Water Science and Engineering, 2013, 6(1): 18-30
doi:10.3882/j.issn.1674-2370.2013.01.002

http://www.waterjournal.cn
e-mail: wse2008@vip.163.com

Goodness-of-fit tests for multi-dimensional copulas: Expanding application to historical drought data

Ming-wei MA[1], Li-liang REN*[1], Song-bai SONG[2], Jia-li SONG[3], Shan-hu JIANG[1]

1. State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing 210098, P. R. China
2. College of Water Resources and Architectural Engineering, Northwest A&F University, Yangling 712100, P. R. China
3. Business School, Hohai University, Nanjing 211100, P. R. China

Abstract: The question of how to choose a copula model that best fits a given dataset is a predominant limitation of the copula approach, and the present study aims to investigate the techniques of goodness-of-fit tests for multi-dimensional copulas. A goodness-of-fit test based on Rosenblatt's transformation was mathematically expanded from two dimensions to three dimensions and procedures of a bootstrap version of the test were provided. Through stochastic copula simulation, an empirical application of historical drought data at the Lintong Gauge Station shows that the goodness-of-fit tests perform well, revealing that both trivariate Gaussian and Student t copulas are acceptable for modeling the dependence structures of the observed drought duration, severity, and peak. The goodness-of-fit tests for multi-dimensional copulas can provide further support and help a lot in the potential applications of a wider range of copulas to describe the associations of correlated hydrological variables. However, for the application of copulas with the number of dimensions larger than three, more complicated computational efforts as well as exploration and parameterization of corresponding copulas are required.

Key words: *goodness-of-fit test; multi-dimensional copulas; stochastic simulation; Rosenblatt's transformation; bootstrap approach; drought data*

1 Introduction

Copulas, initially introduced by Sklar (1959), are functions that join univariate distributions to form their multivariate distribution. They offer the flexibility of modeling multivariate distribution through the choice of margins from different families of univariate distributions and the selection of a suitable dependence structure. Due to their favorable properties, copulas have proved useful in financial applications (Frees et al. 1996; Mendes and Souza 2004). In recent years, copulas have been introduced into analyses of multivariate hydrological extreme events and have become a popular tool for modeling the dependence

This work was supported by the Program of Introducing Talents of Disciplines to Universities of the Ministry of Education and State Administration of the Foreign Experts Affairs of China (the 111 Project, Grant No. B08048) and the Special Basic Research Fund for Methodology in Hydrology of the Ministry of Sciences and Technology of China (Grant No. 2011IM011000).
*Corresponding author (e-mail: *RLL@hhu.edu.cn*)
Received Nov. 22, 2011; accepted Apr. 12, 2012

structures of correlated/non-independent hydrological random variables, e.g., rainfall (Evin and Favre 2008; Wang et al. 2010; Zhang et al. 2012), floods (Grimaldi and Serinaldi 2006; Zhang and Singh 2007; Chowdhary et al. 2011), and droughts (Shiau 2006; Song and Singh 2010; Zhang et al. 2011; Ma et al. 2012).

Considering the availability of excessive copula functions, some criteria (e.g., the Akaike information criterion (AIC), Bayesian information criterion (BIC), and root mean square error (RMSE)) are widely used to select appropriate copulas as well as other multi-dimensional models by estimating their fitting biases. However, relatively small fitting biases do not invariably guarantee a satisfactory representation of the observations. Whether or not a certain copula or a parametric family of copulas is competent for the description of the dependence structures in the historical data can be investigated by applying specialized goodness-of-fit tests for copulas. Although several goodness-of-fit tests have been proposed, there are no general guidelines for selecting the optimal parametric copula. Genest and Rivest (1993) developed an empirical method to identify the best copula in the Archimedean case. Since copulas are invariant under strictly increasing transformations (Nelsen 1999), Diebold et al. (1998, 1999), Berkowitz (2001), and Berg and Bakken (2005) used the probability integral transform (PIT) of the data in the evaluation of copula models. Panchenko (2005) focused on positive definite bilinear forms, while Genest et al. (2006) utilized the Kendall’s process. For a thorough review of contributions to this field, see also Malevergne and Sornette (2003), Breymann et al. (2003), Dobrić and Schmid (2005), Junker and May (2005), and Fermanian (2005).

Dobrić and Schmid (2007) addressed a test for parametric families of bivariate copulas based on Rosenblatt’s transformation, which was also suggested and applied in Breymann et al. (2003). In these applications, bivariate copulas were mainly investigated while the methodology was tested and verified with either financial data or artificial samples. Though Dobrić and Schmid (2007) declared that the computation of the test statistics could be applied to the cases of higher-dimensional copulas, relevant studies exploring multi-dimensional copulas and coping with hydrological data have not been reported so far. In fact, difficulties and special issues are expected to arise in the process of transformations from two dimensions to three dimensions (or even to higher numbers of dimensions). Therefore, the present study aims (1) to propose a goodness-of-fit test for multi-dimensional copulas with parametric expressions based on Rosenblatt’s transformation, and (2) to verify the capability of the test through stochastic simulation of trivariate Gaussian and Student t copulas using historical drought observations.

