Copula-based drought severity-area-frequency curve and its uncertainty, a case study of Heihe River basin, China

Zhanling Li, Quanxi Shao, Qingyun Tian and Louie Zhang

ABSTRACT

Copulas are appropriate tools in drought frequency analysis. However, uncertainties originating from copulas in such frequency analysis have not received significant consideration. This study aims to develop a drought severity-areal extent-frequency (SAF) curve with copula theory and to evaluate the uncertainties in the curve. Three uncertainty sources are considered: different copula functions, copula parameter estimations, and copula input data. A case study in Heihe River basin in China is used as an example to illustrate the proposed approach. Results show that: (1) the dependence structure of drought severity and areal extent can be modeled well by Gumbel; Clayton and Frank depart the most from Gumbel in estimating drought SAF curves; (2) both copula parameter estimation and copula input data contribute to the uncertainties of SAF curves; uncertainty ranges associated with copula input data present wider than those associated with parameter estimations; (3) with the conditional probability decreasing, the differences in the curves derived from different copulas are increasing, and uncertainty ranges of the curves caused by copula parameter estimation and copula input data are also increasing. These results highlight the importance of uncertainty analysis of copula application, given that most studies in hydrology and climatology use copulas for extreme analysis.

Key words | copula, drought, input data, parameter estimation, SPEI, uncertainty

Zhanling Li (corresponding author)
Qingyun Tian
School of Water Resources and Environment, MOE
    Key Laboratory of Groundwater Circulation and
    Environmental Evolution,
China University of Geosciences,
Beijing 100083,
China
E-mail: *zhanling.li@cugb.edu.cn*

Quanxi Shao
Louie Zhang
CSIRO Data 61,
Private Bag 5, Wembley, WA 6913,
Australia

INTRODUCTION

Drought is a stochastic natural phenomenon that arises from considerable deficiency in precipitation. With the global temperature increasing, it becomes more common and has great effects on terrestrial ecosystems (Mishra & Singh 2009; Fang et al. 2018a; Han et al. 2018). The frequency analysis of drought is important since it is capable of quantifying the drought characteristics (Rahmat et al. 2015). Bivariate or even multivariate frequency analysis is more informative than univariate analysis due to the complexity and multivariate features of drought (Yu et al. 2014; Jiang et al. 2015; Rahmat et al. 2015; Amiratee et al. 2018).

Previous literature has been most concerned with the severity and the duration of drought due to their ability to characterize drought from the temporal scale (e.g., Lee & Kim 2011; Beguería et al., 2014; Rahmat et al. 2015; Tan et al. 2015; Afshar et al. 2016; Homdee et al. 2016; Fang et al. 2018b; Hameed et al. 2018). However, the spatial extent of drought is also very important in drought assessments. Recent researches have shown that the drought area has been significantly increasing globally and regionally during the last several decades. The percentage of areas under drought had increased by about 1.74% per decade from 1950 to 2008 globally (Dai 2011). The drought-affected area increased nearly 12-fold from 1950 to 2009 in China (Wang et al. 2012). A significant increase in drought area has been detected in the northwest of China, with an increasing rate of 3.96% per decade since the late 1990s

(Yu *et al.* 2014). Therefore, taking both severity and areal extent into account is very necessary for better understanding drought from both a temporal and spatial scale, and it would be valuable for drought-based decision-making.

Applying the concept of quantifying the frequency analyses of both severity and areal extent based on the copula theory, the drought severity-area-frequency (SAF) curves can be derived. These curves are very useful in drought planning and management since they can provide the probabilities and return periods of the drought areal extent with different severities over a region (Reddy & Ganguli 2013; Amirataee *et al.* 2018). However, uncertainties are usually involved in the estimation of drought SAF curves due to the effects of different copula functions, marginal distributions, input data, parameter estimations, etc.

The former uncertainty sources, copula functions and marginal distributions, are of major concern in previous literatures. For example, Halwatura *et al.* (2016) compared four marginal distribution functions and two types of copulas, and assessed the impacts of these uncertainties on the estimations of recurrence intervals (RIs) of drought events. They stated that no differences were found between the RIs derived from the different marginal distributions for mild drought events, and the effects of different copulas on RIs can nearly be negligible in their study. However, Zhao *et al.* (2017) drew a slightly different conclusion by comparing six marginal distributions and three Archimedean copulas in a case study of the Weihe River basin. They pointed out that the return period of drought varies depending on the selected marginal distributions and copula functions, and the differences become larger with the increase of return periods. Zhang *et al.* (2015) evaluated the influences of marginal distributions to the uncertainty of joint distribution by employing a Bayesian framework, and the results showed that the stronger the tail of marginal distribution, the greater the uncertainty of joint distribution, especially for extreme events. All the above studies made great contributions in enriching the uncertainty analysis in drought frequency curves. However, the uncertainties caused by copula parameter estimations and copula input data are rarely regarded. Sadegh *et al.* (2017) presented a multivariate copula analysis toolbox which has the ability to quantify the uncertainties in drought frequency curves associated with copula parameter estimations and copula

functions based on Bayesian framework; whereas the fixed drop-down menu options in the toolbox and the non-editable display for the outputs restrict its wide application to some extent.

This paper aims to develop a drought SAF curve with copula theory and to evaluate the uncertainties in the curve caused by different copula functions, copula parameter estimations, and copula input data. The following issues will be addressed, namely: (i) to compare and select the copula function with best performance in modeling the dependence structure of drought severity and drought areal extent; (ii) to construct a drought SAF curve based on the conditional probability distribution in the study area; and (iii) to quantify the uncertainties in drought SAF curves from three uncertainty sources.

METHODS

Drought index and its characteristics

A number of indices have been developed so far for drought quantification, monitoring, and analysis, such as PDSI (Palmer Drought Severity Index), SPI (Standardized Precipitation Index), and SPEI (Standardized Precipitation-Evapotranspiration Index), etc. Under the current climate warming conditions, many studies show that, SPEI, developed by Vicente-Serrano *et al.* (2010), could give a better performance in drought assessments due to the fact that it combines the multi-scalar character with the capacity of involvement of temperature effects on droughts (e.g., Tan *et al.* 2015; Homdee *et al.* 2016). Thus, SPEI was selected and 12-monthly scale of SPEI will be analyzed in this study. The temperature-based Thornthwaite method is used to calculate potential evapotranspiration (PET) as it only requires monthly mean temperature data and the latitudinal coordinate of the location, which are readily available at most meteorological stations. SPEI is computed with the assumption that P-PET (precipitation minus potential evapotranspiration) series follow the Log-logistic distributions. Detailed information about SPEI and its calculation can be found in the studies of Vicente-Serrano *et al.* (2010), Beguería *et al.* (2014), and Hameed *et al.* (2018).

Specifically, drought characteristics used here include drought severity (S) and drought areal extent (A). The severity of drought events is defined as the average of SPEI values of all cells in the study area, which is located below the drought threshold (Amirataee *et al.* 2018). It is computed using the equation:

$$S_t = -\frac{\sum_{i=1}^{N} I[SPEI_{i,t} \leq SPEI_{thr}] \times SPEI_{i,t}}{N_{Ct}} \quad \text{for } t = 1, \ldots T$$

(1)

where $I[.]$ is the logical indicator function with the value 1 if its argument is true, and the value 0 otherwise. $SPEI_{i,t}$ is the value of SPEI in cell i and period t, $SPEI_{thr}$ is the threshold of SPEI. N_{Ct} represents the total number of drought cells in period t, N means the total number of cells in the study area, T means the data record length. Positive SPEI values indicate wet periods, while negative values indicate dry periods compared with the normal conditions of the area (Vicente-Serrano *et al.* 2010). Light drought (-0.5 to -0.99), moderate drought (-1.0 to -1.49), severe drought (-1.5 to -1.99), and extreme drought (≤ 2.0) are defined according to SPEI values (Vicente-Serrano *et al.* 2010). Drought events occur when the SPEI is continuously negative and falls below a certain threshold value. A threshold value of -0.5 is selected here. In order to represent the extreme condition and to analyze the associated risk of droughts using the exceedance probability, the negative drought severity is transformed to positive values for fitting to the available distributions (Mishra & Singh 2009; Amirataee *et al.* 2018), thus, there is a negative sign in Equation (1).

