Singh and Zhang *Geosci. Lett.* (2018) 5:6
https://doi.org/10.1186/s40562-018-0105-z

Geoscience Letters

REVIEW

Open Access

Copula–entropy theory for multivariate stochastic modeling in water engineering

Vijay P. Singh[1,2*†] and Lan Zhang[1†]

Abstract

The copula–entropy theory combines the entropy theory and the copula theory. The entropy theory has been extensively applied to derive the most probable univariate distribution subject to specified constraints by applying the principle of maximum entropy. With the flexibility to model nonlinear dependence structure, parametric copulas (e.g., Archimedean, extreme value, meta-elliptical, etc.) have been applied to multivariate modeling in water engineering. This study evaluates the copula–entropy theory using a sample dataset with known population information and a flood dataset from the experimental watershed at the Walnut Gulch, Arizona. The study finds the following: (1) both univariate and joint distributions can be derived using the entropy theory. (2) The parametric copula fits the true copula better using empirical marginals than using fitted parametric/entropy-based marginals. This suggests that marginals and copula may be identified separately in which the copula is investigated with empirical marginals. (3) For a given set of constraints, the most entropic canonical copula (MECC) is unique and independent of the marginals. This allows the universal solution for the proposed analysis. (4) The MECC successfully models the joint distribution of bivariate random variables. (5) Using the "AND" case return period analysis as an example, the derived MECC captures the change of return period resulting from different marginals.

Keywords: Copula theory, Entropy theory, Multivariate stochastic modeling, Probability density function, Most entropic canonical copula, Return period

Introduction

A multitude of processes in water engineering involve more than one random variable. For example, floods are characterized by peak, duration, volume, and inter-arrival time, which are all random in nature. Droughts are described by their severity, duration, inter-arrival time, and areal extent, which are also random. Extreme precipitation events are represented by their intensity, amount, duration, and inter-arrival time, which are all random. Inter-basin water transfer involves transfer of excess water from one basin (say, donor) to a water deficient basin (say, recipient). The transfer involves the volume of water, availability of water in both donor and recipient basins, duration of transfer, rate of transfer, and time interval between water transfers which are all random variables. Water quality entails pollutant load, duration for which the load is higher than the protection limits, and peak pollutant concentration, which are all random variables. Likewise, erosion in a basin may be characterized by sediment yield, number of erosion events, duration of events, intensity of events, and time interval between two consecutive events. These are all random variables. Flooding in a coastal watershed may be caused by the simultaneous occurrence of high precipitation and high tides where both precipitation and tide are random variables. Examples of processes involving more than one random variable abound in hydrologic, hydraulic, environmental, and water resources engineering. There usually exists some degree of dependence among the random variables or at least among some of the variables. Often we are concerned with multivariate stochastic modeling and risk analysis of the systems and processes that involve the derivation of probability distributions of the random

*Correspondence: vsingh@tamu.edu
†Vijay P. Singh and Lan Zhang contributed equally to this work
[1] Department of Biological & Agricultural Engineering, Texas A&M University, College Station, TX 77843-2117, USA
Full list of author information is available at the end of the article

variables considering the dependence structure among them. Nowadays, these stochastic processes can be modeled with the copula–entropy theory that has proven to be more flexible and accurate than the traditional approaches. The objective of this paper therefore is to reflect on some recent advances made in the application of the copula–entropy theory and future challenges.

Methods

Copula–entropy theory

The copula–entropy theory (CET) is an amalgam of the copula theory and the entropy theory. These two theories are now discussed.

Entropy theory

The entropy theory comprises (1) formulation of entropy, (2) principle of maximum entropy (POME), and (3) theorem of concentration (TOC). Entropy can be defined in the real domain or frequency domain. In the real domain, the most famous form of entropy is the Shannon entropy (Shannon 1948), although Tsallis entropy (Tsallis 1988) and Renyi entropy (Renyi 1951) have been receiving much attention in recent years. Another popular formulation of entropy is the cross-entropy or relative entropy due to Kullback and Leibler (1951) which is a generalization of the Shannon entropy. For a continuous random variable X with a probability density function (PDF), $f(x)$, and cumulative probability distribution function (CDF), $F(x)$, the Shannon entropy, $H(X)$ or $H[f(x)]$, can be defined as

$$H(X) = H[f(x)] = -\int_0^\infty f(x)\ln f(x)\,\mathrm{d}x, \tag{1}$$

where $X \in [0,\infty]$ but can also vary from $-\infty$ to $+\infty$ or from a finite lower limit to a finite upper limit. The Shannon entropy can also be defined in an analogous manner for a discrete variable.

The principle of maximum entropy (POME), propounded by Jaynes (1957), states that of all the distributions that satisfy the given constraints, the distribution yielding the maximum entropy is the least-biased distribution and should hence be preferred. If there are no constraints then POME says that the resulting distribution would be a uniform distribution, which is consistent with the Laplacian principle of insufficient reason.

The theorem of concentration states that POME yields the best constrained probability distribution and is the preferred method for inferring this distribution, and this distribution best represents our state of knowledge about the behavior of the system. This is a consequence of Shannon's inequality and the relation between entropy and Chi square statistic.

Copula theory

The foundation of the copula theory is the Sklar theorem (Sklar 1959). The theorem states that the joint (multivariate) probability distribution of two or more random variables is a function of the probability distributions of individual variables (also referred to as marginal distributions which are one-dimensional). In other words, the multivariate distribution is coupled to its marginal distributions. It is implied that these random variables are not independent of each other. The copula theory does not specify the way to derive the marginal distributions and does not lead to a unique copula. There are different ways to construct copulas and different ways to select the best copula.

Methodology for application of copula–entropy theory

The copula–entropy theory can be applied in different ways: (1) the marginal distributions are derived using the entropy theory and the joint distribution using the copula theory (e.g., Hao and Singh 2012; Zhang and Singh 2012). Since there can be more than one joint distribution fitted to the multivariate random variables, the best distribution is then selected from either visual goodness-of-fit plot (e.g. Q–Q plot) or formal goodness-of-fit test statistics (Genest et al. 2009). (2) With the marginal distributions derived using the entropy theory, the best copula is selected as the copula function yielding the maximum entropy. (3) Both marginal and joint distributions are derived using the entropy theory (e.g., Chu 2011; Chen et al. 2013; Aghakouchak 2014). The methodology for application of the copula–entropy theory will depend on the way it is applied. Each of the three ways is now outlined. First, the methodology for application of the entropy theory is outlined, since entropy is needed in all three ways.

Methodology for application of entropy theory

Fundamental to applying the entropy theory is the specification of constraints the derived probability distribution must satisfy. There can be any number of constraints which can be defined in different ways but the easiest way is to define them in terms of moments. Let $g(x)$ be any function of random variable X. Then, the ith constraint, C_i, can be expressed as

$$C_i = \int_0^\infty g_i(x)f(x)\,\mathrm{d}x = E[g_i(x)], \quad i = 1, 2, \ldots, m, \tag{2}$$

where E is the expectation operator. If $g_0(x) = 1$, then Eq. (1) will lead to the total probability as

$$C_0 = \int_0^\infty f(x)\,\mathrm{d}x = 1. \tag{3}$$

The next step is to maximize entropy given by Eq. (1), subject to Eqs. (2) and (3). Entropy maximizing can be done using the method of Lagrange multipliers where the Lagrange function L can be written as

$$L = -\int_0^\infty f(x)\ln f(x)\,dx - (\lambda_0 - 1)\left[\int_0^\infty f(x)\,dx - 1\right] - \sum_{i=1}^{m}\lambda_i\left[\int_0^\infty g_i(x)f(x)\,dx - C_i\right], \tag{4}$$

where $\lambda_i, i = 0, 1, \ldots, m$, are the unknown Lagrange multipliers. Applying the Lagrange–Euler calculus of variation, Eq. (4) leads to the maximum entropy distribution:

$$f(x) = \exp\left[-\sum_{i=0}^{m}\lambda_i g_i(x)\right]. \tag{5}$$

Now the unknown Lagrange multipliers are determined from the known constraints. The multipliers can be determined in two ways: regular entropy method and parameter space expansion method (Singh 1998; Singh and Rajagopal 1986). Substituting Eq. (5) in Eq. (3), we get

$$\exp(\lambda_0) = Z = \int_0^\infty \exp\left[-\sum_{i=1}^{m}\lambda_i g_i(x)\right]dx$$
$$\text{or}$$
$$\lambda_0 = \ln Z = \ln\left\{\int_0^\infty \exp\left[-\sum_{i=1}^{m}\lambda_i g_i(x)\right]dx\right\}, \tag{6}$$

where Z is called the partition function, and

$$ZC_i = \int_0^\infty g_i(x)\exp\left[-\sum_{i=1}^{m}\lambda_i g_i(x)\right]dx, \quad i = 1, 2, \ldots, m. \tag{7}$$

Equation (6) shows that λ_0 is a function of $\lambda_1, \lambda_2, \ldots, \lambda_m$, i.e., $\lambda_0 = \lambda_0(\lambda_1, \lambda_2, \ldots, \lambda_m)$, and the function is convex. Differentiating λ_0 with respect to $\lambda_1, \lambda_2, \ldots, \lambda_m$ individually, we get the relations between Lagrange multipliers. Substituting Eq. (6) into Eq. (7), we obtain

$$C_i = \frac{\int_0^\infty g_i(x)\exp\left[-\sum_{i=1}^{m}\lambda_i g_i(x)\right]dx}{\int_0^\infty \exp\left[-\sum_{i=1}^{m}\lambda_i g_i(x)\right]dx}, \quad i = 1, 2, \ldots, m. \tag{8}$$

Equation (8) shows that C_i, $i = 1, 2, \ldots, m$, are functions of $\lambda_1, \lambda_2, \ldots, \lambda_m$.