2 Methodology

2.1 Rosenblatt’s transformation

Rosenblatt (1952) proposed a transformation mapping a k-variate random vector with a continuous distribution to one with a uniform distribution on the k-dimensional hypercube. The transformation can be used to obtain the residuals for various multivariate probability models, which allows for formal goodness-of-fit testing of these models. A simple description of Rosenblatt's transformation is as follows:

Following the notation of Rosenblatt (1952), let $\boldsymbol{X}=\left(X_1, X_2, \cdots, X_k\right)$ be a random vector with distribution function $F\left(x_1, x_2, \cdots, x_k\right)$. The conditional cumulative distribution functions are defined as

$$
\begin{gathered}
F_1\left(x_1\right)=P\left(X_1 \leq x_1\right) \\
F_2\left(x_2 \mid x_1\right)=P\left(X_2 \leq x_2 \mid X_1=x_1\right) \\
\vdots \\
F_k\left(x_k \mid x_1, x_2, \cdots, x_{k-1}\right)=P\left(X_k \leq x_k \mid X_1=x_1, X_2=x_2, \cdots, X_{k-1}=x_{k-1}\right)
\end{gathered}
$$

Then, Rosenblatt's transformation T is given by $\boldsymbol{z}=\left(z_1, z_2, \cdots, z_k\right)=T\boldsymbol{x}=T\left(x_1, x_2, \cdots, x_k\right)$, where

$$
\begin{gathered}
z_1=F_1\left(x_1\right) \\
z_2=F_2\left(x_2 \mid x_1\right) \\
\vdots \\
z_k=F_k\left(x_k \mid x_1, x_2, \cdots, x_{k-1}\right)
\end{gathered}
$$

If the distribution of $\boldsymbol{X}$ is continuous, the random vector $\boldsymbol{Z}$, given by $\boldsymbol{Z}=T\boldsymbol{X}$, is uniformly distributed on the *k*-dimensional hypercube.

2.2 Mathematical derivation of goodness-of-fit test

Let X, Y, and Z denote three random variables with a joint probability distribution function $F_{X,Y,Z}\left(x, y, z\right)=P\left(X \leq x, Y \leq y, Z \leq z\right)$ for $\left\{x, y, z\right\} \in \mathbf{R}^3$ and the marginal distribution functions $F_X\left(x\right)=P\left(X \leq x\right)$, $F_Y\left(y\right)=P\left(Y \leq y\right)$, and $F_Z\left(z\right)=P\left(Z \leq z\right)$ for $x, y, z \in \mathbf{R}$. Suppose F_X, F_Y, and F_Z are all continuous functions; then, there exists a unique copula $C:\left[0,1\right]^3 \rightarrow\left[0,1\right]$ with

$$
F_{X,Y,Z}\left(x, y, z\right)=C\left(F_X\left(x\right), F_Y\left(y\right), F_Z\left(z\right)\right)
$$

where $C\left(\cdot\right)$, the trivariate copula, denotes the joint distribution function of the variables. Let $U=F_X\left(x\right)$, $V=F_Y\left(y\right)$, and $W=F_Z\left(z\right)$, i.e., $C\left(u, v, w\right)=P\left(U \leq u, V \leq v, W \leq w\right)$ for $\left\{u, v, w\right\} \in\left[0,1\right]^3$, and the conditional distribution function of W at given $U=u$ and $V=v$ can be expressed as

$$
C\left(w \mid u, v\right)=P\left(W \leq w \mid U=u, V=v\right)=\lim_{\Delta u \rightarrow 0, \Delta v \rightarrow 0} P\left(W \leq w \mid u \leq U \leq u+\Delta u, v \leq V \leq v+\Delta v\right)=
$$

$$
\lim_{\Delta u \rightarrow 0, \Delta v \rightarrow 0} \frac{C\left(u+\Delta u, v+\Delta v, w\right)-C\left(u, v, w\right)}{C\left(u+\Delta u, v+\Delta v\right)-C\left(u, v\right)}=\frac{\dfrac{\partial^2 C\left(u, v, w\right)}{\partial u \partial v}}{\dfrac{\partial^2 C\left(u, v\right)}{\partial u \partial v}}
$$

Here, we assume that the second-order partial derivative exists. According to Rosenblatt (1952), the random variables

$$
J_1=C\left(U, V\right)=F_{X,Y}\left(X, Y\right)
$$

and

$$J_2 = C\left(W \mid U, V\right) = C\left(F_Z\left(Z\right) \mid F_{X,Y}\left(X,Y\right)\right)$$

are independent and uniformly distributed in $[0,1]$. Thus, the random variable

$$S\left(J_1, J_2\right) = \left[\Phi^{-1}\left(F_{X,Y}\left(X,Y\right)\right)\right]^2 + \left[\Phi^{-1}\left(C\left(F_Z\left(Z\right) \mid F_{X,Y}\left(X,Y\right)\right)\right)\right]^2$$

has a χ_2^2 distribution, i.e., $S\left(J_{11}, J_{21}\right), S\left(J_{12}, J_{22}\right), \cdots, S\left(J_{1n}, J_{2n}\right)$ is a random sample from a χ_2^2-distributed random variable with a corresponding random sample $\left(X_1, Y_1, Z_1\right)$, $\left(X_2, Y_2, Z_2\right), \cdots, \left(X_n, Y_n, Z_n\right)$ from $\left(X, Y, Z\right)$, where $\Phi^{-1}\left(\cdot\right)$ denotes the inverse of standard normal distribution $\Phi\left(\cdot\right)$. These properties can be used to propose a test for the null hypothesis, i.e., $H_0 : \left(X, Y, Z\right)$ has a copula $C\left(u, v, w\right)$, in the condition that the marginal distribution functions F_X, F_Y, and F_Z are known or given. In this case, the values of $S\left(J_{11}, J_{21}\right), S\left(J_{12}, J_{22}\right), \cdots, S\left(J_{1n}, J_{2n}\right)$ can be calculated and used to test the equivalent null hypothesis, i.e., $H_0^* : S\left(J_1, J_2\right)$ has a χ_2^2 distribution. We reject H_0 if H_0^* is rejected.