The areal extent of drought is defined as a specified percent of total area in period t in which the value of SPEI is below the specified threshold (Reddy & Ganguli 2013; Amirataee *et al.* 2018). It is computed using the following equation:

$$A_t = \frac{\sum_{i=1}^{N} I[SPEI_{i,t} \leq SPEI_{thr}] \times a_i}{\sum_{i=1}^{N} a_i} \quad \text{for } t = 1, \ldots T$$

(2)

where a_i is the area of cell i. Since SPEI is calculated originally at a limited number of sites where observations on climate variables are available, inverse distance weighting (IDW) spatial interpolation technology is generally employed to obtain the spatial distribution of SPEI values.

Copula-based conditional distribution functions

Suppose X and Y are two continuous random variables with marginal distribution functions CDFs u_1 and u_2, then there exists a copula C such that:

$$F_{XY}(x, y) = P(X \leq x, \ Y \leq y) = C(u_1, u_2)$$

(3)

The conditional density functions are related to $C(u_1, u_2; \theta)$ through:

$$C(u_2|u_1; \theta) = P(U_2 \leq u_2|U_1 = u_1) = \frac{\partial^2 C(u_1, u_2; \theta)}{\partial u_1 \partial u_2}$$

(4)

and when it is integrated with respect to u_2,

$$C(u_2|u_1; \theta) = \int_0^{u_2} C(v|u_1; \theta)dv = \int_0^{u_2} \frac{\partial^2 C(u_1, v; \theta)}{\partial u_1 \partial v}dv$$

$$= \frac{\partial}{\partial u_1} \int_0^{u_2} \frac{\partial C(u_1, v; \theta)}{\partial v}dv = \frac{\partial C(u_1, u_2; \theta)}{\partial u_1}$$

(5)

where $(U_1, U_2) \sim C$, and C is a bivariate copula distribution function with parameter θ, representing the bivariate dependence structure of variables X and Y. $C(u_2|u_1; \theta)$ is the numeric vector of the conditional distribution function of the copula with parameter θ evaluated at u_2 given u_1.

Different copulas differ in describing the dependence structures. Comparisons of copulas could be necessary before bivariate drought frequency analysis. Ten copulas are selected and given in Table 1, including not only one-parameter Archimedean families such as Clayton, Gumbel, Frank, and Joe, but also two-parameter copulas, BB1 (Clayton–Gumbel), BB6 (Joe–Gumbel), BB7 (Joe–Clayton), and BB8 (Joe–Frank). The more flexible structures of two-parameter copulas allow for different non-zero lower and upper tail dependence coefficients (Brechmann & Schepsmeier 2013). Many of the above copulas are commonly used in hydrological and climatological applications

Table 1 │ The cumulative density functions and parameter ranges of alternative copulas

No.	Copula family	$C(u_1, u_2)$	Parameter space
1	Normal	$\Phi_\theta(\Phi^{-1}(u_1),\ \Phi^{-1}(u_2))$	$-1 \leq \theta \leq 1$
2	Student-t	$t_{\theta,\delta}(t_\delta^{-1}(u_1),\ t_\delta^{-1}(u_2))$	$\theta > 0$
3	Clayton	$u_1^{-\theta} + u_2^{-\theta} - 1^{-\frac{1}{\theta}}$	$\theta > 0$
4	Gumbel	$\exp\left\{-[(-lnu_1)^\theta + (-lnu_2)^\theta]^{\frac{1}{\theta}}\right\}$	$\theta \geq 1$
5	Frank	$-\dfrac{1}{\theta}ln\left[1 + \dfrac{(e^{-\theta u_1}-1)(e^{-\theta u_2}-1)}{e^{-\theta}-1}\right]$	$\theta \neq 0$
6	Joe	$1 - [(1-u_1)^\theta + (1-u_2)^\theta - (1-u_1)^\theta(1-u_2)^\theta]^{\frac{1}{\theta}}$	$\theta \geq 1$
7	BB1	$\left\{1 + [(u_1^{-\delta}-1)^\theta + (u_2^{-\delta}-1)^\theta]^{\frac{1}{\theta}}\right\}^{-\frac{1}{\delta}}$	$\delta > 0, \theta \geq 1$
8	BB6	$1 - \left\{1 - exp\left[-\{[-\ln(1-(1-u_1)^\theta)]^\delta + [-\ln(1-(1-u_2)^\theta)]^\delta\}^{\frac{1}{\delta}}\right]\right\}^{\frac{1}{\theta}}$	$\delta > 0, \theta \geq 1$
9	BB7	$1 - \left\{1 - [(1-(1-u_1)^\theta)^{-\delta} + (1-(1-u_2)^\theta)^{-\delta} - 1]^{-\frac{1}{\delta}}\right\}^{\frac{1}{\theta}}$	$\delta > 0, \theta \geq 1$
10	BB8	$\dfrac{1}{\delta}\left(1 - \left[1 - \dfrac{(1-(1-\delta u_1)^\theta)(1-(1-\delta u_2)^\theta)}{1-(1-\delta)^\theta}\right]^{\frac{1}{\theta}}\right)$	$\delta > 0, \theta \geq 1$

Note: For Normal, Φ_θ is the bivariate joint normal distribution with linear correlation coefficient θ, and Φ is standard normal marginal distribution.
Formulas in this table are from Chen & Khashanah (2015).

(e.g., Xu *et al.* 2015; Zhang *et al.* 2015, 2017; Zhao *et al.* 2017; Ayantobo *et al.* 2018).

Copula parameter estimation and goodness of fit

The copula parameter θ (or θ and δ) in Table 1 helps to measure the extent of relationship between variables (Ayantobo *et al.* 2018). Its estimation from sample data is not straightforward. One of the commonly used approaches is maximum likelihood estimation (MLE) (Brechmann & Schepsmeier 2013). Although the solution to MLE involves the roots of a cubic equation, and cannot be written in a simple form, the MLE can be easily found using numerical routines. The standard errors for MLE copula parameter estimations are based on inversion of the Hessian matrix (Brechmann & Schepsmeier 2013).

The best copula is selected according to the Akaike and Bayesian information criteria (AIC and BIC, respectively).

The AIC and BIC are defined as:

$$\text{AIC} = 2D - 2\ell \tag{6}$$

$$\text{BIC} = D\ln n - 2\ell \tag{7}$$

where D is the number of parameters in the model, ℓ is the log-likelihood value of the best parameter set, and n represents the sample size. The derivations of both equations appear to be similar. A lower AIC or BIC value associates with a better copula model. The AIC takes into account both the complexity of the model and minimization of model error residuals and provides a more robust measure of quality of model predictions. It avoids the problem of over-parameterization by adding a penalty term based on the number of parameters (Sadegh *et al.* 2017; Sun *et al.* 2018). The BIC equation differs from the AIC equation with respect to its first term that depends on the sample

size. The penalty term of BIC is larger than that of AIC, when the number of samples is too large, it can effectively prevent the model complexity from being too high due to the excessive model accuracy (Brechmann & Schepsmeier 2013).