Differentiating Eq. (6) and using Eqs. (2) and (5), the result is as follows:

$$\frac{\partial\lambda_0}{\partial\lambda_i} = -E[g_i(x)] = -C_i, \quad i = 1, 2, \ldots, m. \tag{9}$$

For obtaining parameters, the derivatives in Eq. (9) are equated to the derivatives obtained from $\lambda_0 = \lambda_0(\lambda_1, \lambda_2, \ldots, \lambda_m)$. Similarly, it can be shown that

$$\frac{\partial^2\lambda_0}{\partial\lambda_i^2} = E\left[g_i^2(x)\right] - \left\{E[g_i(x)]\right\}^2 = \mathrm{Var}[g_i(x)], \quad i = 1, 2, \ldots, m \tag{10}$$

and

$$\frac{\partial^2\lambda_0}{\partial\lambda_i\partial\lambda_j} = E\left[g_i(x)g_j(x)\right] - E[g_i(x)]E\left[g_j(x)\right] = \mathrm{Cov}\left[g_i(x)g_j(x)\right], \quad i, j = 1, 2, \ldots, m,\ i \neq j. \tag{11}$$

The maximum entropy, $H_{\max}$, of the derived POME-based PDF can be expressed as

$$\begin{aligned} H_{\max} &= -\int_0^\infty f(x)\ln f(x)\,dx \\ &= -\int_0^\infty \left[-\lambda_0 - \sum_{i=1}^{m}\lambda_i g_i(x)\right] \\ &\quad \times \exp\left[-\lambda_0 - \sum_{i=1}^{m}\lambda_i g_i(x)\right]dx \\ &= \lambda_0 + \sum_{i=1}^{m}\lambda_i E[g_i(x)] = \sum_{i=0}^{m}\lambda_i C_i. \end{aligned} \tag{12}$$

Equation (12) shows that maximum entropy is a function of Lagrange multipliers and constraints, such that $H_{\max}$ is a concave function. Equation (12) also shows that Lagrange multipliers, $\lambda_1, \lambda_2, \ldots, \lambda_m$, are partial derivatives of $H_{\max}$ with respect to constraints C_i, $i = 1, 2, \ldots, m$, respectively.

If q_i and p_i, $i = 1, 2, \ldots, n$, are the frequencies computed from POME-based and given fitted parametric distributions, respectively, for n class intervals, then we have

$$2N\Delta H = \sum_{i=1}^{n}\frac{(q_i - p_i)^2}{p_i} = \chi^2, \tag{13}$$

where χ^2 is Chi square distributed with s degrees of freedom as

$$s = n - m - 1. \tag{14}$$

With the Chi square distribution as the limiting distribution, it is shown that $2N\Delta H$ is Chi square distributed. Hence, the Chi square statistic may be applied to assess if the fitted parametric distribution is close to the POME-based distribution (i.e., the reference distribution of random variable).

Methodology for application of copula theory

Definition and main properties for copula

As stated by Sklar (1959), copula couples the multivariate distribution to its marginal distributions which are uniformly distributed on [0,1]. In other words, copula is a mapping function as $[0,1]^d \rightarrow [0,1]$. For d-dimensional continuous random variables, there is a unique copula function (C) to represent the joint distribution function (H) as

$$\begin{aligned} &H(x_1, x_2, \ldots, x_d) = C(u_1, u_2, \ldots, u_d); \\ &u_i = F_i(x_i) \sim \text{uniform } (0,1),\ i = 1, \ldots, d. \end{aligned} \tag{15}$$

As shown in Eq. (15), u_i is the CDF of random variable X_i. Representing the joint distribution, the copula function has the following properties:

1. $0 \leq C(u_1, \ldots, u_d) \leq 1$;
2. if any $u_i = 0$, then $C(u_1, \ldots, u_d) = 0$;
3. if all $u_j = 1$, $j = 1, \ldots, d$ and $j \neq i$; then $C(1, \ldots, u_i, \ldots, 1) = u_i$;
4. C is bounded by the Fréchet–Hoeffding bounds as

$$\begin{aligned} &W \leq C \leq M; \quad W = \max\left(1 - d + \sum_{i=1}^{d} u_i, 0\right), \\ &M = \min(u_1, \ldots, u_d) \end{aligned} \tag{16}$$

   In Eq. (16), W represents the perfectly negative dependence, while M represents the perfect positive dependence. For independent random variables, the corresponding copula function is simply given as $\Pi = u_1 u_2 \cdots u_d = F_1(x_1)F_2(x_2)\cdots F_d(x_d)$; and
5. C is d-increasing, that is, the $C(u_1, \ldots, u_d)$ volume for any given d-dimensional interval is non-negative.

Copula families and parameter estimation

The major copula families are Archimedean copulas, meta-elliptical copulas, extreme value copulas, vine copulas, and entropic copulas. The Archimedean copula (2-dimensional) is symmetric and easy to construct through the generating function as

$$C(u, v) = \phi^{-1}(\phi(u) + \phi(v)), \tag{17}$$

where ϕ is the generating function which is non-increasing. Based on the choice of Archimedean copulas, different copulas within the family may cover different ranges of dependence (Nelsen 2006). For example, the Gumbel–Hougaard copula may only model the positive dependence, while Frank copula may model the entire range of dependence structure. Given its easy construction, the Archimedean copulas have been extensively applied in bivariate hydrological frequency analysis (e.g., Sraj et al. 2015; Salvadori and Michele 2015; Requena et al. 2016a, b).

Meta-elliptical copulas (Fang et al. 2002), as the name suggests, is derived from the elliptical joint distribution. The popularly applied meta-elliptical copulas are meta-Gaussian and meta-Student t copulas. Unlike the Archimedean copulas, the meta-elliptical copulas can model the entire range of dependence structure and can be easily applied to high-dimensional multivariate modeling. Comparing the two popularly applied meta-elliptical copulas, there exists the symmetric tail dependence for meta-Student t copula, while no tail dependence exists for meta-Gaussian copula (e.g. Genest et al. 2007; Song and Singh 2010).

The extreme value copula is derived in accordance with the extreme value theory which may be applied to model the rare events. As stated by Gudendorf and Segers (2009) and Joe (2014), the following relation exists:

$$C_F(u_1^{1/n}, \ldots, u_d^{1/n}) \rightarrow C(u_1, \ldots, u_d); \quad \exists\, n \rightarrow \infty. \tag{18}$$

In Eq. (18), C denotes the extreme value copula, and C_F denotes that the copula fulfills the limiting relation.

In other words, the extreme value copula must be max-stable. For the bivariate case, the extreme value copula may be written as

$$C(u, v) = uv \exp\left[A\left(\frac{\log(v)}{\log(uv)}\right)\right], \quad u, v \in [0,1]. \tag{19}$$

In Eq. (19), A denotes the Pickands dependence function (Pickands 1981; Falk and Reiss 2005) that is convex as $A : [0,1] \rightarrow [1/2, 1]$ and $\max(t, 1-t) \leq A(t) \leq 1$ for $t \in [0,1]$.

The Gumbel–Hougaard copula (Archimedean copula family) is the only Archimedean copula that belongs to the extreme value family. Hence, the Gumbel–Hougaard copula has been popularly applied in bivariate flood frequency analysis, storm analysis, drought analysis, etc.

Vine copula is constructed, based on the probability density decomposition. The vine copula is applied for high-dimensional analysis (i.e., $d \geq 3$). It is usually categorized into Canonical (C)-Vine copula, D-Vine copula, and Regular R-Vine copula (Aas et al. 2007). Using 3-dimensional analysis as an example, we can write the joint probability density function as

$$\begin{aligned} f(x_1, x_2, x_3) = &\prod_{i=1}^{3} f_i(x_i) c_{12}(F_1(x_1), F_2(x_2)) c_{23}(F_2(x_2), F_3(x_3)) c_{13|2} \\ &\times \left(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)\right). \end{aligned} \tag{20}$$

In Eq. (20), c denotes the copula density function. As seen in Eq. (20), the vine copula is very flexible, since the

bivariate copula is applied at all the levels. The vine copula has also been applied in high-dimensional hydrological frequency analysis (e.g., Pham et al. 2016; Arya and Zhang 2017; Verneiuwe et al. 2015)

The parameters of the parametric copula functions constructed above may be estimated with one of the following three approaches:

(i) Full-Maximum Likelihood Estimation (Full-MLE): In this method, the parameters of the marginal distributions and copula functions are estimated simultaneously.
(ii) Two-Stage Maximum Likelihood Estimation (Two-Stage MLE): In this method, one first estimates the parameters of marginal distributions and then the parameters of the copula function are estimated using MLE with the marginals computed from the previously fitted marginal distributions.
(iii) Semi-Parametric (or Pseudo) Maximum Likelihood Estimation (Pseudo-MLE): In this method, the parameters of the copula function are estimated from the empirical marginals (i.e., empirical CDF computed from the plotting position formula or kernel density function).

Of the three estimation methods for parametric copula functions, the Pseudo-MLE is considered least impacted by the possible misidentification of marginal distributions. The advantage of Pseudo-MLE is the separate parameter estimation of marginal distributions and the copula function.

The most entropic canonical copula may be derived using the entropy theory, similar to the application of entropy theory to the univariate random variables. The Shannon entropy of the copula function for two variables is written as

$$H(u,v) = -\int_{[0,1]^2} c(u,v)\ln\, c(u,v)\,\mathrm{d}u\mathrm{d}v \tag{21}$$

and the joint density function is given through the copula function as

$$f(x,y) = f_X(x)f_Y(y)c(u,v). \tag{22}$$

Substituting Eq. (22) into Eq. (21), one may conclude, with some simple algebra, that the negative copula entropy [i.e., Eq. (21)] denotes the mutual information of random variables X and Y through the Kullback–Leibler cross-entropy as

$$\begin{aligned} H_C(u,v) &= -\int_{[0,1]^2} c(u,v)\ln[c(u,v)]\,\mathrm{d}u\mathrm{d}v \\ &= -\int\int \frac{f(x,y)}{f_X(x)f_Y(y)}\ln\left(\frac{f(x,y)}{f_X(x)f_Y(y)}\right)f_X(x)f_Y(y)\,\mathrm{d}x\mathrm{d}y \\ &= -\int\int f(x,y)\ln\left(\frac{f(x,y)}{f_X(x)f_Y(y)}\right)\mathrm{d}x\mathrm{d}y \\ &= -KLCE(f_X;f_Y) = -I(X;Y). \end{aligned} \tag{23}$$

According to the information theory, the mutual information [i.e., $I(X;Y)$] is a measure of the total correlation between random variables, that is, the mutual dependence between random variables X and Y. From the copula theory [e.g., Eq. (22) for bivariate random variables], the copula density [i.e., $c(u,v)$] also denotes the mutual dependence between variables X and Y. Thus, the information maintained in the copula function is the mutual information (i.e., total correlation) between X and Y which results in the copula entropy being negative. In other words, a higher absolute value of the copula entropy represents higher mutual dependence (or total correlation) among the random variables.

