6-1-2017

Multivariate Copula Analysis Toolbox (MvCAT): Describing Dependence and Underlying Uncertainty Using a Bayesian Framework

Mojtaba Sadegh
Boise State University

Elisa Ragno
University of California

Amir AghaKouchak
University of California

AGU PUBLICATIONS

Water Resources Research

TECHNICAL REPORTS: METHODS

10.1002/2016WR020242

Key Points:
- Copulas are a powerful tool to analyze the dependence structure of multiple variables
- Traditional copula fitting approaches are prone to local optima and do not provide any estimate of underlying uncertainties
- MvCAT is a toolbox to rigorously and comprehensively analyze the dependence structure of hydrologic variables and beyond

Correspondence to:
A. AghaKouchak,
amir.a@uci.edu

Citation:
Sadegh, M., E. Ragno, and A. AghaKouchak (2017), Multivariate Copula Analysis Toolbox (MvCAT): Describing dependence and underlying uncertainty using a Bayesian framework, *Water Resour. Res.*, *53*, 5166–5183, doi:10.1002/2016WR020242.

Received 6 DEC 2016
Accepted 11 MAY 2017
Accepted article online 24 MAY 2017
Published online 16 JUN 2017

Multivariate Copula Analysis Toolbox (MvCAT): Describing dependence and underlying uncertainty using a Bayesian framework

Mojtaba Sadegh[1,2], Elisa Ragno[1], and Amir AghaKouchak[1,3]

[1]Department of Civil and Environmental Engineering, University of California, Irvine, California, USA, [2]Department of Civil Engineering, Boise State University, Boise, Idaho, USA, [3]Center for Hydrometeorology and Remote Sensing, University of California, Irvine, California, USA

Abstract We present a newly developed Multivariate Copula Analysis Toolbox (MvCAT) which includes a wide range of copula families with different levels of complexity. MvCAT employs a Bayesian framework with a residual-based Gaussian likelihood function for inferring copula parameters and estimating the underlying uncertainties. The contribution of this paper is threefold: (a) providing a Bayesian framework to approximate the predictive uncertainties of fitted copulas, (b) introducing a hybrid-evolution Markov Chain Monte Carlo (MCMC) approach designed for numerical estimation of the posterior distribution of copula parameters, and (c) enabling the community to explore a wide range of copulas and evaluate them relative to the fitting uncertainties. We show that the commonly used local optimization methods for copula parameter estimation often get trapped in local minima. The proposed method, however, addresses this limitation and improves describing the dependence structure. MvCAT also enables evaluation of uncertainties relative to the length of record, which is fundamental to a wide range of applications such as multivariate frequency analysis.

1. Introduction

Hydrological and climatological variables are interdependent, and we often use multivariate methods to understand their interactions and associations. Examples include storm duration and intensity [*De Michele and Salvadori*, 2003; *Salvadori and De Michele*, 2004a], flood peak and volume [*De Michele et al*., 2005], concurrent drought and heat waves [*AghaKouchak et al*., 2014], drought duration and severity [*Lee et al*., 2013; *Mishra and Singh*, 2009], and precipitation and soil moisture [*AghaKouchak*, 2015]. Univariate distributions may not be sufficient to describe hydrological variables (or events) that bear intrinsic multivariate characteristics. For instance, the commonly used univariate frequency/risk assessment methods might not be sufficient to describe the failure probability or recurrence intervals of interdependent extreme events [*Salvadori et al*., 2016; *Grimaldi and Serinaldi*, 2006].

