CrossMark

Multivariate Drought Frequency Analysis using Four-Variate Symmetric and Asymmetric Archimedean Copula Functions

Olusola O. Ayantobo [1,2,3] · Yi Li [1,2] · Songbai Song [1]

Abstract

In drought frequency analysis, as the number of drought variables increases, the joint behavior between these variables needs to be studied. Therefore, this study aims to develop a flexible four-variate joint distribution function of the regional stochastic nature of drought. Using run theory, drought duration, severity, peak, and inter-arrival time were abstracted from the Standardized Precipitation Evapotranspiration Index (SPEI) aggregated at six months, observed in mainland China between 1961 and 2013. As these drought variables showed significant dependence properties and followed different marginal distributions, we employed and compared six four-variate symmetric and asymmetric Archimedean copulas (i.e., Frank, Clayton, Gumbel–Hougaard). The best-fitting model for each region was carefully selected using RMSE, AIC, and BIAS goodness-of-fit tests. Results revealed that the empirical and theoretical probabilities of the symmetric Clayton in regions NE (Northeast), CS (Central and Southern China), EMC (Entire China), and symmetric Frank in regions NC (North China), SC (South China), IM (Inner Mongolia), NW (Northwest), TP (Tibet Plateau) agreed well. Symmetric Frank copula was considered the best-fit for station-based drought analysis in EMC. Based on these copulas, the drought probabilities and return periods for the occurrence of drought events over the next 5, 10, 20, 50, and 100 years in each region were hereby comprehensively explained, and the results shown here could be helpful in the appraisal of the adequacies of water supply systems under drought conditions in all regions. This study showed that a four-variate copula approach is a vital tool for probabilistic interpretation of hydrological and meteorological data in the different climatic region of mainland China.

Keywords Symmetric and asymmetric copula · Four-variate drought analysis · Probabilistic analysis · Return period · Water management · SPEI

✉ Yi Li
liyikitty@126.com

Olusola O. Ayantobo
skollar24k@yahoo.com; ayantobooo@funaab.edu.ng

Extended author information available on the last page of the article

1 Introduction

Drought-affected areas have remarkably increased over the past 50 years in China as a result of variation in precipitation (P) and temperature (Yang et al. 2013; Leng et al. 2015; Chang et al. 2016). In fact, a noticeable extreme winter-spring droughts events occurred in southwest China from August 2009 to May 2010 (Lu et al. 2011; Lu et al. 2012; Zhang et al. 2012; Yang et al. 2013), while in 2011 the middle and lower reaches of the Yellow River were impacted by the spring-summer drought (Lu et al. 2013; Zhao et al. 2013; Chang et al. 2016). Droughts have caused greatest damages in China during 1949–1995, and many damages have resulted to economic losses more than the US $12 billion (Xu et al., 2015a; Qin et al. 2015; Chang et al. 2016). Consequently, droughts are of prominent concern in the outlining and control of water resources. Studying meteorological droughts is necessary by itself, and also because they act as antecedents to the longer-lasting and more significant agricultural and hydrological droughts (Haslinger et al., 2014, Wilhite et al., 2014).

The popular method of assessing drought risk requires the calculation of the probability that a specific value of drought variable will be exceeded, which is equal to the evaluation of the recurrence interval (Prohaska et al. 2008; Chen et al. 2012). This procedure is normally focused on the univariate frequency analysis. In this analysis, the authors have concentrated on some characteristic of drought and assume that the other variables are constant. A direct attempt to simplify the multi-dimensional character of droughts has been investigated by several authors (Shiau, 2006; Ganguli and Reddy, 2014; Huang et al., 2014a). Although the simplistic analysis produces practical results, it does not strictly represent the actual mathematical interactions between the various dimensions of the phenomenon and obviously cannot give accurate results. Since drought variables are usually reliant on one other, a proper frequency analysis of droughts should consider such dependencies within a suitable multivariate framework.

It should be mentioned that the attention of researchers in multivariate modeling has increased considerably in the last decades, due to the application of copulas. The comprehensive theoretical structure proposed by Sklar (1959) can be gathered from Nelsen (2006), Joe (1997) and Salvadori et al. (2007). There have been diverse cases of copula applications in the setting of drought management (Serinaldi et al., 2009; Shiau and Modarres, 2009; Kao and Govindaraju, 2010; Wong et al., 2010; Song and Singh, 2010; Mishra and Singh, 2010; Mirabbasi et al., 2012). This includes return-period estimation (Salvadori and De Michele, 2010), multivariate simulation (AghaKouchak et al. 2010), propose a new drought indicator (Kao and Govindaraju, 2010), and many other theoretical studies of multivariate severe issues (Salvadori and De Michele, 2015; Zhang et al. 2015a; She et al. 2016). There are several types of copula functions, which have been described (e.g. Nelsen, 1999). The regularly applied copula for hydrological analysis belongs to four classes which includes: Archimedean class (AMH, Frank, Clayton, Gumbel, and Joe), elliptical class (student t and normal), extreme value class (Gumbel, Galambos, Tawn, Husler-Reiss, and t-EV), and miscellaneous class (Farlie–Gumbel–Morgenstern and Plackett) (Shiau and Modarres, 2009; Mirabbasi et al. 2012; Lee et al. 2013; Xu et al., 2015b). The Archimedean class of copulas are characterized by their simple structure and strong representativeness (Huang et al., 2014a; Tsakiris et al., 2016).

There are many applications of two-variate copula-based drought studies in many countries. For instance in China, Shiau (2006) built a joint distribution between drought severity and duration in Southern Taiwan while Shiau and Modarres (2009) employed Clayton copula to

model drought severity and duration frequency curves for two climatic regions in Iran. Tsakiris et al. (2016) analyzed drought severity and areal extent utilizing Gumbel-Hougaard copula in Greece. The Frank and Gumbel–Hougaard have been considered the best copulas for illustrating drought occurrence frequency in Canada and Iran, respectively (Lee et al. 2013). Meta-Gaussian copula was applied for the joint modeling of drought variables within Texas (Song and Singh, 2010) while drought duration and severity of Sharafkhaneh in the northwest of Iran was modeled using Galambos copula (Mirakbari et al., 2010). In each of the cases, some bivariate probabilistic properties of droughts were investigated.

Three-variate Archimedean copulas have been employed for the joint distribution of rainfall, drought and flood events (Grimaldi and Serinaldi, 2006; Zhang and Singh 2007). Using asymmetric Archimedean copulas, Serinaldi and Grimaldi (2007) described an inference procedure to carry out a three-variate frequency analysis. Ma et al. (2013) created a joint distribution of drought duration, severity, and peak using elliptical, asymmetric and symmetric Archimedean copulas. Because of the flexibility of meta-elliptical copula (Fan et al. 2016), many meta-elliptical, Clayton, Frank, Gumbel-Hougaard, and Ali-MikhailHaq copulas have been employed to create a joint distribution of drought severity, duration, and interval time, and the best copula was chosen (Song and Singh, 2010).Three-variate Plackett copulas have been adopted for the investigations of extreme rainfall cases (Kao and Govindaraju, 2008) and for modeling a joint distribution of drought duration, severity and inter-arrival time (Song and Singh, 2010). Student's t copula has been utilized to identify three-variate drought events under the control of La-Nina, EL-Nino, and natural climate state in New South Wales, Australia (Wong et al. 2010). Other studies on drought frequency analysis using copula includes Huang et al. (2014a, b), Zhao et al. (2015), Zhang et al. (2015b), Zhang et al. (2017) among others.

Crucial hurdles remain not only in drought characterization but also in the interpretation of drought variables to information relevant for regional monitoring, early warning, and water resources planning. Assuming that the territorial unit affected by drought is the entire region, there are two variables which together characterize each drought event; the severity and the duration. If the peak and interarrival time of drought are considered as an additional variable, a four-dimensional copulas approach may be used in which duration, severity, peak and interarrival time of droughts can be jointly analyzed. Recently, the applications of four-variate copula have been reported. De Michele et al. (2007) proposed a procedure towards building four-variate distributions, stating two copulas for a separate two-variate candidate case and utilizing the approach to give a four-variate nature of sea state.

Employing a four-variate student copula, Serinaldi et al. (2009) built the corresponding joint distributions of drought length, mean, minimum SPI values, and drought mean areal extent to investigate their joint probabilistic characteristics. So far, most of the research focused on parametric two-variate and symmetric three-variate copulas (Ganguli and Reddy, 2014), but four-variate copulas are somewhat limited because their constructions are very complicated. Although there have been some attempts to construct multivariate extensions of two-variate Archimedean copula (Embrechts et al., 2003; Whelan, 2004; Savu and Trede, 2010), however such studies seldom compares the performances of four-variate symmetric and asymmetric copulas in different climatic regions of China, considering probability and return periods of drought events. Further, knowing that droughts are multiplex phenomena, two- and three-variate analysis cannot give an exhaustive appraisal of droughts because the investigation could lead to a deficient drought probabilistic estimation (De Michele et al. 2005; Grimaldi and

Serinaldi, 2006; Chebana and Ouarada, 2011; Ganguli and Reddy, 2014; Xu et al., 2015a; Tsakiris et al. 2016).

Given the challenges as mentioned above in drought characterization, a focused research is required to comprehensively analyze and delineate drought trends and spatiotemporal patterns more flexibly and intuitively. It is, therefore, the objective of this paper to produce a methodology using the four-variate copulas approach for analyzing the multivariate drought frequency of droughts. In particular, we built a joint distribution utilizing drought duration (Dd), severity (Ds), peak (Dp) and inter-arrival time (Di), based on four-variate symmetric (Clayton, Frank, and Gumbel) and asymmetric (Clayton, Frank, and Gumbel) Archimedean Copulas. We showed how it could be used (i) to observe the temporal evolution of a drought event, and (ii) to perform a real-time appraisal of drought in seven climatic regions of mainland China over 1961–2013. The risks of drought events have been described based on joint probability and return periods, which has become a conventional method for a risk-based plan of water resources structures.

The paper is prepared as follows. Section 2 describes details about the study area and data, Standardized Precipitation Evapotranspiration Index (SPEI) calculation and estimation of Dd, Ds, Dp, and Di. Section 3 introduces the four-variate symmetric and asymmetric Archimedean copula. Section 4 compares the performance of the copulas. Section 5 selects the best-fitted copula to build a joint distribution of Dd, Ds, Dp, and Di in seven climatic regions of China. Section 6 contains some concluding remarks.

2 Study Region and Data

Geographically, China lies between 15^{0}N - 60^{0}N and 75°E - 135°E. By its geographical extent, China is endowed with diverse landforms that include hills, mountains, high plateaus and deserts in the western reaches, while in the central and east areas, the land slopes into broad plains and deltas. From the higher elevations in the west, thousands of rivers drain the country eastwards; the most significant are the Yangtze (i.e., the longest river in China and third longest in the world), Heilong (Amur), Mekong, Pearl, and Yellow river. The climate of China fluctuates significantly from one region to another because of its tremendous space, complex topography as well as variation in monthly P and temperature (Zhai *et al.* 2010; Wu *et al.* 2011). The northern China and western are influenced by dry climate while humid and semi-humid predominates over eastern China (Wu *et al.* 2011).