Similar to the POME-based univariate distribution, the common constraints are the constraints of total probability of marginals (i.e., for uniform distributed variable on [0,1]), and a measure of dependence (also called association):

$$\int_{[0,1]^2} c(u,v)\,\mathrm{d}u\mathrm{d}v = 1 \quad \text{(total probability)} \tag{24}$$

$$\int_{[0,1]^2} u^r c(u,v)\,\mathrm{d}u\mathrm{d}v = E(u^r) = \frac{1}{r+1}, \quad r = 1,2,\ldots \text{ (constraints of } u = F_X(x)). \tag{25a}$$

Applying $f(x) = \int\int f(x,y)\,\mathrm{d}y$, we can evaluate Eq. (25a) as

$$\begin{aligned} \int_{[0,1]^2} u^r c(u,v)\,\mathrm{d}u\mathrm{d}v &= \int_0^1 u^r\,\mathrm{d}u\int_0^1 c(u,v)\,\mathrm{d}v \\ &= \int_0^1 u^r f(u)\,\mathrm{d}u = \int_0^1 u^r\,\mathrm{d}u = E(u^r) = \frac{1}{r+1}. \end{aligned} \tag{25b}$$

In Eq. (25b), $f(u) = 1$ since $u \sim$ uniform (0,1). Similarly,

$$\int_{[0,1]^2} v^r c(u,v)\,\mathrm{d}u\mathrm{d}v = E(v^r) = \frac{1}{r+1}, \quad r = 1,2,\ldots \text{ (constraints of } v = F_Y(y)). \tag{26}$$

$$\int_{[0,1]^2} a_j(u,v)c(u,v)\,\mathrm{d}u\mathrm{d}v = E[a_j(u,v)] = \Theta_j, \quad j = 1,2,\ldots \text{ (constraints of dependence measure).} \tag{27}$$

In Eq. (27), Spearman's rho is commonly applied as the constraint to measure the dependence with $a_j(u,v) = uv \Rightarrow E(uv) = \frac{\rho_s+3}{12}$. One can also apply other dependence measures discussed in Nelsen (2006) and Chu (2011).

Using the constraints [Eqs. (24)–(27)], the Lagrange function for the most entropic canonical copula (MECC) can be written as

$$L = -\int_{[0,1]^2} c(u,v)\ln[c(u,v)]\,\mathrm{d}u\mathrm{d}v - (\lambda_0 - 1)\left[\int_{[0,1]^2} c(u,v)\,\mathrm{d}u\mathrm{d}v - 1\right] - \sum_{i=1}^{n}\lambda_i\left[\int_{[0,1]^2} u^i c(u,v)\,\mathrm{d}u\mathrm{d}v - \frac{1}{i+1}\right] - \sum_{i=1}^{n}\gamma_i\left[\int_{[0,1]^2} v^i c(u,v)\,\mathrm{d}u\mathrm{d}v - \frac{1}{i+1}\right] - \sum_{j=1}^{k}\lambda_{n+j}\left[\int_{[0,1]^2} a_j(u,v)c(u,v)\,\mathrm{d}u\mathrm{d}v - \Theta_j\right]. \quad (28)$$

In Eq. (28), $\lambda_0,\ldots,\lambda_n,\gamma_1,\ldots,\gamma_n,\lambda_{n+1},\ldots,\lambda_{n+k}$ are the Lagrange multipliers. More specifically for MECC, $\lambda_r = \gamma_r$, $r = 1,\ldots,n$. The Lagrange multipliers $\lambda_{n+1},\ldots,\lambda_{n+k}$ are pertaining to the constraints in relation to the rank-based dependence measures.

Differentiating Eq. (28) with respect to $c(u,v)$, we have

$$c(u,v) = \frac{\exp\left(-\sum_{i=1}^{n}\lambda_i u^i - \sum_{i=1}^{n}\gamma_i v^i - \sum_{j=1}^{k}\lambda_{n+j}a_j(u,v)\right)}{\int_{[0,1]^2}\exp\left(-\sum_{i=1}^{n}\lambda_i u^i - \sum_{i=1}^{n}\gamma_i v^i - \sum_{j=1}^{k}\lambda_{n+j}a_j(u,v)\right)\mathrm{d}u\mathrm{d}v}. \quad (29)$$

Based on the principle of maximum entropy, maximizing Eq. (21) is equivalent to minimizing the objective function

$$Z(\Lambda) = \ln\left[\int_{[0,1]^2}\exp\left(-\sum_{i=1}^{n}\lambda_i u^i - \sum_{i=1}^{n}\gamma_i v^i - \sum_{j=1}^{k}\lambda_{n+j}a_j(u,v)\right)\mathrm{d}u\mathrm{d}v\right] + \sum_{i=1}^{n}\lambda_i\frac{1}{i+1} + \sum_{i=1}^{n}\gamma_i\frac{1}{i+1} + \sum_{j=1}^{k}\lambda_{n+j}\hat{\Theta}_j. \quad (30)$$

In Eq. (30), $\Lambda = [\lambda_1,\ldots,\lambda_n,\gamma_1,\ldots,\gamma_n,\lambda_{n+1},\ldots,\lambda_{n+k}]$.

The most entropic canonical copula (MECC) may be generalized to most entropic copula (MEC) with respect to a given parametric copula (Chu 2011). In the case of MEC, Eqs. (29)–(30) can be re-written as

$$c(u,v) = \frac{\exp\left(-\sum_{i=1}^{n}\lambda_i u^i - \sum_{i=1}^{n}\gamma_i v^i - \sum_{j=1}^{k}\lambda_{n+j}\,a_j(u,v) - b\tilde{c}(u,v)\right)}{\int_{[0,1]^2}\exp\left(-\sum_{i=1}^{n}\lambda_i u^i - \sum_{i=1}^{n}\gamma_i v^i - \sum_{j=1}^{k}\lambda_{n+j}\,a_j(u,v) - b\tilde{c}(u,v)\right)\mathrm{d}u\mathrm{d}v}, \quad (31a)$$

$$Z(\Lambda) = \ln\left[\int_{[0,1]^2}\exp\left(-\sum_{i=1}^{n}\lambda_i u^i - \sum_{i=1}^{n}\gamma_i v^i - \sum_{j=1}^{k}\lambda_{n+j}\,a_j(u,v) - b\tilde{c}(u,v)\right)\mathrm{d}u\mathrm{d}v\right] + \sum_{i=1}^{n}\lambda_i\frac{1}{i+1} + \sum_{i=1}^{n}\gamma_i\frac{1}{i+1} + \sum_{j=1}^{k}\lambda_{n+j}\,\hat{\Theta}_j. \quad (31b)$$

In Eq. (31), b is a generic constant, $\tilde{c}(u,v)$ is the given reference copula. It is seen that the MECC is obtained by setting $b = 0$. In what follows, we will focus on the application of MECC for bivariate cases through examples.

Copula–entropy for multivariate modeling

Following the discussion of Shannon entropy and copula theory in the previous sections, we will outline the copula–entropy theory for stochastic modeling in this section. In general, we can apply the copula–entropy theory in three ways:

(i) The marginal distributions are derived using the entropy theory, while the joint distribution (i.e., copula function) is modeled through the parametric copula function with its parameter estimated using the Full-MLE, Two-Stage MLE, or Pseudo-MLE. In this approach, the goodness-of-fit of the copula function may be assessed either graphically through the K–K plot or statistically with the formal goodness-of-fit test statistics (Genest et al. 2009).

(ii) The difference of this second approach from (i) above is that the parametric copula function is selected such that it yields the maximum entropy among all copula candidates.
(iii) The approach (iii) takes full advantage of the entropy theory. Both marginal and joint distributions are derived using the entropy theory. The Lagrange multipliers are estimated by maximizing entropy or minimizing the corresponding objective function which is the dual problem of maximizing entropy. The Lagrange multipliers of the MECC (joint distribution) may be optimized from the fitted POME-based marginal distributions or from the empirical marginal distribution. The approach (iii) is further adopted for the applications.

