HYDROLOGICAL PROCESSES
Hydrol. Process. **27**, 1454–1463 (2013)
Published online 23 April 2013 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/hyp.9233

Data-based analysis of bivariate copula tail dependence for drought duration and severity

T. Lee,[1]* R. Modarres[2] and T. B. M. J. Ouarda[2,3]

[1] *Department of Civil Engineering, ERI, Gyeongsang National University, 501 Jinju-daero, Jinju, Gyeongsangnam-do, South Korea*
[2] *INRS-ETE, 490 De La Couronne, Québec, QC, G1K9A9, Canada*
[3] *Masdar Institute of Science and Technology, PO Box 54224, Abu Dhabi, UAE*

Abstract:

In recent decades, copula functions have been applied in bivariate drought duration and severity frequency analysis. Among several potential copulas, Clayton has been mostly used in drought analysis. In this research, we studied the influence of the tail shape of various copula functions (i.e. Gumbel, Frank, Clayton and Gaussian) on drought bivariate frequency analysis. The appropriateness of Clayton copula for the characterization of drought characteristics is also investigated. Drought data are extracted from standardized precipitation index time series for four stations in Canada (La Tuque and Grande Prairie) and Iran (Anzali and Zahedan). Both duration and severity data sets are positively skewed. Different marginal distributions were first fitted to drought duration and severity data. The gamma and exponential distributions were selected for drought duration and severity, respectively, according to the positive skewness and Kolmogorov–Smirnov test. The results of copula modelling show that the Clayton copula function is not an appropriate choice for the used data sets in the current study and does not give more drought risk information than an independent model for which the duration and severity dependence is not significant. The reason is that the dependence of two variables in the upper tail of Clayton copula is very weak and similar to the independent case, whereas the observed data in the transformed domain of cumulative density function show high association in the upper tail. Instead, the Frank and Gumbel copula functions show better performance than Clayton function for drought bivariate frequency analysis. Copyright © 2012 John Wiley & Sons, Ltd.

KEY WORDS bivariate frequency analysis; Clayton copula; distribution tail; drought; joint distribution

Received 15 October 2011; Accepted 2 February 2012

INTRODUCTION AND REVIEW

Drought is a special natural hazard that is observed in both humid and arid regions. The negative effects of drought events have been reported in different countries and regions of the world (Frick *et al.*, 1990; Panu and Sharma, 2002, 2009; IPCC, 2007). Several investigators have tried to characterize many drought characteristics such as intensity, duration and their recurrence intervals as well as their effects on agriculture and water supply (Frick *et al.*, 1990; Sharma, 2000; Shiau and Shen, 2001; Shabbar and Skinner, 2004; Pielke *et al.*, 2005; Arena *et al.*, 2006; Modarres, 2007; Ouarda *et al.*, 2008; Karamouz *et al.*, 2009; Khaliq *et al.*, 2009; Panu and Sharma, 2009).

In the scientific literature, droughts are generally characterized through the study of their probability of occurrence and the resulting risk. The reason is that drought is a stochastic phenomenon in space and time. Several probabilistic approaches have been applied in the past decades to model drought characteristics such as severity and duration. For example, Sen (1980) derived analytical formulations of critical droughts either as longest drought duration or maximum deficit sum by assuming that drought duration and deficit are entirely independent from each other. Lee *et al.* (1986) carried out a univariate drought frequency analysis of multiyear durations of annual streamflow series. The probabilistic behaviour and statistical characteristics of drought severity in Catalonia was investigated by Lana and Burgueno (1998) using the univariate Gumbel frequency distribution.

In addition to meteorological droughts, various univariate frequency distributions have been applied to low flows (Ryu *et al.*, 2011) to identify the magnitude of a hydrologic drought corresponding to a recurrence interval. Several studies have applied different frequency distributions to estimate the probability of occurrence or the return period of hydrologic droughts (Kroll and Vogel, 2002; Zaidman *et al.*, 2003; Modarres, 2008; Ouarda and Shu, 2009). More recently, the frequency distribution of the annual extreme hydrologic dry spell length for southeastern Iran was investigated by Modarres and Sarhadi (2010). The statistical models commonly used for low-flow frequency analysis were briefly discussed by Ouarda *et al.* (2008). Most of these models include univariate statistical frequency distribution functions and disregard the relationship between drought characteristics.

However, in recent years, the bivariate probabilistic approaches, which take into account the relationship between drought properties, have been applied for modelling the multidisciplinary nature of drought events

*Correspondence to: T. Lee, Department of Civil Engineering, ERI, Gyeongsang National University, 501 Jinju-daero, Jinju, Gyeongsangnam-do, South Korea.
E-mail: tae3lee@gnu.ac.kr

(El-Jabi *et al.*, 1997; Nadarajah, 2009a,b; Song and Singh, 2010a,b; Wong *et al.*, 2010; Mishra and Singh, 2011). Shiau (2006) first applied copula functions to bivariate frequency analysis of drought duration and severity.

Copula functions represent appropriate tools for assessing the multivariate nature of droughts (Nelsen, 1999). Copula functions have been applied extensively in hydrological frequency analysis and simulation (De Michele and Salvadori, 2003; Salvadori and De Michele, 2004; Chebana and Ouarda, 2011; De Michele *et al.*, 2007; Kao and Govindaraju, 2008; Lee and Salas, 2011). The bivariate joint distribution of drought characteristics, for example, duration and severity, has been modelled with several copula functions and marginal distributions. Shiau (2006) applied different types of copulas (e.g. Gumbel, Clayton, Gaussian and Frank) for the modelling of the joint distribution of drought duration and severity and fitted the exponential and gamma marginal distributions to drought duration and severity, respectively. The bivariate probabilistic behaviour of drought properties of the Yellow River, China, was studied by Shiau *et al.* (2007) using copula functions. More recently, a joint deficit index was proposed by Kao and Govindaraju (2010) on the basis of copula functions.

For drought bivariate frequency analysis by copula functions, marginal distributions are first fitted to drought severity and duration separately. Several marginal distribution functions, such as gamma, exponential, generalized extreme value, two-parameter log-normal and Pareto distributions, have been suggested in the literature and applied to drought severity and duration. Although the effect of the tail behaviour on fitting frequency distributions to extreme flood events was discussed in the literature (El Adlouni *et al.*, 2008), its effect on drought bivariate frequency analysis and on the selection of the copula function has not yet been fully considered. Also, it is interesting to note that the Clayton copula has been commonly suggested and used for drought duration and severity frequency analysis, but it is applied without comparison with other copula functions (Shiau *et al.*, 2007; Laux *et al.*, 2009; Shiau and Modarres, 2009).

The aim of the present study was the identification of suitable marginal and copula functions for bivariate drought frequency analysis. As the Clayton copula has been commonly suggested for drought duration and severity frequency analysis, we also investigate its suitability and compare its performance with other copula functions. The approach adopted in the present study is based on the tail behaviour of the various copula functions related to drought analysis.

