

Application of copula modelling to the performance assessment of reconstructed watersheds

Alireza Nazemi · Amin Elshorbagy

Published online: 21 April 2011
© Springer-Verlag 2011

Abstract The existence of interdependence among environmental variables has been qualitatively known for centuries. Recent studies have shown that copula modelling can provide a simple, yet powerful framework for modelling interdependence among hydrological data; however, still there are several studies which use outdated and superficial methods to perform this task. By considering the current state of knowledge, this study tries to introduce a pragmatic procedure to perform copula modelling in real-world problems. Our study uses copula modelling to find a notion for conditional quantities of the maximum annual water deficit with respect to the annual cumulative evapotranspiration. Therefore, by having an estimate for the annual cumulative evapotranspiration, the hydrological performance of the reconstructed watershed can be assessed even in nearby ungauged reconstructed watersheds with similar physical characteristics and reclamation strategy. Several competitive models are developed for joint description of these variables in a prototype reclaimed oil-sands mining site in northern Alberta, Canada. The developed joint models are compared and analyzed according to their convergence feasibility, overall accuracy, tail behaviour and a goodness-of-fit test. Our study concludes that copula modelling can be considered as a powerful option in practitioners' toolkit. For

the case under consideration, the Gumbel–Hougaard structure provides the most credible model of dependence. Moreover, our study provides some initial supports for the application of minimum distance methods for copula parameter estimation.

Keywords Reconstructed watersheds · Interdependence · Copulas · Parameter estimation · Falsification framework

1 Introduction

It has been widely warned that implementing simple univariate statistical approach in the case of dependence among variables may result in severe under/over estimations. Traditionally, classical multivariate distributions are used as the main tool for such considerations. If so, the marginal variables (or their transformations) must be described using identical parametric families. Copula modelling provides an alternative for overcoming such limitations based on a theorem given by Sklar in 1959. In its bivariate form, if x and y are two continuous random variables with continuous marginal cumulative distributions of $F(x)$ and $G(y)$, Sklar's theorem implies that the joint cumulative distribution of $H(x,y)$ can be described as:

$$H(x,y) = C(u,v) \quad x,y \in \mathfrak{R} \text{ and } u = F(x), v = G(y) \quad (1)$$

In this case, $C: [0,1]^2 \rightarrow [0,1]$ is the copula function, which *couples* the cumulative marginal distributions and form the cumulative joint distribution. It can be proved that if H is known, then C , F , and G are uniquely determined (Genest and Favre 2007). So technically, the copula approach divides the problem into two mutually independent steps concerning fitting the marginal distributions and finding the dependence structure among

A. Nazemi (✉)

Global Institute for Water Security, University of Saskatchewan,
National Hydrology Research Centre, 11 Innovation Boulevard,
Saskatoon, SK S7N 3H5, Canada
e-mail: aln626@mail.usask.ca

A. Elshorbagy

Centre for Advanced Numerical Simulation (CANSIM),
Department of Civil and Geological Engineering, University of
Saskatchewan, Saskatoon, SK S7N 5A9, Canada
e-mail: amin.elshorbagy@usask.ca

the margins (Wang et al. 2009). This ability can provide a powerful modelling scheme even “when no common probability space can be found for a given set of random variables” (Sklar 1996). Furthermore, Genest and MacKay (1986) showed how copulas can discover dependencies that traditional dependence measures, such as Pearson’s correlation coefficient, cannot capture.

In the most recent decade, the rate of publications in copula modelling and its applications has increased exponentially in various fields. The application of copula modelling in hydrology and environmental modelling was initialized by De Michele and Salvadori (2003) and was followed by Favre et al. (2004), who discussed the advantages of copula modelling in characterizing complex hydrological variables. This recognition was amplified by dedication of a special issue of the *ASCE Journal of Hydrologic Engineering* (2007) to copula modelling and its application in hydrology. So far, the application of copula modelling in hydrology has been mainly cited in the extreme events modelling, such as flood frequency analysis (e.g. Zhang and Singh, 2006; Bénard and Lang 2007; Genest et al. 2007; Karmakar and Simonovic 2008, 2009; Wang et al. 2009; Chebana and Ouarda 2009), flood retention measures (Nijssen et al. 2009), designing flood control systems (e.g. Klein et al. 2008; Osorio et al. 2009; Muhaisen et al. 2009), inland flooding analysis (e.g. Sunyer Pinya et al. 2009), flood damage calculation (e.g. Gartsman et al. 2009), and characterization of drought or low flow conditions (e.g. Shiau et al. 2006, 2007; Dupuis 2007; Serinaldi et al. 2009; Kao and Govindaraju 2010). Recently, the application of copula modelling has been extended to new areas. For example, Bárdossy (2006) discussed the applicability of copula as a geostatistical model for deriving the groundwater quality parameters. Salvadori and De Michele (2007) used copulas for finding the temporal structure of the sequential storms. Chowdhary and Singh (2008) used copula for rainfall frequency analysis. Serinaldi (2008) used copulas for unfolding the dependencies among the rainfall fields. In more advanced examples, Serinaldi (2009a) used copulas as a module embedded in rainfall simulators. Haberlandt et al. (2008) implemented copulas for space–time hybrid hourly rainfall models. Leonard et al. (2008) applied copula for considering the seasonal/annual dependencies among rainfall and streamflow. Maity and Kumar (2008) reported the application of copulas for analyzing the dependencies among the teleconnected hydroclimatic variables. Bárdossy and Pegram (2009) implemented copulas to model the spatial interdependence structure of the rainfall amounts together with the rainfall occurrences. Karamouz et al. (2009) used copulas for deriving the joint probability of supplying both water demand and water quality in a complex reservoir operation system. In a regionalization context, Gargouri-

Ellouze and Bargaoui (2009) showed that the copula parameters derived for describing infiltration can be further linked to the actual physical properties of the watershed. In the same context, Samaniego et al. (2010) performed streamflow prediction in ungauged catchments using copula-based dissimilarity measures. More recently, van den Berg et al. (2010) used copulas for spatial downscaling of rainfall fields. The applications are still growing rapidly (www.stahy.org/Topics/CopulaFunction/tabid/67/Default.aspx).

Nonetheless, it should be noted that copula modelling has received some strong criticism. In a landmark discussion, Mikosch (2006) argued that the use of copula modelling might not be scientifically legitimate and can suffer from lack of rationale in many circumstances (see also the reply by Genest and Rémillard 2006). Regardless of the theoretical/philosophical discussions, it has been also warned by copula practitioners that the method might be still far from a universal tool (Bénard and Lang 2007) and could result in disappointing results in some occasions (Genest et al. 2007). Maity and Kumar (2008) concluded that even though Sklar’s theorem is mathematically coherent, this does not imply that one of the classical copulas can certainly describe the dependence among all possible datasets. As a result, it has been argued by Ashkar (2008) that practitioners should be very careful not to consider copula modelling as the only approach to tackle the statistical dependence. Therefore, copulas should be considered in conjunction with other models of dependence and should only remain as a potential candidate for describing the dependence in a given dataset.

This paper aims at providing a pragmatic framework to perform copula modelling in real-world practical contexts and to present the application of copula modelling in a new practical context—assessing the long-term behaviour of reconstructed watersheds. Long-term behaviour of reconstructed watersheds can be assessed by the aid of maximum annual water deficit, which is a measure of available water for evapotranspiration (Elshorbagy and Barbour 2007). However, estimation of maximum annual water deficit requires the knowledge of soil moisture dynamic within the watershed, which can be only available in limited watersheds with extensive soil moisture monitoring. By recognizing the interdependence between maximum annual soil moisture deficit and annual cumulative evapotranspiration, building the joint description of these variables is attempted. By having the joint model and the estimation for the annual cumulative evapotranspiration, the cumulative distribution of maximum annual water deficit can be estimated even in nearby ungauged reclaimed sites with similar physical characteristics and reclamation strategy. Throughout our study, a brief review on the theory of copula modelling will be presented. Accordingly, three

research questions would be addressed: (1) how to compare and to select among several competing models of dependence; (2) how the efficiency of the dependence model can depend on the method used for parametric estimation; and finally, (3) how the efficiency of model can be changed by altering the structure of the dependence model. The outline of this paper is as follows: Section 2 introduces the considered application area as well as the case study and the data used for the experimentations. In Sect. 3, the dependence modelling is viewed as a top-down sequential procedure, and a brief review of the copula modelling is provided. Section 4 introduces the applied test space and proposes a two-step framework to compare/falsify several competitive models of dependence. In Sect. 5, the results are reported and discussed from several perspectives, and finally Sect. 6 summarizes the paper, provides the conclusions, and highlights some directions for further research.

2 Performance assessment of the reconstructed watersheds

2.1 Context and the problem definition

Oil sands mining industry is one of the main economical resources in Canada. The extraction process of the oil can completely disrupt the natural behaviour of the mining area. Based on Leskiw (2004), the end results of the oil sands mining can be spatially extended to over 100 km² from each individual site. An extensive reclamation work, therefore, is required to reconstruct various functions of the natural landscape. The main reclamation strategy in these sites is to cover the tailing sand remaining of the mining process by one or multiple soil layers that enable the watershed to naturally “store and release the water” (Elshorbagy and Barbour, 2007). This ability can provide the chance for re-planting the mining site and re-establishing the natural conditions in the watershed. In order to evaluate the hydrological efficiency of different options of soil covers, building some prototype watersheds with an extensive monitoring program is essential. The acquired knowledge can result in proposing an optimal cover design regarding the characteristics of the site and the target vegetation. Based on Elshorbagy et al. (2006), by gathering a short-period of data (i.e. few years), a hydrological model can be calibrated with the short-term observations. Then, the verified model can be used to simulate the long-term behaviour of the watershed. Based on the long-term simulation results, Elshorbagy and Barbour (2007) proposed a novel methodology for hydrological performance assessment of the reconstructed watersheds. They suggested the dynamic concept of maximum annual moisture deficit as a

measure to assess the performance of the soil cover. Assuming that the averaged daily soil moisture content (S_i), interflow (I_i), and percolation below the cover depth (P_i) are available through long-term simulation, the daily moisture deficit (D_i) that contributes to evapotranspiration can be calculated as (Elshorbagy and Barbour 2007):

$$D_i = S_i - S_{i+1} - (I_i + P_i) \quad (2)$$

By accumulating the annual daily series of D_i , it would be possible to retrieve the maximum annual soil moisture deficit. The daily values of D_i should be accumulated only over the growing season. The maximum value of the cumulative D_i in each year can be marked as the maximum annual soil moisture deficit. By having a long-term series of maximum D_i values, it would be possible to tackle the long-term behaviour of the reconstructed catchment in a statistical manner.