Application to multivariate data of known population

Here, we will first show the application of copula–entropy theory to the bivariate sample dataset with the known true population. In this sample study, the sample dataset ($N = 1000$) is generated from the known Gumbel–Hougaard copula ($\theta = 4.5$) with the true marginal distributions:

$$X \sim \text{Gamma}\,(10.5, 4.3)\text{: } \frac{1}{10.5\Gamma(4.3)}\left(\frac{x}{10.5}\right)^{3.3} e^{-(x/10.5)}$$

$$Y \sim \text{Lognormal}\,(4, 0.7^2)\text{: } \frac{1}{y(0.7)\sqrt{2\pi}}\exp\left(-\frac{(\ln y - 4)^2}{2(0.7^2)}\right).$$

Study of univariate variates

In Singh (1998), it was shown that $E[X]$, and $E[\ln(X)]$ should be applied as constraints to derive the POME-based gamma distribution; while $E[\ln(x)]$ and $E\left[(\ln x)^2\right]$ are the constraints to derive the POME-based lognormal distribution. Following Singh (1998), we have the following:

Gamma distribution

The POME-based gamma distribution may be written as

$$f(x) = \exp(-\lambda_0 - \lambda_1 x - \lambda_2 \ln x) \tag{32a}$$

$$\frac{\partial \lambda_0}{\partial \lambda_1} = \frac{\lambda_2 - 1}{\lambda_1} = -E(X) \approx -\bar{x} \tag{32b}$$

$$\frac{\partial \lambda_0}{\partial \lambda_2} = \ln \lambda_1 - \Gamma(1-\lambda_2)\psi(1-\lambda_2) = -E[\ln x]; \quad \psi(\mathrm{t}) = \frac{\mathrm{d}\ln[\Gamma(t)]}{\mathrm{d}t}. \tag{32c}$$

The relation of Lagrange multipliers to the parameters of gamma distribution (Singh 1998) is given as

$$\lambda_1 = \frac{1}{a};\ \lambda_2 = 1 - b;\ f(x; a, b) = \frac{1}{a\Gamma(b)}\left(\frac{x}{a}\right)^{b-1}\exp\left(-\frac{x}{a}\right). \tag{32d}$$

Lognormal distribution

The POME-based lognormal distribution may be written as

$$f(x) = \exp\left(-\lambda_0 - \lambda_1 \ln x - \lambda_2 (\ln x)^2\right) \tag{33a}$$

$$\frac{\partial \lambda_0}{\partial \lambda_1} = \frac{\lambda_1 - 1}{2\lambda_2} = -E[\ln X] \tag{33b}$$

$$\frac{\partial^2 \lambda_0}{\partial \lambda_1^2} = \frac{(\lambda_1 - 1)^2}{4\lambda_2^2} - \frac{1}{2\lambda_2} = E\left[(\ln x)^2\right] \tag{33c}$$

$$\lambda_1 = 1 - \frac{\bar{y}}{s_y^2}; \quad \lambda_2 = \frac{1}{2s_y^2}. \tag{33d}$$

In Eq. (33d), $y = \ln(x)$ and s_y^2 represents the sample variance of y.

Using the bivariate data sampled from the true population, Table 1 lists the Lagrange parameters that are estimated for the univariate variables based on both sample moments and population moments.

Besides applying the constraints directly related to the parametric distribution that may be fitted to the observed dataset, one may also directly apply the first three or four monocentral moments [i.e., $E(X)$, $E(X^2)$, $E(X^3)$, $E(X^4)$], given that the moments about the origin govern the shape and mode of the univariate probability density functions (Zellner and Highfield 1988; Cobb et al. 1983). The POME-based distribution so derived is given as

Table 1 Lagrange multipliers estimated from sample dataset and the true population

Lagrange multipliers	*X* ~ gamma			*Y* ~ lognormal		
	λ_0	λ_1	λ_2	λ_0	λ_1	λ_2
Sample	7.0881	0.0505	− 1.3190	16.9052	− 6.8616	0.9810
Population	12.2919	0.0952	− 3.3000	17.4612	− 7.1633	1.0204

Table 2 Lagrange multipliers estimated using the first four moments about origin

	λ_0	λ_1	λ_2	λ_3	λ_4
X_s	1.6011	− 29.2444	101.8716	− 125.7947	57.5913
Y_s	− 1.2604	− 11.6222	103.6613	− 182.3986	101.5606

$$f(x) = \exp(-\lambda_0 - \lambda_1 x - \lambda_2 x^2 - \lambda_3 x^3),$$
if kurtosis is not significantly different from 3 (34a)

$$f(x) = \exp(-\lambda_0 - \lambda_1 x - \lambda_2 x^2 - \lambda_3 x^3 - \lambda_4 x^4),$$
if the kurtosis is significantly different from 3. (34b)

The objective function is written as

$$Z(\Lambda) = \ln\left[\int \exp\left(-\sum_{i=1}^{m}\lambda_i x^i\right)\mathrm{d}x\right] - \sum_{i=1}^{m}\lambda_i a_i;\ m = 3, 4;$$
$$\Lambda = [\lambda_1, \ldots, \lambda_m]. \quad (35)$$

To avoid the possible integration problem, the univariate variable is commonly scaled to [0,1] or [− 1,1] (Hao and Singh 2012; Zhang and Singh 2014). In this study, the univariate variables are scaled to [0,1] to assess its appropriateness. The scaled variable x_s is given as

$$x_s = \frac{x - (1-d)\min(x)}{(1+d)\max(x) - (1-d)\min(x)}. \quad (36)$$

In Eq. (36), d is a small number such that the scaled variable will not reach either the lower limit or the upper

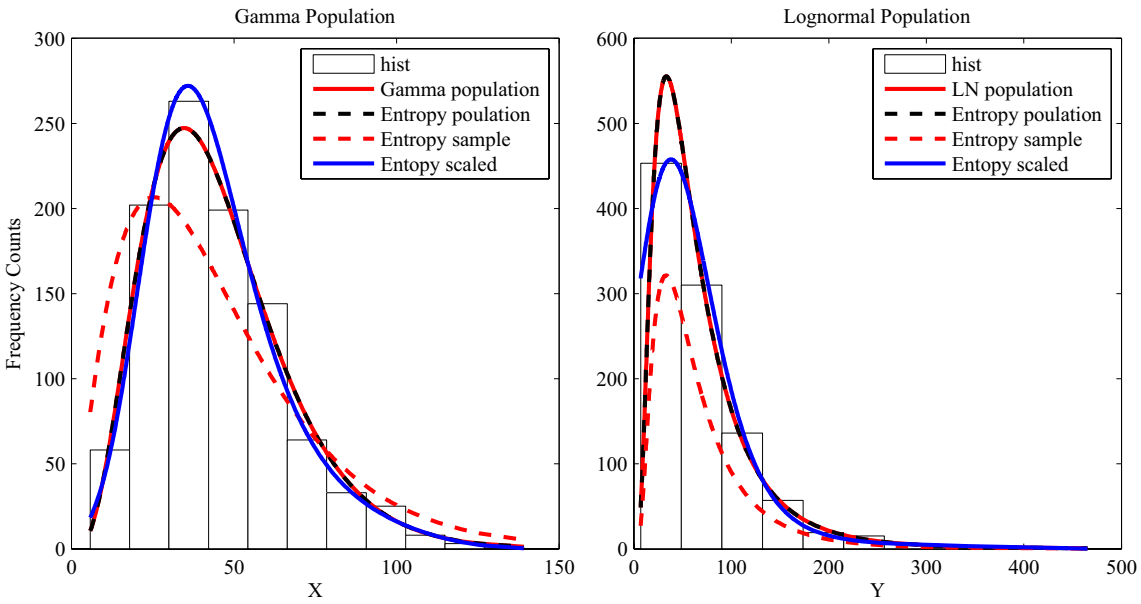

Fig. 1 Comparison of the true population and POME-based distributions from given population or from first four moments about origin (i.e., entropy scaled)

Table 3 Chi square univariate goodness-of-fit results (comparing to the population parameters)

Type	*X*				*Y*			
	S[a]	Cri[b].	*P* value	*df*	*S*	Cri.	*P* value	*df*
Fitted to sample[c]	3.74	15.51	0.88	8	0.92	15.51	1.00	8
Moments about origin[d]	5.94	12.59	0.43	6 6	6.54	12.59	0.37	6 6

[a] Test statistics computed for the Chi square test

[b] Critical value for $\alpha = 0.05$ of the Chi square distribution with certain degrees of freedom

[c] Fitted parametric distribution

[d] POME scaled

limit, i.e., avoiding $P(X \leq \max(x)) = 1$. Here, we chose $d = 0.01$. Equations (34)–(35) may then be re-organized with the use of the scaled variable as

$$f(x) = f(x_s)\left|\frac{dx_s}{dx}\right| = \frac{1}{(1+d)\max(x) - (1-d)\min(x)} \times \exp\left(-\lambda_0 - \sum_{i=1}^{m}\lambda_i x_s^i\right) \quad (37)$$

$$Z(\Lambda) = \ln\left[\int_0^1 \exp\left(-\sum_{i=1}^{m}\lambda_i x_s^i\right) dx\right] - \sum_{i=1}^{m}\lambda_i a_i; \quad m = 3 \text{ or } 4. \quad (38)$$

Now applying the first four moments about origin to the scaled variable X and the first three moments about origin to the scaled variable Y, Table 2 lists the parameters estimated by optimizing the objective function using the first four moments about origin [i.e., Eq. (38)].

The POME-based distributions for the original variables are expressed as

$$f(x) = \frac{1}{153.4792}\exp\left(-1.6011 + 29.2444x_s - 101.8716x_s^2 + 125.7947x_s^3 - 57.5913x_s^4\right) \quad (39a)$$

$$f(y) = \frac{1}{582.2347}\exp\left(1.2604 + 11.6222y_s - 103.6613y_s^2 + 182.3986y_s^3 - 101.5606y_s^4\right). \quad (39b)$$

Furthermore, Fig. 1 plots the relative frequency and the frequency computed from the POME-based distributions. As shown in Fig. 1, the POME-based univariate distributions derived (using the constraints pertaining to certain population, and first four moments about the origin) visually fit the observed data very well. Using the true population from the reference distribution, Table 3 lists the Chi square test for the fitted parametric and POME-based distributions constructed. Results in Table 3 clearly indicate the POME-derived distributions may be applied to model the univariate variables. Thus, it is safe to conclude that one may directly use the moments about origin as the constraints to model the univariate random variables.