The article is organized as follows. In the Mathematical Description section, the fundamental mathematical background related to copula modelling is presented. The two study areas are presented in the Data Description section. The univariate drought analysis for the two case studies is illustrated in the Univariate CDFs of Drought Duration and Severity Analysis section, and the results of the bivariate drought analysis are presented in the Bivariate drought frequency analysis section. Finally, the summary and conclusions are presented in the Summary and Conclusions section.

MATHEMATICAL DESCRIPTION

Background

Univariate distributions. To fit a bivariate distribution to drought duration and severity, a univariate distribution is first fitted to drought characteristics. Several marginal distribution functions have been applied for fitting drought characteristics such as gamma for drought severity and exponential, mixture exponential and geometric for drought duration (Loaiciga and Leipnik, 1996; Salas *et al.*, 2005; Shiau, 2006; Shiau *et al.*, 2007; Shiau and Modarres, 2009).

One of the most common marginal distributions for drought severity is the gamma distribution. The gamma distribution of a random variable X is denoted as

$$f_X(x) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} \exp(-x/\beta) \tag{1}$$

where α and β are shape and scale parameters, respectively, and $\Gamma(.)$ represents the gamma function.

The exponential distribution has been used to model drought duration (Shiau and Modarres, 2009). This distribution is denoted as

$$f_X(x) = \frac{1}{\lambda} \exp(-x/\lambda) \tag{2}$$

where λ is the distribution parameter.

Empirical bivariate distribution analysis. The comparison of the empirical bivariate distribution and copula functions leads to the selection of an appropriate bivariate distribution model. To compare the results of copula models, an empirical bivariate distribution of the observed duration and severity is also used. The estimation procedure of the empirical joint cumulative distribution function (JCDF) assumes a bivariate data represented by X_j *and* Y_j, $j = 1,\ldots, N$, where N is the sample size, which is based on the range of data and the class intervals of X and Y (ΔX and ΔY) (Kottegoda and Rosso, 2008). Using the relative frequency and the cumulative relative frequency for each class will give the empirical probability density function (PDF) and the cumulative distribution function (CDF). In this study, drought severity and its related duration are considered as a bivariate data set for the empirical bivariate distribution analysis.

Drought analysis

Drought definition. For copula-based severity duration frequency analysis, drought characteristics are defined on the basis of the popular standardized precipitation index

(SPI) developed by McKee *et al.* (1993). The SPI is a flexible index, as it can be applied at different time scales. It can also be adapted to different climate regions (Santos *et al.*, 2010) as it requires relatively simple calculations. Various rainfall time scales are used, ranging from 1 to 24 months (e.g. 1, 3, 6, 12 and 24), to represent the short-term and long-term drought condition (NOAA, 2010). The SPI values are calculated by fitting the gamma distribution to rainfall data and by converting the gamma CDF to a standard normal distribution so that an index of 0 (SPI = 0) indicates the median precipitation. The negative and positive values of SPI are considered as dry and wet conditions, respectively. The range at SPI is then adapted to drought conditions from extremely wet conditions (SPI > 2) to extreme dry conditions (SPI < −2).

In this study, the value of SPI = 0 is selected as the threshold for a drought event as suggested by other studies (Shiau, 2006; Shiau *et al.*, 2007). Drought duration, D, is defined as the number of consecutive events with negative SPI, and drought severity, S, is then the cumulative value of the negative SPIs within the duration of the drought. For convenience, drought severity is multiplied by −1 to obtain a positive value. Therefore, we have

$$S = -\sum_{i=1}^{D} \mathrm{SPI}_i \tag{3}$$

Note that the SPI value is a standard normal dimensionless quantile, and therefore S is also dimensionless and indicates the summation of the D number of standard normal quantiles.

Copula-based drought analysis. The modelling of the bivariate relationship between two variables is usually based on Sklar's (1959) theorem, which states that if $F_{X,Y}(x,y)$ is a JCDF of X and Y variables with respective marginal CDFs $F_X(x)$ and $F_Y(y)$, then there exists a copula C such that

$$F_{X,\,Y}(x,\,y) = C(F_X(x),\,F_Y(y)) \tag{4}$$

For the univariate marginal CDFs and any copula function C, the function $F_{X,Y}$ $(x,\ y)$ is a JCDF with marginal CDFs $F_X(x)$ and $F_Y(y)$. Under the assumption that the marginal CDF distributions are continuous with PDFs $f_X(x)$ and $f_Y(y)$, the joint PDF (JPDF) becomes

$$f_{X,Y}(x,y) = c(F_X(x), F_Y(y)) f_X(x) f_Y(y) \tag{5}$$

where c represents the double partial derivative of C over u and v, written as

$$c(u,v) = \frac{\partial^2 C(u,v)}{\partial u \partial v} \tag{6}$$

The reader is referred to Nelsen (1999) for details on copula functions. For drought duration (D) and severity (S), we can give the following copula-based bivariate drought model (Shiau and Modarres, 2009):

$$F_{\mathrm{D,S}}(d,s) = C(F_{\mathrm{D}}(d), F_{\mathrm{S}}(s)\ ) \tag{7}$$

where $F_{\mathrm{D,S}}$ $(d,\ s)$ is the JCDF of drought severity and drought duration, and $F_{\mathrm{D}}(d)$ and $F_{\mathrm{S}}(s)$ are the drought duration and severity CDFs, respectively. Several copula functions have been proposed for the multivariate modelling of hydrological variables. The copula density functions investigated in the present study for bivariate drought severity duration frequency analysis are listed as follows:

- Clayton copula

$$C(u,\ v;\ \theta) = \left(u^{-\theta} + v^{-\theta} - 1\right)^{-1/\theta}, \theta > 0 \tag{8}$$

$$c(u,v) = (\theta+1)\left(u^{-\theta} + v^{-\theta} - 1\right)^{-\frac{1}{\theta}-2}(uv)^{-\theta-1} \tag{9}$$

- Gumbel copula

$$C(u,\ v;\ \theta) = \exp\left\{-\left[-(\ln u)^{\theta} + (-\ln v)^{\theta}\right]^{\frac{1}{\theta}}\right\}, \theta \geq 1 \tag{10}$$

$$c(u,v) = C(u,v)\frac{[(-\ln u)(-\ln v)]^{\theta-1}}{uv} \left[(-\ln u)^{\theta} + (-\ln v)^{\theta}\right]^{\frac{2}{\theta}-2} \cdot\left\{(\theta-1)\left[(-\ln u)^{\theta} + (-\ln v)^{\theta}\right]^{-1/\theta} + 1\right\} \tag{11}$$

- Frank copula

$$C(u,\ v;\ \theta) = \frac{1}{\theta}\ln\left[1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1}\right], \theta \neq 0 \tag{12}$$

$$c(u,v) = \frac{\theta e^{-\theta(u+v)}(e^{-\theta} - 1)}{\left[e^{-\theta(u+v)} - e^{-\theta u} - e^{-\theta v} + e^{-\theta}\right]^2} \tag{13}$$

- Gaussian copula

$$C(u,v;\theta) = \Phi_{\theta}\left(\Phi^{-1}(u), \Phi^{-1}(v)\right), -1 \leq \theta \leq 1 \tag{14}$$

$$c(u,v) = \frac{\phi_{\theta}\left(\Phi^{-1}(u), \Phi^{-1}(v)\right)}{\phi\left(\Phi^{-1}(u)\right)\phi\left(\Phi^{-1}(v)\right)} \tag{15}$$

where $u = F_{\mathrm{D}}(d)$ and $v = F_{\mathrm{S}}(s)$ are the CDFs of drought duration and severity, respectively; θ is the copula parameter used to represent the dependence structure between F_{S} and F_{D}; and Φ and ϕ are the standard normal CDF and PDF, respectively.