Although the concept of maximum annual moisture deficit can provide a reliable dynamic notion of the watershed performance, the calculation of this measure is based on the soil moisture content, which can be only available in limited prototype watersheds with extensive soil moisture measurement scheme. However, given a unique physical characteristics and reclamation strategy, the difference in the maximum annual water deficit between two nearby reconstructed watersheds stems from the difference between their annual cumulative evapotranspiration quantities. Therefore, if the model of dependence between annual cumulative evapotranspiration and maximum annual moisture deficit is known, the cumulative distribution of maximum annual water deficit can be estimated based on the measurements or estimations of annual cumulative evapotranspiration.

2.2 Case study

Located in a continental boreal climate regime, Young Jack Pine (YJP) is a 15-year old planted forest on a reconstructed site on Mildred Lake mine (57°03.8'N, 111°39.8'W, elev. ~310 m). The mine is about 40 km NNW of Fort McMurray, Alberta. The average daily temperature ranges from -18.8 to 16.8°C (Jan.–Jul.), with average annual precipitation of 456 mm. The growing season falls between mid May to early October with 60–70 frost-free days. Most of the precipitation falls within this interval as 313 mm typically falls during day 139–282 of each year. The reconstructed cover is composed of glacial till down to an average depth of 40 cm, placed over a deep layer of sandy soil from the oil sands extraction process. The gathered data from the site includes the daily on-site meteorological data (the micrometeorological data including AET is only available during the growing seasons) as well as 4-h measurements of the soil moisture contents, soil

temperature and soil suctions at five different depths. Three measurements were made in the upper till layer (depths 50, 150 and 300 mm) and the other two in the underlying tailing sand base (depths 500 and 800 mm). In order to measure the turbulent fluxes at the site, an open-path eddy covariance system was used, mounted approximately 4 m above the canopy at height of 8.8 m. Radiant fluxes were measured with instrumentation mounted at a height of 6.3 m. Ground heat flux was obtained using a soil heat flux plate at 5 cm. Precipitation data were measured using a tipping-bucket rain gauge located less than 1 km west of the tower. In total, two years (2007–2008) of daily climate and soil data were available for YJP. Keshta et al. (2009) proposed a Generic System Dynamic Watershed model (GSDW) for hydrological modelling in reconstructed watersheds with similar set of data. GSDW, therefore, has been used for further hydrological modelling in this paper. In brief, the 2007 data of soil moisture and evapotranspiration in YJP were used to calibrate sixteen different conceptualizations of the GSDW model. These conceptualizations were based on different incorporation of soil and/or evapotranspiration data in the model structure. Accordingly, 16 competitive models were calibrated using four automatic calibration techniques. The performances of the competitive modelling alternatives were validated using the 2008 dataset and the most accurate conceptualization/parameterization has been identified. By the aid of this model and the meteorological data of Fort McMurray from the beginning of 1944 to the end of 2000, a notion for long-term behaviour of the watershed was simulated. According to this long-term simulation, 56 pairs of maximum annual moisture deficit and annual cumulative evapotranspiration were produced. This dataset is considered for further experimentation in this paper. Table 1 shows the statistical characteristics of this dataset.

3 Copula approach to modelling interdependence

3.1 Background

Considering the bivariate case, Fréchet-Hoeffding theorem proves that the possible forms of dependence are bounded between two extremes, which can be formulated as the following inequality (Genest and Favre 2007):

$$\max(0, u + v - 1) \leq C(u, v) \leq \min(u, v), \quad u, v \in [0, 1] \quad (3)$$

The right and left bounds of Eq. 3 imply conditions in which x and y are deterministically dependent on each other. In this juncture, the stochastic independence can be verified, if and only if, $C(u, v) = uv$, $u, v \in [0, 1]$. This verification requires setting up both marginal and joint statistical representation of the data and can be described by a top-down framework presented in Fig. 1. Here, the task is to briefly explain the main building blocks of this system.

3.2 Pre-processing and marginal modelling

Nelsen (2006) suggested that the rank-based dependence measures such as Kendall's tau and Spearman's rho can better capture the existence of dependence compared to correlation coefficient. The empirical formulations of these measures can be described respectively as:

$$\tau = \frac{n_c - n_d}{\frac{1}{2}n(n-1)} \quad (4)$$

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n-1)(n+1)} \quad (5)$$

where n_c and n_d are the numbers of concordant pairs and discordant pairs in the data set, respectively; and, d_i is the differences between the ranks of each observation of the two pairs of variables. Genest and Favre (2007) introduced two hypothetical statistical tests for verifying independence according to Kendall's tau and Spearman's rho.

It is better to keep the number of marginal variables as low as possible. This can be achieved either by using the dimension reduction techniques, such as the principal component analysis or considering the possible functional relationships among the marginal variables (e.g. Chowdhary 2008). After selecting the non-redundant margins, the marginal distributions (both empirical and theoretical) should be identified. The empirical cumulative distribution function (CDF) for a continuous random variable x can be calculated as:

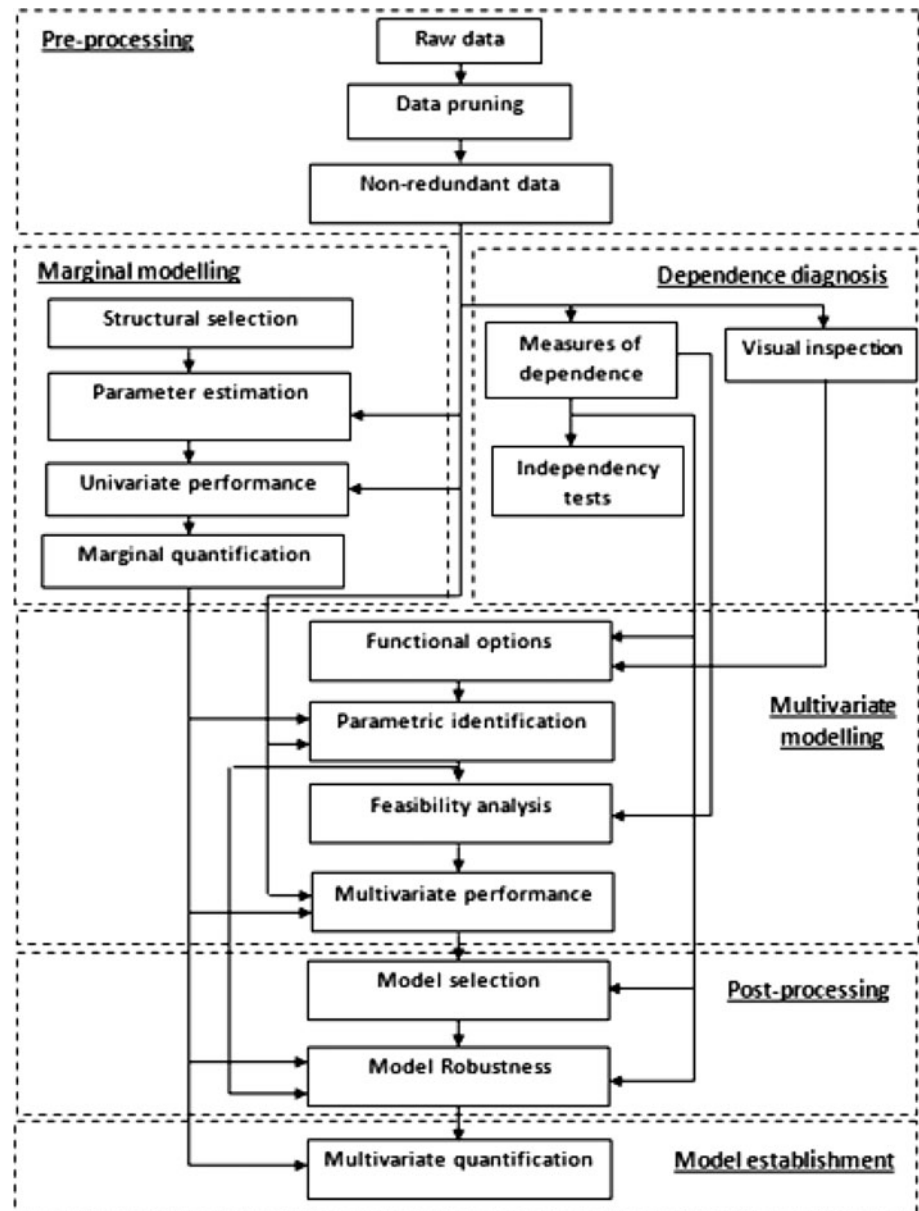
$$F_n(x) = \frac{1}{n+1} \sum_{i=1}^n 1(X_i \leq x) \quad (6)$$

where n is the number of data points. Also, various continuous statistical distributions can be used for

Table 1 The statistical properties of the long-term simulated annual series of maximum water deficit and the annual cumulative evapotranspiration for YJP prototype reclaimed site

	Lower bound	Mean	Median	Upper bound	Standard deviation	Skewness	Kurtosis
Annual maximum water deficit	2.28	56.22	49.30	150.29	38.95	0.61	2.49
Annual cumulative evapotranspiration	235.27	295.13	292.09	343.59	31.73	-0.07	1.82

Fig. 1 A top-down approach to static multivariate modelling



functional representation of the margins, which are well documented in the literature and are widely available through different software packages such as R or MATLAB. Moreover, Zhang and Singh (2006) reported that transforming the original data using transformation techniques, such as Box–Cox transformation, can be also implemented and might result in better identification of marginal CDF. This transformation of the hypothetical variable z can be defined as follows:

$$z^* = \frac{z^\lambda - 1}{\lambda}, \lambda \neq 0 \text{ or } z^* = \ln(z), \lambda = 0 \tag{7}$$

where z^* is the Box-Cox transformation of z and λ is the Box–Cox parameter. After the selection of the functional

families, and applying the possible transformations, the parameters of the considered distribution should be assigned. The well-known maximum likelihood estimation has been frequently used for this purpose and has been well incorporated in most of the available statistical packages.

