cheng Q, Zhang Z, Shi P (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