For the estimation of the parameters, the popular method of inference function for margins (IFMs) is applied. Two separate estimation procedures are involved in the IFM. The parameters are estimated in two stages. In

the first stage, the parameters of the marginal distributions are estimated with the maximum likelihood method. Then, in the second stage, the copula parameters conditioned on the previous marginal distributions are estimated by maximizing the log-likelihood of copula density in Equation 6. The best copula function is then selected on the basis of the maximum log-likelihood of copula density compared with other copulas.

To illustrate the tail behaviour of the copula functions, 1000 pairs of simulated data from each copula function are shown in Figure 1. The parameters of each copula were estimated through the relationship between the Kendall's tau (set to $\tau = 0.7$) and the copula parameters (Nelsen, 1999). The fixed Kendall's tau for the parameters of the copula functions was used to compare them for the same magnitude of association. Figure 1 shows that, for the Clayton copula function (Figure 1a), the degree of association in the upper tail is less than that in the lower tail, whereas this association in the upper tail is higher for the Gumbel function (Figure 1b). The Frank copula (Figure 1c) has the same magnitude of association through the main body and the lower and upper tails. The Gaussian copula (Figure 1d) shows a higher association in both tails.

Return period of drought duration and severity. As drought is a bivariate event characterized by drought duration and severity, the frequency analysis of drought events should consider the joint and conditional properties of return periods. The joint return period is defined for two cases of drought risk corresponding to drought events with [$D \geq d$ and $S \geq s$] and drought events with [$D \geq d$ or $S \geq s$]. The copula-based drought joint return period can be defined as follows (Shiau, 2006):

$$T_{\mathrm{DS}} = \frac{E(L)}{P(D \geq d, S \geq s)} = \frac{E(L)}{1 - F_{\mathrm{D}}(d) - F_{\mathrm{S}}(s) + C(F_{\mathrm{D}}(d), F_{\mathrm{S}}(s))} \quad (16)$$

$$T'_{\mathrm{DS}} = \frac{E(L)}{P(D \geq d \text{ or } S \geq s)} = \frac{E(L)}{1 - C(F_{\mathrm{D}}(d), F_{\mathrm{S}}(s))} \quad (17)$$

where T_{DS} denotes the joint return period for [$D \geq d$ and $S \geq s$], T'_{DS} denotes the joint return period for [$D \geq d$ or $S \geq s$] and $E(L)$ is the expected drought interarrival time. The interarrival time is defined as the time between consecutive arrivals.

DATA DESCRIPTION

The database for this study includes a 3-month SPI time series for two stations in Iran and two stations in Canada. These stations are selected to examine the effect of the distribution's tail and drought characteristics on drought bivariate frequency analysis for two completely different climatic and hydrologic regimes. The two stations selected in Iran are the Anzali and the Zahedan stations, and the two stations selected in Canada are the Grande Prairie and the La Tuque stations.

Canada is considered to possess abundant freshwater resources and is ranked among the top five countries in

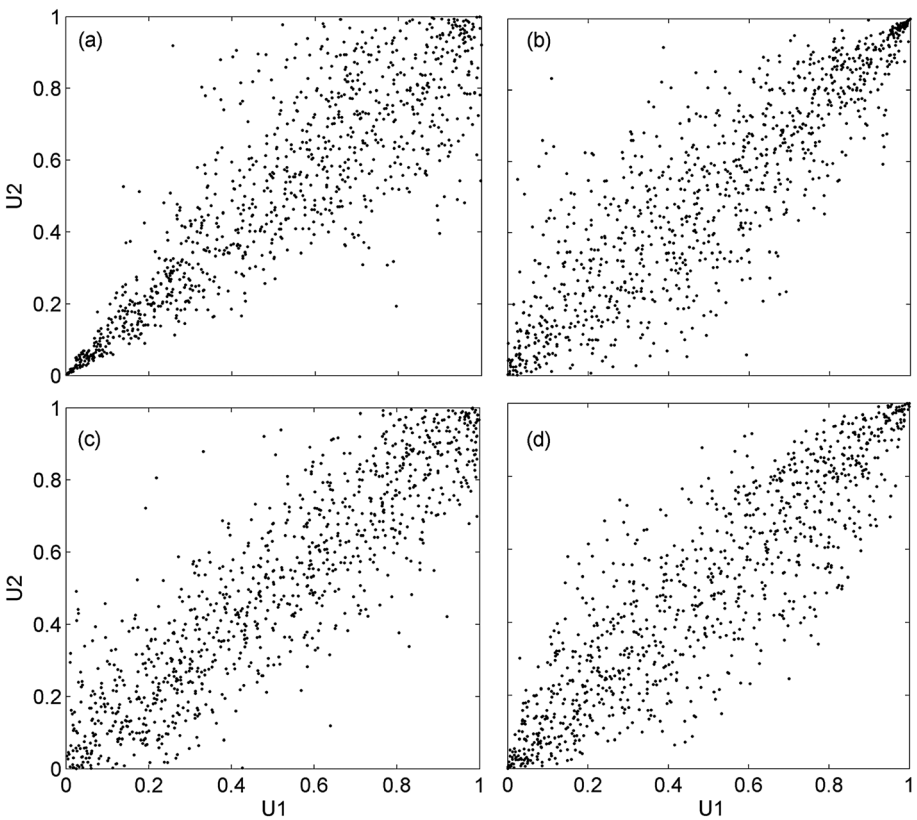

Figure 1. Simulated data from (a) Clayton, (b) Gumbel, (c) Frank and (d) Gaussian copulas

terms of *per capita* water supply. However, drought is becoming a major problem in Canada, especially for the Prairie provinces and for the agriculture sector (Shabbar and Skinner, 2004; Schindler and Donahue, 2006). The first selected station for this study is located in the Prairie region of Canada (Grande Prairie station, Alberta province), and the second one is located in one of the major agricultural regions of Canada (La Tuque station, Quebec province).

On the other hand, drought is a common climate feature in Iran, a country located in the planet's arid and semiarid belt. The natural centre for agricultural drought management of Iran (http://www.ncadm.ir) reported the loss of US$10 billion and US$19 billion during the two recent drought events in Iran, which occurred during 1997–2001 and 2006–2009, respectively. These two drought events have influenced more than the two thirds of the country. The two stations selected from Iran, the Zahedan and Anzali stations, are located in the arid region of southeastern Iran and the most humid region in northern Iran, respectively.