3.3 Copula structures

Copula modelling requires choosing a parametric mathematical structure that can provide the functional equivalent of $H(x,y)$ through copula $C(u,v)$; $u = F(x)$, $v = G(y)$. The current available copula structures can be categorized into four main classes.

3.3.1 Archimedean copulas

These structures are symmetric and associative and can be defined by continuous, strictly decreasing convex generators. One parameter Archimedean copulas have been frequently used in hydrology (e.g. Shiao et al. 2007; Karmakar and Simonovic 2009; Sunyer Pinya et al. 2009; Osorio et al. 2009; Wang et al. 2009). The application of two parameters Archimedean copulas is also reported (e.g. Nijssen et al. 2009).

3.3.2 Metaelliptical copulas

These structures can be formulated based on elliptically contoured distributions (Michiels and De Schepper 2008). Well-known metaelliptical copulas are Gaussian, Student t , Cauchy and Pearson type II (e.g. Fang et al. 2002; Genest et al. 2007; B nard and Lang 2007; Dupuis 2007; Poulin et al. 2007; B rdossy and Pegram 2009). These structures can be further generalized; for instance, Demarta and McNeil (2004) provided a couple of extensions for t copula, called Skewed t and Grouped t copula. In practice, these copulas are more favourable when the joint distribution of three or more variables is concerned.

3.3.3 Extreme value copulas

These structures were initially proposed to deal with the joint probability of extreme events. Different systems have been suggested for constructing the extreme value copulas. Some famous structures such as Gumbel–Hougaard, Galambos, and H sler–Reiss systems, have been frequently implemented in hydrology (e.g. Genest and Favre 2007; Chowdhary and Singh 2008; Maity and Kumar 2008). Salvadori et al. (2007) presented a monograph with several applications of extreme value copulas in geophysical context. Recently, Salvadori and De Michele (2010) presented the application of multi-parameter extreme value copulas.

3.3.4 Other copulas

There are other copula families that cannot be classified according to the three mentioned families. These copulas include some well-known systems, such as Plackett and Farile–Gumbel–Mogenstern (e.g. Shiao 2006; Shiao et al. 2007) as well as some less used copula models, such as Koehler–Symanowski system (Koehler and Symanowski 1995), vine copula (Kurowicka and Cooke 2007) and the contamination families based on Legendre polynomials (Kallenberg 2008).

3.4 Copula parameter estimation

All copula structures contain one or more parameters that should be assigned based on the data under consideration.

The calibration of copula parameters can be classified mainly into five different methods.

3.4.1 Method of moments

In some copula structures, there are mathematical relationships between the copulas' free parameter and the measures of dependence, such as Kendall's tau, Spearman's rho or correlation coefficient. This method of estimation is quite popular in metaelliptical copulas (e.g. B nard and Lang 2007; Genest et al. 2007) and one-parameter Archimedean families (e.g. Zhang and Singh 2006; Maity and Kumar 2008; Sunyer Pinya et al. 2009; Osorio et al. 2009; Chebana and Ouarda 2009). The main limitation of this method is the fact that the closed-form formulas linking the copula parameters to measures of dependence do not exist for all copulas.

3.4.2 The maximum pseudolikelihood estimator

This method, also known as canonical maximum likelihood, is probably the most favourable method for assigning the copula parameters when identifying the model for joint dependence is the main aim of copula modelling. This method is based on maximizing the log-likelihood function of the copula model, which can be described for the bivariate case as follows:

$$l(\theta) = \sum_{i=1}^n \log[c_{\theta}\{F(X_i), G(Y_i)\}] \quad (8)$$

In maximum pseudolikelihood estimator, the marginal distributions $F(X_i)$ and $G(Y_i)$ are replaced by their empirical representation within the sample (Eq. 6). This method can provide a general rule for estimating the copula parameters and mainly requires implementing a numerical procedure. The maximum pseudolikelihood formulation has been used frequently in hydrology (e.g. Dupuis 2007; Poulin et al. 2007; Wang et al. 2009).

3.4.3 Inference from margins (IFM)

In practical circumstances, it might be preferred to use the full likelihood estimator (e.g. Joe 1997; Shiao 2006; Shiao et al. 2006, 2007; Chowdhary and Singh 2008; Karamouz et al. 2009). As a result, the empirical distributions are substituted by the parametric distribution of the marginal variables. Kim et al. (2007) argued that the success of IFM is closely related to the goodness of marginal distributions. Therefore, the estimated copula parameters have the risk of being misidentified because of inappropriate choice of the marginal distributions or their parameters.

3.4.4 Minimum distance method

Recently, some authors proposed new procedures by minimizing the distance between the copula and the empirical joint distribution. The objective function in the bivariate case can be described as (Foscolo et al. 2008):

$$L(\theta) = \sum_{i=1}^n \left[c_{\theta}(\hat{F}_n(x_i), \hat{G}_n(y_i)) - \hat{C}_n(x_i, y_i) \right]^2 \quad (9)$$

where $\hat{F}_n(x_i)$ and $\hat{G}_n(y_i)$ are the marginal distributions (either empirical or theoretical) and $\hat{C}_n(x_i, y_i)$ denotes the rank-based empirical copula (Genest and Favre 2007):

$$\hat{C}_n(u, v) = \frac{1}{n} \sum_{i=1}^n 1 \left(\frac{R_i}{n+1} \leq u, \frac{S_i}{n+1} \leq v \right) \quad (10)$$

where (R_i, S_i) are the pair of observations' rank within the sample. According to the authors' knowledge, there is still no application of this estimation method in the context of hydrology. Genest and Favre (2007) argued that this method results in sub-optimal solutions compared to the pseudolikelihood estimator. In contrast, Foscolo et al. (2008) concluded that when the copula structure is not specified correctly, the minimum distance estimator can provide a better parameterization. These arguments will be further explored in this paper.

3.4.5 Other estimation methods

There are other estimation methods reported in the literature. As an illustration, for the Student metaelliptical copula, Demarta and McNeil (2004) proposed a two-stage approach in which the scale matrix is first estimated by using the relationship with Kendall's tau, and then the pseudo-likelihood function is maximized with respect to the degree-of-freedom using the estimate of the scale matrix. Tsukahara (2005) used the estimating-equation approach to find the copula parameters. Some authors have used Bayesian approach to estimate copula parameters (e.g. Huard et al. 2006; Silva and Lopes 2008). There is also a growing body of literature that implements kernel methods to estimate a smooth representation of a copula without assuming any specific parametric family (e.g. Fermanian and Scaillet 2003).

4 Experimental study

4.1 Rationale and computational platform

Considering that the *true* model of dependence is unknown prior to the experiment, here the aim is to identify the most

credible model of dependence for our dataset based on the top-down system presented in Fig. 1. In each modelling step, several competitive options should be examined in a rigorous falsification framework and the non-falsified options should pass to the next modelling step. The final delivery of this procedure would be a model of dependence that can be used for deriving the conditional probability of maximum annual water deficit based on the given values of annual cumulative evapotranspiration.

Computational platforms for copula modelling are gradually appearing (e.g. Yan 2007). In this study, MATLAB modelling environment (MATLAB 2010) is used. MATLAB Statistics Toolbox contains 20 statistical distributions for continuous univariate description of marginal variables. Also by applying Box–Cox transformation, the Normal distribution can be further extended to the transformed data. The parameters of marginal distributions are estimated using the maximum likelihood estimation. MATLAB also contains several multivariate distributions as well as copula-based model of dependence. By coding more copula-based multivariate descriptions, optimization-based parametric estimation methods, graphical inspections and goodness-of-fit tests, a unified computational package was assembled to conduct the experiments in this study.

4.2 Test space

Michiels and De Schepper (2008) discussed that the functional structure for describing multivariate data should be diverse and relevant. Table 2 shows different modelling options used in this study. Both methods of maximum likelihood and minimum distance are implemented in their full (considering theoretical marginal distributions) and pseudo (with the empirical marginal distributions) forms for identifying the models' parameters. The method of moments was also used for the structures in which the model parameters have direct relationship with the measures of dependence.

4.3 Initial ranking of the dependence models

Assume that the parametric copula structure C_{θ} is fitted to the bivariate data $(X_1, Y_1), \dots, (X_n, Y_n)$ with the marginal distribution of $F_X(x)$ and $G_Y(y)$ and the empirical copula \hat{C} . In this juncture, the goodness-of-fit measures can be calculated in two different fashions. In the pseudo measures, the empirical marginal distributions are considered, therefore, the task is to check the ability of the model of dependence in coupling the marginal variables. On the other hand, in the full goodness-of-fit assignment, the marginal variables are considered with their fitted theoretical distributions $F_X(x)$ and $G_Y(y)$. So, the task is to