2.3 Procedures of bootstrap version for trivariate copulas

According to Dobrić and Schmid (2007), Genest et al. (2009), Song and Singh (2010), and Ma and Song (2010), the procedures of goodness-of-fit tests for trivariate copulas using a bootstrap approach are as follows:

(1) The empirical marginal distribution functions $\hat{F}_X$, $\hat{F}_Y$, and $\hat{F}_Z$ are estimated using the Gringorten plotting position formulas:

$$\hat{F}_X\left(x\right) = \frac{\sum_{k=1}^{n} 1\left\{X_k \le x\right\} - 0.44}{n + 0.12}, \quad \hat{F}_Y\left(y\right) = \frac{\sum_{k=1}^{n} 1\left\{Y_k \le y\right\} - 0.44}{n + 0.12}, \text{ and } \hat{F}_Z\left(z\right) = \frac{\sum_{k=1}^{n} 1\left\{Z_k \le z\right\} - 0.44}{n + 0.12}$$

(2) The joint probability distribution of $\left(X_i, Y_i\right)$ is estimated using a chosen bivariate copula:

$$\hat{F}_{X,Y}\left(X_i, Y_i\right) = C\left(\hat{F}_X\left(x_i\right), \hat{F}_Y\left(y_i\right)\right)$$

(3) $\hat{S}_i$, which has a χ_2^2 distribution, is computed:

$$\hat{S}_i = \hat{S}\left(J_{1i}, J_{2i}\right) = \left[\Phi^{-1}\left(\hat{F}_{X,Y}\left(X_i, Y_i\right)\right)\right]^2 + \left[\Phi^{-1}\left(C\left(\hat{F}_Z\left(Z_i\right) \mid \hat{F}_{X,Y}\left(X_i, Y_i\right)\right)\right)\right]^2$$

for $i = 1, 2, \cdots, n$.

(4) The Anderson-Darling statistic A_n^2 is computed:

$$A_n^2 = -n - \frac{1}{n}\sum_{j=1}^{n}\left\{\left(2j-1\right)\left[\ln\left(F_0\left(S_{(j)}\right)\right) + \ln\left(1 - F_0\left(S_{(n-j+1)}\right)\right)\right]\right\}$$

where $S_j = \hat{S}_i$ for $j = i$, $S_{(1)} \le S_{(2)} \le \cdots \le S_{(n)}$ are in an increasing order, and $F_0\left(\cdot\right)$ denotes the distribution function of a χ_2^2-distributed variable.

(5) Parameter $\hat{\theta}$ of a trivariate copula is estimated from the original observations $\left(x_1, y_1, z_1\right), \left(x_2, y_2, z_2\right), \cdots, \left(x_n, y_n, z_n\right)$.

(6) Independent and identically distributed samples $\left(x_1^*, y_1^*, z_1^*\right), \left(x_2^*, y_2^*, z_2^*\right), \cdots,$ $\left(x_n^*, y_n^*, z_n^*\right)$ are simulated from the corresponding copula with parameter $\hat{\theta}$, and parameter $\hat{\theta}^*$ is then estimated from $\left(x_1^*, y_1^*, z_1^*\right), \left(x_2^*, y_2^*, z_2^*\right), \cdots, \left(x_n^*, y_n^*, z_n^*\right)$.

(7) $\hat{S}_i$ is computed and sorted in an increasing order to obtain $S_{(j)}$, and then the statistic $A_n^{2(*)}$ of the Anderson-Darling test is computed.

(8) Steps (6) and (7) are repeated N_{B} times, with $N_{\mathrm{B}} = 5\,000$ in this study being the number of bootstrap repetitions. The desired critical value is then determined as the (1-α)-quantile of the values $A_n^{2(*1)}, A_n^{2(*2)}, \cdots, A_n^{2(*N_{\mathrm{B}})}$ (where α is the significance level).

Note that some other test statistics, such as Kolmogorov-Smirnov's D_n, can also be applied to the goodness-of-fit tests by replacing Anderson-Darling's A_n^2 with D_n in the procedures above.

3 Copulas simulation

The modeled samples necessary for goodness-of-fit tests resort to copula simulation (step (6) in the above-proposed procedures). Therefore, procedures for Gaussian and Student t copulas as well as a case study are provided below to illustrate goodness-of-fit tests for trivariate copulas.

3.1 Trivariate Gaussian and Student t copulas

The parametric expression of trivariate Gaussian copula derived from Fang et al. (2002) and Žežula (2009) is

$$C\left(u_1,u_2,u_3;\boldsymbol{\Sigma}\right)=\int_{-\infty}^{\Phi^{-1}(u_1)}\int_{-\infty}^{\Phi^{-1}(u_2)}\int_{-\infty}^{\Phi^{-1}(u_3)}\frac{1}{(2\pi)^{\frac{3}{2}}\left|\boldsymbol{\Sigma}\right|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}\boldsymbol{w}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{w}\right)\mathrm{d}\boldsymbol{w} \tag{1}$$

where $u_1 = F_{X_1}\left(x_1\right)$, $u_2 = F_{X_2}\left(x_2\right)$, and $u_3 = F_{X_3}\left(x_3\right)$, taking values in [0,1] as the marginal distributions of random variables X_1, X_2, and X_3; $\boldsymbol{\Sigma} = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{bmatrix}$ is the symmetrical correlation matrix with $-1 \le \rho_{ij} \le 1$ ($i, j = 1,2,3$); and $\boldsymbol{w}$ represents the corresponding integral variables, and $\boldsymbol{w} = \left(w_1, w_2, w_3\right)^{\mathrm{T}}$.

According to Fang et al. (2002) and Demarta and McNeil (2005), the trivariate Student t copula can be parametrically expressed as

$$C\left(u_1,u_2,u_3;\boldsymbol{\Sigma},\upsilon\right)=\int_{-\infty}^{T_\upsilon^{-1}(u_1)}\int_{-\infty}^{T_\upsilon^{-1}(u_2)}\int_{-\infty}^{T_\upsilon^{-1}(u_3)}\frac{\Gamma\left(\frac{\upsilon+3}{2}\right)}{\Gamma\left(\frac{\upsilon}{2}\right)}\frac{1}{(\pi\upsilon)^{\frac{3}{2}}\left|\boldsymbol{\Sigma}\right|^{\frac{1}{2}}}\left(1+\frac{\boldsymbol{w}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{w}}{\upsilon}\right)^{-\frac{\upsilon+3}{2}}\mathrm{d}\boldsymbol{w} \tag{2}$$

where $T_\upsilon^{-1}\left(\cdot\right)$ denotes the inverse of Student t distribution $T_\upsilon\left(\cdot\right)$ with the degree of freedom of υ, and $\Gamma\left(\cdot\right)$ denotes the gamma function.