The four stations selected in this study are located in two completely different climate regions, a high latitude cold region and an arid and semiarid region with different rainfall seasonality conditions. The stations located in Canada receive rainfall mostly in summer time while winter rainfall is dominant in Iran. The SPI time series of these four stations are illustrated in Figure 2. Some basic statistics of the observed drought duration and severity series are also given in Table I, which shows that the Anzali and the Zahedan stations receive the highest and the lowest mean annual rainfall, respectively, among the four stations. The Zahedan station shows also the highest coefficient of variation among all stations. This is probably because it is located in an arid region. One can see that Zahedan station is clearly different from the other stations. The lowest values of the maximum observed drought duration and severity can be seen in Zahedan station as 9 and 11.69, respectively (Table I). The drought characteristics of the other three stations (Anzali, Grand Prairie and La Tuque) are not distinctively different.

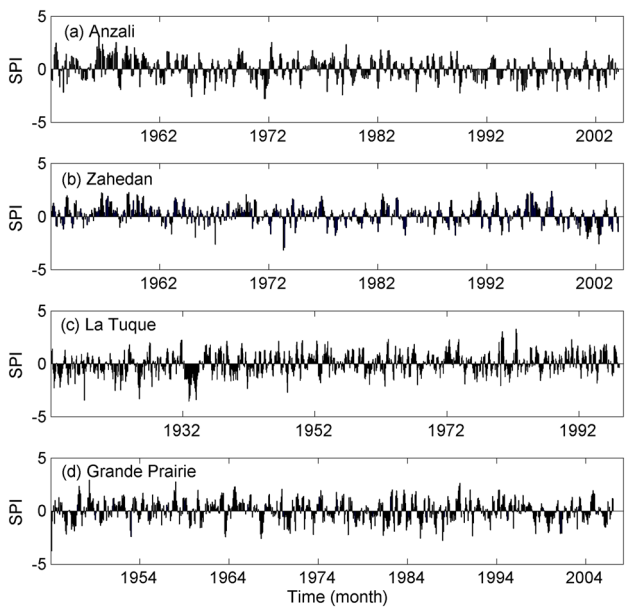

Figure 2. The 3-month SPI time series for the selected stations (i.e. Anzali and Zahedan stations in Iran as well as Grande Prairie and La Tuque stations in Canada)

It is interesting to mention an extraordinary drought event during the 1930s (during the dust bowl extraordinary drought event in North America) at La Tuque station for which the observed drought duration and severity are $D = 27$ months and $S = 47$, respectively. This implies that the average severity for each month is equal to -1.74 ($-47/27$). This value is equivalent to $p < 0.04$ for the gamma precipitation CDF. This rare value occurs consecutively during 27 months for this drought event.

UNIVARIATE CDFS OF DROUGHT DURATION AND SEVERITY ANALYSIS

As indicated in Table I and Figure 3, the marginal distributions of drought duration and severity have a highly positive skewness for all stations. The exceptionally high skewness in La Tuque is induced by the extraordinary drought event of the 1930s.

In this study, different marginal distributions are examined to identify the most suitable distribution for drought duration and severity. These include the gamma, Gumbel, Generalized Extreme Values (GEV), Geometric, Exponential and Generalized Pareto (GP) distribution functions. These distributions cover a wide range of different tail behaviours, including heavily tailed or highly skewed distributions of drought characteristics. The Kolmogorov–Smirnov (KS) test was used to test the goodness-of-fit of each marginal distribution function. Among other marginal distributions, the Gamma distribution was selected because the KS test (Table II) indicates that this distribution is appropriate for all stations and because it is commonly used for drought severity modelling in bivariate drought frequency analysis (Shiau *et al.*, 2007; Shiau and Modarres, 2009). The exponential distribution is selected for drought duration for all stations (see Equation 2) for the same reasons as drought severity. The parameters of the marginal distributions estimated by the maximum likelihood method and the results of the KS test are presented in Table II.

BIVARIATE DROUGHT FREQUENCY ANALYSIS

Copula selection and its parameter estimation

Bivariate drought frequency analysis for the selected stations is carried out by using the four copula functions presented in the Drought Analysis section. The estimated parameters of the copula functions with the IFM method are presented in Table II. The values of the log-likelihood and the parameters for each station and copula function are given in Table III. This table shows that the best copula function for the drought duration-severity joint distribution is the Frank copula for the Anzali, Zahedan

Table I. Basic statistics of annual rainfall and drought events for the selected stations

Station			Anzali	Zahedan	La Tuque	Gr. Prairie
Annual rainfall	Record length		1952–2003	1952–2003	1913–1998	1944–2006
	Mean (mm)		1855.8	91	674.6	514.1
	SD		413.9	46	115	152.4
	CV (%)		22.3	50.5	17.04	29.64
Drought	No. drought events		85	79	131	100
	Duration (months)	Mean	3.69	3.06	3.97	3.75
		SD	2.66	2.32	3.56	3.31
		Skewness	1.33	1.06	2.75	1.51
		Maximum	13	9	27	16
		Minimum	1	1	1	1
	Severity	Mean	2.97	2.18	3.13	3.04
		SD	2.82	2.36	4.94	3.43
		Skewness	1.42	1.89	6.09	1.45
		Maximum	13.59	11.69	47.78	14.14
		Minimum	0.07	0.01	0.02	0.08

Drought severity is dimensionless. CV indicates coefficient of variation.

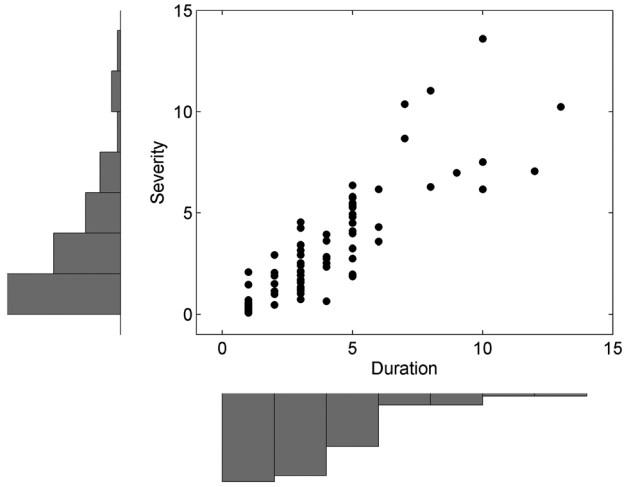

Figure 3. Scatter plot and histograms of the observed duration and severity for Anzali station

Table II. Parameters of marginal distributions (see Equation 1 and 2) and copulas (see Equations 10, 12 and 14) for the selected stations

		Anzali	Zahedan	La Tuque	Grande Prairie
Severity (gamma)	α	0.95	0.81	0.73	0.78
	β	3.11	2.7	4.3	3.89
	P value[a]	0.79	0.7	0.94	0.41
Duration (exponential)	λ	3.69	3.06	3.97	3.75
	P value[a]	0.37	0.6	0.98	0.23
Copula	Selected	Frank	Frank	Frank	Gumbel
	θ	10.51	11.09	12.77	3.1

[a] The P value is estimated from the KS test.

and La Tuque stations and the Gumbel copula for the Grand Prairie station.