Table 2 The multivariate test space implemented in this study

Family	Structure	Reference	Platform	Estimation
Classical multivariate	Normal	MATLAB (2010)	MATLAB built-in	Maximum likelihood (full) Minimum distance (full)
	t	MATLAB (2010)	MATLAB built-in	Maximum likelihood (full) Minimum distance (full)
Metaelliptical copulas	Gaussian	Bénard and Lang (2007)	MATLAB built-in	Maximum likelihood (full, pseudo) Minimum distance (full, pseudo) Method of moment
	t -Student	Demarta and McNeil (2004)	MATLAB built-in	Maximum likelihood (full, pseudo) Minimum distance (full, pseudo) Method of moment
Archimedean copulas	Clayton	Shiau (2006)	MATLAB built-in	Maximum likelihood (full, pseudo) Minimum distance (full, pseudo) Method of moment
	Ali-Mikhail-Haq	Shiau (2006)	Coded function	Maximum likelihood (full, pseudo) Minimum distance (full, pseudo)
	Frank	Shiau (2006)	MATLAB built-in	Maximum likelihood (full, pseudo) Minimum distance (full, pseudo) Method of moment
	BB1	Genest and Favre (2007)	Coded function	Maximum likelihood (full, pseudo) Minimum distance (full, pseudo)
Extreme-value copulas	Gumbel–Hougaard	Genest and Favre (2007)	MATLAB built-in	Maximum likelihood (full, pseudo) Minimum distance (full, pseudo) Method of moment
	Galambos	Genest and Favre (2007)	Coded function	Maximum likelihood (full, pseudo) Minimum distance (full, pseudo)
	Hüsler–Reiss	Genest and Favre (2007)	Coded function	Maximum likelihood (full, pseudo) Minimum distance (full, pseudo)
Other copulas	Farlie–Gumbel–Morgenstern	Shiau (2006)	Coded function	Maximum likelihood (full, pseudo) Minimum distance (full, pseudo)
	Plackett	Shiau (2006)	Coded function	Maximum likelihood (full, pseudo) Minimum distance (full, pseudo)

check the complete goodness-of-fit for the parametric structure describing the dataset. Intuitively, copula models can be ranked based on their overall goodness-of-fit and the error contribution from their marginal quantification. The overall goodness-of-fit measure can be simply defined as the average value of the full and pseudo goodness-of-fit measures. The difference between the corresponding full and pseudo goodness-of-fit measures shows how much of the total error is contributed by the error in the theoretical marginal distributions. In this study, the BIC measure has been chosen for such analysis. BIC provides better distinction between modelling options and puts an emphasis on the parametric parsimony by considering the number of model parameters. BIC can be formulated as follows:

$$BIC = n \log(MSE) + [(no\ of\ fitted\ parameters) \times \log(n)] \quad (11)$$

where n is the number of observations and MSE is the mean square of error of the copula model. By applying this procedure to an ensemble of copula models, the potential candidates can be pruned by eliminating the highly dominated options, i.e., the models that strictly result in poorer goodness-of-fit measures compared to the other competing options.

4.4 Secondary goodness-of-fit test and tail analysis

There is a recent trend in the literature trying to propose formal (also known as blanket or omnibus) hypothetical

tests for goodness-of-fit evaluation of copula models (Ferminan 2005). The goodness-of-fit tests can be further extended and used as a way to test the equality between two copulas (e.g. Rémillard and Scaillet 2009). The main formal approach is to compute the proper p -values through bootstrapping procedures (Genest et al. 2006). These tests are either based on the Rosenblatt transformation (e.g. Berg 2009) to convert the multivariate problem into a univariate hypothetical test or the use of empirical copula (e.g. Genest et al. 2009) to calculate Cramér–von Mises and Kolmogorov–Smirnov statistics. Tests based on Rosenblatt transformation often require double bootstrapping and high-level computational resources, which might not be widely available in practice. Genest et al. (2009) argued that the tests based on empirical copula are more objective and only require parametric bootstrapping if the analytic expression of the copula structure exists. Here, we implement the formal test introduced by Genest et al. (2009) based on the distance between the empirical and the null hypothesis copula. The Cramér–von Mises statistics for bivariate case can be described as the following:

$$\begin{aligned} \hat{T}_n &= n \int_{[0,1]^2} \{ \hat{C}(u, v) - C_\theta(u, v) \}^2 d\hat{C}(u, v) \\ &= \sum_{j=1}^n \{ \hat{C}(u_j, v_j) - C_\theta(u_j, v_j) \}^2 \end{aligned} \tag{12}$$

For the complete bootstrapping algorithm refer to Genest et al. (2009) or Berg (2009). Acceptance or rejection of the copula models are based on the p -values estimated through bootstrapping: Small values suggest discarding the model whereas large values supporting its suitability (e.g. Mesfioui et al. 2009; Durante and Salvadori 2010). However, as argued by Serinaldi (2009b), there might be circumstances where several models of dependence cannot be falsified through goodness-of-fit tests. In these circumstances, exploring the tail dependence of non-rejected copulas can be useful. Tail dependence is a measure for describing the dependence in the upper tail or the lower tail joint space. These values for copula $C(u, v)$ can be defined as follows (Joe 1997; Poulin et al. 2007):

$$\begin{aligned} \lambda_U &= \lim_{t \rightarrow 1^-} \frac{1 - 2t + C(t, t)}{1 - t} \\ \lambda_L &= \lim_{t \rightarrow 0^+} \frac{C(t, t)}{t} \end{aligned} \tag{13}$$

where λ_U and λ_L represent the upper and the lower tail dependence, respectively. Several authors have come with empirical tail dependence estimators (e.g. Schmidt 2005; Schmidt and Stadtmuller 2006; Serinaldi 2008). One of the most frequently used empirical measures of the upper tail

has been suggested by Capéraà et al. (1997) and can be expressed as:

$$\hat{\lambda}_U^{CFG} = 2 - 2 \exp \left[\frac{1}{n} \sum_{i=1}^n \log \left\{ \frac{\sqrt{\frac{\log 1}{u_i} \log 1}}{v_i} \right\} \right] \tag{14}$$

The comparison between copula based upper tail (λ_U) and the empirical measure (Eq. 14) can also provide a means to investigate the performance of the applied parametric copula (e.g. Demarta and McNeil 2004; B enard and Lang 2007; Charpentier and Segers 2009).

5 Results and discussion

Figure 2 shows the scatter and scatter-rank plots for the 56 pairs of maximum annual moisture deficit and annual cumulative evapotranspiration produced for YJP site. The measures of dependence for the site are 0.60, 0.42, and 0.60 for the correlation coefficient, Kendall’s tau, and Spearman’s rho, respectively. Both hypothetical tests based on Kendall’s tau and Spearman’s rho measures have confirmed that the assumption of independence can be rejected at 5% level.

5.1 Marginal quantification

Twenty one different univariate distributions were considered for marginal quantification of the annual maximum water deficit and the annual cumulative evapotranspiration. For the former variable, the Box–Cox transformed Normal distribution has been selected as the theoretical description of the data. This distribution was the only distribution that passed both chi-squared and KS tests at 5% level. For the latter variable however, both Beta distribution and Box–Cox transformed Normal distribution could satisfy one of the applied hypothetical tests. Table 3 shows the characteristics of the fitted marginal distributions and their corresponding goodness-of-fit measures. Figure 3 shows the fitted and empirical marginal distributions.

5.2 Multivariate modelling

Considering the non-falsified marginal distributions, two different approaches can be taken for building up the multivariate description of the data. In the first approach, Box–Cox transformed Normal distribution can be considered as the univariate description of both marginal variables. As a result, the classical multivariate distributions can be also implemented for quantifying the bivariate

Fig. 2 The scatter plot for maximum annual water deficit and annual cumulative evapotranspiration; **a** original scale, **b** rank scale

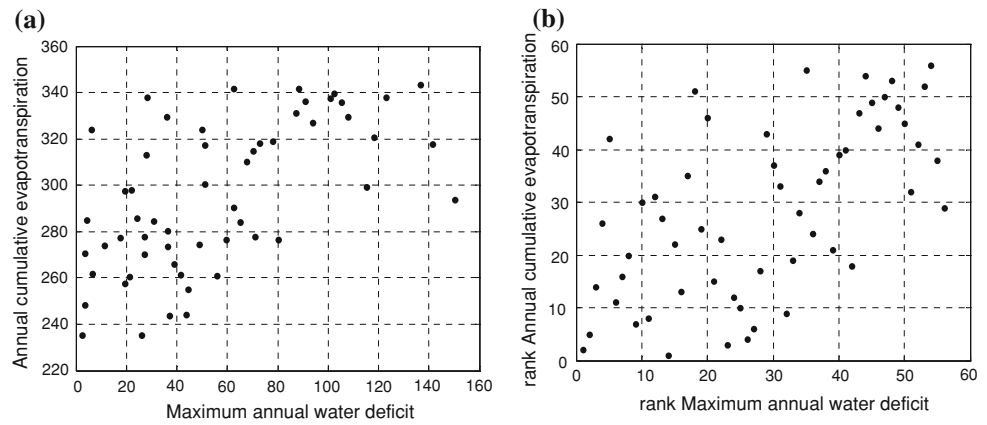
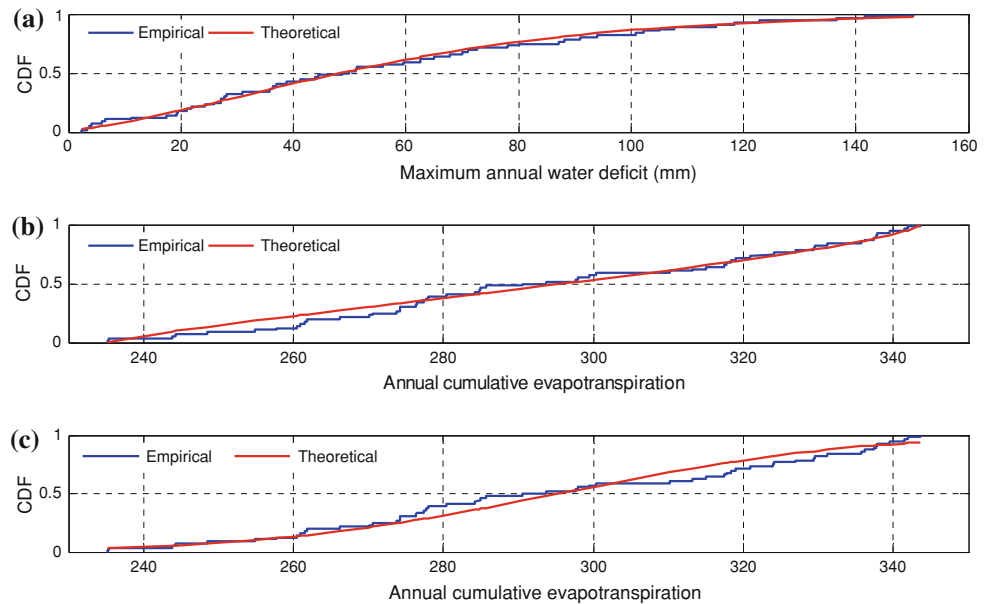