3.2 Gaussian copula simulation

(1) Simulate the independent and uniformly distributed random variables v_1, v_2, and v_3.

(2) Set $u_1 = v_1$.

(3) Set $F\left(u_2 | u_1\right)=v_2$.

Let $g\left(b_1, y\right)=\frac{1}{2\pi|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\boldsymbol{w}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{w}\right)$, $b_1=\Phi^{-1}\left(u_1\right)$, $b_2=\Phi^{-1}\left(u_2\right)$, $\varphi(t)=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-t^2}$, and $\boldsymbol{w}=\left(b_1, y\right)^{\mathrm{T}}$. Using composite function derivative rules, we can obtain $F\left(u_2 | u_1\right)=\frac{\partial C\left(u_1, u_2\right)}{\partial u_1}=\frac{1}{\varphi\left(b_1\right)}\int_{-\infty}^{b_2} g\left(b_1, y\right)\mathrm{d}y$. After algebraic simplifications, we obtains $v_2=\frac{\partial C\left(u_1, u_2\right)}{\partial u_1}=\Phi\left(\frac{b_2-\rho b_1}{\sqrt{1-\rho^2}}\right)$, where ρ is the correlation coefficient between b_1 and b_2. Taking the inverse of v_2, we finally have

$$u_2=\Phi\left(b_2\right)=\Phi\left[\Phi^{-1}\left(v_2\right)\sqrt{1-\rho^2}+\rho b_1\right] \tag{3}$$

(4) Set $F\left(u_3 | u_1 u_2\right)=\frac{\frac{\partial^2 C\left(u_1, u_2, u_3\right)}{\partial u_1 \partial u_2}}{\frac{\partial^2 C\left(u_1, u_2\right)}{\partial u_1 \partial u_2}}=v_3$.

Let $g\left(b_1, b_2, z\right)=\frac{1}{(2\pi)^{\frac{3}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\boldsymbol{w}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{w}\right)$, where $\boldsymbol{w}=\left(b_1, b_2, z\right)^{\mathrm{T}}$, $b_3=\Phi^{-1}\left(u_3\right)$, and $\boldsymbol{\Sigma}^{-1}=\begin{bmatrix}\rho_{11}^{-1} & \rho_{12}^{-1} & \rho_{13}^{-1} \\ \rho_{12}^{-1} & \rho_{22}^{-1} & \rho_{23}^{-1} \\ \rho_{13}^{-1} & \rho_{23}^{-1} & \rho_{33}^{-1}\end{bmatrix}$ is the inverse matrix of $\boldsymbol{\Sigma}$. Using similar mathematical operations, we have

$$\frac{\partial^2 C\left(u_1, u_2, u_3\right)}{\partial u_1 \partial u_2}=\frac{1}{\varphi\left(b_1\right)\varphi\left(b_2\right)}\int_{-\infty}^{b_3} g\left(b_1, b_2, z\right)\mathrm{d}z=\Phi\left(\sqrt{\rho_{33}^{-1}}b_3+\frac{\rho_{13}^{-1}b_1+\rho_{23}^{-1}b_2}{\sqrt{\rho_{33}^{-1}}}\right)\cdot$$
$$\frac{\exp\left[-\frac{1}{2}\left(\rho_{11}^{-1}b_1^2+2\rho_{12}^{-1}b_1 b_2+\rho_{22}^{-1}b_2^2\right)+\frac{1}{2}\left(\frac{\rho_{13}^{-1}b_1+\rho_{23}^{-1}b_2}{\sqrt{\rho_{33}^{-1}}}\right)^2+\frac{1}{2}\left(b_1^2+b_2^2\right)\right]}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}\left(\rho_{33}^{-1}\right)^{\frac{1}{2}}} \tag{4}$$

Since $\frac{\partial^2 C\left(u_1, u_2\right)}{\partial u_1 \partial u_2}=c\left(u_1, u_2\right)$ can be obtained from the probability density function (PDF) of Eq. (1), one can get u_3 by solving the nonlinear equation $\frac{\frac{\partial^2 C\left(u_1, u_2, u_3\right)}{\partial u_1 \partial u_2}}{\frac{\partial^2 C\left(u_1, u_2\right)}{\partial u_1 \partial u_2}}=v_3$.

3.3 Student t copula simulation

(1) Simulate independent uniformly distributed random variables v_1, v_2, and v_3.

(2) Set $u_1 = v_1$.

(3) Set $F(u_2|u_1) = v_2$.

Let $g(b_1, y) = \dfrac{\Gamma\left(\dfrac{\upsilon+2}{2}\right)}{\Gamma\left(\dfrac{\upsilon}{2}\right)} \dfrac{1}{(\pi\upsilon)|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \left(1 + \dfrac{\boldsymbol{w}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{w}}{\upsilon}\right)^{-\frac{\upsilon+2}{2}}$, $b_1 = T_\upsilon^{-1}(u_1)$, $b_2 = T_\upsilon^{-1}(u_2)$, $f_\upsilon(t) = \dfrac{\Gamma\left(\dfrac{\upsilon+1}{2}\right)}{\Gamma\left(\dfrac{\upsilon}{2}\right)} \dfrac{1}{(\pi\upsilon)^{\frac{1}{2}}} \left(1 + \dfrac{t^2}{\upsilon}\right)^{-\frac{\upsilon+1}{2}}$, and $\boldsymbol{w} = (b_1, y)^{\mathrm{T}}$. Then, we have $F(u_2|u_1) =$