Tail behaviour of copula models for the bivariate drought frequency analysis

The relationship between drought duration and severity can be seen by plotting the histogram of the observed drought duration and severity events. As an example, the observed drought duration and severity at Anzali Station are presented in Figure 3. Note that the histograms of drought duration and severity show a positive skewness as expected from Table I. It is observed that the association between drought duration and severity becomes weak as the values of both duration and severity increase. However, a different behaviour can be observed when CDFs of drought duration and severity are plotted (Figure 4). Higher association is observed for large values of drought duration and severity in all the selected stations. It is noticeable that there seems to be a discrepancy between the association in the original domain and the CDF domain (e.g. Figures 3 and 4a for Anzali station).

Table III. Log-likelihood and copula parameters for the selected stations

		Anzali	Zahedan	La Tuque	Grande Prairie
Log-likelihood	Clayton	24.88	21.48	52.13	43.12
	Gumbel	*51.84*	*48.25*	*91.80*	**70.48**
	Frank	**53.21**	**50.35**	**95.78**	65.23
	Gaussian	47.30	42.42	91.30	*70.20*
Parameter of the copula (θ)	Clayton	1.69	1.84	2.46	2.43
	Gumbel	2.89	2.98	3.28	3.10
	Frank	10.51	11.09	12.77	10.54
	Gaussian	0.90	0.87	0.91	0.90

The bold indicates the best performance.

To explain this discrepancy, the example of the gamma distribution is presented in Figure 5 with α and β equal to 2.0 (Equation 1), which is intentionally set to have the same positive skewness as the observed drought data (refer to Table I). The PDF and the CDF of the gamma distribution are shown in the left and right side of Figure 5, respectively. Two pairs of CDF values [$F(x_1) = 0.92$ and $F(x'_1) = 0.96$ versus $F(x_2) = 0.60$ and $F(x'_2) = 0.64$] are selected and presented in the abscissa of Figure 5. The quantiles corresponding to the previously

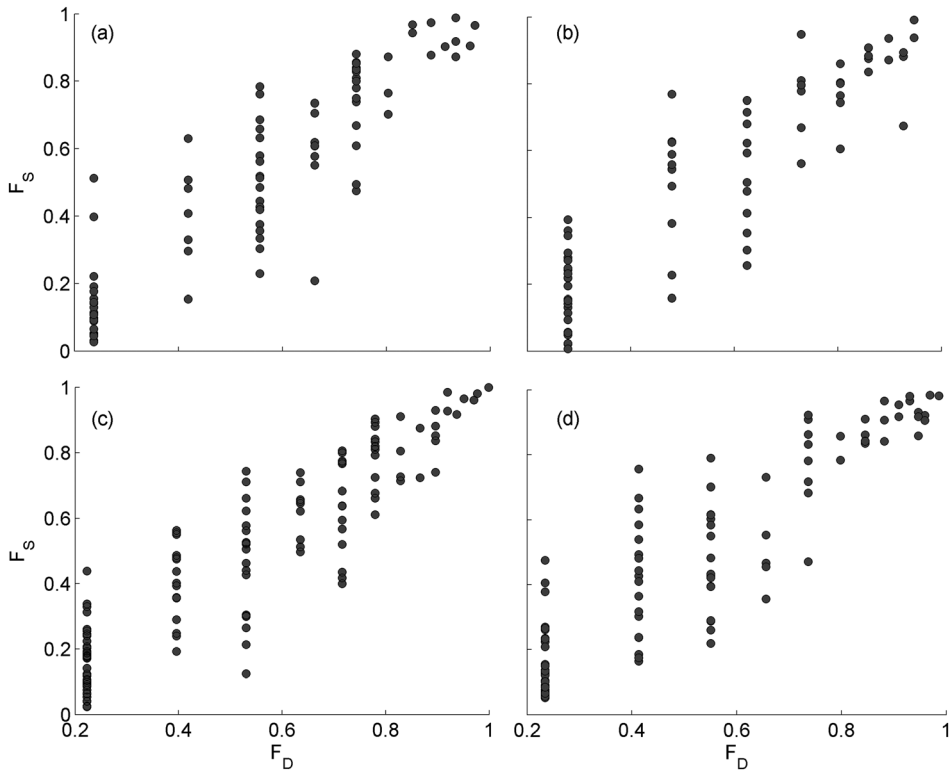

Figure 4. Scatter plots of the CDF of the drought duration $F(d)$ and severity $F(s)$ for (a) Anzali, (b) Zahedan, (c) La Tuque and (d) Grande Prairie stations

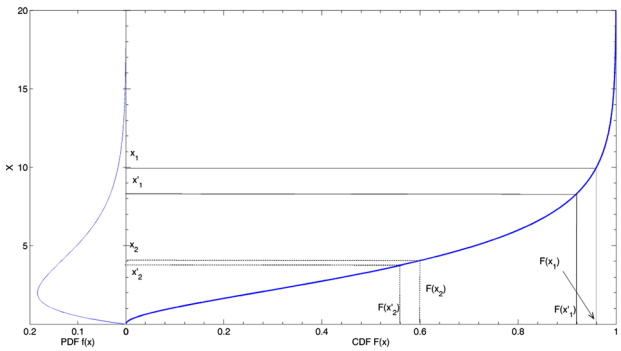

Figure 5. Example of the gamma PDF (Equation 1) and CDF with α and β equal to 2.0

mentioned CDFs are also shown in the ordinate of Figure 5 (x_1 and x'_1 versus x_2 and x'_2). One can see that the difference between the quantiles on the upper tail (x_1 and x'_1) is much more important than that of the quantiles in the middle of the CDF (x_2 and x'_2). In other words, during the CDF transformation (i.e. from Figure 3 to Figure 4a), the large differences corresponding to high quantiles (X) decrease in the CDF values for a positively skewed marginal distribution. This implies that even if in the real domain the observed data illustrates weak association for high values, the CDF values might show a strong association for high values in the transformed domain. Here, the key point is that the copula function is determined by the association in the 'CDF domain' (Figure 4a) rather than that in the real domain (Figure 3) referring to Equation 7. Therefore, the implication from the real domain is not necessarily appropriate to select a suitable copula function.

If one chooses a copula function from the intuition of the observed domain (e.g. Figure 3), the Clayton copula is seemingly an appropriate function (Figure 1a). However, the Gumbel copula (Figure 1b) is a better option because the representation of the copula is based on the CDF transformed domain.

The comparison of the observed drought duration and severity with the simulated copula functions (Figures 1 and 4) indicates a higher dependence between F_D and F_S in the upper tail. This implies that the Clayton copula function is not an appropriate function for bivariate drought frequency analysis. On the other hand, the other copula functions seem to be suitable for drought bivariate analysis for the selected stations according to the upper tail characteristics. Table III illustrates the comparative likelihood values for the various copula functions. Note that a higher likelihood function implies a better fit. The Clayton copula function has the lowest log-likelihood value among the four functions for the selected stations. The Frank copula function has the highest value of the log-likelihood for Anzali, Zahedan and La tuque stations, and the Gumbel function is selected for Grande Prairie station. Table III also confirms the role of the tail behaviour of drought duration and severity distribution functions, or the degree of dependence of the duration and severity of observed droughts, in fitting and selecting the appropriate copula function.

The effect of tail behaviour on the selection of a copula function for bivariate drought analysis can also be illustrated by drawing the JCDFs of different copulas.