Table 3 The non-falsified marginal distributions for annual maximum water deficit and annual cumulative evapotranspiration along with their specifications

	Box–Cox parameter	Distribution parameters	RMSE	MAE	MSE	CORR	AIC	BIC
Annual maximum water deficit								
Box–Cox, Normal	1.695	48.80–19.31	0.022	0.018	0.000	0.998	–431.18	–410.92
Annual cumulative evapotranspiration								
Beta	NA	0.20–1.09	0.040	0.034	0.002	0.993	–363.08	–350.91
Box–Cox, Normal	–10.561	0.09–1.3e–6	0.056	0.048	0.003	0.987	–324.18	–303.92

Fig. 3 Fitted versus empirical CDFs of marginal variables; **a** Annual maximum water deficit fitted with Box–Cox, Normal distribution; **b** annual cumulative evapotranspiration fitted with Beta distribution; **c** annual cumulative evapotranspiration fitted with Box–Cox, Normal distribution



dependence within the data. However, in the second approach only copula-based models of dependence are used, since the marginal variables are described using different theoretical distributions. Based on the possible options for marginal distributions, multivariate structures, and parameter estimation methods (Table 2), an ensemble of 106 parametric structures were developed for describing the interdependence between the annual maximum water deficit and the annual cumulative evapotranspiration. The

characteristics of the fitted models should be carefully investigated before further implementation. Large convergence limits can resemble the lack of parametric robustness and should be taken as a reason for falsifying the fitted model. Also, association with poor goodness-of-fit measures can be interpreted as a sign of model insufficiency. Based on this logic, 55 models were falsified because of the lack of appropriate convergence or very poor goodness-of-fit measures. Tables 4 and 5 show the 51 non-falsified

multivariate models and their associated goodness-of-fit measures. Table 4 refers to the models in which maximum annual water deficit and the annual cumulative evapotranspiration are described by Box–Cox transformed Normal and Beta distributions respectively; whereas in Table 5, both margins are fitted using the Box–Cox transformed Normal distribution.

Considering the BIC goodness-of-fit measures for both full and pseudo representation of the data, Fig. 4 illustrates different distinctive clusters of the 51 models. In general, three clusters can be observed. According to these clusters, two general arguments can be made. First, considering Table 5, it can be observed that classical multivariate options are strictly dominated by copula-based models. Therefore, they can be further eliminated from the potential solutions. It is also worthwhile mentioning that the majority of the feasible options (in particular in the best cluster) are either estimated through the method of moments or the minimum distance criterion. Only less than 13% of the feasible options in the best cluster are derived using maximum likelihood methods, and they are only related to metaelliptical copulas (*t*-copula and Gaussian options reported in rows 2, 4, and 6 of Table 4). This observation might be a hint for further use of the minimum distance estimators in copula parameter identification. This issue will be more investigated in the next section.

5.3 Model selection

39 copula-based models located in the best modelling cluster (shown in Fig. 4) were considered for initial ranking using the procedure described in Sect. 4.3. Figure 5 shows the locations of the solutions in the surface defined by efficiency criteria. The Pareto candidates are highlighted in Fig. 5 as well as Tables 4 and 5 by the star (*) sign. These eleven models have been used for secondary goodness-of-fit test and tail analysis based on the procedure given in Sect. 4.4. Table 6 provides the name, parameter value, the method of parameter estimation of the non-dominated models along with their upper tail limit, T_n statistics, and *p*-values resulted from bootstrapping with 10,000 random series with the same length as the original data. According to the *p*-values, Gaussian and Gumbel–Hougaard options stand superior to other options. However, only Gumbel–Hougaard options can provide close upper tail limit compared to the empirical tail limit. Based on Eq. 14, the empirical upper tail limit is 0.50; however, Gaussian copula theoretically implies tail independence. Among Gumbel–Hougaard non-dominated options, the structure calibrated by the method of moment is preferred based on the estimated *p*-values and closer upper tail limit compared to the original sample-based measure. This copula model is selected for quantifying the conditional

Table 4 The feasible multivariate models developed for describing the maximum annual water deficit (fitted with Box-Cox, Normal distribution) and the annual cumulative evapotranspiration (fitted with Beta distribution)

Structure	Estimation method	$\hat{\theta}$	RMSE		AIC		BIC		
			Pseudo	Full	Pseudo	Full	Pseudo	Full	
1	Gaussian	Method of moments	0.59	0.026	0.031	-403.6	-370.0	-397.6	-333.7
2		Maximum pseudo likelihood	0.60	0.025	0.031	-409.1	-374.5	-403.1	-344.3*
3		Minimum pseudo distance	0.60	0.025	0.031	-409.1	-374.5	-403.1	-344.3*
4		IFM	0.61	0.025	0.031	-410.9	-374.9	-404.8	-344.6*
5		Minimum full distance	0.61	0.025	0.031	-410.9	-374.9	-404.8	-344.6*
6	<i>t</i> -Student	IFM	[0.61, 5]	0.025	0.031	-407.3	-369.7	-395.2	-333.4
7	Gumbel–Hougaard	Method of moments	1.68	0.024	0.030	-416.4	-376.0	-410.3	-339.7*
8		Minimum pseudo distance	2.06	0.018	0.031	-448.4	-375.2	-442.3	-344.9*
9	Frank	Method of moments	4.44	0.024	0.030	-414.9	-373.8	-408.8	-337.5
10		Minimum pseudo distance	5.78	0.021	0.032	-430.7	-370.9	-424.7	-340.7*
11	Clayton	Method of moments	0.98	0.040	0.041	-356.9	-338.8	-350.8	-302.5
12		Minimum pseudo distance	2.23	0.032	0.041	-383.8	-342.4	-377.8	-312.2
13		Minimum full distance	1.48	0.034	0.039	-376.1	-348.2	-370.1	-318.0
14	Ali–Mikhail–Haq	Maximum pseudo likelihood	-1.00	0.107	0.101	-247.5	-241.1	-241.5	-210.9
15		Minimum pseudo distance	1.00	0.040	0.041	-357.9	-342.5	-351.8	-312.2
16	Plackett	Minimum pseudo distance	11.85	0.021	0.032	-431.2	-369.1	-425.1	-338.8
17	Farlie–Gumbel–	Maximum pseudo likelihood	-1.00	0.113	0.108	-241.0	-234.2	-234.9	-204.0
18	Morgenstern	Minimum pseudo distance	1.00	0.047	0.045	-338.8	-333.1	-332.8	-302.9*
19	BB1	Minimum pseudo distance	[0.0, 2.06]	0.018	0.031	-445.4	-372.0	-433.3	-335.7

The asterisks indicate the non-dominated models

Table 5 The feasible multivariate models developed for describing the maximum annual water deficit and the annual cumulative evapotranspiration when both marginal variables are fitted with Box–Cox, Normal distribution

	Structure	Estimation method	$\hat{\theta}$	RMSE		AIC		BIC	
				Pseudo	Full	Pseudo	Full	Pseudo	Full
1	Gaussian	Method of moments	0.59	0.03	0.04	-403.6	-340.8	-397.6	-298.5
2		Maximum pseudo likelihood	0.60	0.03	0.04	-409.1	-348.5	-403.1	-318.3
3		Minimum pseudo distance	0.60	0.03	0.04	-409.1	-348.5	-403.1	-318.3
4		IFM	0.61	0.03	0.04	-410.0	-348.7	-403.9	-318.5
5		Minimum pseudo distance	0.61	0.03	0.04	-410.0	-348.7	-403.9	-318.5
6	Normal	Maximum pseudo likelihood	[11.36, 686.3, 26.9, 267.6, 7525]	0.03	0.04	-256.2	-148.9	-248.7	-132.4
7		Minimum pseudo distance		0.03	0.04	-256.2	-148.9	-248.7	-132.4
8		IFM		0.03	0.04	-256.2	-148.9	-248.7	-132.4
9		Minimum full distance		0.03	0.04	-256.2	-148.9	-248.7	-132.4
10	t	Minimum pseudo distance	[0.6, 2]	0.37	0.37	-106.3	-94.0	-94.2	-57.7
11		Minimum full distance	[0.6, 2]	0.37	0.37	-106.3	-94.0	-94.2	-57.7
12	Gumbel–Hougaard	Method of moments	1.68	0.02	0.04	-416.4	-329.5	-410.3	-287.2
13		Minimum pseudo distance	2.06	0.02	0.04	-448.4	-338.7	-442.3	-308.4
14		Minimum full distance	1.95	0.02	0.04	-445.7	-339.0	-439.7	-308.8
15	Frank	Method of moments	4.44	0.02	0.04	-414.9	-347.3	-408.8	-304.9
16		Minimum pseudo distance	5.78	0.02	0.04	-430.7	-356.0	-424.7	-325.7
17		Minimum full distance	5.39	0.02	0.04	-429.4	-356.4	-423.4	-326.1
18	Clayton	Method of moments	0.98	0.04	0.04	-356.9	-337.1	-350.8	-294.7
19		Minimum pseudo distance	2.23	0.03	0.04	-383.8	-358.9	-377.8	-328.6*
20		Minimum full distance	1.88	0.03	0.04	-382.5	-360.2	-376.4	-329.9*
21	Ali–Mikhail–Haq	Maximum pseudo likelihood	-1.00	0.11	0.10	-247.5	-243.3	-241.5	-213.1
22		Minimum pseudo distance	1.00	0.04	0.04	-357.9	-344.2	-351.8	-313.9
23		IFM	-1.00	0.11	0.10	-247.5	-243.3	-241.5	-213.1
24		Minimum full distance	1.00	0.04	0.04	-357.9	-344.2	-351.8	-313.9
25	Plackett	Minimum pseudo distance	11.85	0.02	0.04	-431.2	-349.6	-425.1	-319.3
26		Minimum full distance	9.52	0.02	0.04	-427.6	-350.6	-421.6	-320.3
27	Farlie–Gumbel–Morgenstern	Maximum pseudo likelihood	-1.00	0.11	0.11	-241.0	-236.5	-234.9	-206.2
28		Minimum pseudo distance	1.00	0.05	0.05	-338.8	-318.0	-332.8	-287.7
29		IFM	-1.00	0.11	0.11	-241.0	-236.5	-234.9	-206.2
30		Minimum full distance	1.00	0.05	0.05	-338.8	-318.0	-332.8	-287.7
31	BB1	Minimum pseudo distance	0.00	0.02	0.04	-445.4	-335.5	-433.3	-299.2
32		Minimum full distance	2.48	0.04	0.03	-369.0	-357.9	-356.9	-321.6*

The asterisks indicate the non-dominated models

probability of maximum annual water deficit with respect to the annual cumulative evapotranspiration.