Study of dependence

As previously discussed, one may apply three different approaches to study the dependence using the copula–entropy theory. Hereafter, each approach is evaluated. Within the objective of the study, the Gumbel–Hougaard, Clayton, Frank and meta-t copulas (Nelsen 2006) were applied as parametric copulas. The MECC copula was derived with the constraints of $E(U), E(U^2), E(V), E(V^2)$ and $E(UV)$. According to the discussion in "Univariate analysis of peak discharge and flood volume" section for

Table 4 Parameters, Log*L*, and entropy estimated from parametric copula

	Copula	GH	Clayton	Frank	*T*	
POME marginals	Parameter	4.8534	4.2251	17.2725	0.9474	$v = 4.4479$
	Log*L*	*1098.3*	712.6209	995.3770	1061.7	
	Entropy[a]	− 1.0983	− 0.7126	− 0.9954	− 1.0617	
Empirical marginals	Parameter	4.5732	3.1897	16.0426	0.9356	$v = 4.1301$
	Log*L*	*1106.2*	653.8323	973.6298	1040.8	
	Entropy	− 1.1062	− 0.6538	− 0.9736	− 1.0408	

Italic values indicate the best fitted copula function under each condition

[a] Entropy computed from the parametric copula using Eq. (40). The copula with the largest absolute value is the best copula candidate

Table 5 Parameters estimated for MECC copula

	λ_0	λ_1	λ_2	γ_1	γ_2	λ_3
With sample Spearman's rho as the constraints						
POME marginal	− 1.7581	1.2443	35.7275	1.2443	35.7275	− 73.9435
Empirical marginal	− 1.7581	1.2443	35.7275	1.2443	35.7275	− 73.9435
With true Spearman's rho as the constraints (from the true GH-copula)						
	− 1.7628	1.2356	36.3731	1.2356	36.3731	− 75.2173

univariate analysis, we will simply apply the POME-based distribution derived using the moments about the origin with the use of scaled variables.

POME-based marginals with parametric copulas

In this approach, the copula parameters were estimated with the use of POME-based marginals and by maximizing the log-likelihood function (it may be also called Two-Stage MLE). Table 4 lists the estimated parameters as well as the corresponding log-likelihood. In this approach, the copula yielding the largest log-likelihood was selected for further analysis. As seen in Table 4, the Gumbel–Hougaard copula was the best candidate. It was in agreement with the sample data actually generated from the Gumbel–Hougaard copula discussed earlier in the section.

POME-based marginals with parametric copulas selected based on the entropy

In this approach, the parameters of copulas again were estimated by Two-Stage MLE. The difference is that the copula–entropy was computed for the fitted parametric copula. Here, the copula–entropy was estimated using

$$\begin{aligned} H_C &= -E[\ln c(u,v;\theta)] \\ &= -\frac{1}{n}\sum_{i=1}^{n} \ln\, c(u,v;\theta); \qquad n\text{: sample size.} \end{aligned} \tag{40}$$

The computed entropy is also listed in Table 4. From the computed entropy using Eq. (40), it is seen that the Gumbel–Hougaard copula yielded the highest mutual information (the absolute value of the copula entropy) among all the copula candidates.

Parametric copulas estimated using Pseudo-MLE

In this approach, the parameters of the copula were directly estimated using the empirical distribution (e.g., empirical distribution using the Weibull plotting position formula) which is listed in Table 4. It is seen that with the Pseudo-MLE, the Gumbel–Hougaard copula again yielded the largest MLE and the highest mutual information.

Most entropic canonical copula with POME-based marginals (or empirical marginals)

In this approach, copulas were derived from the entropy theory with the constraints of $E(U) = E(V) = \frac{1}{2}$, $E(U^2) = E(V^2) = \frac{1}{3}$; $\rho_{\text{spearman}} = 0.9287$ (sample Spearman's rho) with the parameters listed in Table 5. In Eqs. (24)–(31), it is seen that the dependence structure (i.e., Spearman's rho in this sample study) was the controlling factor to optimize the objective function of MECC. Thus, it did not matter how the marginals were handled, the MECC would not change for given Spearman's rho which is shown in Table 5. To further evaluate the impact of Spearman's rho correlation coefficient, we changed Spearman's rho to population Spearman's rho as the constraint (i.e., $\rho = 0.9298$ from the true Gumbel–Hougaard copula). As seen in Table 5, there was a significant difference in the Lagrange multipliers estimated for MECC.

To assess whether the MECC so derived fulfilled the fundamental properties of copula function:

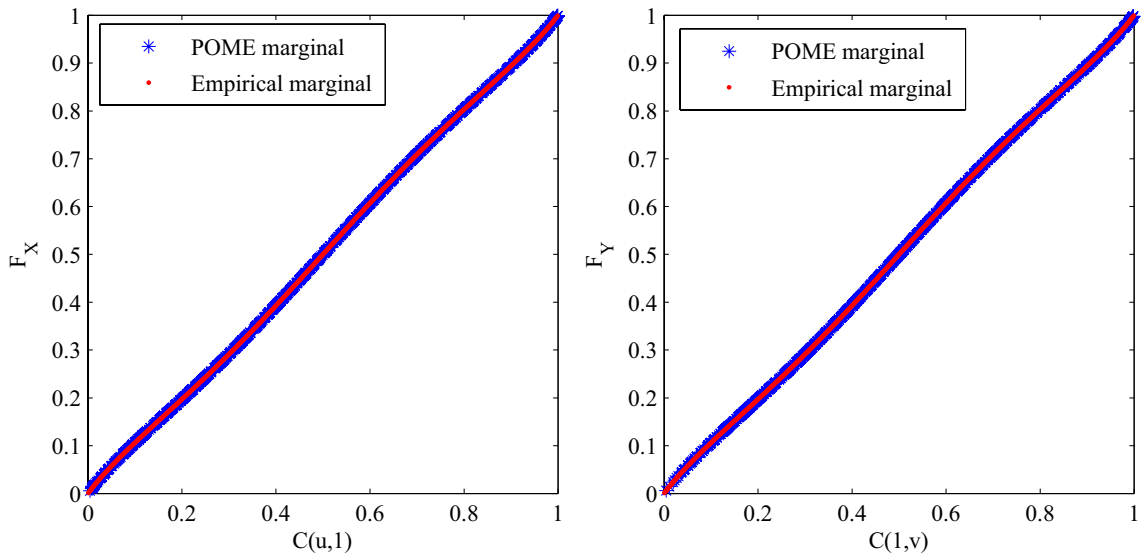

Fig. 2 Comparison of empirical and POME-univariate marginals to computed $C(u,1)$ and $C(1,v)$ computed from MECC

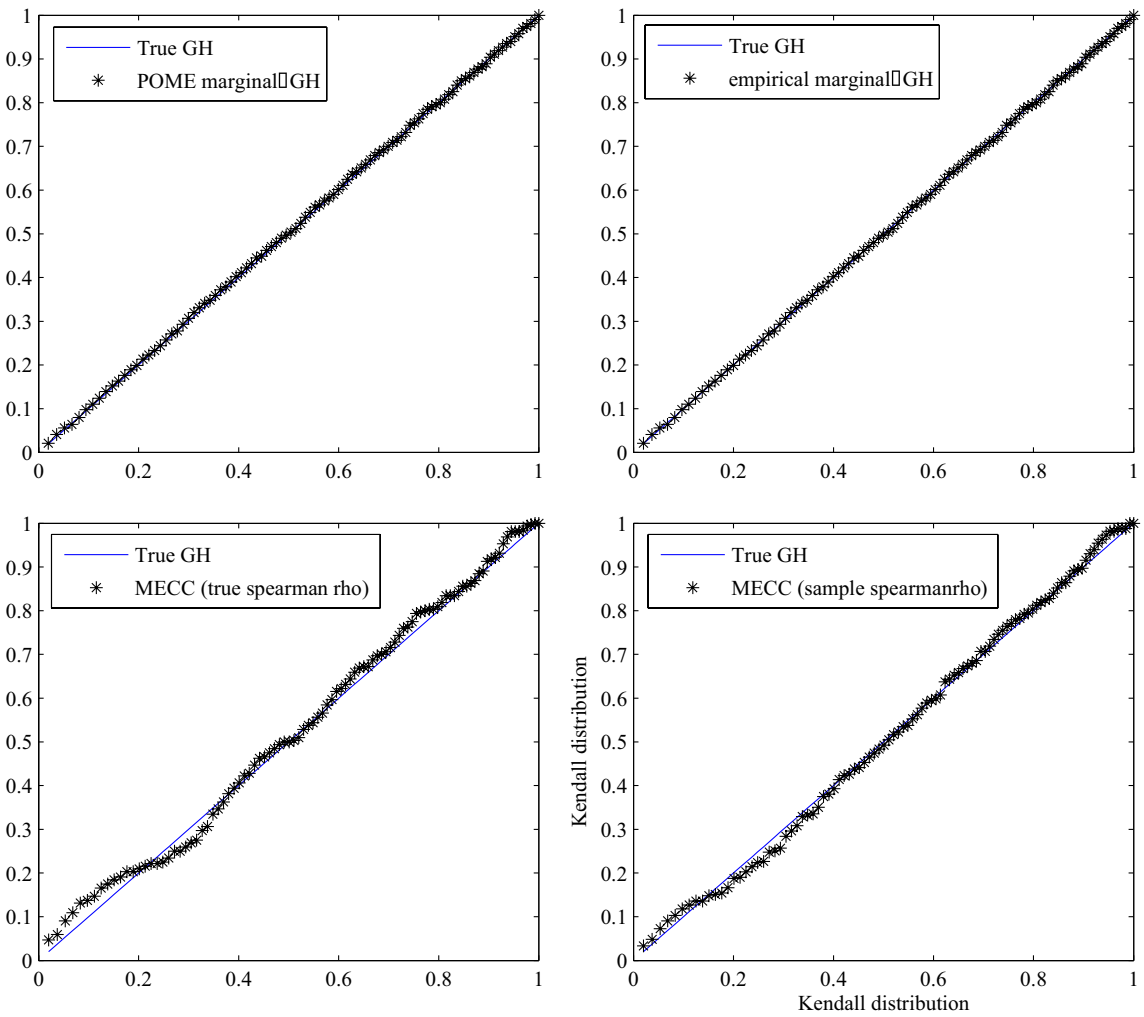

Fig. 3 Comparison of Kendall distribution from fitted Gumbel–Hougaard copula and MECC to that of the true Gumbel–Hougaard copula

$$C(u,1) = u; \quad C(1,v) = v, \tag{41}$$

Figure 2 compares the marginal variables computed using Eq. (41) from the MECC with both empirical and POME-based univariate marginals. As seen in Fig. 2, $C(u,1)$ and $C(1,v)$ were in good agreement with their empirical and POME-based univariate marginals. This also implied the appropriateness of POME-based univariate marginals derived using the first four and three non-central moments for random variables X and Y, respectively.