As an example, the JCDFs of drought duration and severity for Anzali and Grande Prairie stations are presented in Figures 6 and 7. In addition to the copula functions, the JCDFs of the empirical and independent cases for drought duration and severity are also given in Figures 6a and 6b and 7a and 7b. It can be seen that the curves of Frank and Gumbel copula functions are more similar to the empirical JCDF than other functions for both Anzali and Grande Prairie stations especially for high duration and severity values. A similar behaviour of JCDF is observed for the Zahedan and La Tuque stations (data not shown).

Bivariate return period of drought duration and severity

A comprehensive discussion of multivariate quantile and return period curves is presented in Chebana and Ouarda (2011). The bivariate return periods of drought duration and severity are presented in Figures 8 and 9 for the two cases, T_{DS} and T'_{DS} (see Equations 16 and 17). From Figures 8 and 9, it can be seen that the shape and the variation of the return period curves for both T_{DS} (left panels) and T'_{DS}(right panels) rely heavily on the type of the copula function. Figure 8, illustrating the results of the Anzali station, indicates that curves corresponding to T_{DS} and T'_{DS} for the Clayton copula (b-1 and b-2) are similar to the ones corresponding to the independent case (a-1 and a-2). In other words, for the Clayton drought function, the relationship between drought duration and severity seems to be insignificant.

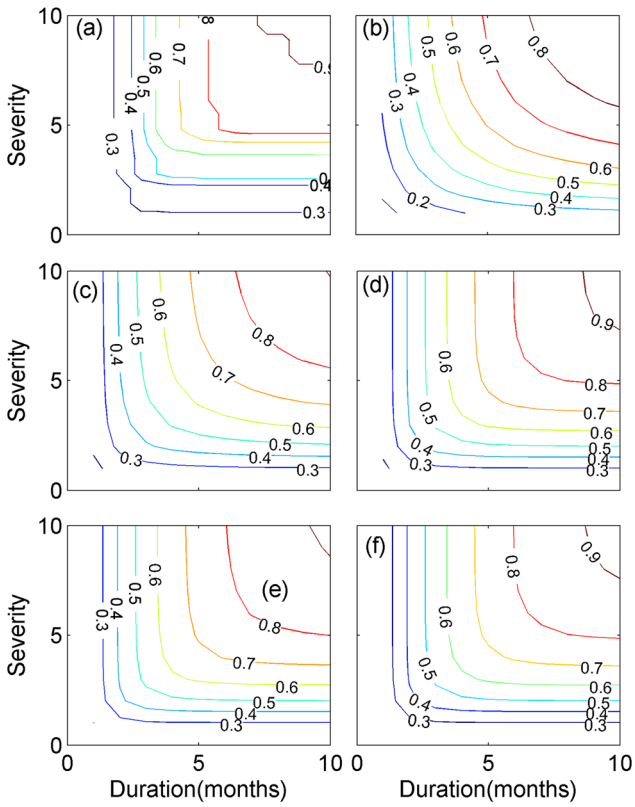

Figure 6. JCDF of drought duration and severity for the (a) empirical, (b) independent, (c) Clayton, (d) Gumbel, (e) Frank and (f) Gaussian copula functions for Anzali station

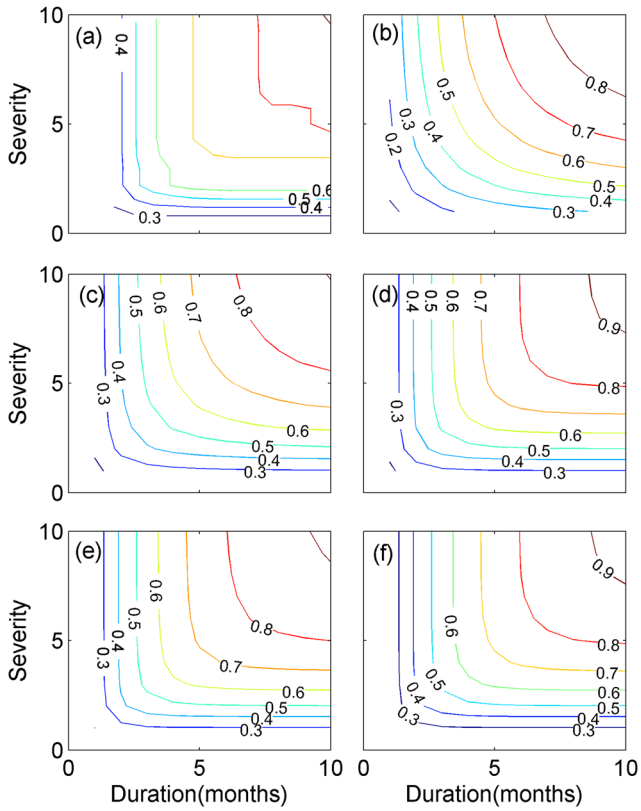

Figure 7. The same as Figure 6 but for Grand Prairie station

As an example, assume that one needs to define the drought event such that $D \geq 5$ and $S \geq 5$. The return period of this drought event in the independent case in Figure 8 (a-1) is approximately 14 years, whereas it is 7 years for the Clayton copula and approximately 4 years for the other three copulas. However, in the case of T'_{DS}, the effect of using different copulas is less significant than that in the case of T_{DS} as shown in the right panels of Figure 8. On the other hand, the return period of drought event $D \geq 5$ or $S \geq 5$ is between 1 and 2 years, according to the independent and the Clayton copula function, whereas it is more than 2 years on the basis of other copula functions.

Finally, the return period plots corresponding to the three remaining stations with the selected copula functions are given in Figure 9, which reveals that the two stations (La Tuque and Grande Prairie) present similar return periods whereas the Zahedan station has higher return periods for the same events.

Therefore, it is concluded that the selection of the appropriate copula function is a critical task in bivariate drought frequency analysis. The underestimation of the return period for drought duration or severity might increase the risk of failure of drought control structures and drought management measures, both for at-site and regional scales.

SUMMARY AND CONCLUSIONS

A bivariate drought analysis study based on copula functions was carried out in this article. Different marginal distributions were examined to fit drought