The concept of copula, as one approach from the cohort of several multivariate analysis methods, is widely used to model the dependence structure of two (or more) random variables. While the rise of this concept dates back to the 1950s, copula gained popularity in hydrology and climatology after the early works of *De Michele and Salvadori* [2003], *Favre et al*. [2004], *Salvadori and De Michele* [2004a,b], and *De Michele et al*. [2005]. Multivariate methods and copulas have been used in many fields of study including finance [*Rachev*, 2003], insurance [*Charpentier and Segers*, 2007], integrated risk management [*Embrechts et al*., 2001], drought monitoring [*Hao and AghaKouchak*, 2013], flood risk analysis [*Jongman et al*., 2014], frequency analysis [*Parent et al*., 2014], rainfall simulation and analysis [*Li et al*., 2013; *Vernieuwe et al*., 2015], dependence analysis of hyetographs and hydrographs [*Serinaldi and Kilsby*, 2013], and extreme value analysis [*Renard and Lang*, 2007; *Ribatet and Sedki*, 2013]. Recent studies have striven to open new frontiers to address real-world problems using copulas. For example, *Wahl et al*. [2015] have examined, using copulas, the cooccurrence probabilities of heavy precipitation and storm surge in the contiguous US and documented that the compounding risks of such events pose a higher risk than each of them in isolation. Other studies have compared regional flood peak-volume dependence structure using the concept of copula [*Szolgay et al*., 2016], have used copula for multivariate analysis of return period [*Serinaldi*, 2015], and have attempted to

transfer the dependence structure information from one watershed to another using copulas [*Grimaldi et al*., 2016]. For a detailed reference list of copula literature in the field of hydrology, refer to stahy.org.

There are more than a dozen copula families that have been used in the hydrological and climatological literature. However, most studies are limited to a handful of copulas (i.e., Gaussian, t-, Frank, Gumbel, and Clayton). Copula families differ in their dependence structure and complexity (i.e., the number of parameters ranging from one to three). Copula parameters reflect the strength of mutual dependence between two (or more) variables [*Balakrishna and Lai*, 2009] and are generally estimated through (1) a theoretical relationship (if exists) between the parameter and empirical dependence measures such as Kendall's τ and Spearman's ρ and/or (2) inference from empirical multivariate probability distribution of data. The most common practice for parameter estimation is using local optimization algorithms, such as Newton-Raphson method [*Salvadori and De Michele*, 2004b; *Bárdossy*, 2006; *Gräler et al*., 2013; *Ribatet and Sedki*, 2013]. Local optimization approaches benefit from efficient (mostly gradient-based) search algorithms, but suffer from susceptibility to getting trapped in local optima [*Duan et al*., 1992]. In a similar line of study, *Brahimi et al*. [2015] introduced a new *L*-moment-based approach to estimate the copula parameters, which outperforms the traditional approaches in terms of bias and computational costs, and is less sensitive to data outliers. However, these methods are not usually capable of characterizing the underlying uncertainties. In recent years, global optimization approaches and Bayesian analysis have also been explored for inferring copula parameters [*Pitt et al*., 2006; *Min and Czado*, 2010; *Smith et al*., 2012; *Parent et al*., 2014; *Kwon and Lall*, 2016].

Despite the recent progress, there are still challenges and research gaps that we address in this paper, namely, (a) lack of a comprehensive and generalized framework for estimating the underlying uncertainties of a wide range of copulas, (b) lack of a robust and efficient algorithm to estimate the posterior distribution of copula parameters, and (c) extending the commonly used copulas in hydrology and climatology through a generalized software package that allows not only parameter inference but also different ways of ranking the best choice of copula for the underlying data. In this paper, we introduce the Multivariate Copula Analysis Toolbox (MvCAT) that employs Markov Chain Monte Carlo (MCMC) simulation within a Bayesian framework to estimate copula parameters and the underlying uncertainties. MCMC simulation estimates the posterior distribution of parameter values, which are then translated into uncertainty ranges for the copula probability isolines. The MCMC simulation searches for the region of interest with multiple chains running in parallel. Chains share information on the fly, characterize the posterior region (even in the presence of multimodality), and estimate the global optimum.