With the fundamental properties fulfilled with the use of MECC derived, one may want to further evaluate the goodness-of-fit of the derived MECC. Here, the Kendall distribution plots were generated and compared to the Kendall distribution of the underlying true Gumbel–Hougaard copula:

$$K_\theta(t) = \frac{t(\theta - \ln\ t)}{\theta}. \tag{42}$$

The Kendall distribution $[K_{\lambda}(t)]$ for MECC may be approximated following the procedure discussed in Genest et al. (2009) as follows:

1. Generate random variables $[U_1,U_2]$ with sample N from the MECC derived, where N is greater than the sample size of the observed dataset.
2. Approximate $[K_{\lambda}(t)]$ using:

$$V_i = \frac{1}{N}\sum_{j=1}^{N} 1\left(U_{1j} \le U_{1i} \cap U_{2j} \le U_{2i}\right); \quad i = 1,2,\ldots,N \tag{43a}$$

$$K^{\text{approx}}(t) = \frac{1}{N}\sum_{i=1}^{N} (V_i \le t); \quad t \in [0,1] \tag{43b}$$

Comparisons shown in Fig. 3 conclude that (i) both fitted Gumbel–Hougaard (GH) copula and MECC derived may properly represent the true GH ($\theta = 4.5$) population through the comparison of the Kendall distribution plots; (ii) visually, there is minimal difference of GH fitted with empirical marginals (b) and the MECC derived based on sample or population Spearman's rho (c and d); (iii) the reason for the minimal difference is due to the rank-based empirical distribution does not impose any external bias on parameter estimation for the parametric copulas; and the MECC derived here does not rely on the actual marginal values, but the population moments about origin for the uniformly distributed variables; and (iv) though the POME-based marginals well represent the univariate random variables, they do introduce external bias to the estimation of parametric copulas (a).

Overall, from the bivariate analysis of sample data, MECC may be directly applied to model the dependence structure of the random variables. In the case of the MECC application, the impact of the marginal distributions is eliminated. In the next section, we will use the real watershed data as a case study to further illustrate the copula–entropy theory as well as risk analysis.

Case study with real watershed data

Collected from Flume 1 at Walnut Gulch Watershed in Arizona, the annual maximum flood data [i.e., peak discharge (Q) and flood volume (V)] from 1957 to 2012 were considered for the case study. Based on the findings from analysis of sample data, the case study proceeded as follows: (i) the POME-based univariate distribution was applied to model the univariate peak discharge and flood volume; and (ii) the MECC was applied to model the dependence between peak discharge and flood volume.

Univariate analysis of peak discharge and flood volume

As discussed in "Univariate analysis of peak discharge and flood volume" section, the moments about origin for the scaled variables were considered as constraints to capture the shape and mode of the univariate flood variables. Choosing $d = 0.1$ in Eq. (36), Table 6 lists the sample statistics for the scaled variables. In Table 6, T and P denote the test statistic and the corresponding P value to evaluate whether kurtosis was significantly different from 3 using

$$\gamma_2^{\text{ex}} = \gamma_2 - 3 \tag{44a}$$

$$G_2 = \frac{n-1}{(n-2)(n-3)}((n+1)\gamma_2 + 6) \tag{44b}$$

$$T = \frac{G_2}{\text{SEK}}; \quad \text{SEK} = 2\sqrt{\frac{6n(n-1)^2}{(n-2)(n+5)(n^2-9)}}. \tag{44c}$$

In Eqs. (44a)–(44c), γ_2 and γ_2^{ex} denote the sample kurtosis and excessive kurtosis; n is the sample size; SEK is the standard error of kurtosis; and T is the test statistic with the underlying distribution of standard normal distribution.

Results in Table 6 indicate that the first three moments about origin were necessary to derive the POME-based distribution for the scaled peak discharge and flood volume variables. The Lagrange multipliers were optimized and listed in Table 7. Figure 4 compares the POME-based probability density to the histogram, as well as the POME-based CDF to the empirical CDF. Comparisons confirmed the appropriateness of the POME-based univariate distribution.

Bivariate flood frequency analysis with MECC

Let U and V represent the univariate marginals for peak discharge and flood volume, the same constraints to construct MECC for sample data [i.e., $E(U), E(U^2), E(V), E(V^2), E(UV)$] were applied to model the dependence of peak discharge and flood volume. The Lagrange multipliers were optimized by minimizing the objective function of Eq. (31a) with $b = 0$.

With the observed data, sample Spearman's rho was computed as $\hat{\rho}_{\text{spearman}} = 0.9419$, we approximated $E(UV)$ from sample Spearman's rho as $E(UV) = \frac{\rho_{\text{spearman}}+3}{12} \approx 0.3285$. With these constraints, the copula density function for the MECC was obtained to model bivariate flood frequency as

$$c(u,v) = \exp(-1.8194 - 1.1352u - 44.9511u^2 - 1.1352v - 44.9511v^2 + 92.1726uv). \tag{45}$$

Using Eq. (45), Fig. 5 compares (a) the $C(u,1), C(1,v)$ to the corresponding empirical and POME-based marginals, and (b) the approximated parametric Kendall distribution for MECC to the empirical Kendall distribution. Comparisons in Fig. 5 indicated that (a) the MECC constructed successfully fulfilled the copula properties of $C(u,1) = u, C(1,v) = v$; and (b) there was a good agreement between the empirical and parametric (i.e., MECC) Kendall distributions, which indicated the appropriateness of the MECC constructed. Applying the POME-based univariate distribution, Fig. 6 plots the simulated random variates versus the observed random variables. Figure 6 shows the dependence structure was well preserved with the application of MECC and POME-based marginals. To further compare the MECC with the empirical copula, Fig. 7 compares the copula and the survival copula with different contour levels. The plot on the left is for the copula function, while the plot on the right is for the survival copula. As shown in Fig. 7, there is good agreement between the contours obtained from the empirical copula (and its

Table 6 Sample statistics for scaled peak discharge and flood volume

Variable	$E(X)$	$E(X^2)$	$E(X^3)$	$E(X^4)$	T	P
Peak discharge	0.1499	0.1712	2.5921	12.0061	12.7843	$\ll 0.05$
Flood volume	0.2004	0.1988	1.4922	5.8259	5.0802	$\ll 0.05$

Table 7 Lagrange multipliers for POME-based univariate distribution

Variable	λ_0	λ_1	λ_2	λ_3	λ_4
Peak discharge	− 1.9340	5.8624	8.3878	− 10.5178	0.0004
Flood volume	− 1.6668	5.8557	− 1.7289	0.2827	0.0003

survival copula) and the MECC (and its survival copula) for different probability levels. This finding further assured the appropriateness of the MECC.

Risk analysis

With the assurance to apply the MECC to model the dependence of annual flood sequences (i.e., peak discharge and flood volume), one may proceed to study the associated risk measure, which may be used as an engineering design criterion. In hydrology and water engineering, risk has commonly been assessed through the return period. In what follows, the joint return period of "AND" case was applied for risk analysis. The joint return period of "AND" case is given as

$$T_{\text{and}} = \frac{\mu}{P(X \geq x, Y \geq y)} = \frac{\mu}{1 - F_X(x) - F_Y(y) + C\left(F_X(x), F_Y(y)\right)}. \quad (45)$$

To assess the return period "AND" case, the peak discharge and flood volume with univariate CDF of $P = 0.8$, 0.9, 0.96, and 0.98 were used. Given the limitation of the sample size ($n = 56$), $P = 0.99$ was not chosen for the study for comparison purposes. Table 8 lists the joint