$\dfrac{\partial C(u_1, u_2)}{\partial u_1} = \dfrac{1}{f_\upsilon(b_1)} \int_{-\infty}^{b_2} g(b_1, y)\mathrm{d}y = T_{\upsilon+1}\left(\dfrac{b_2 - \rho b_1}{\sqrt{\dfrac{\upsilon + b_1^2}{\upsilon+1}(1-\rho^2)}}\right)$, which yields

$$u_2 = T_\upsilon(b_2) = T_\upsilon\left[T_{\upsilon+1}^{-1}(v_2)\sqrt{\frac{\upsilon + b_1^2}{\upsilon+1}(1-\rho^2)} + \rho b_1\right] \tag{5}$$

(4) Set $F(u_3|u_1u_2) = \dfrac{\dfrac{\partial^2 C(u_1, u_2, u_3)}{\partial u_1 \partial u_2}}{\dfrac{\partial^2 C(u_1, u_2)}{\partial u_1 \partial u_2}} = v_3$.

Let $g(b_1, b_2, z) = \dfrac{\Gamma\left(\dfrac{\upsilon+3}{2}\right)}{\Gamma\left(\dfrac{\upsilon}{2}\right)} \dfrac{1}{(\pi\upsilon)^{\frac{3}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \left(1 + \dfrac{\boldsymbol{w}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{w}}{\upsilon}\right)^{-\frac{\upsilon+3}{2}}$, where $\boldsymbol{w} = [b_1 \quad b_2 \quad z]^{\mathrm{T}}$, $b_3 = T_\upsilon^{-1}(u_3)$, and $\boldsymbol{\Sigma}^{-1} = \begin{bmatrix} \rho_{11}^{-1} & \rho_{12}^{-1} & \rho_{13}^{-1} \\ \rho_{12}^{-1} & \rho_{22}^{-1} & \rho_{23}^{-1} \\ \rho_{13}^{-1} & \rho_{23}^{-1} & \rho_{33}^{-1} \end{bmatrix}$ is the inverse matrix of $\boldsymbol{\Sigma}$. Using similar mathematical operations, we have

$$\frac{\partial^2 C(u_1, u_2, u_3)}{\partial u_1 \partial u_2} = \frac{1}{f_\upsilon(b_1)\, f_\upsilon(b_2)} \int_{-\infty}^{b_3} g(b_1, b_2, z)\mathrm{d}z =$$

$$\frac{1}{f_\upsilon(b_1)\, f_\upsilon(b_2)} \frac{\sqrt{\upsilon+2}}{\pi|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \left[\rho_{33}^{-1}(\upsilon+2)\right]^{-\frac{\upsilon+3}{2}} \delta^{-\upsilon-2} \upsilon^{\frac{\upsilon}{2}} \frac{\Gamma\left(\dfrac{\upsilon+2}{2}\right)}{\Gamma\left(\dfrac{\upsilon}{2}\right)} T_{\upsilon+2}\left(\frac{b_3 + \mu}{\delta}\right) \tag{6}$$

where $\mu = -\dfrac{\rho_{13}^{-1}b_1 + \rho_{23}^{-1}b_2}{\rho_{33}^{-1}}$ and $\delta = \sqrt{\dfrac{\upsilon + \rho_{11}^{-1}b_1^2 + 2\rho_{12}^{-1}b_1b_2 + \rho_{22}^{-1}b_2^2}{\rho_{33}^{-1}} - \left(\dfrac{\rho_{13}^{-1}b_1 + \rho_{23}^{-1}b_2}{\rho_{33}^{-1}}\right)}$. Since $\dfrac{\partial^2 C(u_1, u_2)}{\partial u_1 \partial u_2} = c(u_1, u_2)$ can be obtained from the PDF of Eq. (2), one can get u_3 by solving the nonlinear equation $\dfrac{\dfrac{\partial^2 C(u_1,u_2,u_3)}{\partial u_1 \partial u_2}}{\dfrac{\partial^2 C(u_1,u_2)}{\partial u_1 \partial u_2}} = v_3$.

4 Case study

4.1 Data

The historical drought data from the Lintong Gauge Station in the Weihe Basin, China, were used to illustrate this proposed approach for goodness-of-fit tests of trivariate copulas. Monthly precipitations covering a period from 1959 to 2008 were used to define droughts based on the theory of runs. All the data were obtained from the National Climate Center of the China Meteorological Administration and are complete data. Using the Mann-Kendall method, the data do not show obvious trends and can be accepted as temporally homogeneous. As illustrated in Fig. 1 (where *t* is time, X_t is the observed precipitation time series, and X_0 is a given threshold), a drought event is defined as a period when precipitation is equal to or less than the predetermined threshold. Drought characteristics, i.e., duration (*D*), severity (*S*), and peak (*P*) were extracted for each drought event using the averages of monthly precipitation as truncation levels, and some basic statistics of these three components are shown in Table 1. The correlation coefficients of Pearson's r_n, Spearman's ρ_n, and Kendall's τ_n given in Table 2 show that the observed drought duration, severity, and peak are highly correlated with one another, with a maximum correlation coefficient exceeding 0.9. The results were confirmed by the Chi-plots described in Fig. 2 (for a thorough review and more details about Chi-plots, see Fisher and Switzer (1985, 2001), Ma et al. (2012), and references therein). Most of the empirical points fall outside the confidence band ($\alpha = 0.05$) in the Chi-plots, which indicates that apparent dependent relationships exist among drought duration, severity, and peak. While significantly positive dependent relationships between bivariate drought variables are revealed both by the results of the correlation coefficients and Chi-plots, the degree of dependence between the drought duration and severity is larger than that between the drought duration and peak, and is less than that between the drought severity