Uncertainty analysis

According to MLE, the optimal value of copula parameter can be derived, together with its standard errors which can be used to represent the parameter uncertainty ranges (without considering the uncertainty in marginal distributions in severity and areal extent). The effects of copula parameter estimations to drought SAF curves are quantified by random generating copula parameters based on the asymptotic normality first, then calculating the drought curves according to the new generating copula parameter values, and finally identifying the $\alpha/2$th and $(1-\alpha)/2$th percentiles of drought curves which are defined as the lower and upper limits of confidence intervals with given probabilities. These confidence intervals are the uncertainty ranges of drought curves caused by copula parameter estimations.

To quantify the effects of copula input data uncertainty to drought SAF curves, the non-parametric bootstrap method, which has usually been used in quantifying the sample uncertainty in previous studies (e.g., Hu *et al.* 2015; Vergni *et al.* 2017), is selected. The idea behind this method is that since the distribution from which the sample has been taken is unknown, the values in that sample are the best guide to the true distribution. The theoretical details of the bootstrap approach can be found in Efron (1979). The copula input data in this case consists of n sets of drought severity (S) and drought areal extent (A), denoted as (S_1, A_1), (S_2, A_2), …, (S_i, A_i), …, (S_n, A_n), in which n means the data length. The S_i and A_i series are fitted by theoretical probability distributions and the best distribution is selected based on goodness of fit tests. The specifics of bootstrap procedure are as follows:

1. In the jth bootstrap, resampling the original n set values of drought variables with replacement, and generating a bootstrapped sample of the same size n as ((S_1^{j*}, A_1^{j*}), (S_2^{j*}, A_2^{j*}), ……, (S_i^{j*}, A_i^{j*}), ……, (S_n^{j*}, A_n^{j*}).

2. Using the probability distribution type which has been pre-specified according to the original severity and areal extent series, to fit the S^{j*} and A^{j*} for obtaining their optimal parameter estimates and the corresponding marginal distributions, CDF(S^{j*}) and CDF(A^{j*}).

3. Using CDF(S^{j*}), CDF(A^{j*}) and the best candidate copula pre-selected according to the original marginal distributions, to estimate the copula parameters $\hat{\theta}_i^{j*}$ ($i = 1$, or $i = 1, 2$).

4. Calculating drought severities at given drought areal extents according to the conditional probability functions, and then obtaining the drought curve SAF^{j*}.

5. The above procedure is repeated m times to generate m bootstrap samples, m estimations of marginal distributions for severity and areal extent, m sets of copula parameters $\hat{\theta}_i^{j*}$, and m drought curves SAF^{j*} ($j = 1, 2, …, m$).

6. From drought curves SAF^{j*} ($j = 1, …, m$), identifying the $\alpha/2$th and $(1-\alpha)/2$th percentiles of drought curves which is defined as the lower and upper limits of confidence intervals with given probabilities, and then analyzing the contributions of copula input data to the uncertainty of drought curves.

The main steps of this study are as follows. The values of 12-month SPEI at the stations are calculated first. Then, the drought severity and drought areal extent are determined according to IDW spatial interpolation technique. The two drought variables are fitted by a variety of theoretical probability distributions. After that, copulas are used to compare the performances of capturing the dependence structures of the two drought variables with goodness of fit tests. Drought SAF curves are then extracted with the best copula. Finally, the uncertainties in SAF curves caused by different copulas, copula parameter estimations, and copula input data, are investigated.

STUDY AREA AND DATA

The Heihe River basin, which falls between 97.1°–102.0°E and 37.7°–42.7°N, is the second largest typical inland river basin in the northwest of China with a controlled catchment area of 143,000 km² (Figure 1). It is one of the most drought-
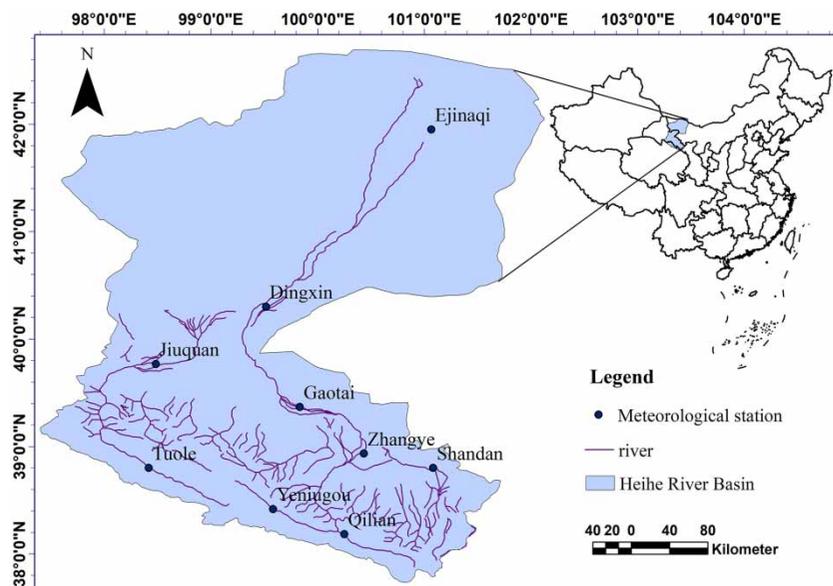
Figure 1 | Locations of the meteorological stations in Heihe River basin.

prone river basins in China (Wang *et al.* 2016; Ayantobo *et al.* 2018; Niu *et al.* 2019) with an annual precipitation of approximately 150 mm across the basin, an annual average air temperature of 6–8 °C, and annual potential evaporation of 1,000–2,000 mm. The runoff of the basin mainly comes from the upper reach, and consumed by large areas of irrigation in the middle reach and oasis ecosystem in the lower reach. Severe drought events would lead to inadequate water supply and thus affect the local agriculture and vulnerable oasis ecosystem. In this study, monthly precipitation and temperature data spanning from 1960 to 2015 were collected from nine meteorological stations within the basin.

The locations and the detailed information of stations are shown in Figure 1 and Table 2.

RESULTS AND ANALYSIS

Drought characteristics and their dependence structures

Drought severity and areal extent series over the study area determined by Equations (1) and (2) are plotted in Figure 2. It is clear that, before 1997, the drought was mainly

Table 2 | Climatic and geographical characteristics of stations used in the study during 1960–2015

| No.	Station	Longitude	Latitude	Elevation (mm)	Annual precipitation (mm)	Annual mean temperature (°C)
1	Tuole	98.25	38.48	3,367	298	−2.5
2	Yeniugou	99.36	38.26	3,320	418	−2.8
3	Qilian	100.15	38.11	2,787	409	1.1
4	Shandan	101.05	38.48	1,765	201	6.6
5	Zhangye	100.17	39.05	1,483	128	7.5
6	Jiuquan	98.29	39.46	1,477	86	7.6
7	Gaotai	99.50	39.22	1,332	108	7.9
8	Dingxin	99.31	40.18	1,177	55	8.5
9	Ejinaqi	101.04	41.57	941	35	9.0
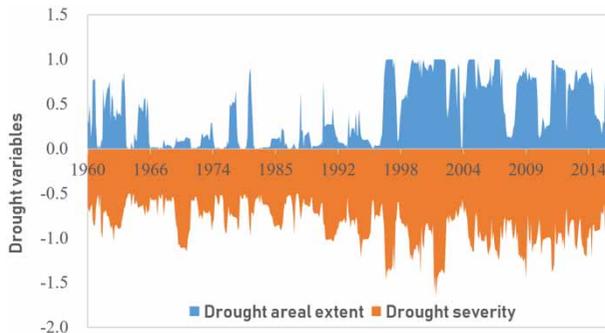
Figure 2 | Drought characteristics of Heihe River basin during the period of 1960–2015.

characterized by light drought conditions. However, several moderate or even severe droughts were captured after 1997, and furthermore, drought events with stronger severity and larger areal extent occurred more frequently than before.