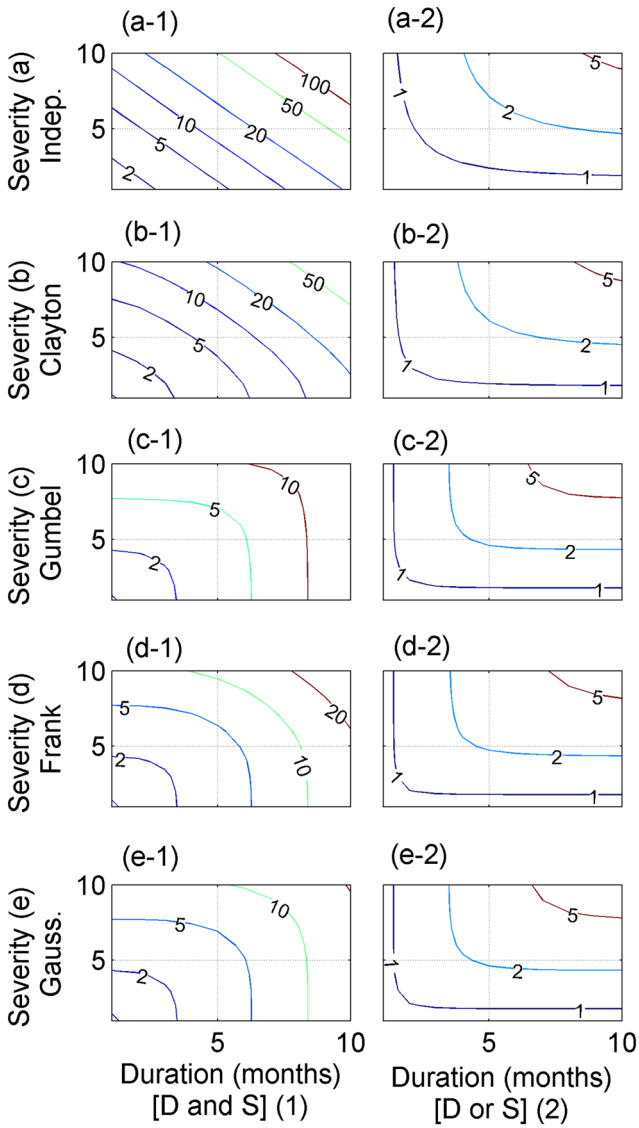

Figure 8. Return period (years) of (1) DS in Equation 16, left column and (2) DS' in Equation 17, right column with (a) independent, (b) Clayton, (c) Gumbel, (d) Frank and (e) Gaussian copula for Anzali station

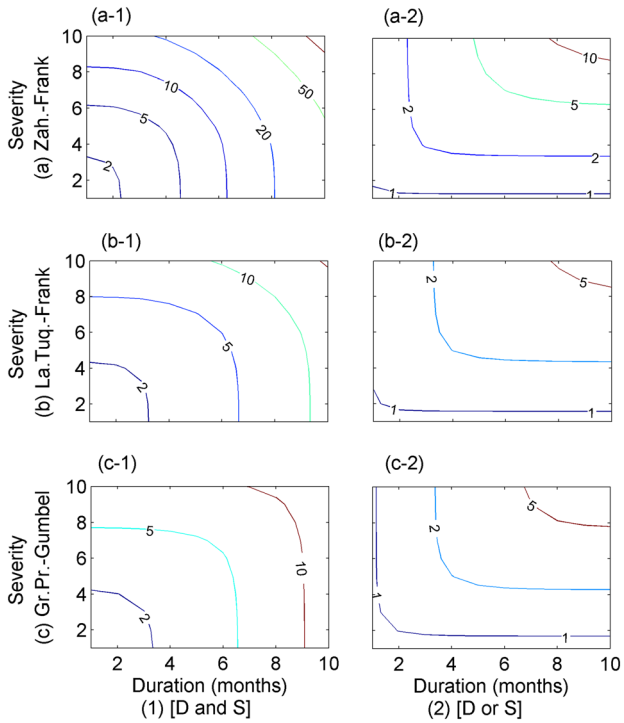

Figure 9. Return period of (11) DS in Equation (16), left column and 2) DS' in Equation (17), right column for (a) Zahedan–Frank, (b) La Tuque–Frank and (c) Grande Prairie–Gumbel

duration and severity data for four stations from Canada and Iran. It was illustrated that the gamma and exponential distributions represent the most adequate marginal distributions to fit the observed drought data with positive skewness.

It was observed that the association between drought duration and severity is notably different in the real domain and in the CDF transformed domain. These differences are especially significant when the marginal distributions are highly skewed. The examples considered in the present study represent positively skewed drought observations, and the weaker association between duration and severity in the upper tail in the real domain is not sustained in the CDF transformed domain. This would be the main reason for the weak performance of the Clayton copula function compared with the other copula functions (i.e. Frank, Gumbel and Gaussian) for bivariate drought frequency analysis.

Clayton copula has traditionally been selected for bivariate drought frequency analysis in the literature without an elaborate selection process due to the relational presentation between drought severity and duration in the original domain (see Figure 3). It was observed in the present study that the Clayton copula is not capable of modelling the association between drought duration and severity. The comparison of the performance of four copula functions with the empirical JCDF of the observed data supports this conclusion. The results indicate that the shape of the upper tail of the empirical JCDF for the four stations is different from the shape of the Clayton copula function but similar to the shape of the Frank and Gumbel copula functions. Furthermore, it was observed that the empirical return period curves of DS and DS' (Equations 16 and 17) are significantly different from those corresponding to the Clayton copula model. The use of the Clayton copula induces a considerable overestimation of the return period corresponding to certain drought events especially in the case of DS (Equation 16). Therefore, highly frequent drought events in reality may be estimated as low-frequency drought events according to the Clayton copula function. Eventually, this could lead to the failure of drought risk management programs and to inefficient disaster mitigation measures.

The results of this data-based analysis indicate that the Clayton copula is not an appropriate choice for bivariate drought frequency analysis. In fact, the use of the Clayton copula function does not lead to significantly more information than when drought duration and severity are assumed to be independent. Alternatively, Frank or Gumbel copulas can be selected for bivariate drought frequency analysis because these copulas reproduce well

the upper tail dependence structure between drought duration and severity.

Future work dealing with the effect of tail dependence on bivariate drought frequency analysis can examine simulated data sets to consider a wide range of distribution characteristics for the selection of the proper copula functions The investigation of other copula functions can shed more light on the importance of the tail dependence in bivariate drought frequency analysis and could lead to the reduction of the risk of wrongly estimating drought return periods. Finally, the development of a regional bivariate drought frequency analysis model would be an important step as drought usually occurs on a regional scale.

REFERENCES

Arena C, Cannarozzo M, Mazzola MR. 2006. Multi-year drought frequency analysis at multiple sites by operational hydrology—A comparison of methods. *Physics and Chemistry of the Earth* **31**: 1146–1163.

Chebana F, Ouarda TBMJ. 2011. Multivariate quantiles in hydrological frequency analysis. *Environmetrics* **22**: 63–78.

De Michele C, Salvadori G. 2003. A Generalized Pareto intensity-duration model of storm rainfall exploiting 2-Copulas. *Journal of Geophysical Research-Atmospheres* **108**: 1–11.

De Michele C, Salvadori G, Passoni G, Vezzoli R. 2007. A multivariate model of sea storms using copulas. *Coastal Engineering* **54**: 734–751.

El-Jabi N, Ashkar F, Issa M. 1997. Stochastic application of the low-water phenomenon. *Canadian Journal of Civil Engineering* **24**: 191–200.

El Adlouni S, Bobee B, Ouarda TBMJ. 2008. On the tails of extreme event distributions in hydrology. *Journal of Hydrology* **355**: 16–33.

Frick DM, Bode D, Salas JD. 1990. Effect of Drought on Urban Water-Supplies. 2. Water-Supply Analysis. *Journal of Hydraulic Engineering ASCE* **116**: 754–764.

IPCC. 2007. Climate Change 2007: Working Group I: The Physical Science Basis, Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL (eds).