Figure 6 compares the empirical cumulative bivariate distribution (Eq. 10) with the selected copula-based distribution in both pseudo and full representations. Comparison between pseudo and full subfigures can explicitly show the error contribution initiating by the misspecification in the marginal theoretical distributions.

Before extracting further information from the fitted joint model, it is worthwhile to compare the performance of different estimation methods. In order to perform this task, a bootstrapping study has been conducted to compare the robustness of the three Gumbel–Hougaard models

calibrated using method of moment, minimum distance and the maximum likelihood, respectively. The specifications of the first two models have been reported in Table 6. The third model has the same margins as the first two models; however, it has been calibrated using the maximum likelihood. This model has been falsified in our preliminary inspection, due to the wide distance between the lower/upper bounds of parametric convergence. For each model, 1000 bivariate samples have been extracted using the fitted structure and the model parameters have been re-estimated using the generated samples and each of the estimation methods. Figure 7 shows the results of this experimentation. The horizontal line shows the Gumbel–Hougaard

Fig. 4 Different clusters of the feasible models in terms of their pseudo and full BIC goodness-of-fit measures. The *dots* and *triangles* represent the feasible solutions reported in Tables 4 and 5, respectively

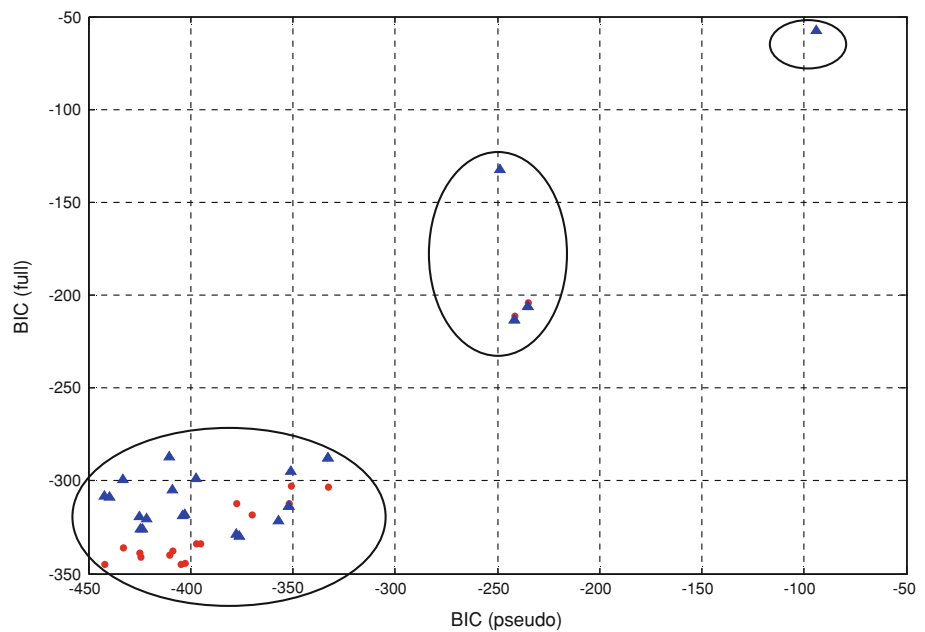
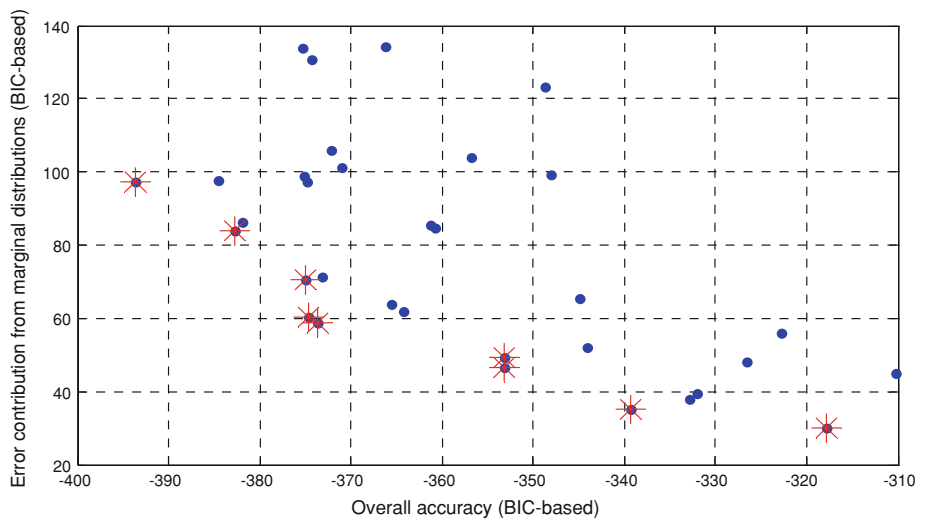


Fig. 5 The location of solutions within the best modelling cluster (Fig. 4) in the surface defined by measures introduced for overall accuracy and the error contribution from marginal variables. The *stars* refer to the non-dominated solutions



parameter estimated using the original data set and Kendall’s tau. As can be easily investigated, the method of maximum likelihood considerably suffers from lack of robustness comparing with both method of moment and minimum distance.

Based on the Gumbel–Hougaard copula fitted using the method of moment, the conditional probability can be estimated based on the following equation:

$$C_{\theta}^u(v) = P\{V \leq v | U = u\} = \frac{\partial}{\partial u} C(u, v) \tag{15}$$

Figure 8 provides graphical presentation of the conditional probabilities. Based on this figure, and given a value for annual cumulative evapotranspiration, the possible quantities and the corresponding CDF for maximum annual water deficit can be identified. Figure 9

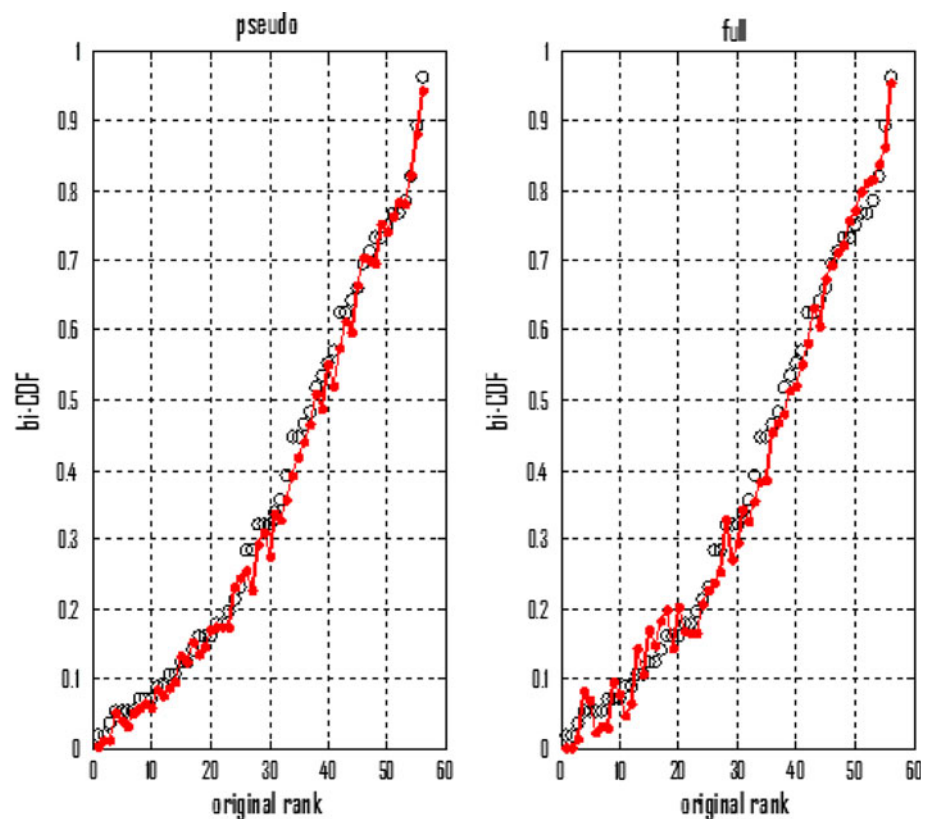
compares the CDF of maximum annual water deficit when annual cumulative evapotranspiration estimations are at 10, 50 and 90 percentiles. The joint model not only capture the uncertainties in maximum annual water deficit with respect to the estimations of annual cumulative evapotranspiration; but also, can be further used in nearby reconstructed watersheds with similar physical characteristics and soil cover, in which the soil-moisture measurements are not available to directly assess the performance of the reclamation soil cover.