It is necessary to evaluate the dependency between variables prior to the application of copula when analyzing the bivariate joint probability distribution of such random variables (Amirataee *et al.* 2018). For this purpose, Pearson's, Spearman's, and Kendall's correlation coefficients are applied to explore the dependency between drought severity and drought areal extent. These correlation coefficients are often used in measuring the monotone association. The first belongs to the parametric method and sensitive to outliers, and the latter two belong to the non-parametric methods and usually suggested for non-normally distributed data (Shong 2010).

The correlation coefficients between the two drought variables are 0.678, 0.722, and 0.549, respectively, and all the corresponding *p*-values are less than the significance level of 0.05. It is concluded that there is a significant correlation between the variables of drought severity and drought areal extent at 95% confidence level.

Marginal distributions and copula goodness of fit

Determination of marginal distribution for drought variables

The first step in copula fitting consists of the identification of appropriate marginal distributions for the two drought variables. Since they are typically unknown, we first fit them by using several theoretical distributions. According to

previous studies (e.g., Loukas & Vasiliades 2004; Serinaldi *et al.* 2009; Reddy & Ganguli 2013; Afshar *et al.* 2016; Amirataee *et al.* 2018), ten univariate probability distributions (Exponential, Normal, Lognormal, Logistic, Loglogistic, Gamma, Weibull, Log-Pearson III, Burr, and GEV distribution) are tested for fitting drought severity (*S*). Since the drought areal extent has definite ranges from 0 to 1, the bounded distributions, Beta, Power function, Pert, and Uniform, are tested as its candidate margins. Amirataee *et al.* (2018) stated that Beta distribution performed satisfactorily in fitting the drought area extent in their study. Kolmogorov–Smirnov (K-S), Anderson–Darling (A-D), and Chi-squared (C-S) are used to test the goodness of fit.

Results are shown in Table 3. The critical values for the three tests are 0.061, 2.502, and 15.507 at 0.05 significance level, respectively. As shown, GEV is selected finally as the marginal distribution for drought severity, because it not only passes the K-S, A-D, and C-S tests at 0.05 significance levels (all the statistic values are lower than the critical values for the three tests), but shows the best performance with the smallest statistic values. This distribution was also proved to be suitable in fitting the drought severity in many other studies (e.g., Lee & Kim 2011; Reddy & Ganguli 2013). As to the variable of drought areal extent, none of the distributions passes the tests. Then, we perform the fitting process based on its pseudo observations (Ghoudi & Rémillard 2004). In light of the references of Ghoudi & Rémillard (2004) and Kojadinovic *et al.* (2010), it is acceptable to replace the unknown marginal by their empirical counterparts obtained from the pseudo observations.

Goodness of fit of copula functions

The results of modeling the joint probability distribution of the two drought variables by using ten alternative copula functions are summarized in Table 4, including the statistic values of goodness of fit tests (AIC, BIC, and log-likelihood), the optimal parameter estimations, and their standard errors. Clearly, nearly all the copulas have close AIC, BIC, and log-likelihood values except for Frank which shows the greatest departure from the others, indicating that most of the selected copulas have similar performance in capturing the

Table 3 │ Goodness of fit of theoretical probability distributions for drought severity (S) and drought areal extent (A)

Drought variable	No.	Probability distribution	Kolmogorov–Smirnov (K-S)	Anderson–Darling (A-D)	Chi-squared (C-S)
Drought severity (S)	1	Exponential	0.4475	125.74	810.61
	2	Normal	0.0680	3.87	10.14
	3	Lognormal	0.0434	1.55	10.50
	4	Logistic	0.0776	5.28	24.32
	5	Loglogistic	0.0580	2.93	24.79
	6	Gamma	0.0452	1.44	9.48
	7	Weibull	0.0841	7.45	23.22
	8	Log-Pearson III	0.0431	1.60	12.14
	9	Burr	0.0481	2.10	17.74
	10	GEV	*0.0387*	*1.30*	*8.79*
Drought areal extent (A)	1	Beta	0.1806	35.46	101.81
	2	Power function	*0.1266*	38.91	112.88
	3	Pert	0.1424	37.16	188.62
	4	Uniform	0.1767	*16.43*	*48.12*

Note: The minimum statistic values for K-S, A-D, and C-S tests are in bold italics.

Table 4 │ Goodness of fit of different copulas and their optimal parameter estimations

No.	Copula	Akaike information criteria (AIC)	Bayesian information criteria (BIC)	Log likelihood	Optimal parameter (standard error)
1	Normal	−465.60	−461.38	233.80	0.79 (0.006)
2	Student-t	−465.43	−460.00	234.72	0.79, 7.98 (0.019, 6.082)
3	Clayton	−442.62	−438.41	222.31	2.01 (0.415)
4	Gumbel	*−473.45*	*−469.24*	*237.73*	2.27 (0.114)
5	Frank	−364.49	−360.27	183.24	6.24 (1.637)
6	Joe	−433.61	−429.40	217.81	2.77 (0.190)
7	BB1	−490.71	−482.28	247.36	0.92, 1.60 (0.138, 0.096)
8	BB6	−471.45	−463.01	237.72	1.00, 2.27 (0.230, 0.350)
9	BB7	−514.65	−506.21	259.33	2.09, 1.82 (0.127, 0.151)
10	BB8	−435.22	−426.79	219.61	2.89, 0.99 (0.145, 0.004)

Note: The minimum AIC, BIC and log likelihood values are in bold italics.

dependence structures of drought severity and drought areal extent in this study case. Comparatively, Gumbel, among the one-parameter copula functions, provides the best fit with the smallest AIC (−473.45) and BIC (−469.24) values. Among the two-parameter copulas, BB7 performs the best. Student-t and BB8 copulas even perform inferior to Gumbel according to the corresponding AIC and BIC values.

To better understand the performance of each copula in modeling the drought variables in this study, the generated samples derived from the estimated theoretical copula by Monte Carlo technique are compared with the observed data for checking whether they have similar features. A set

of 5,000 samples are generated, and the scatter plot of generated samples, together with the observed datasets are shown in Figure 3. Different copula functions perform a little differently. The Normal, Student-t, and Frank are symmetric and can be used to model symmetric correlation patterns of variables. Gumbel and Clayton are asymmetric and can model the asymmetrical tail correlation patterns. Gumbel exhibits greater correlation in the upper tail, whereas Clayton exhibits greater correlation in the lower tail (Aas *et al.* 2009).