Based on topography and climate, China is divided into seven climatic regions (Zhao, 1983; Ayantobo *et al.* 2017). Northeast humid/semi-humid warm region (NE, 72 stations), North China humid/semi-humid temperate zone (NC, 104 stations), Central and Southern China humid subtropical zone (CS, 165 stations), South China humid tropical zone (SC, 57 stations), Inner Mongolia steppe zone (IM, 44 stations), Northwest desert areas (NW, 61 stations), and Qinghai-Tibet Plateau (TP, 49 stations) (Fig. 1). Regions NE, NC, CS, SC are associated with Eastern Monsoon. NW is an arid region, and the replenishment of water in this region is mostly from melting glacial and perennial frozen soil, not from P.

The dataset employed in this research to appraise drought events are daily and monthly weather data from 552 national basic meteorological stations in mainland China from 1961 to 2013. These 53-year records were obtained from the China Meteorological Data Sharing Network. From Fig. 1, the meteorological stations are not uniformly spread as more are situated within the east than western regions particularly the Qinghai-Tibet Plateau. The
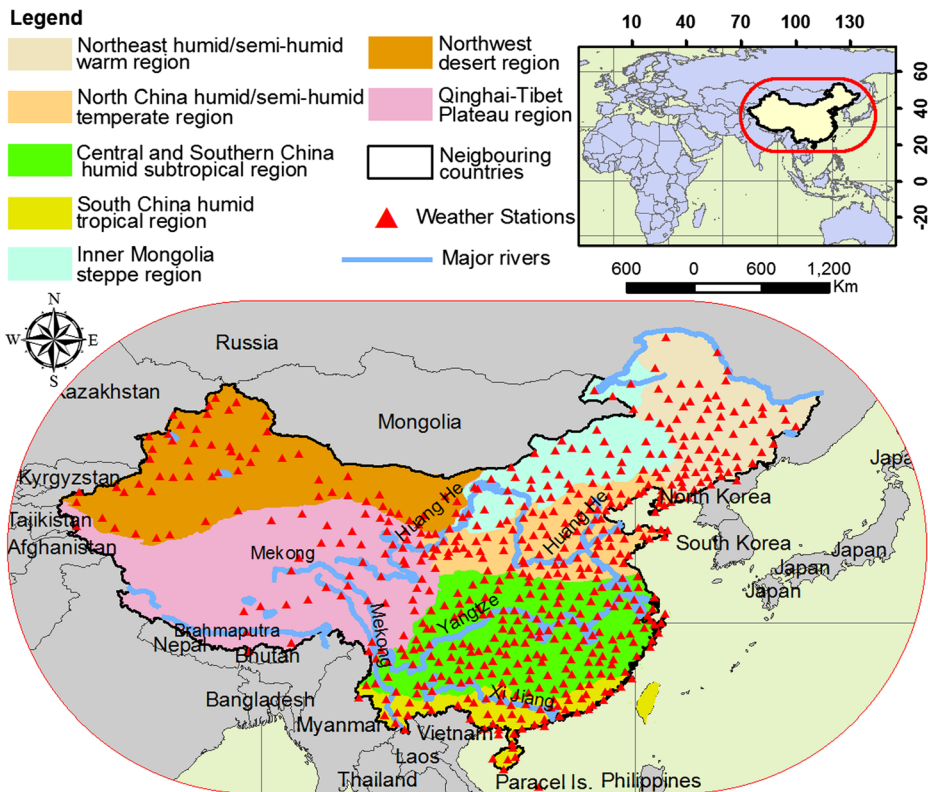
Fig. 1 Hydrometerological map of China showing the climatic regions, major rivers and national weather stations

non-parametric tests were used to cross-check the reliability and quality of the climatic data. According to Helsel and Hirsch (1992), Kendall autocorrelation test, Mann-Kendall trend test, and Mann-Whitney homogeneity tests for mean and variance were employed to test randomness, homogeneity, and absence of trends. The results showed that randomness and stationarity of the weather data were fixed between the critical points (5% statistical significance level).

3 Methods

3.1 Drought Indices and Drought Identification

Several indices have been suggested to identify and observe droughts, and there is a detailed article on modeling drought events utilizing these indices (Mishra and Singh, 2010; Zargar et al. 2011). Indices based exclusively on precipitation, do not consider the complexity of land surface processes and cannot directly account for the impacts of evapotranspiration on soil moisture. This may be a particularly serious disadvantage under warming conditions or other changes in regional hydrology. Based on these considerations, since SPEI have been widely applied in various copula-based drought-related studies in China (Liu et al. 2015; Xu et al.,

2015b; Wang et al. 2015; Zhang et al. 2015b; Hao et al. 2017), it is therefore employed for this study. The SPEI is a multi-scalar drought index, recommended as an enhancement of the SPI (Vicente-Serrano et al. 2010). Firstly, in SPEI estimation, the simple water balance is calculated, which is described as a deviation in P and potential evapotranspiration (PET) for i^{th} month as follows:

$$D_i = P_i - PET_i \tag{1}$$

PET is estimated according to the FAO-56 Penman-Monteith equation (Allen et al. 1998). The World Meteorological Organization commonly approves this approach, and the calculation is based on the minimum and maximum air temperature, wind speed, and sunshine duration. The D_i values are synthesized in each time scale as:

$$D_n^k = \sum_{i=0}^{k-1}(P_{n-1} - PET_{n-1}) \tag{2}$$

where k is the time scales of synthesis, and n is the month used for calculation.

Next, D_i is fitted with the three-parameter log-logistic distribution. Finally, the SPEI is obtained as standardized values, and details of the estimation can be found in Vicente-Serrano et al. (2010). The mean value of the SPEI is 0, and the standard deviation is 1. In this study, SPEI at the 6-month timescale, denoted by SPEI-6 is estimated for each station and seven regions. SPEI-6 is most useful for describing the shallow soil moisture available to crops (Reddy and Singh, 2014; Abdi et al. 2016).

Using the theory of runs (Fig. 2), drought identification is achieved by considering drought years as a continuous period where SPEI-6 is below 0 (Shiau, 2006). Next, significant drought variables which include Dd, Ds, Dp, and Di, are extracted. The definitions of Dd, Ds, and Dp, can be found in Shiau (2006), Mishra and Singh (2010) and Ayantobo et al. (2017) while Di is described as the interval between the commencement of a drought to the start of the subsequent drought (Song and Singh, 2010). Water manager needs to be aware of the risk of having drought scenarios of different Dd, Ds, Dp, and Di during drought scenarios. Therefore, a four-variate distribution must be constructed using copula functions.
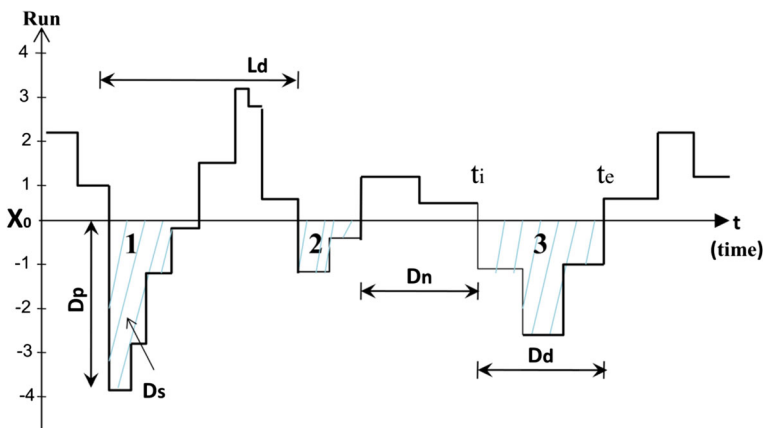

Fig. 2 Definition sketch of drought characteristics showing three drought events (labeled as 1, 2 and 3), on the basis of Run Theory. **Note:** X_O; Truncation level, Dd; Drought duration, Ds; Drought severity, Dp; Drought peak, Ld; Inter arrival time, Dn; Non-drought duration, t_i; initiation time, t_e; termination time

3.2 Marginal Distribution of Drought Variable

The univariate distribution forms the framework of four-variate analysis using copulas. The marginal distributions for Dd, Ds, and Dp for different climatic regions across mainland China have been documented (Ayantobo et al. 2017). In the case of Di, ten marginal distributions which include exponential, two-parameter exponential, Weibull, three-parameter Weibull, generalized extreme value, inversion Gaussian, three-parameter inversion Gaussian, generalized Pareto, gamma, and three-parameter gamma are compared to choose the best distribution to fit Di.

The Kolmogorov–Smirnov (KS) and Anderson-darling (AD) tests are employed as a goodness-of-fit test, and the principle that the threshold value should be as small as possible to preserve the largest sample is employed (Shiau and Modarres, 2009; Song and Singh, 2010; Abdi et al. 2016). The fitted distributions of Di are compared against the empirical non-exceedance probabilities estimated utilizing Gringorten's plotting position function as follows (Gringorten, 1963; Cunnane, 1978; Song and Singh, 2010):

$$\hat{F}(x_i) = \frac{1}{n+1} \sum_{m=1}^{i} N_m \tag{3}$$

Here, n is sample size, N_m is the amount of x_i regarded as $x_j \le x_i$, $i = 1, \ldots, n$, $1 \le j \le i$. Utilizing the observed data (x), the marginal parameters are calculated employing the maximum likelihood estimation.

3.3 Empirical Four-Variate Distribution of Drought Variables

The empirical four-variate joint probability distribution of Dd, Ds, Dp, and Di can be modified from Song and Singh (2010):

$$\hat{F}(x_i, y_i, z_i, k_i) = \frac{1}{n+1} \sum_{m=1}^{i} \sum_{n=1}^{i} \sum_{l=1}^{i} \sum_{p=1}^{i} N_{mnlp} \tag{4}$$

Here, n is size of the sample. N_{mnlp} is amount of (x_i, y_i, z_i, k_i) regarded as $x_j \le x_i$, $y_j \le y_i$, $z_j \le z_i$, and $k_j \le k_i$, $i = 1, \ldots, n$, $1 \le j \le i$.

3.4 Joint Cumulative Probability Distribution of Drought Variables

Assuming four correlated random variables of Dd, Ds, Dp, and Di are represented using X_1, X_2, X_3, and X_4; and $u_1 = F_{x_1}(x_1)$, $u_2 = F_{x_2}(x_2)$, $u_3 = F_{x_3}(x_3)$, and $u_4 = F_{x_4}(x_4)$ denote their cumulative distribution function (CDF), respectively. Therefore, their joint CDF could be represented as (Song and Singh, 2010):

$$F(x_1, x_2, x_3, x_4) = P(X_1 \le x_1, X_2 \le x_2, X_3 \le x_3, X_4 \le x_4) = \int_{-\infty}^{x_4} \int_{-\infty}^{x_3} \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} f(x, y, z) \mathrm{d}x \mathrm{d}y \mathrm{d}z \mathrm{d}k \tag{5}$$

3.5 Modeling Four-Variate Drought Variables Using Copulas

For an explicit hydrological application of copulas, readers are referred to Joe (1997), Nelsen (2006), Genest and Favre (2007) and Salvadori et al. (2007). By way of definition, copulas are functions that combines marginal CDF's to form multivariate CDF. This function allows one to

characterize dependence properties more flexibly by separating effects of dependence from effects of margins. Assuming $F_{1, 2, \ldots, n}(x_1, x_2, \ldots, x_n)$ is many variables CDF of n associated arbitrary variables of $X_1, X_2, \ldots, X_n$ with the corresponding candidate CDF $F_1(x_1), F_2(x_2), \ldots, F_n(x_n)$, next the n-dimensional CDF including univariate distributions of $F_1(x_1), F_2(x_2), \ldots, F_n(x_n)$ could be written as shown below (Sklar 1959; Nelsen, 2006; Ganguli and Reddy 2014):

$$H(x_1, x_2, \ldots x_n) = C\{F_1(x_1), F_2(x_2), \ldots, F_n(x_n)\} = C(u_1, \ldots \ldots u_n) \tag{6}$$

Here, C is a d-dimensional copula in the form: $[0, 1]$ d $\rightarrow [0, 1]$, with association parameter θ; $F_k(x_k) = u_k$ for $k = 1, \ldots, n$; $U_k \sim U(0, 1)$.