MvCAT includes 26 copula families, discussed in section 2.1, including a number of copula families that have not been fully explored in the hydrological literature. MvCAT includes both the commonly used local optimization method and a state-of-the-art MCMC framework. If MCMC is selected, MvCAT automatically plots posterior parameter distributions of the chosen copula(s), as well as the fitted (and empirical) probability isolines. Moreover, a summary report is automatically generated that ranks the performance of selected copulas based on Maximum Likelihood, Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). Summary report also details on the best and 95% uncertainty ranges of parameters of each copula, and their best performance in terms of root mean square error (RMSE) and Nash-Sutcliff efficiency (NSE). In this toolbox, we strive to analyze the existing dependence structure between the hydrologic variables and beyond. Generating synthetic multivariate data sets, however, is beyond the scope of the current study and interested readers are referred to *Nelsen* [2007] for that purpose.

The remainder of this paper is organized as follows: section 2 briefly describes the concept of copulas and multivariate dependence analysis (section 2.1) and introduces the copula families used in MvCAT. The paper then discusses the model inference problem in general terms (section 2.2) and continues with the definition of Bayesian analysis (section 2.3) for updating a prior belief about a hypothesis when new information came to light. Section 2.4 introduces a hybrid-evolution MCMC algorithm that numerically and iteratively estimates the posterior solution of the Bayes' equation using differential evolution, snooker update, and adaptive metropolis within Gibbs sampling. Next, we discuss the measures of goodness of fit (AIC, BIC, RMSE, and NSE) to compare the performance of different copula models (section 2.5), and then introduce the Multivariate Copula Analysis Toolbox (MvCAT) to perform dependence analysis using a rigorous and comprehensive approach (section 2.6). We proceed with the results section (section 3) that discusses a drought analysis study based on bivariate analysis of precipitation and soil moisture, and a flood frequency analysis based on flood peak and volume joint distribution. Section 4 concludes this paper with a summary of the

multivariate dependence analysis and provides some conclusions derived from the uncertainty analysis of presented case studies.

2. Materials and Methods

2.1. Copulas

Copulas are mathematical functions that "join" or "couple" two or more **time-independent** variables [*Nelsen*, 2003], regardless of their univariate distributions [*Genest and Favre*, 2007]. They are a systematic way of studying the underlying dependence structure and providing a basis for constructing families of bivariate (multivariate) distributions [*Fisher*, 1997]. A copula can be informally defined as a mapping tool from $\mathbf{I}^2$ (F, G) to $\mathbf{I}$ (H) [*Nelsen*, 2003], when $[F(x), G(y), H(x,y)]$ is a point in $\mathbf{I}^3$ ($\mathbf{I} \in [0,1]$). X and Y are then continuous random variables with distribution functions $F(x)=P(X \leq x)$ and $G(y)=P(Y \leq y)$, and $H(x,y)=P(X \leq x, Y \leq y)$ is a function that describes their joint distribution.

Sklar [1959] first used the word copula in a statistical and mathematical context and introduced a theorem that is named after him [*Nelsen*, 2003],

"Let H be a joint cumulative distribution function with marginal univariate distributions F and G. Then there exists a copula C so that $H(x,y)=C[F(x), G(y)]$. This copula C is unique, if F and G are continuous."

Similarly, if there exists a joint distribution H with continuous marginals F and G, $u=F(x)$ and $v=G(y)$, one can establish the associated copula as $C(u,v)=H[F^{-1}(u), G^{-1}(v)]$. One of the most widely used copula families in the literature is the Gaussian copula defined as,

$$C_\theta(u,v)=\Phi_\Sigma\left[\phi^{-1}(u), \phi^{-1}(v)\right], \tag{1}$$

where Φ_Σ denotes the joint cumulative distribution function of a bivariate normal vector with zero means and covariance matrix Σ, and ϕ^{-1} notifies inverse of a standard Gaussian distribution. In this paper we employ 26 parametric models of copula families which are explained in detail in Table 1. We have selected copula families with simple closed form mathematical formulation, which are amenable for model inference, and represent different forms of dependence structures. A much larger pool of copula families is available at http://www.maths.manchester.ac.uk/~saralees/chap20.pdf.