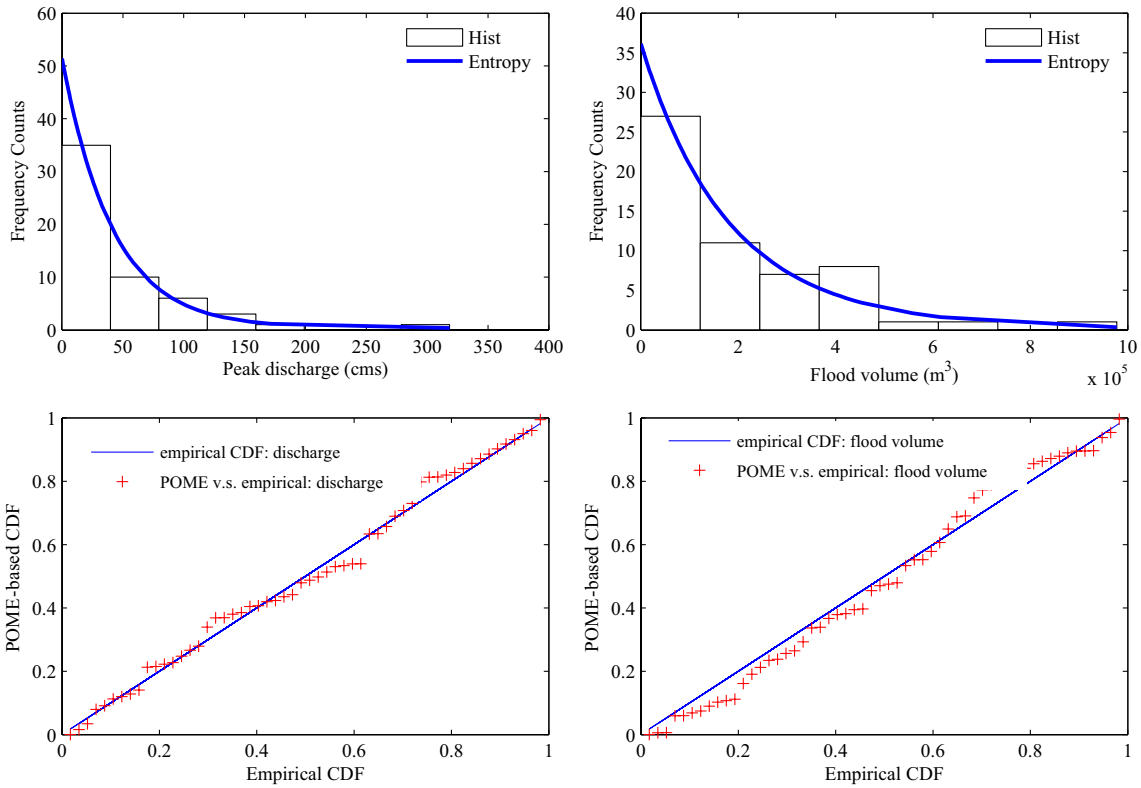

Fig. 4 Comparison of POME-based probability density function to the histogram

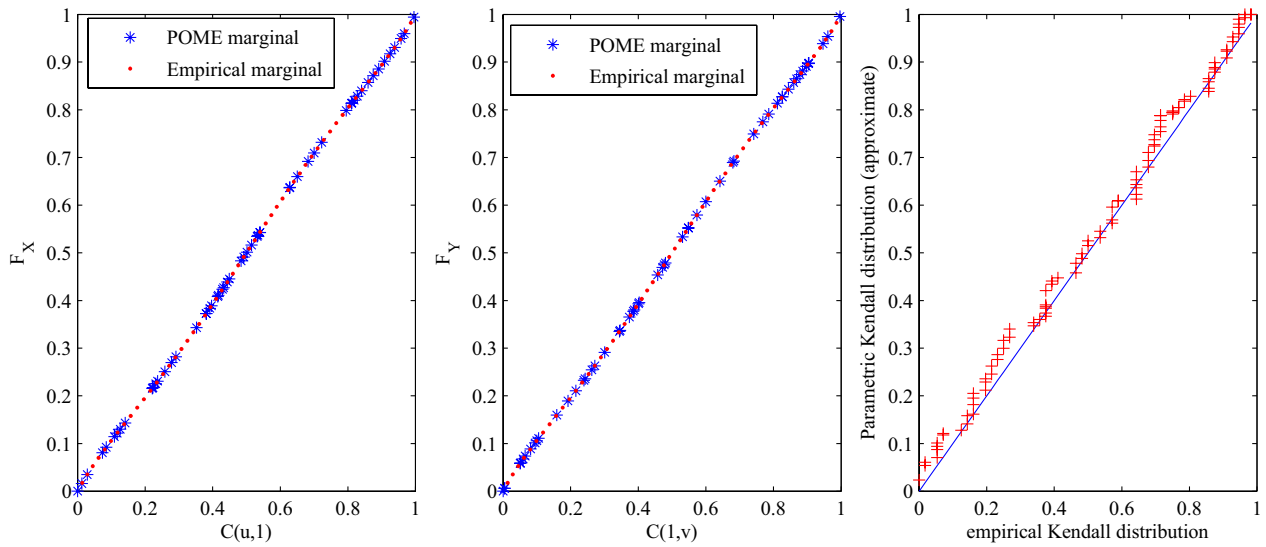

Fig. 5 Comparison of $C(u,1)$, $C(1,v)$, and the Kendall distribution

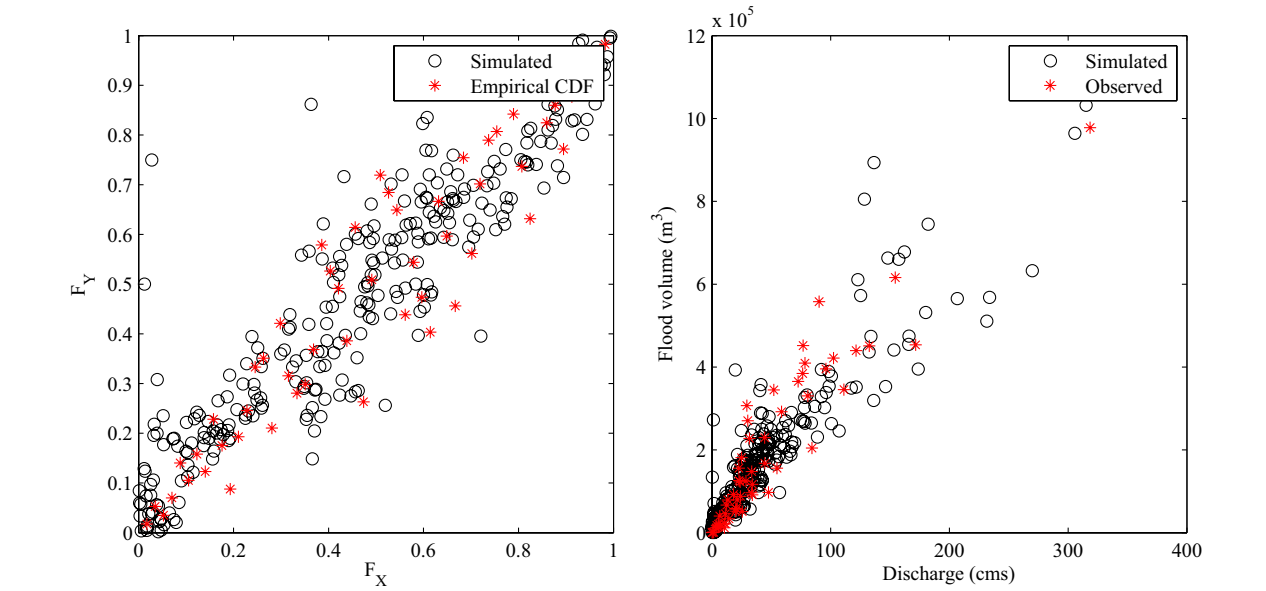

Fig. 6 Simulation results

return period estimated from both empirical copula and MECC. Results in Table 8 indicated the following:

(i) There was a small difference between the joint CDFs computed from empirical copula and the MECC. The absolute relative difference was in the range from 0.96% for $C(0.8,0.8)$ to 2.17% for $C(0.8,0.9)$. Thus, in regard to the joint CDF, the differences were insignificant.

(ii) Though the difference with joint CDF estimated may not be significant, it resulted in larger differences in regard to the "AND" case return period. It is seen that with the increased marginal probability, the discrepancy also increased between the T_{and} estimated from empirical copula and the MECC.

(iii) There was an interesting finding which was in agreement with T_{and} estimated from empirical copula and MECC. Using volume $= 6.44 \times 10^5$

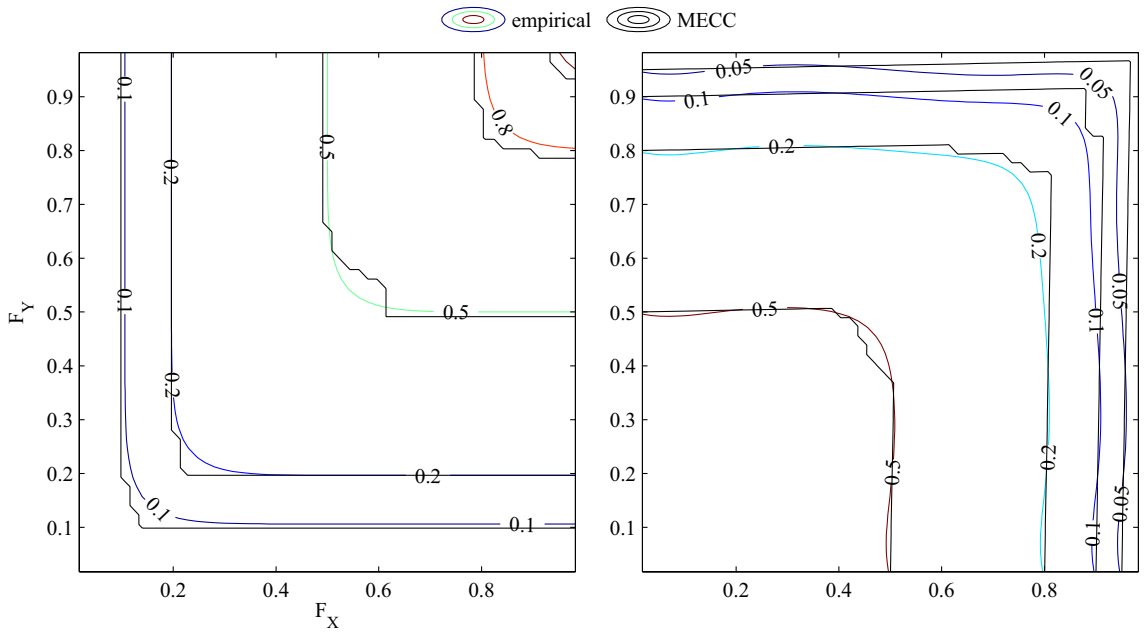

Fig. 7 Comparison of empirical to copula to MECC

Table 8 Joint CDF and T_{and} estimated from the empirical copula and MECC

C(*u*,*v*)		*P* (discharge)			
	P **(volume)**	***P*** **= 0.8**	***P*** **= 0.9**	***P*** **= 0.96**	***P*** **= 0.98**
Empirical	$P = 0.8$	0.7500	0.7857	0.8036	0.8036
	$P = 0.9$	0.8036	0.8750	0.9107	0.9107
	$P = 0.96$	0.8036	0.9107	0.9464	0.9643
	$P = 0.98$	0.8036	0.9107	0.9643	0.9821
MECC	$P = 0.8$	0.7503	0.7861	0.7940	0.7953
	$P = 0.9$	0.7861	0.8649	0.8946	0.9013
	$P = 0.96$	0.7940	0.8946	0.9429	0.9557
	$P = 0.98$	0.7953	0.9013	0.9557	0.9709
T_{and} **(years)**		**Discharge (cms)**			
	Volume (m^3)	**73.20**	**109.90**	**170.66**	**230.59**
Empirical	3.18×10^5	6.6667	11.6667	22.9508	42.4242
	4.60×10^5	9.6552	13.3333	19.7183	32.5581
	6.44×10^5	22.9508	19.7183	37.8378	41.1765
	7.71×10^5	42.4242	32.5581	41.1765	45.1613
MECC	3.18×10^5	6.6535	11.6146	29.4142	65.3461
	4.60×10^5	11.6146	15.3998	28.9062	46.8990
	6.44×10^5	29.4142	28.9062	43.6433	63.4992
	7.71×10^5	65.3462	46.8990	63.4992	91.8599

m^3 corresponding to $P = 0.96$ as an example, the joint return period computed from smaller peak discharge (e.g., $Q = 73.2$ cms corresponding to $P = 0.8$) was less than that computed with larger peak discharge (e.g., $Q = 109.9$ cms). This was true in reality, since it was more likely for ($Q \geq 109.9$

cms and $V \geq 6.44 \times 10^5$ m^3) to occur simultaneously compared to that for ($Q \geq 73.2$ cms and $V \geq 6.44 \times 10^5$ m^3). This finding was also in the agreement that higher discharge was most likely associated higher flood volume. This scenario also happened for large flood volume with relatively low peak discharge and vice versa.