Archimedean copulas have been utilized because the functions can be conveniently built, and can capture a wide range of dependence structures with many satisfactory properties. According to Joe (1997), Nelsen (1999), Embrechts et al. (2003), Whelan (2004), and Grimaldi and Serinaldi (2006) the generalization of n-dimensional Archimedean copula could be represented as (Chen et al. 2012):

$$C(u_1, u_2, \ldots, u_n) = C_1(u_n, C_2(u_{n-1}, \ldots, C_{n-1}(u_2, u_1) \ldots))$$
$$= \varphi_1^{[-1]}\left(\varphi_1(u_n) + \varphi_1\left(\varphi_2^{[-1]}\left(\varphi_2(u_{n-1}) + \cdots + \varphi_{n-1}^{[-1]}(\varphi_{n-1}(u_2) + \varphi_{n-1}(u_1)) \ldots\right)\right)\right) \tag{7}$$

Here, φ is the copula generator. As observed, $\varphi_1, \ldots, \varphi_{n-1} = \varphi$, Eq. (7) becomes Archimedean n-copula and can be rewritten as (Chen et al. 2012):

$$C(\mathbf{u}) = \left[\varphi^{-1}\sum_{k=1}^{n}\varphi(u_k)\right] \tag{8}$$

Many different parametric families of copulas are commonly used in dependence analysis. The two most frequently used are elliptical copulas and the Archimedean copulas. In this study, we consider three families of four-variate symmetric and asymmetric Archimedean copulas, paying particular attention to Clayton, Frank, and Gumbel (Table 1), because they are simple and easy to generate (Nelsen, 2006; Chen et al. 2012). These copulas are compared to ascertain the excellent copula for modeling drought variables across various climatic regions of mainland China.

3.6 Copula Parameters Estimation

The parameters of Clayton, Frank and Gumbel copulas are calculated by employing the curve fitting method (CFM). The CFM is the minimum standards of squared residuals and determined as follows (Ganguli and Reddy, 2014):

$$CFM = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_c(i) - X_0(i))^2} \tag{9}$$

Where n, $x_c(i)$, $x_0(i)$ are sample size, i^{th} calculated value, i^{th} observed value respectively.

3.7 Selection of Appropriate Copula Function

The goodness-of-fit assessment indices (i.e., Root mean square error (RMSE), Akaike information criterion (AIC) and BIAS) are employed to select the best-fitted copula function (Akaike, 1974; Zhang and Singh, 2006, 2007; Genest and Favre, 2007).

Table 1 Description of the four-variate symmetric and asymmetric Archimedean copula

Copula Family	$C(u_1, u_2, u_3, u_4)$
CLAYTON (Symmetric)	$\left(u_1^{-\theta} + u_2^{-\theta} + u_3^{-\theta} + u_4^{-\theta} - 3\right)^{\frac{-1}{\theta}}$
CLAYTON (Asymmetric)	$\left(u_4^{-\theta_1} + \left(\left(u_1^{-\theta_3} + u_2^{-\theta_3} - 1\right)^{\frac{\theta_2}{\theta_3}} + u_3^{-\theta_2} - 1\right)^{\frac{\theta_2}{\theta_3}} - 1\right)^{\frac{\theta_2}{\theta_3}}$
FRANK (Symmetric)	$\frac{-1}{\theta_1}\ln\left(1 + \frac{-1}{\theta_1}\right)$
FRANK (Asymmetric)	$\frac{-1}{\theta_1}\ln\left(1 + \frac{-1}{\theta_1}\left(e^{-\theta_1 u_4} - 1\right)\cdot\left(\left(1 - \frac{-1}{\theta_1}\cdot\left(1 - \left(1 - \frac{-1}{\theta_1}\cdot\left(1 - e^{-\theta_3 u_1}\right)\left(1 - e^{-\theta_3 u_2}\right)\right)^{\frac{-1}{\theta_1}}\right)\cdot\left(1 - e^{-\theta_2 u_3}\right)\right)^{\frac{-1}{\theta_1}} - 1\right)\right)$
GUMBEL (Symmetric)	$\exp\left(-[(-\ln u_1)^{\theta} + (-\ln u_2)^{\theta} + (-\ln u_3)^{\theta} + (-\ln u_4)^{\theta}]^{\frac{1}{\theta}}\right)$
GUMBEL (Asymmetric)	$\exp\left\{-\left[(-\ln u_4)^{\theta_1} + \left(\left((-\ln u_1)^{\theta_3} + (-\ln u_2)^{\theta_3}\right)^{\frac{\theta_2}{\theta_3}} + (-\ln u_3)^{\theta_2}\right)^{\frac{\theta_2}{\theta_3}}\right]^{\frac{\theta_2}{\theta_3}}\right\}$

u_1, u_2, u_3, u_4 are marginal distributions of Dd, Ds, Dp and Di, respectively; θ is symmetric copula parameter such that $\theta \geq 1$; θ_1, θ_2, θ_3 are asymmetric copula parameters such that $1 \leq \theta_1 \leq \theta_2 \leq \theta_3$

$$\text{RMSE} = \sqrt{E\left[(x_c - x_0)^2\right]} = \left\{\frac{1}{n-k}\sum_{i=1}^{n}[x_c(i) - x_0(i)]^2\right\}^{0.5} \tag{10}$$

$$\text{BIAS} = \sum_{i=1}^{n}\frac{X_c(i) - X_0(i)}{X_c} \tag{11}$$

$$\text{AIC} = n\log(\text{MSE}) + 2k \tag{12}$$

Where $E\left[\bullet\right]$ = Mean square error (MSE) is the expectation operator; n, $x_c(i), x_0(i)$ are sample size, i^{th} calculated value, i^{th} observed value respectively and k is the number of parameters utilized in getting the computed value. The copula with least values of AIC and RMSE is selected as the best-fitted fuction.

Also, using the observed data (u), the empirical copula, $C_n(u)$, and fitted parametric copula, $C_p(u)$, are compared to select the best-fitted function. Constructing $C_n(u)$ requires: (1) ranking Dd, Ds, Dp and Di variables considering their pseudo-observations ($U_{i,1}$, $U_{i,2}$, $U_{i,3}$, $U_{i,4}$), (2) the empirical CDFs are constructed based on the ranks and (3) the empirical four-variate copula are calculated using the empirical CDFs (Nelsen, 2006; Genest and Favre, 2007):

$$C_n(u) = \frac{1}{n}\sum_{i=1}^{n}I\left\{\hat{U}_{i,1} \leq u_1, \hat{U}_{i,2} \leq u_2, \hat{U}_{i,3} \leq u_3, \hat{U}_{i,4} \leq u_4\right\}$$
$$= \frac{1}{n}\sum_{i=1}^{n}I\left\{\frac{R_{i,1}}{n+1} \leq u_1, \frac{R_{i,2}}{n+1} \leq u_2, \frac{R_{i,3}}{n+1} \leq u_3, \frac{R_{i,4}}{n+1} \leq u_4\right\} u = \{u_1, u_2, u_3, u_4\} \in [0,1]^4 \tag{13}$$

Where $I(A)$ is the pointer variable of function A which gives a number of 1 if A is valid and 0 if A is invalid. Ranks of ith Dd, Ds, Dp and Di value are given as $R_{i,1}$, $R_{i,2}$, $R_{i,3}$ and $R_{i,4}$ respectively (Mirabbasi et al., 2012). The Quantile-Quantile (Q-Q) plot is then used to compare the closeness between $C_n(u)$ and $C_p(u)$. graphically.

3.8 Four-Variate Joint Drought Frequency Analysis

According to Salvadori and De Michele (2004), in a multivariate situation such that $X_1, X_2, \ldots, X_d$ exceeds their corresponding thresholds $(X_1 > x_1, \ldots, X_d > x_d)$, the joint probability and return period of all drought events can be computed. In this study, the four-variate joint probabilities, P_{DSPI}, of $(Dd \geq d, Ds \geq s, Dp \geq p, Di \geq i)$ can be calculated utilizing the copula-based procedure modified from Shiau (2006), Ganguli and Reddy (2014):

$$P_{DSPI} = 1 - F_{DSPI}(d, s, p, i) = 1 - C_{DSPI}(d, s, p, , i) \tag{14}$$

Analogous to return periods, the four-variate joint return periods, T_{DSPI}, of $(Dd \geq d, Ds \geq s, Dp \geq p, Di \geq i)$ can be estimated as follows:

$$T_{DSPI} = \frac{\zeta}{1 - F_{DSPI}(d, s, p, i)} = \frac{\zeta}{1 - C_{DSPI}(d, s, p, i)} \tag{15}$$

In the equations above, $\zeta = N/q$, N = period of SPEI time-series (years), q = amount of drought events in N years. $C_{DSPI}(d, s, p, i)$ = four-variate joint distributions, P_{DSPI} and T_{DSPI}, expresses the joint probability and return periods of occurrence of Dd or Ds or Dp or Di exceeding a particular value of d, s, p and i, respectively.

4 Results and Discussions

4.1 Drought Properties and Joint Dependence

The statistical description of Di extracted from SPEI-6 time series are shown in Table 2. The average regional values of Di across NE, NC, CS, SC, IM, NW, TP, and EMC during 1961–2013 were 8.9, 11.7, 12.4, 10.1, 10.1, 14.7, 9.9, and 12.1 months, respectively. Further, the qualitative joint dependence of Di against Di, Ds, and Dp was examined by employing a graphical tool in the form of a scatter plot matrix (Salvadori et al. 2007; Genest and Favre, 2007). According to Serinaldi et al. (2009), the scatter plot helps to synthesize information on dependence structure and marginals. The joint dependence between Di, Ds and Dp have been studied (Ayantobo et al. 2017) and would not repeat again here. The joint dependence of Di against Di, Ds, and Dp in regions NE, NC, CS, SC, IM, NW, TP, and EMC are shown in Fig. 3. Figure 3 gave a visual judgment on the pair-wise dependence between drought variables (Genest and Favre, 2007). Although some regions showed the collection of points in the lower left corner, an uphill pattern relationship was observed from left to right. For each pair of variables, a growing positive linear relationship existed.