From the scatterplot, the observed data in the two opposite corners appear clustering, seemingly displaying both upper- and lower-tail dependence, while the dependence
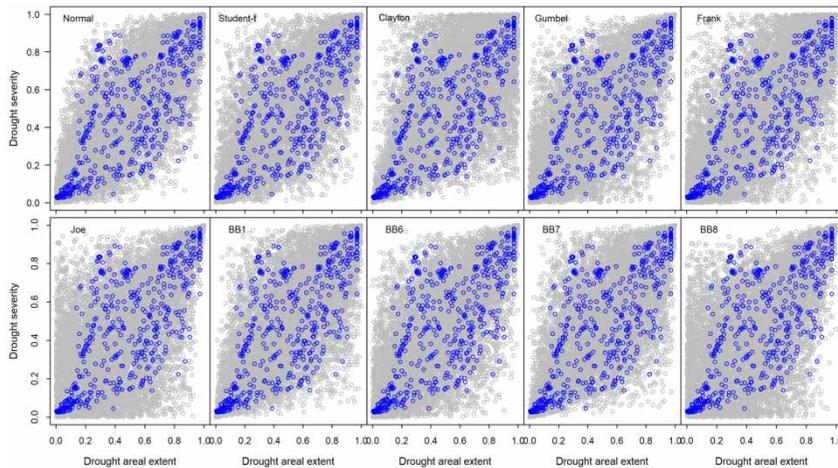
Figure 3 | Generated and observed CDF for drought severity and drought areal extent based on different copula functions. Gray dots present the Monte Carlo simulations from each copula and blue dots present the observed data. Please refer to the online version of this paper to see this figure in color: http://dx.doi.10.2166/nh.2020.173.

structures in the tails appear slightly asymmetric and a little skewed. Therefore, asymmetric copulas may be a good choice. Combined with the results of goodness of fit presented in Table 3, Gumbel is then selected as the benchmark copula in this study. The two-parameter copulas, BB1, BB6, and BB7, are not selected herein since, on one hand, they show similar performance with Gumbel and, on the other hand, the higher model complexity (with more parameters) would make a stimulation over conditioning of the model to some extent (Sadegh et al. 2017).

Actually, the optimality of copula functions could be dataset- and case-specific (Afshar et al. 2016). For example, Reddy & Ganguli (2013) evaluated three copula families in modeling the joint dependence between the drought severity and drought areal extent in western Rajasthan of India, and found that the Gumbel–Hougaard copula best represents the drought properties in their study. Xu et al. (2015) constructed a trivariate joint distribution based on the copula theory and concluded that the Joe and Gumbel copulas are more suitable to estimate the joint distribution of drought duration, affected area, and severity in southwest China. Amirataee et al. (2018) regarded the Frank copula as the most appropriate in modeling the joint probability of drought severity and percent of area in the Lake Urmia basin, Iran. In our case, with considering the results of goodness of fit and the greater dependence in upper tails of the data, the one-parameter Gumbel is finally selected as the benchmark copula.

Drought SAF curves

The drought SAF curve, developed by using conditional probability function based on Gumbel copula, is plotted in Figure 4. The horizontal axis represents the percentage of drought areal extent and the vertical axis represents the drought severity. Only droughts with conditional probability of less than 0.2 (corresponds to return period of more than five years) are considered in the analysis. It can be seen that the drought event with larger areal extent is generally
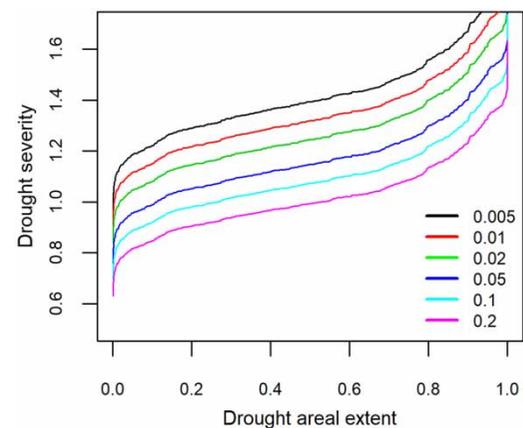

Figure 4 | Drought severity-area-frequency (SAF) curves derived from Gumbel copula over the study area with six conditional probabilities (0.005, 0.01, 0.02, 0.05, 0.1, and 0.2).

associated with higher severity. For example, with a conditional probability of 0.1 (at the return period of ten years), when the drought areal extent amounts to 20%, the average drought severity is 0.98, belonging to light drought; while when the drought areal extent increases to 60% and 80%, the severity increases to 1.17 and 1.30, resulting in moderate droughts.

The drought SAF curve based on the conditional probability can allow a water planner to assess the probability of a drought with a certain areal extent and severity occurring simultaneously. For example, for a drought areal extent reaching 40%, the occurrence probability is 0.05 when the severity is 1.12, and 0.02 when the severity increases to 1.21. This information cannot be obtained from the univariate frequency analysis of droughts, in which only the probability of univariate occurrence can be derived. The copula-based drought SAF curve can provide decision-makers with probability maps of drought severity and useful information on drought areal extent in the study area, yielding important information for water management purposes.

Drought SAF curve uncertainties

Uncertainty caused by different copula functions

In this subsection, whether different copulas lead to significantly different assessments of SAF curves is investigated. Figure 5 shows the conditional probability curves derived from different copula functions. Conditional probabilities of 0.005, 0.01, 0.02, 0.05, 0.1, and 0.2 are compared. Gumbel copula is selected as the benchmark as stated above.

With a conditional probability of 0.2, most of the curves are close to each other, and only slight differences are found between the benchmark and the alternatives, except for the two curves derived from Clayton and Frank. As the conditional probability decreases (conditional return period increases), these differences increase in general, being consistent with the findings of Zhao *et al.* (2017), who pointed out that with an increase of the return period, the drought differences from different copula functions become larger. Such differences are not very large in this study as a whole, except for those from Clayton and Frank.

As to Clayton and Frank, large differences are detected. Interestingly, it is also noted that the AIC and BIC values for

Clayton are −442.62 and −438.41 (Table 3), very close to those for Gumbel (−473.45 and −469.24). However, as shown in Figure 3, the drought SAF curves derived from these two copulas have a considerably larger difference, and Clayton gives larger possibilities of drought severity than Gumbel.

BB6 gives quite close drought severity estimations with Gumbel no matter how much the conditional probability is. BB1, BB7, and Student-t copulas give similar curve shapes with Gumbel, whereas they overestimate the drought severity overall. A similar conclusion can be drawn from Normal when the drought areal extent is higher than 10%, overestimating the drought severities comparatively. Joe and BB8 overestimate the drought severity when the drought area is lower than about 10%, but underestimate the drought severity when the drought areal extent becomes higher.

In summary, the drought SAF curves are influenced by copula choices, and the curves have different conditional probabilities of occurrence depending on which copula is used. BB6 gives quite close estimations in drought severity to Gumbel. Normal, Student-t, BB1, BB7 overestimate the drought severities, while Joe and BB8 underestimate the drought severities as a whole. Clayton and Frank copulas depart the most, and overestimate the drought severity much more. As the conditional probability decreases, the differences in drought severity estimations from different copulas increase, although most of the differences are not large. It implies that the choice of copula should be implemented with caution since different copulas show different performances in tail behaviors, especially when the focus is on extreme events (with small probabilities).