Discussion and conclusions

In this study, we investigate the copula–entropy theory in bivariate analysis. Using the sample data with the known univariate populations (i.e., gamma and lognormal) and known dependence (Gumbel–Hougaard), it is concluded that the POME-based distribution derived may model the univariate distribution well. There is minimal difference for POME-based distribution based on the moment of the observed variable and that derived based on the scaled variable (i.e., scaling the observed variable to [0,1]). To avoid the improper integrals, the scaled variable is suggested to derive the POME-based distribution. Comparing to the true Gumbel–Hougaard copula, the MECC derived using the constraints of $E(U)$, $E(U^2)$, $E(V)$, $E(V^2)$, and $E(UV)$ can properly model the dependence structure of the sample data. The MECC constructed successfully fulfills the fundamental properties of the copula, i.e., $C(u,1) = u$; $C(1,v) = v$. In addition, the derived MECC can well present the true dependence structure represented with the Gumbel–Hougaard copula.

Using the real watershed data (i.e., Flume 1 at Walnut Gulch, Arizona), the case study shows the appropriateness of POME-univariate distribution of scaled variable to model the univariate distribution for the observed variates. With the constraints $E(U)$, $E(U^2)$, $E(V)$, and $E(V^2)$ converging to the population moments of the uniform distributed variables as $E(U^i) = E(V^i) = 1/(i+1)$; the MECC constructed only depends on the rank-based dependence measure (in this case, Spearman's rho). The derived MECC properly models the dependence of annual peak discharge and flood volume, which is independent of the marginal distributions (non-parametric or parametric). The evaluation of the flood risk (using "AND" case return period) indicates that the MECC copula reasonably represents the change of the return period of "AND" case.

Overall, the study concludes as follows:

(i) For the bivariate random variables investigated, the MECC may be easily and efficiently applied to model the dependence structure. Unlike other copulas, the MECC is uniquely defined for a given set of constraints. Its uniqueness allows one universal solution for the proposed frequency analysis.

(ii) Similar to other copula families (e.g., Archimedean copulas, meta-elliptical copulas, vine copulas, etc.), the MECC may be applied for multivariate analysis in hydrology and water engineering, including multivariate rainfall analysis, multivariate drought analysis, spatial analysis of drainage networks, and spatial analysis of water quality as few examples.

(iii) The bivariate MECC copula may be easily extended to higher dimensions. For example, for the d-dimensional variables $[X_1, X_2, \ldots, X_d]$ with the marginals of $U_i = F_i(X_i), i = 1, 2, \ldots, d$; the MECC may be constructed using the set of constraints, i.e., marginal $E(U_i^r) = 1/(r+1), \quad i = 1, 2, \ldots, d$ and pair-wise $E(U_i U_j); \quad i, j \in [1, d], i \neq j$ estimated from rank-based Spearman's coefficient of correlation. The same optimization procedure applied for the bivariate case may be applied to construct the MECC for dependence structure in higher dimensions.

Authors' contributions
VPS conceptualized the paper, helped with data interpretation and crafting the manuscript. LZ did analysis, processed the data, constructed all the graphs and wrote the first draft. Both authors read and approved the final manuscript.

Author details
[1] Department of Biological & Agricultural Engineering, Texas A&M University, College Station, TX 77843-2117, USA. [2] Zachry Department of Civil Engineering, Texas A&M University, College Station, TX 77843-2117, USA.

Acknowledgements
No applicable.

Competing interests
The authors declare that they have no competing interests.

Availability of data and materials
The data used is in public domain and is available for anyone to use.

Consent for publication
We consent for publication.

Ethics approval and consent to participate
Not applicable.

Funding
No funding source is available.

Publisher's Note

Received: 7 September 2017 Accepted: 5 February 2018
Published online: 24 February 2018

References

Aas K, Czado C, Frigessi A, Bakken H (2007) Pair-copula constructions of multiple dependence. Insur Math Econ. https://doi.org/10.1016/j.insmatheco.2007.02.001

Aghakouchak A (2014) Entropy–copula in hydrology and climatology. J Hydrometeorol 15:2176–2189. https://doi.org/10.1175/jhm-d-13-0207.1

Arya FK, Zhang L (2017) Copula-based Markov process for forecasting and analyzing risk of water quality time series. J Hydrol Eng 22(6):04017005. https://doi.org/10.1061/(asce)he.1943-5584.00001494

Chen L, Singh VP, Guo S (2013) Measure of correlation between river flows using entropy–copula theory. J Hydrol Eng 18(12):1591–1608. https://doi.org/10.1061/(asce)he.1943-5584.0000714
Chu B (2011) Recovering copulas from limited information and an application to asset allocation. J Bank Finance 35:1824–1842. https://doi.org/10.1016/j.jbankfin.2010.12.011
Cobb L, Koppstein P, Chen NH (1983) Estimation and moment recursion relations for multimodal distributions of the exponential family. J Am Stat Assoc 78:124–130
Falk M, Reiss R-D (2005) On pickands coordinates in arbitrary dimensions. J Multivar Anal 92:426–453
Fang HB, Fang KT, Kotz S (2002) The meta-elliptical distributions with given marginals. J Multivar Anal 82:1–16
Genest C, Favre A-C, Béliveau 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, Remillard B, Beaudoin D (2009) Goodness-of-fit tests for copulas: a review and a power study. Insur Math Econ 44(2):199–213. https://doi.org/10.1016/j.insmatheco.2007.10.005
Gudendorf G, Segers J (2009) Extreme-value copulas. arXiv:0911.1015v2
Hao Z, Singh VP (2012) Entropy–copula method for single-site monthly streamflow simulation. Water Resour Res 48:W06604. https://doi.org/10.1029/wr011419
Jaynes ET (1957) Information theory and statistical mechanics. Phys Rev 106:620–630
Joe H (2014) Dependence modeling with copulas. CRC Press, Boca Raton
Kullback S, Leibler RA (1951) On information and sufficiency. Ann Math Stat 22:79–86
Nelsen RB (2006) An introduction to copulas, 2nd edn. Springer Science + Business Media, Inc., Berlin
Pham MT, Vernieuwe H, Baets BD, Willems P, Verhoest NEC (2016) Stochastic simulation of precipitation-consistent daily reference evapotranspiration using vine copulas. Stoch Environ Res Risk Assess 30:2197–2214. https://doi.org/10.1007/s00477-015-1181-7
Pickands J (1981) Multivariate extreme value distribution. Bull Int Stat Inst 49:859–878
Renyi A (1951) On measure of entropy and information. In: Proceedings, 4th Berkeley symposium, mathematics, statistics, and probability, Berkeley, California, pp 547–561
Requena AI, Chebana F, Mediero L (2016a) A complete procedure for multivariate index-flood model application. J Hydrol 535:559–580. https://doi.org/10.1016/j.jhydrol.2016.02.004
Requena AI, Flores I, Mediero L, Garrote L (2016b) 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/10.1007/200477-015-1138-x
Salvadori G, Michele CD (2015) Multivariate real-time assessment of droughts via copula-based multi-site hazard trajectories and fans. J Hydrol 526:101–115. https://doi.org/10.1016/j.jhydrol.2014.11.056
Shannon CE (1948) A mathematical theory of communication. Bell Syst Technol J 27:379–423
Singh VP (1998) Entropy-based parameter estimation in hydrology. Springer, Dordrecht
Singh VP, Rajagopal AK (1986) A new method of parameter estimation for hydrologic frequency analysis. Hydrol Sci Technol 2(3):33–40
Sklar M (1959) Fonctions de repartition an dimensions et leurs marges. Universite Paris, Paris, p 8
Song S, Singh VP (2010) Meta-elliptical copulas for drought frequency analysis of periodic hydrologic data. Stoch Environ Res Risk Assess 24(3):425–444. https://doi.org/10.1007/s00477-009-0331-1
Sraj M, Bezak N, Brilly M (2015) Bivariate flood frequency analysis using the copula function: a case study of the Litija station on the Sava River. Hydrol Process 29:225–238. https://doi.org/10.1002/hyp.10145
Tsallis C (1988) Possible generalizations of Boltzmann–Gibbs statistics. J Stat Phys 52(1/2):479–487
Verneiuwe H, Vandenberghe S, Baets BD, Verhoest NEC (2015) A continuous rainfall model based on vine copulas. Hydrol Earth Syst Sci 19:2685–2699. https://doi.org/10.5194/hess-19-2685-2015
Zellner A, Highfield RA (1988) Calculation of maximum entropy distribution and approximation of marginal posterior distributions. J Econom 37:95–209
Zhang L, Singh VP (2012) Bivariate rainfall and runoff analysis using entropy and copula theories. Entropy 14:1784–1812. https://doi.org/10.3390/e14091784
Zhang L, Singh VP (2014) Joint conditional probability distributions of runoff depth and peak discharge using entropy theory. J Hydrol Eng 19(6):1150–1159