The Kendall's k, Spearman's p, and Pearson's classical correlation coefficients r are applied to test the strength of the relationship of drought variables, and their respective values in regions NE, NC, CS, SC, IM, NW, TP and EMC are shown (Table 3). Positive and significant associations between drought variables were observed. The correlations between Dd and Ds as well as Ds and Dp were high while the correlations between Dd and Dp were small. The dependence analysis demonstrated it in Fig. 3 and Table 3 that the mutual dependence of drought variables are tremendous and hence suitable for building a regional joint distribution using copula functions.

Table 2 Statistical description of drought inter-arrival time across different regions during 1961–2013

Statistic	Climatic region							EMC
	NE	NC	CS	SC	IM	NW	TP	
Event	68	52	51	63	63	43	64	53
Range	22	34	32	30	38	60	35	25
Mean	8.97	11.71	12.42	10.07	10.05	14.74	9.92	12.10
Variance	27.24	67.21	65.06	51.21	56.51	164.54	61.66	57.70
Std. Deviation	5.22	8.19	8.07	7.16	7.52	12.83	7.85	7.60
Coef. Variation	0.58	0.70	0.65	0.71	0.75	0.87	0.79	0.63
Std. Error	0.64	1.15	1.14	0.92	0.95	1.98	0.99	1.05
Skewness	0.82	1.09	0.58	1.10	1.51	1.64	1.28	0.52
Kurtosis	0.98	1.17	−0.23	0.78	3.23	3.25	1.29	−0.82

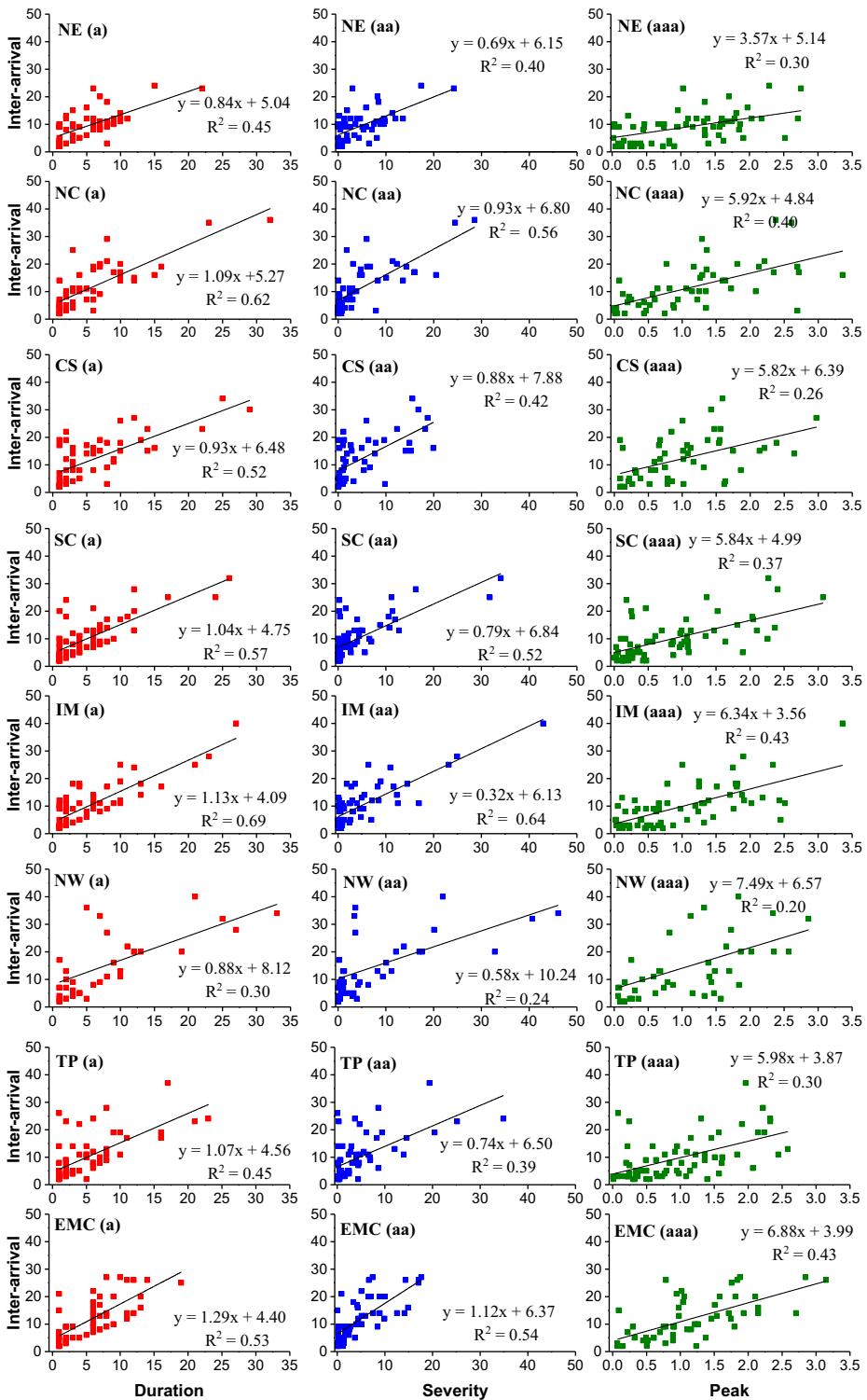## 4.2 Marginal Distribution for Inter-Arrival Time

Different marginal distributions have been recommended as a uniform procedure to fit Dd, Ds, and Dp (Shiau and Modarres, 2009; Song and Singh, 2010; Wong et al. 2010; Lee et al. 2013). The marginal distributions and associated parameters for Dd, Ds, and Dp in different regions of mainland China have been documented (Ayantobo et al. 2017). These distributions showed that the empirical and theoretical distributions had an excellent agreement. In the case of Di, many marginal distributions were also compared and the goodness of fit evaluated. According to Table 4, the KS and AD values were compared for each distribution, and their respective ranks in each region are shown. It was demonstrated that the KS was better than the AD test. Therefore, based on the KS test, the generalized Pareto was the best for regions NE, NC, CS, SC, NW, TP and EMC while gamma was the best for region IM because they had the smallest values. As given in Table 5, the parameters of the distributions were calculated using the MLE and after that used to fit the drought data.

4.3 Copula-Based Four-Variate Joint Distributions

Table 6 presents the copula parameters, AIC, RMSE and BIAS values in regions NE, NC, CS, SC, IM, NW, TP and EMC used to compare the performances of the symmetric and asymmetric copulas. It could be observed that the copula parameters are positive; connoting that four-variate dependence existed between variables of Dd, Ds, Dp, and Di. Generally, it was shown that symmetric copulas fit better than the asymmetric copulas because they had the lowest RMSE and AIC values in most regions. In the symmetric class, the RMSE and AIC values of Clayton and Frank were more or less the same but gave different fittings in the different region. Clayton copula was best for regions NE, CS and EMC while Frank copula was the best for NC, SC, IM, NW, and TP because they had the smallest values of RMSE and AIC.

The observed and theoretical probabilities of the symmetric and asymmetric copula functions are compared in Fig. 4. The figure showed that the theoretical probabilities for the Clayton in regions NE, CS, EMC, and Frank copula in NC, SC, IM, NW, TP fit the observed values very well.

Further, the theoretical joint probabilities of the occurrence orders of the observed values were evaluated. As shown in Fig. 5, the theoretical joint probabilities were sketched against

NE (a) $y = 0.84x + 5.04$ $R^2 = 0.45$
NE (aa) $y = 0.69x + 6.15$ $R^2 = 0.40$
NE (aaa) $y = 3.57x + 5.14$ $R^2 = 0.30$

NC (a) $y = 1.09x + 5.27$ $R^2 = 0.62$
NC (aa) $y = 0.93x + 6.80$ $R^2 = 0.56$
NC (aaa) $y = 5.92x + 4.84$ $R^2 = 0.40$

CS (a) $y = 0.93x + 6.48$ $R^2 = 0.52$
CS (aa) $y = 0.88x + 7.88$ $R^2 = 0.42$
CS (aaa) $y = 5.82x + 6.39$ $R^2 = 0.26$

SC (a) $y = 1.04x + 4.75$ $R^2 = 0.57$
SC (aa) $y = 0.79x + 6.84$ $R^2 = 0.52$
SC (aaa) $y = 5.84x + 4.99$ $R^2 = 0.37$

IM (a) $y = 1.13x + 4.09$ $R^2 = 0.69$
IM (aa) $y = 0.32x + 6.13$ $R^2 = 0.64$
IM (aaa) $y = 6.34x + 3.56$ $R^2 = 0.43$

NW (a) $y = 0.88x + 8.12$ $R^2 = 0.30$
NW (aa) $y = 0.58x + 10.24$ $R^2 = 0.24$
NW (aaa) $y = 7.49x + 6.57$ $R^2 = 0.20$

TP (a) $y = 1.07x + 4.56$ $R^2 = 0.45$
TP (aa) $y = 0.74x + 6.50$ $R^2 = 0.39$
TP (aaa) $y = 5.98x + 3.87$ $R^2 = 0.30$

EMC (a) $y = 1.29x + 4.40$ $R^2 = 0.53$
EMC (aa) $y = 1.12x + 6.37$ $R^2 = 0.54$
EMC (aaa) $y = 6.88x + 3.99$ $R^2 = 0.43$

Duration   Severity   Peak

◀ **Fig. 3** Scatter-plots of the pair-wise drought variables for different regions across mainland China during 1961–2013

empirical probabilities, which showed that there were significant differences between empirical and theoretical joint probabilities. The plots of all the copulas deviate from the 45^0 diagonal line uniquely. The maximum deviations from the diagonal line were observed for Gumbel plots, showing that the functions might not be suitable for building joint dependence. The symmetric Clayton in regions NE, CS and EMC, and symmetric Frank in regions NC, SC, IM, NW, and TP exhibited good agreement between the theoretical and empirical probabilities, and seemed very satisfying in modeling drought variables. Therefore, these copulas were thereafter selected for regional drought frequency analysis.

In the same way, the selected copulas were compared in each station, and the RMSE and AIC goodness of fit test results are presented in Fig. 6. It could be observed that the average AIC (Fig. 6a) and RMSE values (Fig. 6b) of all the 552 stations for symmetric Frank copula were smaller than that of Clayton and Gumbel copulas. Therefore, the four-variate symmetric Frank copula was selected for the station-based drought frequency analysis across EMC.