Kao SC, Govindaraju RS. 2008. Trivariate statistical analysis of extreme rainfall events via the Plackett family of copulas. *Water Resources Research* **44**: 1–19.

Kao SC, Govindaraju RS. 2010. A copula-based joint deficit index for droughts. *Journal of Hydrology* **380**: 121–134.

Karamouz M, Rasouli K, Nazif S. 2009. Development of a hybrid Index for drought prediction: Case study. *Journal of Hydrologic Engineering* **14**: 617–627.

Khaliq MN, Ouarda TBMJ, Gachon P. 2009. Identification of temporal trends in annual and seasonal low flows occurring in Canadian rivers: The effect of short- and long-term persistence. *Journal of Hydrology* **369**: 183–197.

Kottegoda NT, Rosso R. 2008. Applied Statistics for Civil and Environmental Engineers. Wiley-Blackwell: West Sussex, UK; 736

Kroll CN, Vogel RM. 2002. Probability distribution of low streamflow series in the United States. *Journal of Hydrologic Engineering* **7**: 137–146.

Lana X, Burgueno A. 1998. Spatial and temporal characterization of annual extreme droughts in Catalonia (Northeast Spain). *International Journal of Climatology* **18**: 93–110.

Laux P, Wagner S, Wagner A, Jacobeit J, Bardossy A, Kunstmann H. 2009. Modelling daily precipitation features in the Volta Basin of West Africa. *International Journal of Climatology* **29**: 937–954.

Lee KS, Sadeghipour J, Dracup JA. 1986. Approach for Frequency Analysis of Multiyear Drought Durations. *Water Resources Research* **22**: 655–662.

Lee T, Salas JD. 2011. Copula-based stochastic simulation of hydrological data applied to Nile River flows. *Hydrology Research* **42**: 318–330.

Loaiciga HA, Leipnik RB. 1996. Stochastic renewal model of low-flow streamflow sequences. *Stochastic Hydrology and Hydraulics* **10**: 65–85.

McKee TB, Doesken NJ, Kleist J. 1993. The relationship of drought frequency and duration to time scales. *Proceedings of the Eighth Conference on Applied Climatology*, American Meteorological Society, Boston; 179–184.

Mishra AK, Singh VP. 2011. Drought modeling—A review. *Journal of Hydrology* **403**: 157–175.

Modarres R. 2007. Streamflow drought time series forecasting. *Stochastic Environmental Research and Risk Assessment* **21**: 223–233.

Modarres R. 2008. Regional frequency distribution type of low flow in North of Iran by L-moments. *Water Resources Management* **22**: 823–841.

Modarres R, Sarhadi A. 2010. Frequency distribution of extreme hydrologic drought of southeastern semiarid region, Iran. *Journal of Hydrologic Engineering* **15**: 255–264.

Nadarajah S. 2009a. A bivariate distribution with gamma and beta marginals with application to drought data. *Journal of Applied Statistics* **36**: 277–301.

Nadarajah S. 2009b. A bivariate pareto model for drought. *Stochastic Environmental Research and Risk Assessment* **23**: 811–822.

Nelsen RB. 1999. An Introduction to Copulas. Springer-Verlag: New York; 217.

NOAA. 2010. Climate of 2010—US Standardized Precipitation Index.

Ouarda TBMJ, Charron C, St-Hilaire A. 2008. Statistical models and the estimation of low flows. *Canadian Water Resources Journal* **33**: 195–206.

Ouarda TBMJ, Shu C. 2009. Regional low-flow frequency analysis using single and ensemble artificial neural networks. *Water Resources Research* **45**: 1–16.

Panu US, Sharma TC. 2002. Challenges in drought research: Some perspectives and future directions. *Hydrological Sciences Journal-Journal Des Sciences Hydrologiques* **47**: S19–S30.

Panu US, Sharma TC. 2009. Analysis of annual hydrological droughts: The case of northwest Ontario, Canada. *Hydrological Sciences Journal* **54**: 29–42.

Pielke RA, Doesken N, Bliss O, Green T, Chaffin C, Salas JD, Woodhouse CA, Lukas JJ, Wolter K. 2005. Drought 2002 in Colorado: An unprecedented drought or a routine drought? *Pure and Applied Geophysics* **162**: 1455–1479.

Ryu JH, Lee JH, Jeong S, Park SK, Han K. 2011. The impacts of climate change on local hydrology and low flow frequency in the Geum River Basin, Korea. *Hydrological Processes* **25**: 3437–3447.

Salas JD, Fu CJ, Cancelliere A, Dustin D, Bode D, Pineda A, Vincent E. 2005. Characterizing the severity and risk of drought in the Poudre River, Colorado. *Journal of Water Resources Planning and Management-ASCE* **131**: 383–393.

Salvadori G, De Michele C. 2004. Analytical calculation of storm volume statistics involving Pareto-like intensity-duration marginals. *Geophysical Research Letters* **31**: 1–4.

Santos JF, Pulido-Calvo I, Portela MM. 2010. Spatial and temporal variability of droughts in Portugal. *Water Resources Research* **46**: 1–13.

Schindler DW, Donahue WF. 2006. An impending water crisis in Canada's western prairie provinces. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 7210–7216.

Sen Z. 1980. Statistical analysis of hydrologic critical droughts. *Journal of the Hydraulics Division-ASCE* **106**: 99–115.

Shabbar A, Skinner W. 2004. Summer drought patterns in Canada and the relationship to global sea surface temperatures. *Journal of Climate* **17**: 2866–2880.

Sharma KD. 2000. Water resource development in jamnagar district, Gujarat, India. *Annals of Arid Zone* **39**: 145–150.

Shiau JT. 2006. Fitting drought duration and severity with two-dimensional copulas. *Water Resources Management* **20**: 795–815.

Shiau JT, Feng S, Nadaraiah S. 2007. Assessment of hydrological droughts for the Yellow River, China, using copulas. *Hydrological Processes* **21**: 2157–2163.

Shiau JT, Modarres R. 2009. Copula-based drought severity-duration-frequency analysis in Iran. *Meteorological Applications* **16**: 481–489.

Shiau JT, Shen HW. 2001. Recurrence analysis of hydrologic droughts of differing severity. *Journal of Water Resources Planning and Management-ASCE* **127**: 30–40.

Sklar M. 1959. Fonctions de r´epartition `a n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris* **8**: 229–231.

Song S, Singh VP. 2010a. Frequency analysis of droughts using the Plackett copula and parameter estimation by genetic algorithm. *Stochastic Environmental Research and Risk Assessment* **24**: 783–805.

Song S, Singh VP. 2010b. Meta-elliptical copulas for drought frequency analysis of periodic hydrologic data. *Stochastic Environmental Research and Risk Assessment* **24**: 425–444.

Wong G, Lambert MF, Leonard M, Metcalfe AV. 2010. Drought analysis using trivariate copulas conditional on climatic states. *Journal of Hydrologic Engineering* **15**: 129–141.

Zaidman MD, Keller V, Young AR, Cadman D. 2003. Flow-duration-frequency behaviour of British rivers based on annual minima data. *Journal of Hydrology* **277**: 195–213.