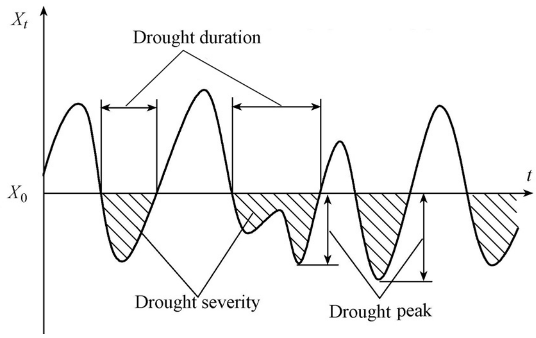

Fig. 1 Definition of drought using theory of runs

Table 1 Basic statistics of drought variables

Statistic parameter	D (month)	S (mm)	P (mm)	Statistic parameter	D	S	P
Mean	2.4	49.6	30.4	Coefficient of variation	0.73	0.97	0.71
Standard deviation	1.8	48.0	21.5	Coefficient of skewness	1.53	1.99	0.42
Minimum	1.0	0.2	0.2				
Maximum	9.0	271.0	80.2				

Table 2 Correlation coefficients of drought variables

Variables	Pearson's r_n	Spearman's ρ_n	Kendall's τ_n
D, S	0.659 0	0.562 5	0.436 0
D, P	0.294 3	0.272 8	0.212 1
S, P	0.795 9	0.919 1	0.772 4

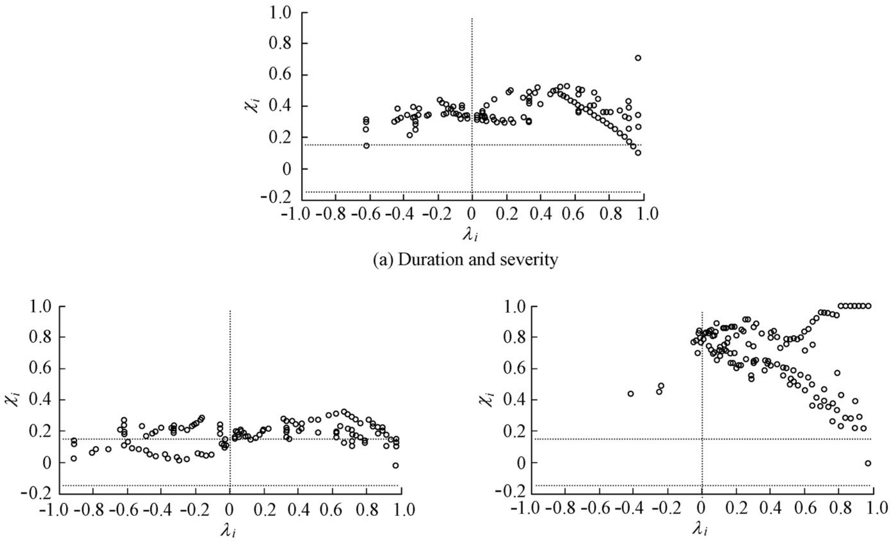

Fig. 2 Chi-plots for drought duration, severity, and peak

and peak. However, distributions of the points in the Chi-plots also indicate different dependence structures of drought components: for duration-severity and duration-peak they are similar (almost symmetric), but they are strictly distinct (extremely asymmetric) for severity-peak.

Assuming that the drought duration, severity, and peak are continuous variables, a variety of univariate cumulative distribution functions (CDFs) were used to fit the observed drought data first. Two criteria (AIC and RMSE) and various goodness-of-fit techniques (the Chi-square, Kolmogorov-Smirnov, Cramer-von Mises, Anderson-Darling, and modified weighted Durbin-Watson tests) were adopted to select margins. The exponential distribution, Weibull distribution, and generalized Pareto distribution, respectively, were eventually chosen as the optimal marginal distributions for drought duration, severity, and peak. The maximum likelihood (ML) method was applied to estimate parameters of the exponential distribution for the drought duration, while parameters of the Weibull distribution for the drought severity and the generalized Pareto distributions for the drought peak were estimated using the probability weight-moment method (PWM). Dependence structures of drought duration, severity, and peak were then modeled with the trivariate Gaussian and Student t copulas to obtain their multivariate joint distribution. Parameters of the Gaussian and Student t copulas were computed using the maximum pseudo-likelihood estimation method (Nadarajah 2006; Song and Singh 2010) and are shown in Table 3.

Table 3 Parameters of Gaussian and Student t copulas

Copula	ρ_{12}	ρ_{13}	ρ_{23}
Gaussian	0.608 2	0.304 2	0.904 2
Student t	0.425 6	0.161 1	0.916 2

Note: In this case, the degree of freedom of Student t copula is $\upsilon = 6.624\ 2$.

4.2 Results and discussion

According to the procedures described in Section 2.3, the Kolmogorov-Smirnov and Anderson-Darling statistics of the Gaussian and Student t copulas were numerically computed and are shown in Tables 4 and 5, respectively. Given the significance level $\alpha = 0.05$, it was found that all test statistics based on the observed drought duration, severity, and peak were less than the corresponding critical values, which indicates that neither Gaussian copula nor Student t copula can be rejected at the significance level $\alpha = 0.05$. In other words, the null hypothesis H_0^* as well as H_0 is accepted, i.e., both of the Gaussian and Student t copulas are acceptable for describing the dependence structures of the drought duration, severity, and peak as well as for modeling their trivariate joint probability distribution.