Table 3 Values of classical, Kendall and Spearman's correlation coefficients for drought variables in different regions between 1961 and 2013

Region	Variable	Correlation Coefficient								
		Classical			Kendall			Spearman		
		r	t (obs)	t (crit)	t	t (obs)	t (crit)	p	t (obs)	t (crit)
NE	Dd vs. Di	0.67	7.27	1.997	0.69	7.70	1.997	0.67	7.35	1.997
	Ds vs. Di	0.63	6.59		0.50	4.69		0.59	5.92	
	Dp vs. Di	0.51	4.82		0.45	4.13		0.53	5.01	
NC	Dd vs. Di	0.78	8.93	2.009	0.70	6.87	2.009	0.78	8.73	2.009
	Ds vs. Di	0.75	7.99		0.61	5.43		0.77	8.40	
	Dp vs. Di	0.62	5.62		0.53	4.43		0.68	6.59	
CS	Dd vs. Di	0.72	7.33	2.010	0.62	5.49	2.010	0.65	5.96	2.010
	Ds vs. Di	0.65	5.92		0.53	4.32		0.64	5.80	
	Dp vs. Di	0.51	4.17		0.42	3.25		0.51	4.12	
SC	Dd vs. Di	0.76	9.05	2.000	0.72	8.15	2.000	0.73	8.32	2.000
	Ds vs. Di	0.72	8.10		0.57	5.43		0.68	7.24	
	Dp vs. Di	0.61	6.01		0.51	4.57		0.62	6.19	
IM	Dd vs. Di	0.83	11.76	2.000	0.72	8.13	2.000	0.74	8.67	2.000
	Ds vs. Di	0.80	10.52		0.58	5.59		0.71	7.83	
	Dp vs. Di	0.66	6.84		0.50	4.46		0.63	6.35	
NW	Dd vs. Di	0.53	3.97	2.020	0.58	4.57	2.020	0.65	5.40	2.020
	Ds vs. Di	0.49	3.61		0.54	4.13		0.67	5.72	
	Dp vs. Di	0.45	3.23		0.46	3.29		0.60	4.75	
TP	Dd vs. Di	0.67	7.15	1.999	0.69	7.49	1.999	0.70	7.72	1.999
	Ds vs. Di	0.62	6.24		0.53	4.97		0.63	6.32	
	Dp vs. Di	0.54	5.08		0.46	4.11		0.55	5.13	
EMC	Dd vs. Di	0.73	7.52	2.008	0.68	6.64	2.008	0.73	7.61	2.008
	Ds vs. Di	0.74	7.81		0.64	5.89		0.77	8.54	
	Dp vs. Di	0.66	6.21		0.56	4.79		0.69	6.75	

t (obs) is greater than t (crit) in the entire region

Table 4 Comparison of selected marginal distributions for drought inter-arrival time across various regions between 1961 and 2013

Region	Statistic	Marginal Distributions									
		Exponential	Exponential (2P)	Gamma	Gamma (3P)	Gen. Extreme Value	Gen. Pareto	Inv. Gaussian	Inv. Gaussian (3P)	Weibull	Weibull (3P)
NE	K-S Test	0.262h	0.261g	0.206e	0.293j	0.174b	0.171a*	0.237f	0.199d	0.193c	0.281i
	A-D Test	5.964f	32.273i	2.326d	23.388h	1.630a*	13.303g	5.753e	1.908c	1.689b	32.72j
NC	K-S Test	0.157h	0.110f	0.082c	0.178i	0.081b	0.069a*	0.131g	0.104e	0.083d	0.266j
	A-D Test	2.150f	15.614h	0.389e	20.448i	0.423d	0.327a*	2.189g	0.549e	0.365b	23.549j
CS	K-S Test	0.156g	0.164h	0.138e	0.219j	0.119b	0.089a*	0.188i	0.150f	0.129d	0.125c
	A-D Test	2.393e	10.904i	1.202d	18.117j	0.669b	4.198f	4.468g	1.013c	0.626a*	9.105h
SC	K-S Test	0.180i	0.107f	0.091b	0.199j	0.101d	0.077a*	0.137g	0.106e	0.093c	0.169h
	A-D Test	2.638g	18.195j	0.438b	8.137h	0.533c	0.340a*	1.943f	0.715e	0.589d	11.73i
IM	K-S Test	0.180h	0.138e	0.101a*	0.250j	0.114d	0.109c	0.172g	0.157f	0.105b	0.195i
	A-D Test	2.609g	18.641i	0.713b	14.031h	0.873d	0.651a*	2.354f	1.291e	0.832c	23.380j
NW	K-S Test	0.127i	0.071b	0.081c	0.178j	0.082d	0.063a*	0.108h	0.087g	0.083e	0.084f
	A-D Test	0.793f	6.764i	0.309e	5.177h	0.365d	0.188a*	1.003g	0.300b	0.516c	11.189j
TP	K-S Test	0.183j	0.118c	0.119d	0.153i	0.134f	0.092a*	0.152h	0.114b	0.138g	0.133e
	A-D Test	2.095g	12.806i	0.864b	7.275h	1.054d	0.593a*	1.824f	0.957c	1.309e	19.441j
EMC	K-S Test	0.166i	0.135g	0.112f	0.301j	0.092e	0.061a*	0.159h	0.092d	0.077b	0.089c
	A-D Test	2.751 g	10.317i	0.755f	23.964j	0.541c	0.277a	2.965 h	0.603d	0.398b	0.750e

Note: letters a − j represents the K-S and A-D ranks of the marginal distributions in each region

Table 5 The best-fitted marginal distributions with estimated parameters for different regions across mainland China between 1961 and 2013

Regions	Estimated Parameters for Marginal Distributions			
		β	α	λ
NE	Generalized Pareto	−0.609	11.90	1.576
NC	Generalized Pareto	−0.305	13.493	1.363
CS	Generalized Pareto	−0.554	18.168	0.732
SC	Generalized Pareto	−0.205	10.351	1.473
IM	Gamma	1.787	5.623	
NW	Generalized Pareto	−0.007	13.490	1.349
TP	Generalized Pareto	−0.063	9.183	1.286
EMC	Generalized Pareto	−0.499	16.267	1.249

*Note:β; shape parameter, α; scale parameter, λ; location parameter

4.4 Four-Variate Joint Drought Frequency Analysis

4.4.1 Regional Four-Variate Probabilities of Drought Events

The average historical drought events for regions NE, NC, CS, SC, IM, NW, TP, and EMCas well as their corresponding 5-, 10-, 20-, 50- and 100-years univariate return periods defined by separate d, s, p and i were estimated and presented in Fig. 7. Next, the P_{DSPi}in each of these regions were obtained using Eq. (14). Fig. 7 shows the P_{DSPI}, representing the historical, 5-, 10-, 20-, 50- and 100-year return periods to provide adequate information about the potential drought risk associated with Dd, Ds, Dp and Di across different regions. Overall, as the return period increased, the trend of the joint probability decreased. Hence, the simultaneous existence of higher degree drought events in different regions was less frequent as the year increases. For example, in sub-region I, considering 20- years univariate return period, the values of d, s, p, and i, were 13.56 months, 17.46, 2.44 and 20.73 months, respectively. The $P_{DSPI}(Dd \geq 13.56, Ds \geq 17.46, Dp \geq 2.44, Di \geq 20.73)$was 0.124.

These results could provide useful hints in appraising the drought risk in different regions. For instance, regions CS and NE had the highest probability of about 0.57, indicating that drought management and planning are needed within these regions. It was shown that regions SC and IM had a low probability of 0.48, which meant a low combination risk. The joint probability not only confirmed the occurrence of regional drought events but also gave a quantitative approach to analyze the probability of drought under varying d, s, p and i situations.

4.4.2 Regional Four-Variate Return Period of Drought Events

The ζ values were 0.78, 1.02, 1.04, 0.84, 0.84, 1.23, 0.83 and 1.00 months for regions NE, NC, CS, SC, IM, NW, TP and EMC, respectively. Using the average historical and univariate return

Table 6 Copula parameter, RMSE, AIC and BIAS of four-variate symmetric and asymmetric Archimedean copulas for different regions between 1961 and 2013

Copula	Statistic	Climatic region							
		NE	NC	CS	SC	IM	NW	TP	EMC
CLAYTON (Symmetric)	θ	3.790	4.940	3.680	5.300	6.280	5.940	5.490	7.340
	RMSE	0.063	0.033	0.027	0.044	0.042	0.040	0.037	0.039
	AIC	-161.214	-152.218	-157.561	-168.865	-171.183	-118.484	-180.949	-147.231
	BIAS	-9.478	-0.001	-10.130	-13.091	-9.750	-7.791	-12.854	0.719
CLAYTON (Asymmetric)	θ_3	6.950	8.760	7.740	16.620	12.480	21.460	16.670	156.070
	θ_2	1.840	5.340	2.280	3.850	4.480	4.240	4.070	3.810
	θ_1	0.530	0.710	0.620	0.900	0.890	0.670	0.740	0.900
	RMSE	0.065	0.086	0.088	0.062	0.068	0.069	0.076	0.061
	AIC	-130.961	-110.744	-106.368	-150.084	-146.655	-98.276	-142.033	-126.961
	BIAS	-0.009	1.671	-1.672	-4.602	-10.242	-1.817	-4.679	4.024
FRANK (Symmetric)	θ	9.420	10.730	9.400	12.700	13.710	13.290	13.090	17.510
	RMSE	0.069	0.033	0.030	0.041	0.042	0.039	0.033	0.041
	AIC	-156.164	-152.441	-153.424	-172.250	-171.224	-118.767	-187.520	-144.509
	BIAS	7.382	7.027	4.839	1.522	4.878	2.392	-0.474	6.744
FRANK (Asymmetric)	θ_3	6.480	5.810	6.920	4.650	7.560	5.380	4.440	5.920
	θ_2	2.160	1.937	2.307	1.550	2.520	1.793	1.480	1.973
	θ_1	0.040	0.010	0.050	0.010	0.050	0.030	0.010	0.020
	RMSE	0.069	0.088	0.067	0.063	0.074	0.064	0.069	0.066
	AIC	-114.474	-81.448	-86.125	-108.598	-104.214	-73.069	-107.608	-89.746
	BIAS	3.494	-9.436	4.851	-25.608	-7.349	-6.830	-28.379	-12.356
GUMBEL (Asymmetric)	θ	3.060	3.510	3.170	3.910	4.420	4.100	4.210	5.600
	RMSE	0.071	0.034	0.033	0.042	0.046	0.041	0.037	0.042
	AIC	-154.504	-150.787	-149.110	-171.622	-166.834	-117.616	-181.509	-143.919
	BIAS	3.752	5.474	1.436	-1.373	-0.750	-0.186	-3.009	4.119
GUMBEL (Asymmetric)	θ_3	25.750	10.150	21.730	10.330	7.880	15.080	12.830	16.010
	θ_2	12.840	0.860	1.030	0.690	11.870	0.820	0.800	0.850
	θ_1	0.170	0.150	0.210	0.150	0.150	0.190	0.160	0.190
	RMSE	0.332	0.348	0.295	0.356	0.351	0.331	0.342	0.347
	AIC	-86.922	-68.057	-84.822	-75.979	-76.923	-66.505	-80.373	-73.817
	BIAS	13.294	5.844	-8.053	15.578	14.691	-8.184	14.230	1.133