6 Summary and conclusion

Majority of hydrologic variables are interdependent. Considering the joint characteristic in the case of dependence

Table 6 The upper tail dependence and the goodness-of-fit test results for the Pareto candidates shown in Fig. 5; p -values are extracted by 10,000 bootstrap samples

Structure	Estimation method	$\hat{\theta}$	λ_U	T_n	p -value (%)
Gaussian	Maximum pseudo likelihood	0.60	0	0.015	57.7
Gaussian	Minimum pseudo distance	0.60	0	0.015	57.7
Gaussian	IFM	0.61	0	0.015	59.6
Gaussian	Minimum full distance	0.61	0	0.015	59.6
Gumbel–Hougaard	Method of moments	1.68	0.49	0.014	69.5
Gumbel–Hougaard	Minimum pseudo distance	2.06	0.60	0.017	26.7
Frank	Minimum pseudo distance	5.78	0	0.019	19.3
Farlie–Gumbel–Morgenstern	Minimum pseudo distance	1.00	0	0.043	1.2
Clayton	Minimum pseudo distance	2.23	0	0.038	2.9
Clayton	Minimum full distance	1.88	0	0.034	4.7
BB1	Minimum full distance	2.48	−0.29	0.037	3.3

Fig. 6 The empirical (*empty dots*) versus fitted joint cumulative distribution (*lines*) for maximum annual water deficit and annual cumulative evapotranspiration in both pseudo and full representation based on the selected Gumbel–Hougaard model

not only provides more realistic representation of the involved variables, but also offers the means of quantifying one variable with respect to the others. In this study, the interdependence between annual cumulative evapotranspiration and the maximum annual water deficit in reconstructed watersheds was studied. Assigning the maximum annual water deficit requires extensive measurements of soil moisture in different depths, which can be only available in very few prototype watersheds. On the other hand, estimating the annual cumulative evapotranspiration

can be much easier in practice. Therefore, if the joint model of interdependence is known, the maximum annual water deficit can be approximated based on the estimated annual cumulative evapotranspiration. The current study has applied the copula framework for such a problem. A diverse test space has been considered for setting up the marginal distributions, describing the multivariate joint model and estimating the models' parameters. A simple framework for initial ranking of the copula models was proposed. A goodness-of-fit test and tail analysis was

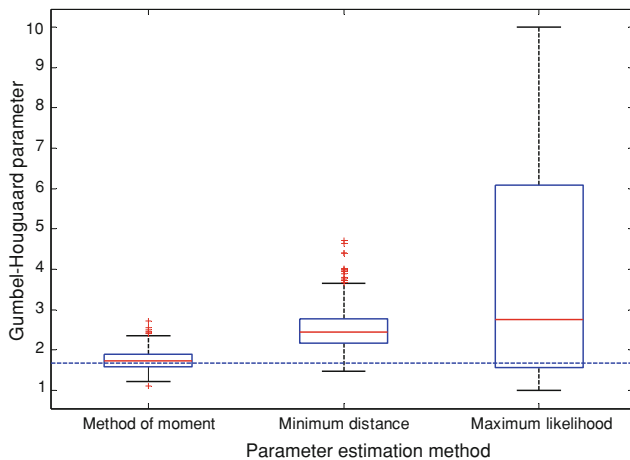


Fig. 7 The variation in the sampled copula parameters in the three Gumbel–Hougaard models calibrated using the method of moment, minimum distance and maximum likelihood, respectively

applied to select the most credible copula model among an ensemble of potential candidates. It has been concluded that Gumbel–Hougaard copula provides the most credible model of dependence for the considered dataset. Moreover, it seems that the method of moments can provide the most reliable way for adjusting the copula parameters. However, this requires a direct relationship between the measures of dependence and the model parameters, which is not available for many copula structures. Comparing the general rules of copula parameter estimation, i.e. maximum likelihood versus minimum distance, the results of this study supported the superiority of the latter technique; however, by no means this can be extended to other problems and/or datasets. Further research in this direction is suggested.

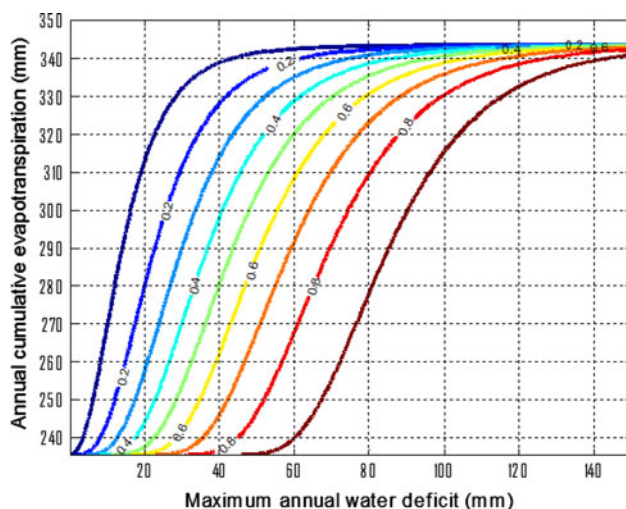


Fig. 8 Conditional CDF for maximum annual water deficit given the annual cumulative evapotranspiration derived for YJP prototype site

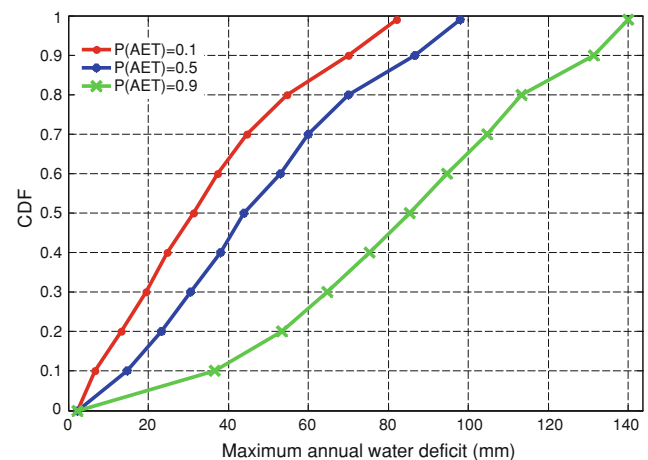


Fig. 9 Different CDFs for maximum annual water deficit given the annual cumulative evapotranspiration in its 10, 50 and 90% of exceedance probability derived for YJP prototype site

The application of copula modelling can be potentially capable in various fields within the scope of hydrology and environmental modelling. However it should be noted that copula modelling is valid, if and only if, the considered variables and their margins are continuous random variables as Embrechts (2009) warned that discrete margins can cause severe problems. The works of Genest and Nelehova (2007) as well as Meisar and Komorníková (2009) on discrete copulas and copulas based on partial knowledge can be promising. Also copula modelling, at least in the field of hydrology and environmental modelling, has been mainly considered for constructing static stochastic distributions rather than describing multivariate stochastic processes. Describing the dependence within dynamic datasets is still fairly untouched in the field and, indeed, is a very interesting topic for future investigations. Regarding the application context reported in this paper, one can question the uncertainties propagated into the copula modelling because of the involved uncertainties in the applied field measurements and/or hydrologic model. This would be the next step in our investigation.

Acknowledgment The authors thank Dr. Sean Carey for providing some of the data of the case study, and they also acknowledge the financial support of NSERC through its Discovery and CRD Grants Program, and CEMA. The authors would like to greatly appreciate the inputs from Editor-in-Chief and two anonymous referees which have been extremely useful to increase the quality of the paper.

References

Ashkar F (2008) Discussion of “Everything you always wanted to know about copula modeling but were afraid to ask” by C. Genest and A.-C. Favre. *J Hydrol Eng*. doi:10.1061/(ASCE)1084-0699(2007)12:4(347)