Uncertainty caused by copula parameter estimations

After estimating the copula parameter(s) with the MLE method, 1,000 copula parameters are randomly generated from the Normal distribution with the mean and standard error derived from MLE. The estimated optimal parameter value and standard error for Gumbel copula is 2.27 and 0.114 (see Table 3). Such range of standard error in parameter estimation is then translated to drought SAF curves, rendering the analysis of uncertainty. Combining with the conditional probability function and the generated copula parameters, a set of conditional probability isolines
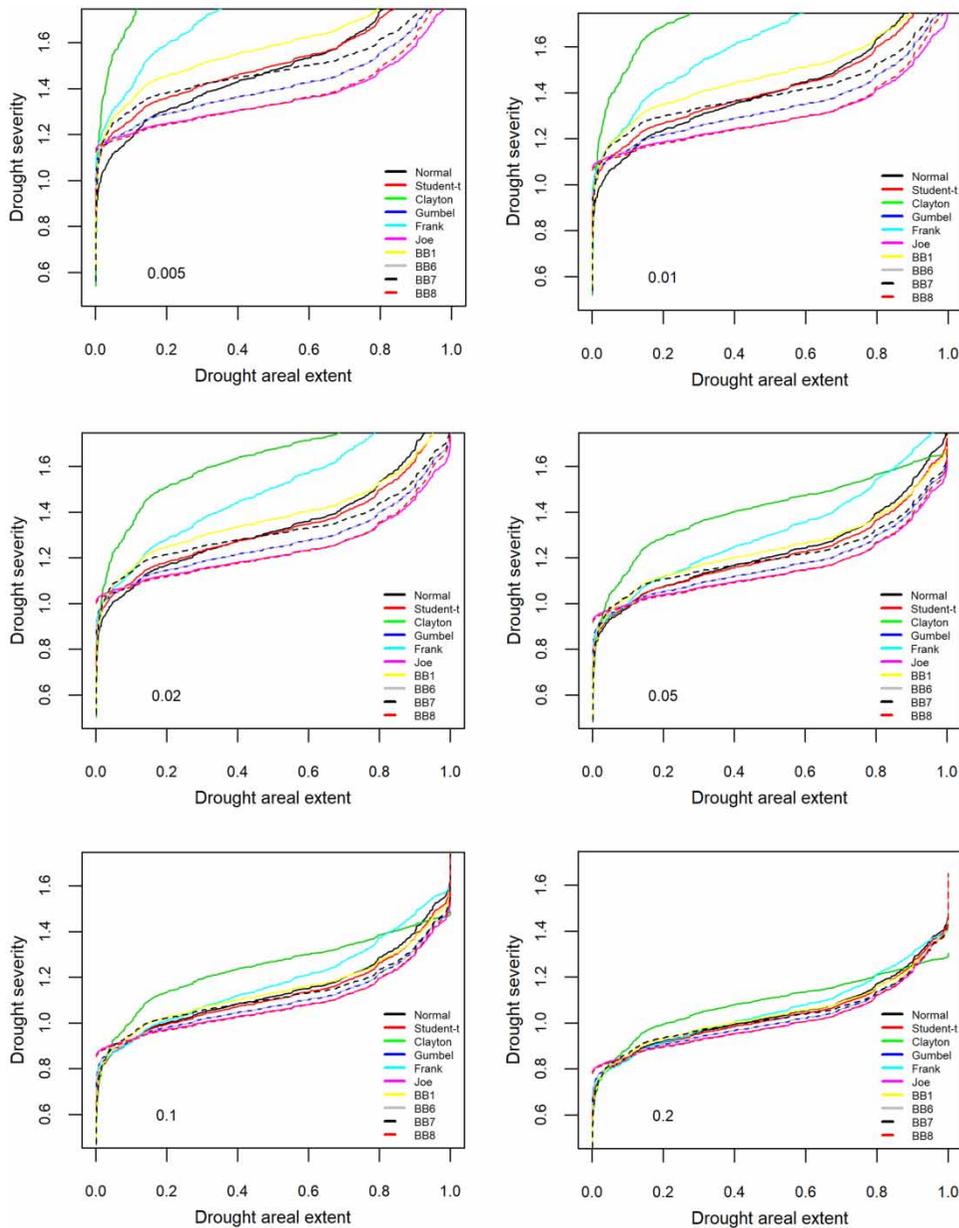
Figure 5 │ Drought severity-area-frequency (SAF) curves derived from different copula functions over the study area with six conditional probabilities (0.005, 0.01, 0.02, 0.05, 0.1, and 0.2).

are then derived, which will be used to analyze the SAF curve uncertainties caused by copula parameter estimations.

Figure 6 presents the 95% CIs of conditional probability isolines associated with the uncertain parameter estimations of Gumbel. To avoid the curve overlapping, the six conditional probabilities are separated into two plots. Obviously, the uncertainty in copula parameter estimation has a certain effect on the conditional probability isolines. The uncertainty ranges of drought

SAF curves become wider with the conditional probability decreasing. This is because the lower the probability of the drought event, the less the number of the observed data, and the less information available for the analysis.

Kendall's tau method is also used to estimate the parameter in Gumbel, yielding an optimal parameter value of 2.22 with a standard error of 0.130, which does not deviate much from the results of the MLE method. The corresponding
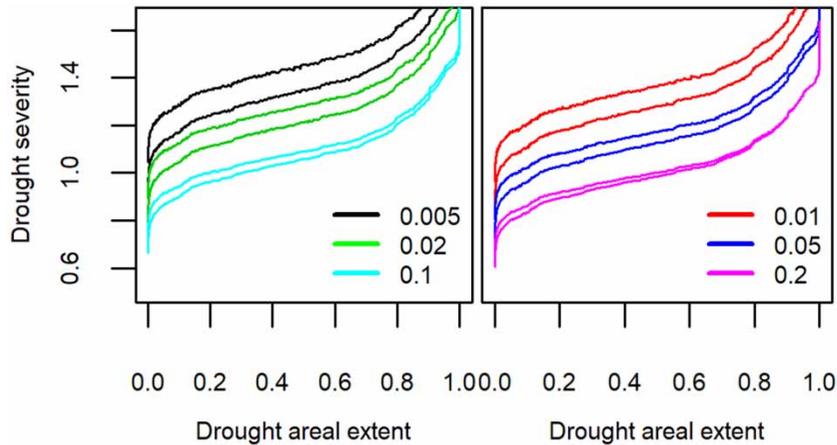
Figure 6 | Uncertainty ranges of drought severity-area-frequency (SAF) curve derived from copula parameter estimations over the study area with six conditional probabilities (0.005, 0.01, 0.02, 0.05, 0.1, and 0.2).

uncertainty ranges of conditional probability isolines are also quite close to the results from the MLE method.

Uncertainty caused by copula input data

Using bootstrap technique to generate 1,000 set of drought samples $((S_1^{j^*}, A_1^{j^*}), (S_2^{j^*}, A_2^{j^*}), \ldots, (S_i^{j^*}, A_i^{j^*}), \ldots, (S_{501}^{j^*}, A_{501}^{j^*})$ $(j = 1, \cdots, 1,000))$. To help visual comparisons of dependence structures among the original and the generated copula input data, the marginal distributions of the original samples and bootstrapped samples of drought severity and drought areal extent are plotted in Figure 7. The density of points in any area represents the density of the copula. It is observed that the bootstrapped samples well preserve the dependence structure between the drought severity and drought areal extent.

Uncertainty bands in Figure 8 reveal the effects of copula input data uncertainty to the estimation of conditional drought severity. Obviously, with the conditional probability decreasing, the uncertainty bands are increasing, similar findings as those from the effects of copula parameter uncertainty. Table 5 presents the lower (2.5% percentile) and upper (97.5% percentile) limits and the ranges of estimated drought severity at a certain drought areal extent owing to both input data uncertainty and parameter estimation uncertainty. The ranges resulting from both uncertainty sources grow remarkably as the conditional probability decreases. Compared with those caused by copula parameter, the uncertainty bands caused
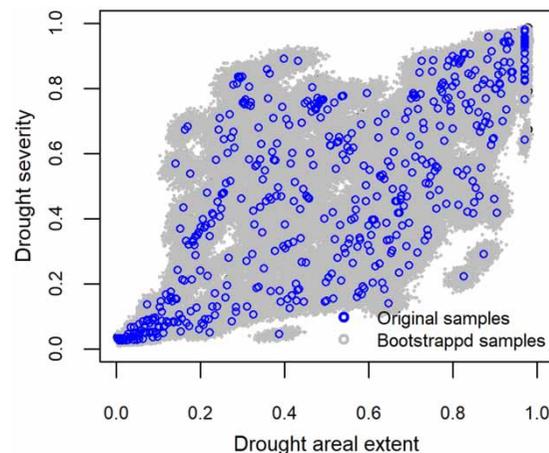

Figure 7 | Dependence structures between drought severity and drought areal extent from 1,000 bootstrapped samples and the original samples.

by copula input data present more widely. This reflects that the uncertainty of input data brings more uncertainty to drought SAF curves than that of copula parameter estimations. In other words, to improve the quality of copula input data can reduce the uncertainty of drought SAF curves more effectively.