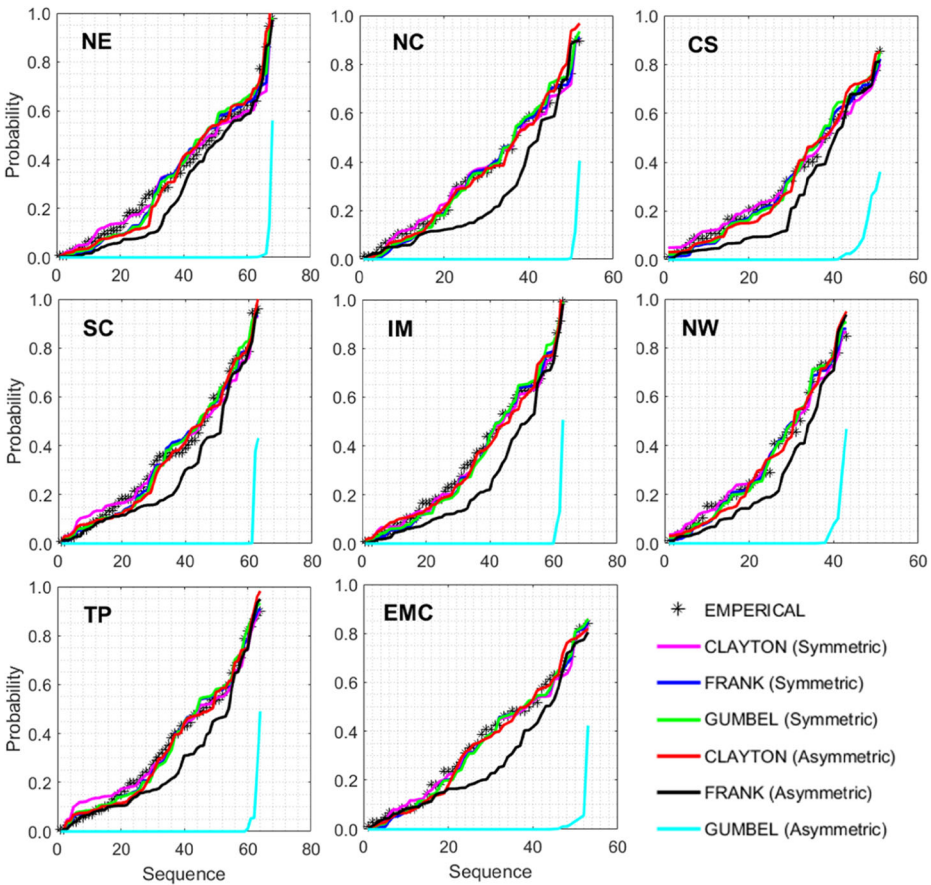*Note: Values in bold represents the best fitted copula in each region

Fig. 4 Comparison between empirical and theoretical probabilities for symmetric and asymmetric Archimedean copulas in different regions of mainland China during 1961–2013

periods of d, s, p and i estimated for 5-, 10-, 20-, 50- and 100-year, the T_{DSPI} were obtained for each region. Figure 7 shows the T_{DSPI} for the historical, 5-, 10-, 20-, 50- and 100-year return periods. The graph showed that the T_{DSPI} trends increased with the year.

For example, in region NE, considering 20- years univariate return period, the values of d, s, p, and i, were 13.56 months, 17.46, 2.44 and 20.73 months, respectively. The T_{DSPI}, ($Dd \geq$ 13.56, $Ds \geq 17.46$, $Dp \geq 2.44$, $Di \geq 20.73$) was 6.27 years. The T_{DSPI} were lowest in regions NE and CS. This implied that drought frequency would be higher in these regions compared to the other regions. The return period of drought variables obtained from univariate frequency analysis is higher than those by joint distribution for the AND case (T_{DSPI}). This showed that the univariate analysis does not furnish satisfactory knowledge about the drought risks associated with the four-variates. If engineers design hydraulic structures based on the results from univariate frequency analysis, the drought variables may be overestimated and this will lead to an increased cost of the structure. These results could be valuable in risk evaluation of water resources operations under severe and extreme drought situations.

The P_{DSPI} and T_{DSPI} showed that the result computed agreed with the actual data and confirmed that the selected copula functions for each region fitted the data well. Figure 7
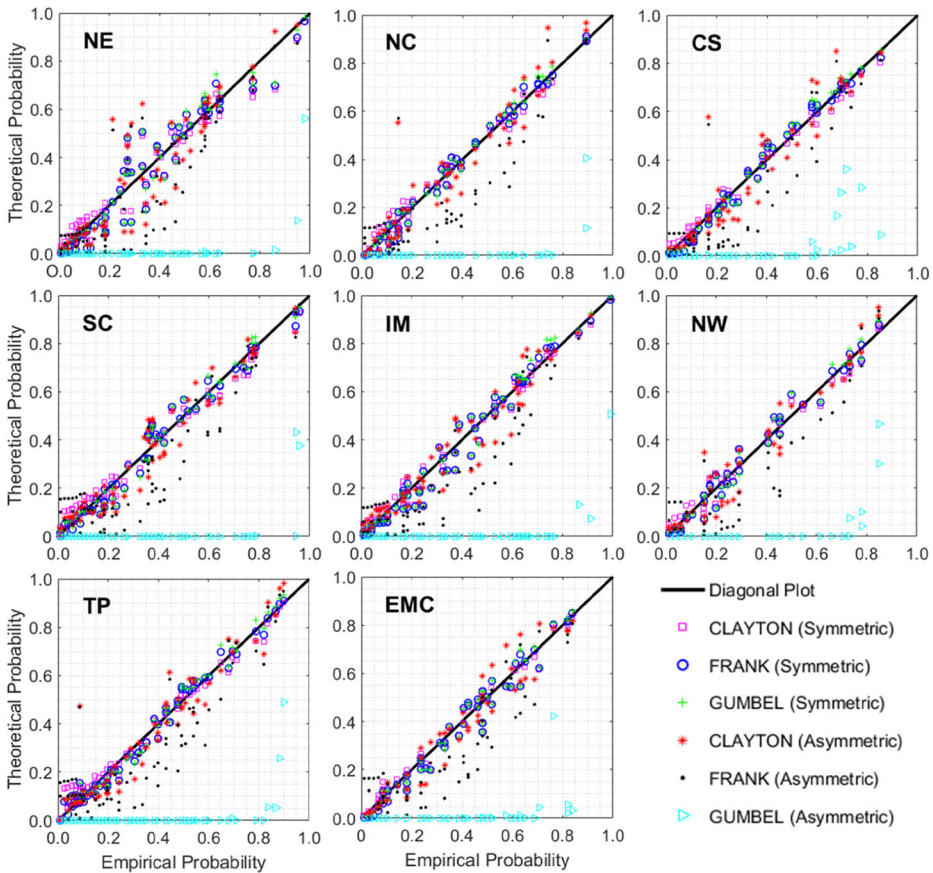
Fig. 5 Quantile-Quantile (Q-Q) plots of symmetric and asymmetric Archimedean copulas in different regions of mainland China during 1961–2013. The solid line indicates the 45^0 line

further explained that for a particular recurrence interval, the P_{DSPI} showed a decreased trend while the T_{DSPI} showed an increased trend. For example, given the occurrence of d, s, p, and i in region NE, the estimated P_{DSPI} of 5- and 100-year plan drought are 0.332 and 0.029, respectively. Considering these illustrations, the calculated results seemed feasible. Therefore, from Fig. 7, the P_{DSPI} and T_{DSPI} of any other drought events in regions NE, NC, CS, SC, IM, NW, TP, and EMC can be obtained directly through interpolation.

4.4.3 Spatial Distribution of Four-Variate Probability and Return Periods

The average d, s, p, and i for EMC were 5.81 months, 4.96, 1.15 and 11.1 months respectively. Similar to the regional drought analysis, the station-based P_{DSPI} and T_{DSPI} of drought events exceeding these specific values, (i. e., $Dd \geq 5.81$, $Ds \geq 4.96$, $Dp \geq 1.15$, $Di \geq 11.1$) were estimated for each station using the symmetric Frank copula. The spatial distribution of the P_{DSPI} and T_{DSPI} are mapped in Fig. 8.

The P_{DSPI} varied from 0.34 to 0.75. In this case, most of northern China experienced higher probability spreading from the northwest to northeast China. Notably, it seems that the P_{DSPI} were very high in NW and some stations in NC and TP. Based on the spatial pattern of drought
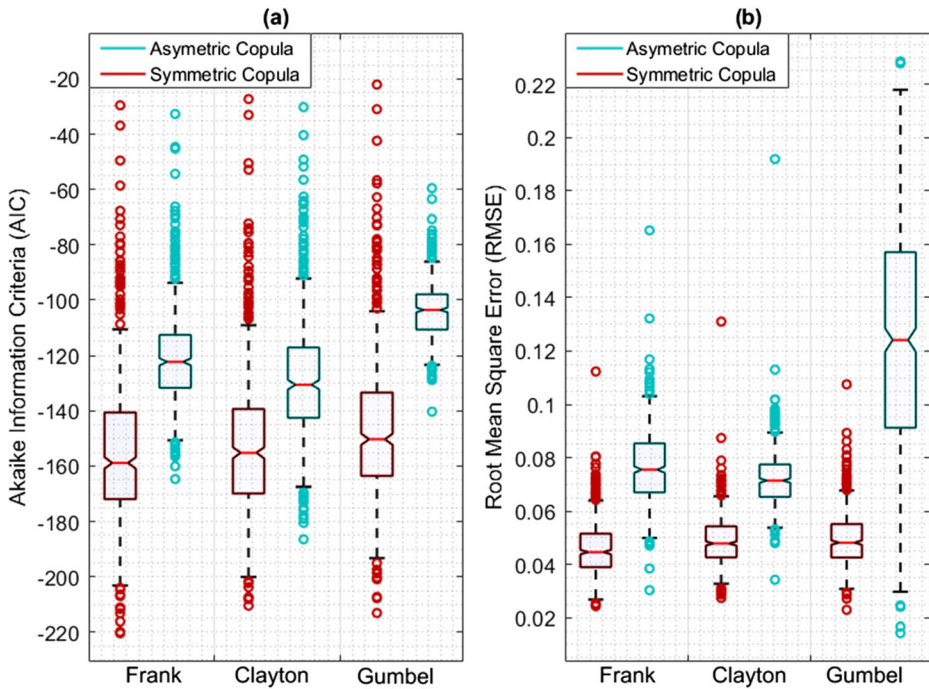
Fig. 6 Station-based comparison of (a) AIC and (b) RMSE for symmetric and asymmetric Archimedean copulas in different regions of mainland China during 1961–2013

events, the drought risk resulted from these regions are consistent. On the other hand, the results for southern China including SC and CS were slightly different as relatively lower P_{DSPI} was observed. Accordingly, counter measures of drought hazards need to be set for proactive actions especially in regions that had greater drought risks.

The T_{DSPI} ranged from 1.02 to 6.79 years. Long T_{DSPI} were found around NC while short T_{DSPI} dominated parts of CS, NW, and TP. Medium T_{DSPI} were found in eastern China, most especially NE, NC, CS. The spatial pattern of P_{DSPI} and T_{DSPI} suggested tremendous variations within different regions. In most cases, regions with high P_{DSPI} are often associated with short T_{DSPI} and vice versa. Severe droughts are most liable to eventuate in most of the northwestern and southwestern regions because of short T_{DSPI}. Also, because these areas are commercially advanced with huge populations, severe drought events could place grave danger on water resources in these regions. This requisite information is needed by water managers and other government agencies for adequate planning as well as management of water resources under severe drought situations. In this study, four-variate distribution was determined and successfully used.