- Bárdossy A (2006) Copula-based geostatistical models for ground-water quality parameters. *Water Resour Res* 42:W11416. doi:10.1029/2005WR004754
- Bárdossy A, Pegram G (2009) Copula based multisite model for daily precipitation simulation. *Hydrol Earth Syst Sci Discuss* 6:4485–4534
- Berg D (2009) Copula goodness-of-fit testing: an overview and power comparison. *Eur J Finance* 15(7 & 8):675–701
- Bénaud B, Lang M (2007) Use of a Gaussian copula for multivariate extreme value analysis: some case studies in hydrology. *Adv Water Resour* 30:897–912
- Capéraà P, Fougères A-L, Genest C (1997) A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika* 84(3):567–577
- Charpentier A, Segers J (2009) Tails of multivariate Archimedean copulas. *J Multivar Anal* 100:1521–1537
- Chebana F, Ouarda TBMJ (2009) Index flood-based multivariate regional frequency analysis. *Water Resour Res* 45. doi:10.1029/2008WR007490
- Chowdhary H (2008) Discussion of “Gumbel-Hougaard copula for trivariate rainfall frequency analysis” by L. Zhang and Vijay P. Singh. *J Hydrol Eng.* doi:10.1061/(ASCE)1084-0699(2007)12:4(409)
- Chowdhary H, Singh VP (2008) Gains from copula-based multivariate distribution functions for rainfall processes. In: Proceedings of environmental and water resources congress 2008 Ahupua’a. doi:10.1061/40976(316)571, pp 1–10
- De Michele C, Salvadori G (2003) A generalized Pareto intensity-duration model of storm rainfall exploiting 2-copulas. *J Geophys Res* 108(D2):4067
- Demarta S, McNeil AJ (2004) The t copula and related copulas. <http://www.ma.hw.ac.uk/~mcneil/ftp/tCopula.pdf>. Retrieved March 2010
- Dupuis DJ (2007) Using copulas in hydrology: benefits, cautions, and issues. *J Hydrol Eng* 12(4):381–393
- Durante F, Salvadori G (2010) On the construction of multivariate extreme value models via copulas. *Environmetrics* 21:143–161
- Elshorbagy A, Barbour L (2007) A probabilistic approach for design and hydrologic performance assessment of reconstructed watersheds. *J Geotech Geoenviron Eng ASCE* 133(9):1110–1118
- Elshorbagy A, Barbour SL, Qualizza C (2006) Chapter 14: Multi-criterion decision making approach to assess the performance of reconstructed watersheds. In: Castelletti A, Sessa RS (eds) Topics on system analysis and integrated water resource management. Elsevier, Amsterdam, pp 257–269
- Embrechts P (2009) Copulas: a personal view. *J Risk Insur* 76(3): 639–650
- Fang HB, Fang KT, Kotz S (2002) The meta-elliptical distributions with given marginals. *J Multivar Anal* 82:1–16
- Favre A-C, El Adlouni S, Perreault L, Thiémonge N, Bobée B (2004) Multivariate hydrological frequency analysis using copulas. *Water Resour Res* 40:W01101
- Fermanian J-D, Scaillet O (2003) Nonparametric estimation of copulas for time series. *J Risk* 5(4):25–54
- Fermanian J-D (2005) Goodness-of-fit for copulas. *J Multivar Anal* 95(1):119–152
- Foscolo E, Ayyad C, Porcu E, Mateu J (2008) Some new estimators for copula-based models. In statistics for spatio-temporal modelling. In: Cocchi D, Mateu J, Montes F, Otranto E, Porcu E, Usai A (eds) Proceedings of the fourth international workshop on spatio-temporal modelling (METMA4). Democratica Sarda, pp 217–228. ISBN: 88-6025-098-6
- Gargouri-Ellouze E, Bargaoui Z (2009) Investigation with Kendall plots of infiltration index—maximum rainfall intensity relationship for regionalization. *Phys Chem Earth* 34:642–653
- Gartsman B, van Nooyen R, Kolechkin A (2009) Implementation issues for total risk calculation for groups of sites. *Phys Chem Earth* 34:619–625
- Genest C, Favre A-C (2007) Everything you always wanted to know about copula modelling but were afraid to ask. *J Hydrol Eng* 12(4):347–368
- Genest C, MacKay J (1986) The joy of copulas: bivariate distributions with uniform marginals. *Am Stat* 40(4):280–283
- Genest C, Neslehova J (2007) A primer on copulas for count data. *ASTIN Bull* 37:475–515
- Genest C, Rémillard B (2006) Discussion of “Copulas: tales and facts by T. Mikosch”. *Extremes* 9(1):27–36
- Genest C, Beaudoin D, Rémillard B (2006) Omnibus goodness-of-fit tests for copulas: a review and a power study. *Les Cahiers du GERAD, G-2006-74, Copyright © 2006 GERAD*
- Genest C, Favre A-C, Béliveau J, Jacques C (2007) Metaelliptical copulas and their use in frequency analysis of multivariate hydrological data. *Water Resour Res* 43. doi:10.1029/2006WR005275
- Genest C, Rémillard B, Beaudoin D (2009) Goodness-of-fit tests for copulas: a review and a power study. *Insur Math Econ* 44:199–213
- Haberlandt U, Ebner von Eschenbach A-D, Buchwald I (2008) A space-time hybrid hourly rainfall model for derived flood frequency analysis. *Hydrol Earth Syst Sci Discuss* 5:2459–2490
- Huard D, Evin G, Favre A-C (2006) Bayesian copula selection. *Comput Stat Data Anal* 51(2):809–822
- Joe H (1997) Multivariate models and dependence concepts. Chapman and Hall, London
- Kallenberg WCM (2008) Estimating dependence, model selection of copula. <http://eprints.eemcs.utwente.nl/12121/01/memo1866.pdf>. Retrieved March 2010
- Kao S-C, Govindaraju RS (2010) A copula-based joint deficit index for droughts. *J Hydrol* 380:121–134
- Karamouz M, Ahmadi A, Moridi A (2009) Probabilistic reservoir operation using Bayesian stochastic model and support vector machine. *Adv Water Resour* 32:1588–1600
- Karmakar S, Simonovic SP (2008) Bivariate flood frequency analysis using copula with parametric and non-parametric marginals. In: Proceedings of the 4th international symposium on flood defence: managing flood risk, reliability and vulnerability, Toronto, Ontario, Canada, May 6–8
- Karmakar S, Simonovic SP (2009) Bivariate flood frequency analysis. Part 2: a copula-based approach with mixed marginal distributions. *J Flood Risk Manage* 2:32–44
- Keshta N, Elshorbagy A, Carey S (2009) A generic system dynamics model for simulating and evaluating the hydrological performance of reconstructed watersheds. *Hydrol Earth Syst Sci* 13:1–17
- Kim G, Silvapulle MJ, Silvapulle P (2007) Comparison of semiparametric and parametric methods for estimating copulas. *Comput Stat Data Anal* 51(6):2836–2850
- Klein B, Pahlow M, Hundecha Y, Schumann A (2008) Probability analysis of hydrological loads for the design of flood control systems using copulas. *J Hydrol Eng.* doi:10.1061/(ASCE)HE.1943-5584.0000204
- Koehler KJ, Symonowski JT (1995) Constructing multivariate distributions with specified marginal distributions. *J Multivar Anal* 55(2):261–282
- Kurowicka D, Cooke RM (2007) Sampling algorithms for generating joint uniform distributions using the vine-copula method. *Comput Stat Data Anal* 51:2889–2906
- Leonard M, Metcalfe A, Lambert M (2008) Frequency analysis of rainfall and streamflow extremes accounting for seasonal and climatic partitions. *J Hydrol* 348:135–147

- Leskiw LA (2004) Land capability classification for forest ecosystems in the oil sands. Paragon Soil and Environmental Consulting Inc., Alberta
- Maity R, Kumar DN (2008) Probabilistic prediction of hydroclimatic variables with nonparametric quantification of uncertainty. *J Geophys Res* 113. doi:[10.1029/2008JD009856](https://doi.org/10.1029/2008JD009856)
- MATLAB (2010) MATLAB statistics toolbox™ 7 user's guide. The MathWorks, Inc., Natick
- Mesfioui M, Quesy J-F, Toupin M-H (2009) On a new goodness-of-fit process for families of copula. *Can J Stat* 37(1):80–101
- Mesiar R, Komorníková M (2009) Copulas: An approach how to Model the dependence structure of random vectors. *Acta Polytech Hung* 6(1):5–19
- Michiels F, De Schepper A (2008) A copula test space model: how to avoid wrong copula choice. *Kybernetika* 44(6):864–878
- Mikosch T (2006) Copulas: Tales and facts. *Extremes* 9(1):3–20
- Muhaisen OS, Osorio F, García PA (2009) Two-copula based simulation for detention basin design. *Civ Eng Environ Syst* 26(4):355–366
- Nelsen RB (2006) An introduction to copulas. Springer, New York
- Nijssen D, Schumann A, Pahlow M, Klein B (2009) Planning of technical flood retention measures in large river basins under consideration of imprecise probabilities of multivariate hydrological loads. *Nat Hazards Earth Syst Sci* 9:1349–1363
- Osorio F, Muhaisen O, García PA (2009) Copula-based simulation for the estimation of optimal volume for a detention basin. *J Hydrol Eng*. doi:[10.1061/\(ASCE\)HE.1943-5584.0000124](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000124)
- Poulin A, Huard D, Favre A-C, Pugin S (2007) Importance of tail dependence in bivariate frequency analysis. *J Hydrol Eng* 12(4):394–403
- Rémillard B, Scaillet O (2009) Testing for equality between two copulas. *J Multivar Anal* 100:377–386
- Salvadori G, De Michele C (2007) On the use of copulas in hydrology: theory and practice. *J Hydrol Eng* 12(4):369–380
- Salvadori G, De Michele C (2010) Multivariate multiparameter extreme value models and return periods: a copula approach. *Water Resour Res* 46:W10501. doi:[10.1029/2009WR009040](https://doi.org/10.1029/2009WR009040)
- Salvadori G, De Michele C, Kottegoda NT, Rosso R (2007) Extremes in nature: an approach using copulas. Springer, New York
- Samaniego L, Bárdossy A, Kumar R (2010) Streamflow prediction in ungauged catchments using copula-based dissimilarity measures. *Water Resour Res* 46:W02506. doi:[10.1029/2008WR007695](https://doi.org/10.1029/2008WR007695)
- Schmidt R (2005) Tail dependence. In: Cizek P, Hardle W, Weron R (eds) *Statistical tools in finance and insurance*. Springer, New York
- Schmidt R, Stadtmuller U (2006) Non-parametric estimation of tail dependence. *Scand J Stat* 33:307–335
- Serinaldi F (2008) Analysis of inter-gauge dependence by Kendall's τ_k , upper tail dependence coefficient, and 2-copulas with application to rainfall fields. *Stoch Env Res Risk Assess* 22:671–688
- Serinaldi F (2009a) A multisite daily rainfall generator driven by bivariate copula-based mixed distributions. *J Geophys Res*. doi:[10.1029/2008JD011258](https://doi.org/10.1029/2008JD011258)
- Serinaldi F (2009b) Copula-based mixed models for bivariate rainfall data: an empirical study in regression perspective. *Stoch Env Res Risk Assess* 23:677–693
- Serinaldi F, Bonaccorso B, Cancelliere A, Grimaldi S (2009) Probabilistic characterization of drought properties through copulas. *Phys Chem Earth* 34:596–605
- Shiau J-T (2006) Fitting drought duration and severity with two-dimensional copulas. *Water Resour Manage* 20:795–815
- Shiau J-T, Wang H-Y, Tsai C-T (2006) Bivariate frequency analysis of flood floods using copulas. *J Am Water Resour Assoc*, 42, pp 1549–1564
- Shiau J-T, Feng S, Nadarajah S (2007) Assessment of hydrological droughts for the Yellow River, China, using copulas. *Hydrol Process* 21:2157–2163
- Silva RS, Lopes HF (2008) Copula, marginal distributions and model selection: a Bayesian note. *Stat Comput* 18(3):313–320
- Sklar A (1996) Random variables, distribution functions, and copulas: a personal look backward and forward. *Lect Notes Monogr Ser* 28:1–14
- Sunyer Pinya MA, Madsen H, Rosbjerg D (2009) Assessment of the risk of inland flooding in a tidal sluice regulated catchment using multi-variate statistical techniques. *Phys Chem Earth* 34:662–669
- Tsukahara H (2005) Semiparametric estimation in copula models. *Can J Stat* 33(3):357–375
- van den Berg M, Vandenberghe S, De Baets B, Verhoest N (2010) Copula-based spatial downscaling of rainfall fields. In: *Proceedings of the international workshop in advances in statistical hydrology*, May 23–25, Taormina, Italy
- Wang C, Ni-Bin Chang N-B, Yeh G-T (2009) Copula-based flood frequency (COFF) analysis at the confluences of river systems. *Hydrol Process* 23:1471–1486
- Yan J (2007) Enjoy the joy of copulas: with a package copula. *J Stat Softw* 21(4):1–21
- Zhang L, Singh VP (2006) Bivariate flood frequency analysis using the copula method. *J Hydrol Eng* 11(2):150–164