Table 4 Critical values of D_n for Gaussian and Student t copulas

Copula	D_n	Critical values at various significance levels α				
		0.20	0.15	0.10	0.05	0.01
Gaussian	0.109 4	0.111 1	0.118 5	0.127 9	0.141 4	0.167 3
Student t	0.077 9	0.129 5	0.137 0	0.146 8	0.164 2	0.196 3

Table 5 Critical values of A_n^2 for Gaussian and Student t copulas

Copula	A_n^2	Critical values at various significance levels α				
		0.20	0.15	0.10	0.05	0.01
Gaussian	2.000 4	2.781 1	3.311 8	4.032 8	5.203 1	7.762 0
Student t	1.007 8	3.880 9	4.492 8	5.325 2	6.781 6	10.334 8

Throughout the limited current applications of copula-based methods to multivariate drought issues, Archimedean copulas (many of which are, generally, valid for roughly identical and symmetric dependence structures among the considered multi-variables) seem to have been most commonly used (Ma et al. 2012). Nevertheless, in reality, chances are that most of the multi-contributing variables in hydrological or meteorological processes (e.g., rainfall, floods, and especially droughts) possess various dependence structures and degrees of associations, which are asymmetric and unbalanced. For instance, the markedly heterogeneous dependences of drought duration, severity, and peak reflected in the Chi-plots (Fig. 2) are better modeled by a selected meta-elliptical family of copulas. The fitting efficiencies of trivariate Gaussian and Student t copulas are shown in Fig. 3, which can be naturally confirmed by the results of goodness-of-fit tests, and this indicates that the Gaussian and Student t copulas both produce a satisfactory representation of the historical drought observations. Thus, the dependence structures of drought duration, severity, and peak can be readily modeled using the Gaussian and Student t copulas in order to obtain corresponding multivariate characteristics (such as joint probabilities and return periods) of drought events. These potential messages are useful and essential for drought risk management as well as for practical design and planning; since the drought duration, severity, and peak can be considered in total, it is possible to obtain various combinations of different drought components for several purposes in hydrological practices.

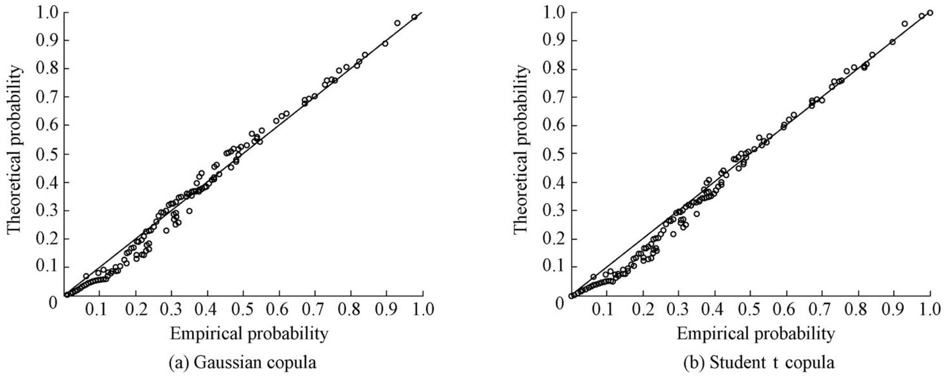

Fig. 3 Comparison of multivariate empirical and theoretical distributions

5 Conclusions

Rosenblatt's transformation can be applied to copulas in order to propose a test of fit for them and this technique of goodness-of-fit testing can in principle be used for every parametric family of copulas. Mathematical foundations of the goodness-of-fit test for trivariate copulas and corresponding procedures of a bootstrap approach were provided. Using the Gaussian and Student t copulas as an example, we demonstrate through copula simulation that the observed historical drought data at the Lintong Gauge Station with a trivariate meta-elliptical copula are acceptable at certain significance levels. As copulas are increasingly used to describe dependences of correlated random variables, the methodologies of goodness-of-fit testing for multi-dimensional copulas can provide strong support and help a lot in the further applications of a wide variety of copulas as useful tools for exploring the dependency relationships and subsequent multivariate joint probability distributions of non-independent hydrological variables with different dependence structures and degrees of associations.

Although, in theory, the methods of goodness-of-fit tests for trivariate copulas described in this paper could be extended to have higher numbers of dimensions, more complicated computational efforts are surely required. Besides, as we pointed out in the beginning, the existing framework and methods remain ineffective for non-parametric families of copulas (whereas there are many of them in potential applications); and exploration of analytical formulas and estimation of parameters for multi-dimensional copulas can also be better addressed with more efforts in the future.

References

Berg, D., and Bakken, H. 2005. *A Goodness-of-fit Test for Copulae based on the Probability Integral Transform*. Oslo: Department of Mathematics, University of Oslo.

Berkowitz, J. 2001. Testing density forecasts, with applications to risk management. *Journal of Business and Economic Statistics*, 19(4), 465-474. [doi:10.1198/07350010152596718]

Breymann, W., Dias, A., and Embrechts, P. 2003. Dependence structures for multivariate high-frequency data in finance. *Quantitative Finance*, 3(1), 1-14. [doi:10.1080/713666155]

Chowdhary, H., Escobar, L. A., and Singh, V. P. 2011. Identification of suitable copulas for bivariate frequency analysis of flood peak and flood volume data. *Hydrology Research*, 42(2-3), 193-216. [doi:10.2166/nh.2011.065]

Demarta, S., and McNeil, A. J. 2005. The t copula and related copulas. *International Statistical Review*, 73(1), 111-129. [doi:10.1111/j.1751-5823.2005.tb00254.x]

Diebold, F. X., Gunther, T. A., and Tay, A. S. 1998. Evaluating density forecasts with applications to financial risk management. *International Economic Review*, 39(4), 863-883. [doi:10.2307/2527342]

Diebold, F. X., Hahn, J., and Tay, A. S. 1999. Multivariate density forecast and calibration in financial risk management: High-frequency returns on foreign exchange. *The Review of Economics and Statistics*, 81(4), 661-673. [doi:10.1162/003465399558526]

Dobrić, J., and Schmid, F. 2005. Testing goodness of fit for parametric families of copulas: Application to financial data. *Communications in Statistics — Simulation and Computation*, 34(4), 1053-1068. [doi: 10.1080/03610910500308685]

Dobrić, J., and Schmid, F. 2007. A goodness of fit test for copulas based on Rosenblatt's transformation. *Computational Statistics and Data Analysis*, 51(9), 4633-4642. [doi:10.1016/j.csda.2006.08.012]

Evin, G., and Favre, A. C. 2008. A new rainfall model based on the Neyman-Scott process using cubic copulas. *Water Resources Research*, 44, W03433. [doi:10.1029/2007WR006054]

Fang, H. B., Fang, K. T., and Kotz, S. 2002. The meta-elliptical distributions with given marginals. *Journal of Multivariate Analysis*, 82(1), 1-16. [doi:10.1006/jmva.2001.2017]

Fermanian, J. D. 2005. Goodness of fit tests for copulas. *Journal of Multivariate Analysis*, 95(1), 119-152. [doi:10.1016/j.jmva.2004.07.004]

Fisher, N. I., and Switzer, P. 1985. Chi-plots for assessing dependence. *Biometrika*, 72(2), 253-265.