5 Summary and Conclusions

Most meteorological phenomena are multiplicative, and the advantages of using a multivariate assessment strategy are evident in this study. In particular, drought events are characterized by Dd, Ds, Dp, and Di and the joint drought risk, which reveals their probability at the same time, is vital for management and planning of water resources under drought situations. Using the
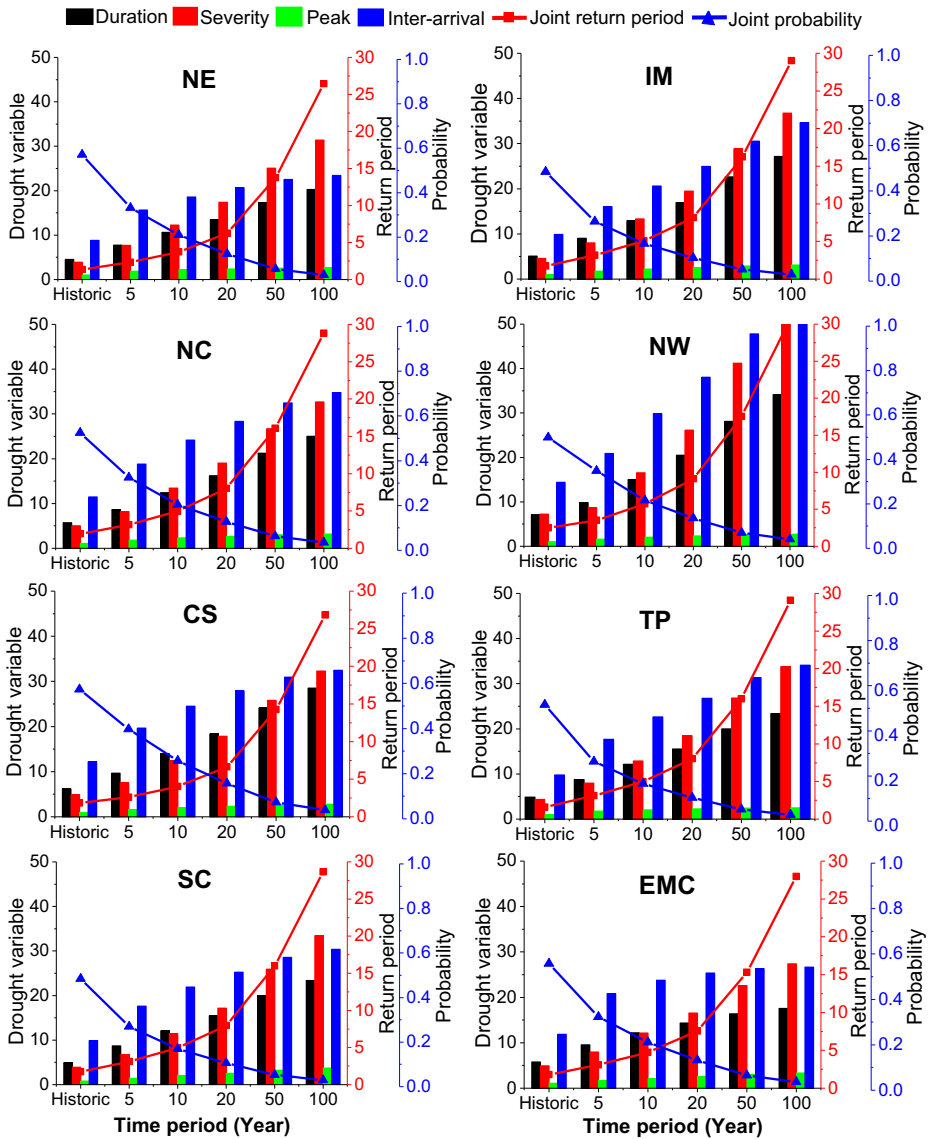
Fig. 7 Four-variate joint probability and return periods of *Dd*, *Ds*, D*p*, and *Di* in different regions across mainland China during 1961–2013

symmetric and asymmetric Archimedean copula functions, annual droughts were studied as four-dimensional phenomena incorporating these variables to build the joint four-variate distributions. Next, the P_{DSPI} for different T_{DSPI} in each region and EMC were then estimated.

The foremost outcomes of this research were summed as follows: (1) For D*i*, the generalized Pareto was the best distribution for regions NE, NC, CS, SC, NW, TP, and EMC while gamma was the best distribution for region IM. (2) The RMSE and AIC values were utilized to choose the suitable copula. We employed Clayton copula in regions NE, CS, and EMC, and Frank copula in regions NC, SC, IM, NW, and TP. Symmetric Frank copula was used for the
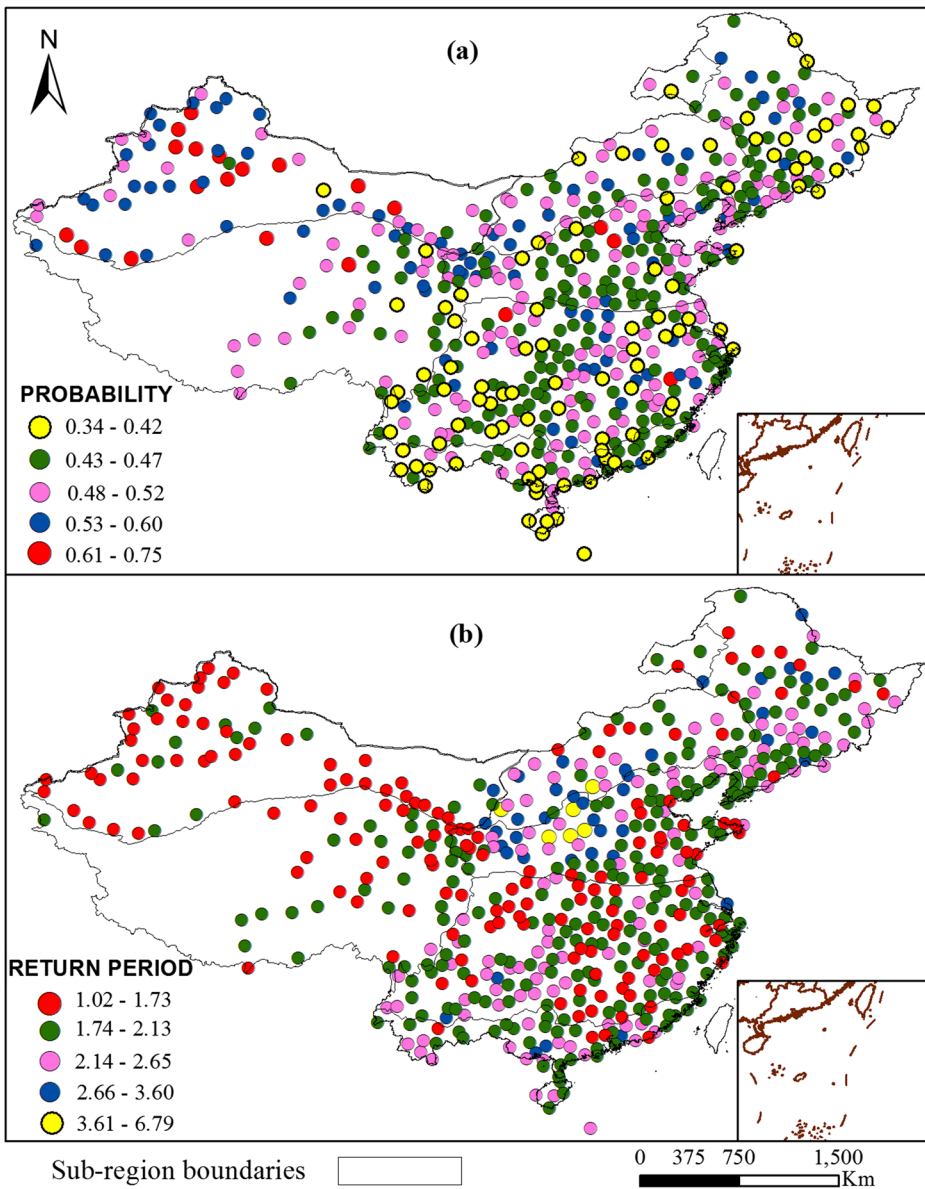
Fig. 8 Spatial pattern of four-variate (a) joint probability and (b) joint return periods (years) of Dd, Ds, Dp, and Di in different regions across mainland China during 1961–2013

station-based drought analysis. The selected copula functions provided the best fit for the dependence structures of Dd, Ds, Dp, and Di, and consequently, it was adopted for the joint risk drought analysis. (3) The risk of drought event is determined based on P_{DSPI} and T_{DSPI}, which furnished vital information for drought analysis. For instance, by using only the univariate information provided by either Dd, Ds, Dp, and Di, may yield under or overestimates of the actual drought state, and, in turn, of the corresponding risk. (4) Analyzing the probabilities and return periods of four-variate drought events, the trends or changes in drought

events were then assessed over time. The concept of return period used in this study is a familiar concept in the community of water resources professionals. However, for a deep understanding of drought event, the ultimate evaluation was based on the hazard which the event possesses, and the vulnerability of this region to the drought event. Therefore, based on the results of our analysis, it is reasonable to increase drought control concerns especially in regions with very high P_{DSPI} and low T_{DSPI}. As a recommendation, a drought return period of below ten years seems to indicate that the drought is entering an alert or a dangerous state.

This study added to a better valuable knowledge in the sphere of disaster management, especially concerning the appraisal of drought issues and the performance of full drought-risk investigations. Our insights on the real-time evaluation of meteorological droughts of different regions may provide valuable information about the possible evolution of drought episode and also help the water manager to plan effective mitigation strategies. As a conclusion, in this paper, a multivariate frequency analysis of meteorological droughts in seven climatic regions is addressed using Copulas. Such an approach is flexible, comprehensive and offered other several advantages over previous definitions of multivariate frequency analysis (Salvadori et al., 2015). The copula-derived drought analysis considers complete interdependencies between the drought variables over a geographical region. Finally, in this paper, no consideration has been included for the climate change influence on the above analysis although climate change will affect the estimation of the return period of droughts. For this aspect, rigorous and systematic attempts should be made for estimating the anticipated changes in the meteorological parameters which directly or indirectly affect the Dd, Ds, Dp, and Di and ultimately the frequency of drought events.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

References

Abdi A, Hassanzadeh Y, Talatahari S, Fakheri-Fard A, Mirabbasi R 2016 Regional bivariate modeling of droughts using L-components and copulas. Stoch. Environ. Res. Risk. Assess
AghaKouchak A, Bardossy A, Habib E (2010) Conditional simulation of remotely sensed rainfall fields using a non-Gaussian transformed copula. Adv Water Resour 33(6):624–634
Akaike H 1974 A new look at the statistical model identification. IEEE Trans Autom Control AC 19(6):716–723
Allen RG, Pereira LS, Raes D, Smith M 1998 Crop evapotranspiration: guidelines for computing crop requirements, irrigation, and drainage paper 56. FAO, Roma

Ayantobo OO, Li Y, Song S, Yao N (2017) Spatial comparability of drought characteristics and related return periods in mainland China over 1961-2013. J Hydrol 550:549–567

Begueria S, Vicente-Serrano SM, Reig F, Latorre B (2013) Standardized precipitation evapotranspiration index (SPEI) revisited: parameter fitting, evapotranspiration models, tools, datasets and drought monitoring. Int J Climatol 34:3001–3023