CONCLUSIONS

In this study, 12-month SPEI is calculated over the Heihe River basin in northwest China, the drought SAF curve is developed, and then the uncertainties in the curve
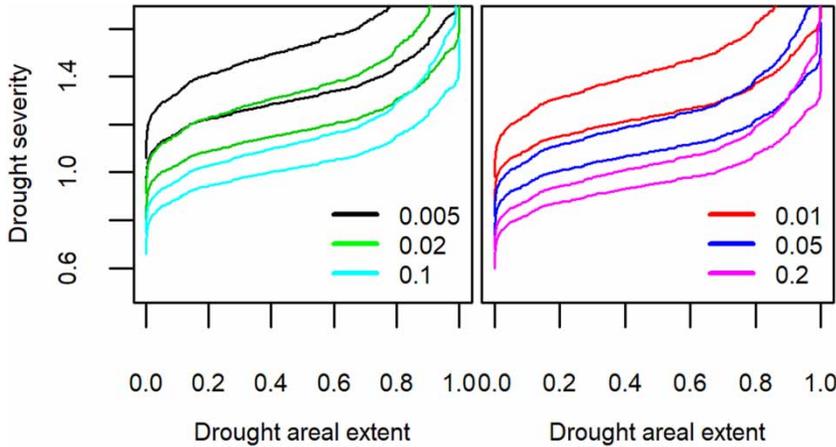
Figure 8 | Uncertainty ranges of drought severity-area-frequency (SAF) curve derived from copula input data over the study area with six conditional probabilities (0.005, 0.01, 0.02, 0.05, 0.1, and 0.2).

Table 5 | Uncertainty bands of the estimated drought severity at given drought areal extent over the study area

	Conditional probability					
Areal extent	0.005	0.01	0.02	0.05	0.1	0.2
20%-A	0.158 [1.088, 1.245]	0.133 [1.021, 1.155]	0.112 [0.964, 1.077]	0.088 [0.883, 0.971]	0.073 [0.820, 0.893]	0.059 [0.755, 0.814]
20%-B	0.108 [1.096, 1.204]	0.098 [1.030, 1.128]	0.083 [0.970, 1.053]	0.066 [0.888, 0.954]	0.050 [0.828, 0.879]	0.037 [0.763, 0.800]
40%-A	0.161 [1.117, 1.278]	0.140 [1.048, 1.188]	0.117 [0.991, 1.108]	0.088 [0.910, 0.999]	0.072 [0.848, 0.919]	0.058 [0.781, 0.839]
40%-B	0.114 [1.125, 1.240]	0.104 [1.060, 1.163]	0.087 [0.996, 1.083]	0.062 [0.917, 0.978]	0.049 [0.857, 0.906]	0.034 [0.791, 0.825]
60%-A	0.162 [1.134, 1.296]	0.141 [1.065, 1.206]	0.120 [1.006, 1.126]	0.090 [0.926, 1.016]	0.073 [0.863, 0.936]	0.058 [0.796, 0.854]
60%-B	0.109 [1.143, 1.252]	0.097 [1.079, 1.176]	0.081 [1.017, 1.098]	0.064 [0.933, 0.997]	0.043 [0.873, 0.915]	0.031 [0.807, 0.837]
80%-A	0.166 [1.148, 1.314]	0.146 [1.079, 1.224]	0.123 [1.019, 1.143]	0.093 [0.939, 1.032]	0.074 [0.877, 0.950]	0.059 [0.810, 0.868]
80%-B	0.119 [1.158, 1.277]	0.092 [1.093, 1.185]	0.082 [1.032, 1.114]	0.062 [0.949, 1.011]	0.042 [0.888, 0.930]	0.030 [0.822, 0.852]

Note: A and B mean the estimated drought severity uncertainties caused by copula input data (A) and copula parameter estimations (B); numbers in the brackets indicate the lower (2.5% percentile) and upper (97.5% percentile) limits, and those outside mean the ranges of 95% CIs.

associated with different copula functions, copula parameter estimations, and copula input data are investigated.

Results show that GEV is finally selected as the marginal distribution for drought severity, and pseudo observations for drought areal extent. Gumbel shows the best performance in capturing the dependence structure of drought severity and drought areal extent and thus is selected as the benchmark copula. The estimations of drought SAF curves are influenced by copula choices. BB6 is capable of giving close drought severity as Gumbel; Normal, Student-t, BB1, BB7 overestimate the drought severities; Joe and BB8 underestimate the drought severities when the drought areal extents become higher; Clayton and Frank produce the most deviation in drought SAF curves although Clayton

even presents quite close AIC and BIC values with Gumbel. The differences in drought SAF curves caused by different copulas are increasing with the conditional probability decreasing. As to the parameter estimations and input data, both of them result in the drought SAF curves uncertainty. The uncertainty ranges caused by copula input data are larger than those caused by copula parameter estimations, and consequently, to improve copula input data quality is recommended for developing the reliability of such frequency analysis. Results also show that the uncertainty bands in the drought SAF curves increase as the conditional probability decreases. It indicates that larger uncertainties exist in the frequency analysis of extreme values (with small probability). This highlights the importance of

uncertainty analysis in frequency study, given that most studies in hydrology and climatology focus on extreme values. Couple-based frequency analysis of drought with considering the uncertainty issues is capable of providing more beneficial probability information to water managers, and aiding in the development of mitigation strategies and disaster preparedness plans for different levels of drought.

ACKNOWLEDGEMENTS

REFERENCES

Aas, K., Czado, C., Frigessi, A. & Bakken, H. 2009 Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics* **44** (2), 182–198.

Afshar, M. H., Sorman, A. U. & Yilmaz, M. T. 2016 Conditional copula-based spatial-temporal drought characteristics analysis – a case study over Turkey. *Water* **8** (10), 426.

Amirataee, B., Montaseri, M. & Rezaie, H. 2018 Regional analysis and derivation of copula-based drought severity-area-frequency curve in Lake Urmia basin, Iran. *Journal of Environmental Management* **206**, 134–144.

Ayantobo, O. O., Li, Y., Song, S., Javed, T. & Yao, N. 2018 Probabilitic modelling of drought events in China via 2-dimensional joint copula. *Journal of Hydrology* **559**, 373–391.

Beguería, S., Vicente-Serrano, S. M., Reig, F. & Latorre, B. 2014 Standardized precipitation evapotranspiration index (SPEI) revisited: parameter fitting, evapotranspiration models, tools, datasets and drought monitoring. *International Journal of Climatology* **34** (10), 3001–3023.

Brechmann, E. C. & Schepsmeier, U. 2013 Modeling dependence with C- and D-Vine copulas: the R package CDVine. *Journal of Statistical Software* **52** (3), 1–27.

Chen, K. H. & Khashanah, K. 2015 Analysis of systemic risk: a dynamic Vine copula-based ARMA-EGARCH model. In: *Transactions on Engineering Technologies, World Congress on Engineering and Computer Science* (S.-I. Ao, H. K. Kim & M. A. Amouzegar, eds). Springer Nature, Singapore.