Fisher, N. I., and Switzer, P. 2001. Graphical assessment of dependence: Is a picture worth 100 tests? *American Statistician*, 55(3), 233-239.

Frees, E. W., Carriere, J., and Valdez, E. 1996. Annuity valuation with dependent mortality. *The Journal of Risk and Insurance*, 63(2), 229-261.

Genest, C., and Rivest, L. P. 1993. Statistical inference procedures for bivariate Archimedean copulas. *Journal of the American Statistical Association*, 88(423), 1034-1043. [doi:10.1080/01621459.1993.10476372]

Genest, C., Quessy, J. F., and Rémillard, B. 2006. Goodness-of-fit procedures for copula models based on the probability integral transformation. *Scandinavian Journal of Statistics*, 33(2), 337-366. [doi:10.1111/j.1467-9469.2006.00470.x]

Genest, C., Rémillard, B., and Beaudoin, D. 2009. Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics*, 44(2), 199-213. [doi:10.1016/j.insmatheco.2007.10.005]

Grimaldi, S., and Serinaldi, F. 2006. Asymmetric copula in multivariate flood frequency analysis. *Advances in Water Resources*, 29(8), 1155-1167. [doi:10.1016/j.advwatres.2005.09.005]

Junker, M., and May, A. 2005. Measurement of aggregate risk with copulas. *Econometrics Journal*, 8(3), 428-454. [doi:10.1111/j.1368-423X.2005.00173.x]

Ma, M. W., and Song, S. B. 2010. Elliptical copulas for drought characteristics analysis of Xi'an gauging station. *Journal of China Hydrology*, 30(4), 36-42. (in Chinese)

Ma, M. W., Song, S. B., Ren, L. L., Jiang, S. H., and Song, J. L. 2012. Multivariate drought characteristics using trivariate Gaussian and Student t copulas. *Hydrological Processes*, published online at http://onlinelibrary.wiley.com/doi/10.1002/hyp.8432/abstract on April 17, 2012 [doi:10.1002/hyp.8432]

Malevergne, Y., and Sornette, D. 2003. Testing the Gaussian copula hypothesis for financial asset dependences. *Quantitative Finance*, 3(4), 231-250. [doi:10.1088/1469-7688/3/4/301]

Mendes, B. V. M., and Souza, R. M. 2004. Measuring financial risks with copulas. *International Review of Financial Analysis*, 13(1), 27-45. [doi:10.1016/j.irfa.2004.01.007]

Nadarajah, S. 2006. Fisher information for the elliptically symmetric Pearson distributions. *Applied Mathematics and Computation*, 178(2), 195-206. [doi:10.1016/j.amc.2005.11.037]

Nelsen, R. B. 1999. *An Introduction to Copulas*. New York: Springer.

Panchenko, V. 2005. Goodness-of-fit test for copulas. *Physica A*, 355, 176-182. [doi:10.1016/j.physa.2005.02.081]

Rosenblatt, M. 1952. Remarks on a multivariate transformation. *The Annals of Mathematical Statistics*, 23(3), 470-472.

Shiau, J. T. 2006. Fitting drought duration and severity with two-dimensional copulas. *Water Resources Management*, 20(5), 795-815. [doi:10.1007/s11269-005-9008-9]

Sklar, A. 1959. Distribution functions of *n* dimensions and margins. *Publications of the Institute of Statistics of the University of Paris*, 8, 229-231. (in French)

Song, S. B., and Singh, V. P. 2010. Meta-elliptical copulas for drought frequency analysis of periodic hydrologic data. *Stochastic Environmental Research and Risk Assessment*, 24(3), 425-444. [doi:10.1007/s00477-009-0331-1]

Wang, X. J., Gebremichael, M., and Yan, J. 2010. Weighted likelihood copula modeling of extreme rainfall events in Connecticut. *Journal of Hydrology*, 390(1-2), 108-115. [doi:10.1016/j.jhydrol.2010.06.039]

Žežula, I. 2009. On multivariate Gaussian copulas. *Journal of Statistical Planning and Inference*, 139(11), 3942-3946. [doi:10.1016/j.jspi.2009.05.039]

Zhang, L., and Singh, V. P. 2007. Trivariate flood frequency analysis using the Gumbel-Hougaard copula. *Journal of Hydrologic Engineering*, 12(4), 431-439. [doi:10.1061/(ASCE)1084-0699(2007)12:4(431)]

Zhang, Q., Chen, Y. Q., Chen, X. H., and Li, J. F. 2011. Copula-based analysis of hydrological extremes and implications of hydrological behaviors in the Pearl River basin, China. *Journal of Hydrologic Engineering*, 16(7), 598-607. [doi:10.1061/(ASCE)HE.1943-5584.0000350]

Zhang, Q., Li, J. F., and Singh, V. P. 2012. Application of Archimedean copulas in the analysis of the precipitation extremes: Effects of precipitation changes. *Theoretical and Applied Climatology*, 107(1-2), 255-264. [doi:10.1007/s00704-011-0476-y]

(Edited by Yun-li YU)