Chang J, Li Y, Wang Y, Yuan M (2016) Copula-based drought risk assessment combined with an integrated index in the Wei River Basin, China. J Hydrol 540:824–834

Chebana F, Ouarada T (2011) Multivariate quantiles in hydrological frequency analysis. Environmetrics 22(1): 63–78

Chen L, Singh VP, Shenglian G, Hao Z, Li T 2012 Flood coincidence Risk Analysis using Multivariate Copula Functions. Journal of Hydrologic Engineering 17, 6

Cunnane C (1978) Unbiased plotting positions—a review. J Hydrol 37(3–4):205–222

De Michele C, Salvadori G, Canossi M, Petaccia A, Rosso R (2005) Bivariate statistical approach to check adequacy of dam spillway. J Hydrol Eng 10(1):50–57

De Michele C, Salvadori G, Passni G, Vezzoli R (2007) A multivariate model of sea storms using copulas. Coast Eng 54(10):734–751

Embrechts P, Lindskog F, McNeil A (2003) Modeling dependence with copulas and applications to risk management. In: Handbook of Heavy Tailed Distributions in Finance. Elsevier Sci, New York, pp 329–384

Fan, L., Qian, Z., 2016. Probabilistic modelling of flood events using the entropy copula in China. Advances in Water Resources. 97, 233-240

Ganguli P, Reddy MJ (2014) Evaluation of trends and multivariate frequency analysis of droughts in three meteorological subdivisions of western India. Int J Climatol 34(3):911–928

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

Grimaldi S, Serinaldi F (2006) Asymmetric copula in multivariate flood frequency analysis. Adv Water Resour 29(8):1115–1167

Gringorten II (1963) A plotting rule for extreme probability paper. J Geophys Res 68(3):813–814

Haslinger K, Koffler D, Schöner W, Laaha G (2014) Exploring the link between meteorological drought and streamflow: effects of climate-catchment interaction. Water Resour Res 50:2468–2487

Hao C, Zhang J, Yao F (2017) Multivariate drought frequency estimation using copula method in Southwest China. Theor Appl Climatol. 127:977

Helsel DR, Hirsch RM (1992) Statistical Methods in Water Resources. Elsevier, Amsterdam

Huang SZ, Chang JX, Huang Q, Chen YY et al (2014b) Spatio-temporal changes and frequency analysis of drought in the Wei River Basin, China. Water Resour Manag 28(10):3095–3110

Huang SZ, Hou BB, Chang JX, Huang Q, Chen YT (2014a) Copulas-based probabilistic characterization of the combination of dry and wet conditions in the Guanzhong Plain, China. J Hydrol 519:3204–3213

Joe H (1997) Multivariate models and dependence concepts. Chapman and Hall, New York

Kao SC, Govindaraju RS (2010) A copula-based joint deficit index for droughts. J Hydrol 380:121–134

Kao SC, Govindaraju RS (2008) Trivariate statistical analysis of extreme rainfall events via the Plackett family of copulas. Water Resour Res 44(2):W02415

Lee T, Modarres R, Quarda TBMJ (2013) Data-based analysis of bivariate copula tail dependence for drought duration and severity. Hydrol Process 27(10):1454–1463

Leng GY, Tang QH, Rayburgc S (2015) Climate change impacts on meteorological, agricultural and hydrological droughts in China. Glob Planet Chang 126:23–34

Liu XF, Wang SX, Zhou Y, Wang FT, Li WJ, Liu WL 2015 Regionalization and Spatiotemporal Variation of Drought in China Based on Standardized Precipitation Evapotranspiration Index (1961–2013). Advances in Meteorology, 2015, 18 pgs

Lu E, Cai WY, Jiang ZH, Zhang Q, Zhang CJ, Higgins RW, Halpert MS (2013) The day-to-day monitoring of the 2011 severe drought in China. Clim Dyn 43(1–2):1–9

Lu E, Luo YL, Zhang RH, Wu QX, Liu LP 2011 Regional atmospheric anomalies responsible for the 2009–2010 severe drought in China. J. Geophys.- Atmos. 116

Lu J, Ju J, Ren J, Gan W (2012) The influence of the Madden-Julian Oscillation activity anomalies on Yunnan's extreme drought of 2009–2010. Sci China Earth Sci 55:98–112

Ma M, Song S, Ren L, Jiang S, Song J (2013) Multivariate drought characteristics using trivariate Gaussian and Student t copulas. Hydrol Process 27(8):1175–1190

Mirabbasi R, Fakheri-Fard A, Dinpashoh Y (2012) Bivariate drought frequency analysis using the copula method. TheorAppl Climatol 108:191–206

Mirakbari M, Ganji A, Fallah SR (2010) Regional bivariate frequency analysis of meteorological droughts. J Hydrol Eng 15(12):985–1000

Mishra AK, Singh VP (2010) A review of drought concepts. J Hydrol 391(1–2):202–216

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

Nelsen, R. B., (1999). An introduction to copulas: Springer Verlag, New York, 215p

Prohaska S, llic A, Majkic B 2008 Multiple-coincidence of flood waves on the main river and its tributaries. IOP Conf. Series: Earth and Environmental Science, IOP Publishing, Bristol, UK

Qin Y, Yang DW, Lei HM, Xu K, Xu XY (2015) Comparative analysis of drought based on precipitation and soil moisture indices in Haihe basin of North China during the period of 1960–2010. J Hydrol 526:55–67

Reddy MJ, Singh VP (2014) Multivariate modeling of droughts using copulas and meta-heuristic method. Stoch Env Res Risk A 28(3):475–489

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

Salvadori G, De Michele C (2010) Multivariate multiparameter extreme value models and return periods: a copula approach. Water Resour Res 46:W10501

Salvadori G, De Michele C (2015) Multivariate real-time assessment of droughts via copula-based multi-site Hazard Trajectories and Fans. J Hydrol 526(SI):101–115

Salvadori G, De Michele C, Kottegoda NT, Rosso R (2007) Extremes in nature. An approach using copulas. Springer, Dordrecht

Savu C, Trede M (2010) Hierarchies of Archimedean copulas. Quant Finan 10(3):295–304

Serinaldi F, Bonaccoroso B, Cancelliere A, Grimaldi S (2009) Probabilistic characterization of drought properties through copulas. Phys Chem Earth Parts A/B/C 34(10–12):596–605

Serinaldi F, Grimaldi S (2007) Fully nested 3-copula: Procedure and application on hydrological data. J Hydrol Eng 12(4):420–430

She DX, Mishra AK, Xia J, Zhang LP, Zhang X (2016) Wet and dry spell analysis using copulas. Int J Climatol 36(1):476–491

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

Shiau JT, Modarres R (2009) The copula-based drought severity-duration frequency analysis in Iran. J Appl Meteorol 16(4):481–489

Sklar A (1959) Fonctions de répartition à n dimensions et leursmarges. Publ Inst Statist Univ Paris 8:229–231

Song S, Singh VP (2010) Meta-elliptical copulas for drought frequency analysis of periodic hydrologic data. Stoch Env Res Risk A 24:425–444

Tsakiris G, Kordalis N, Tigkas D, Tsakiris V, Vangelis H (2016) Analysing drought severity and areal extent by 2D Archimedean copulas. Water Resour Manag 30:1–13

Vicente-Serrano SM, Begueria S, Lopez-Moreno JI (2010) A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index. J Clim 23(7):1696–1718

Wang H, Chen Y, Pan Y (2015) Characteristics of drought in the arid region of northwestern China. Clim Res 62: 99–113

Whelan N (2004) Sampling from Archimedean copulas. Quant Finan 4(3):339–352

Wilhite DA, Sivakumar MV, Pulwarty R (2014) Managing drought risk in a changing climate: the role of national drought policy. Weather Clim Extrem 3:4–13

Wong G, Lambert MF, Leonard M, Metcalfe AV (2010) Drought analysis using trivariate copulas conditional on climatic states. J Hydrol Eng 15(2):129–141

Wu ZY, Lu GH, Wen L, Lin CA (2011) Reconstructing and analyzing China's fifty-nine year (1951–2009) drought history using hydrological model simulation. Hydrol Earth Syst Sci Discuss 8(1):1861–1893

Xu K, Yang D, Yang H, Li Z, Qin Y, Shen Y (2015b) Spatio-temporal variation of drought in China during 1961–2012: A climatic perspective. J Hydrol 526:253–264

Xu K, Yang DW, Xu XY, Lei HM (2015a) Copula-based drought frequency analysis considering the spatiotemporal variability in Southwest China. J Hydrol 527:630–640

Yang P, Xiao Z, Yang J, Liu H (2013) Characteristics of clustering extreme drought events in China during 1961–2010. Acta Meteor Sin 27(2):186–198

Zargar A, Sadiq R, Naser B, Khan FI (2011) A review of drought indices. Environ Rev 19:333–349

Zhai J, Su B, Gao C, Jiang T (2010) Spatial variation and trends in PDSI and SPI indices and their relation to streamflow in 10 large regions of China. J Clim 23:649–663

Zhang D, Chen P, Zhang Q, Li X (2017) Copula-based probability of concurrent hydrological drought in the Poyang lake-catchment-river system (China) from 1960 to 2013. J Hydrol 553:773–784

Zhang L, Singh VP (2006) Bivariate flood frequency analysis using the copula method. J Hydrol Eng 11(2):150–164

Zhang L, Singh VP (2007) Gumbel-Hougaard copula for trivariate rainfall frequency analysis. J Hydrol Eng 12(4):409–419

Zhang L, Xiao J, Li J, Wang K, Lei L, Guo H (2012) The 2010 spring drought reduced primary productivity in southwestern China. Environ Res Lett 7(4):045706

Zhang Q, Qi T, Singh VP et al (2015b) Regional Frequency Analysis of Droughts in China: A Multivariate Perspective. J Water Resour Manag 29(6):1767–1787

Zhang Q, Xiao MZ, Singh VP (2015a) Uncertainty evaluation of copula analysis of hydrological droughts in the East River basin, China. Glob Planet Chang 129:1–9

Zhao CH, Deng XZ, Yuan YM, Yan HW, Liang HM (2013) Prediction of drought risk based on the WRF model in Yunnan province of China. Adv Meteorol 2013

Zhao J, Huang Q, Chang JX, Liu DF, Huang SZ, Shi XY (2015) Analysis of temporal and spatial trends of hydro-climatic variables in the Wei River Basin. Environ Res 139:55–64

Zhao S 1983 A new scheme for comprehensive physical regionalization in China. Acta GeographicaSinica, 1–10. (In Chinese with English Abstract)

Affiliations

Olusola O. Ayantobo [1,2,3] · **Yi Li** [1,2] · **Songbai Song** [1]

✉ Yi Li
liyikitty@126.com

[1] College of Water Resources and Architectural Engineering, Northwest Agriculture & Forestry University, Yangling 712100 Shaanxi, China

[2] Institute of Water-Saving Agriculture in Arid Areas of China, Northwest Agriculture & Forestry University, Yangling 712100 Shaanxi, China

[3] Department of Water Resources Management and Agricultural-Meteorology, Federal University of Agriculture, Abeokuta PMB 2240, Nigeria