Dai, A. 2011 Characteristics and trends in various forms of the palmer drought severity index during 1900–2008. *Journal of Geophysical Research Atmospheres* **16**. https://doi.org/10.1029/2010JD015541.

Efron, B. 1979 Bootstrap methods: another look at the jackknife. *Annals of Statistics* **7**, 1–26.

Fang, Q. Q., Wang, G. Q., Xue, B. L., Liu, T. X. & Kiem, A. 2018a How and to what extent does precipitation on multi-temporal scales and soil moisture at different depths determine carbon flux responses in a water- limited grassland ecosystem? *Science of the Total Environment* **635**, 1255–1266.

Fang, Q. Q., Wang, G. Q., Liu, T. X., Xue, B. L. & Yinglan, A. 2018b Controls of carbon flux in a semi-arid grassland ecosystem experiencing wetland loss: vegetation patterns and environmental variables. *Agricultural and Forest Meteorology* **259**, 196–210.

Ghoudi, K. & Rémillard, B. 2004 Empirical processes based on pseudo-observations II: the multivariate case. *Asymptotic Methods in Stochastics* **44**, 381–406.

Halwatura, D., Lechner, A. M., McIntyre, N. & Arnold, S. 2016 Uncertainties in estimating design droughts. *8th International Congress on Environmental Modelling and Software. International Environmental Modelling and Software Society (iEMSs)*, Toulouse, France. http://www.iemss.org/society/index.php/iemss-2016-proceedings.

Hameed, M., Ahmadalipour, A. & Moradkhani, H. 2018 Apprehensive drought characteristics over Iraq: results of a multidecadal spatiotemporal assessment. *Geosciences* **8**, 58. doi:10.3390/geosciences8020058.

Han, D. M., Wang, G. Q., Liu, T. X., Xue, B. L., Kuczera, G. & Xu, X. Y. 2018 Hydroclimatic response of evapotranspiration partitioning to prolonged droughts in semiarid grassland. *Journal of Hydrology* **563**, 766–777.

Homdee, T., Pongput, K. & Kanae, S. 2016 A comparative performance analysis of three standardized climatic drought indices in the Chi River basin, Thailand. *Agriculture and Natural Resources* **50** (3), 211–219.

Hu, Y., Liang, Z., Liu, Y., Wang, J., Yao, L. & Ning, Y. 2015 Uncertainty analysis of SPI calculation and drought assessment based on the application of Bootstrap. *International Journal of Climatology* **35** (8), 1847–1857.

Jiang, C., Xiong, L., Xu, C. & Guo, S. 2015 Bivariate frequency analysis of nonstationary low-flow series based on the time-varying copula. *Hydrological Processes* **29** (6), 1521–1534.

Kojadinovic, I., Yan, J., Leeuw, J. D. & Zeileis, A. 2010 Modeling multivariate distributions with continuous margins using the copula R package. *Journal of Statistical Software* **34** (9), 1346–1352.

Lee, J. H. & Kim, C. J. 2011 Derivation of drought severity-duration-frequency curves using drought frequency analysis. *Journal of Korea Water Resources Association* **44** (11), 889–902.

Loukas, A. & Vasiliades, L. 2004 Probabilistic analysis of drought spatiotemporal characteristics in Thessaly region, Greece. *Natural Hazards & Earth System Sciences* **4** (5/6), 719–731.

Mishra, A. K. & Singh, V. P. 2009 Analysis of drought severity-area-frequency curves using a general circulation model and scenario uncertainty. *Journal of Geophysical Research Atmospheres* **114**(D6). https://doi.org/10.1029/2008JD010986.

Niu, J., Kang, S., Zhang, X. & Fu, J. 2019 Vulnerability analysis based on drought and vegetation dynamics. *Ecological Indicators* **105**, 329–336. https://doi.org/10.1016/j.ecolind.2017.10.048.

Rahmat, S. N., Jayasuriya, N. & Bhuiyan, M. 2015 Development of drought severity-duration-frequency curves in Victoria, Australia. *Australian Journal of Water Resources* **19** (1), 31–42.

Reddy, M. J. & Ganguli, P. 2013 Spatio-temporal analysis and derivation of copula-based intensity-area-frequency curves for droughts in western Rajasthan (India). *Stochastic Environmental Research and Risk Assessment* **27** (8), 1975–1989.

Sadegh, M., Ragno, E. & Aghakouchak, A. 2017 Multivariate copula analysis toolbox (MvCAT): describing dependence and underlying uncertainty using a Bayesian framework. *Water Resources Research* **53**, 5166–5183.

Serinaldi, F., Bonaccorso, B., Cancelliere, A. & Grimaldi, S. 2009 Probabilistic characterization of drought properties through copulas. *Physics & Chemistry of the Earth* **34** (10–12), 596–605.

Shong, N. 2010 *Pearson's Versus Spearman's and Kendall's Correlation Coefficients for Continuous Data*. Master of Science Thesis, University of Pittsburg, Pittsburg, PA, USA. http://d-scholarship.pitt.edu/8056/1/Chokns_etd2010.pdf

Sun, P., Zhang, Q., Yao, R., Singh, V. P. & Song, C. 2018 Low flow regimes of the Tarim River Basin, China: probabilistic behavior, causes and implications. *Water* **10** (4), 470.

Tan, C., Yang, J. & Li, M. 2015 Temporal-spatial variation of drought indicated by SPI and SPEI in Ningxia Hui Autonomous Region, China. *Atmosphere* **6** (10), 1399–1421.

Vergni, L., Lena, B. D., Todisco, F. & Mannocchi, F. 2017 Uncertainty in drought monitoring by the standardized precipitation index: the case study of the Abruzzo region (central Italy). *Theoretical and Applied Climatology* **128** (1), 13–26.

Vicente-Serrano, S. M., Beguería, S. & López-Moreno, J. I. 2010 A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index. *Journal of Climate* **23** (7), 1696–1718.

Wang, X. J., Zhang, J. Y., Shahid, S., Elmahdi, A., He, R. M., Bao, Z. X. & Ali, M. 2012 Water resources management strategy for adaptation to droughts in China. *Mitigation and Adaptation Strategies for Global Change* **17** (8), 923–937.

Wang, Y., Xu, Z., Zhang, B. & Li, Q. 2016 Monitoring the meteorological drought in the middle reaches of Heihe River basin based on TRMM precipitation data. *2016 IEEE International Geoscience and Remote Sensing Symposium*, Beijing, China, pp. 4313–4316. doi: 10.1109/IGARSS.2016.7730124.

Xu, K., Yang, D., Xu, X. & Lei, H. 2015 Copula based drought frequency analysis considering the spatio-temporal variability in Southwest China. *Journal of Hydrology* **527**, 630–640.

Yu, M., Li, Q., Hayes, M. J., Svoboda, M. D. & Heim, R. R. 2014 Are droughts becoming more frequent or severe in China based on the standardized precipitation evapotranspiration index: 1951–2010? *International Journal of Climatology* **34** (3), 545–558.

Zhang, Q., Xiao, M. & Singh, V. P. 2015 Uncertainty evaluation of copula analysis of hydrological droughts in the East River basin, China. *Global and Planetary Change* **129**, 1–9.

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 1960–2013. *Journal of Hydrology* **553**, 773–784.

Zhao, P., Lü, H., Fu, G., Zhu, Y., Su, J. & Wang, J. 2017 Uncertainty of hydrological drought characteristics with copula functions and probability distributions: a case study of Weihe River, China. *Water* **9** (5), 334.