In this paper, we have concentrated our attention on bivariate copulas, since bivariate analysis is commonly used in the hydrological and climatological literature. While some copula families are limited in the number of dimensions they can handle, others can be extended to higher dimensions. Interested readers can refer to *Nelsen* [2007]; *Joe* [2014] for more information about copulas and their properties.

2.2. Model Inference

In a modeling analysis, an unknown process, $\mathcal{K}$, links observations, $\tilde{\mathbf{Y}}$, to "true" parameters, $\boldsymbol{\theta}^*$, through

$$\tilde{\mathbf{Y}}=\mathcal{K}(\boldsymbol{\theta}^*)+\varepsilon, \tag{2}$$

in which ε denotes a vector of measurement errors [*Sadegh and Vrugt*, 2013]. A model hypothesis, $\mathcal{M}$, simulates the system response, $\tilde{\mathbf{Y}}$, given a $d\times 1$ vector of parameter values, $\boldsymbol{\theta}=\{\theta_1, \theta_2, \cdots, \theta_d\}$, and forcing, $\tilde{\mathbf{I}}$ [*Vrugt and Sadegh*, 2013],

$$\mathbf{Y}=\mathcal{M}(\boldsymbol{\theta}, \tilde{\mathbf{I}}). \tag{3}$$

A vector $\mathbf{e}=\tilde{\mathbf{Y}}-\mathbf{Y}$ then characterizes error residuals. $\mathbf{e}=\{e_1, e_2, \cdots, e_n\}$ (n: number of observations) includes the effects of model structural errors ($\mathcal{M}$ systematically deviating from $\mathcal{K}$), as well as calibration data, input and parameter errors [*Sadegh and Vrugt*, 2014]. In the case of a copula, model structural error arises from the deviation between the copula formulation and the true dependence structure of the variables, calibration data error stems from the uncertainties in the empirical estimates of the joint probability distribution, and input error is due to the measurement uncertainties of the underlying variables that are translated into the uniform marginals.

For a model, $\mathcal{M}$, we wish to estimate the parameters, $\boldsymbol{\theta}$, by tuning them such that model simulations, $\mathbf{Y}$, fit the observations, $\tilde{\mathbf{Y}}$, given the forcing, $\tilde{\mathbf{I}}$ [*Sadegh et al.*, 2015]. In the case of copula modeling, $\tilde{\mathbf{Y}}$ represents the joint probability of observed variables and $\mathbf{Y}$ signifies the copula predicted probability values. Empirical

Table 1. Copula Families and Their Closed-Form Mathematical Description

Name	Mathematical Description[a]	Parameter Range	Reference
Gaussian	$\int_{-\infty}^{\phi^{-1}(u)}\int_{-\infty}^{\phi^{-1}(v)}\frac{1}{2\pi\sqrt{1-\theta^2}}\exp\left(\frac{2\theta xy-x^2-y^2}{2(1-\theta^2)}\right)dxdy$[b]	$\theta\in[-1,1]$	*Li et al.* [2013]
t	$\int_{-\infty}^{t_{\theta_2}^{-1}(u)}\int_{-\infty}^{t_{\theta_2}^{-1}(v)}\frac{\Gamma((\theta_2+2)/2)}{\Gamma(\theta_2/2)\pi\theta_2\sqrt{1-\theta_1^2}}\left(1+\frac{x^2-2\theta_1xy+y^2}{\theta_2}\right)^{(\theta_2+2)/2}dxdy$[c]	$\theta_1\in[-1,1]$ and $\theta_2\in(0,\infty)$	*Li et al.* [2013]
Clayton	$\max\left(u^{-\theta}+v^{-\theta}-1,0\right)^{-1/\theta}$	$\theta\in[-1,\infty)\backslash 0$	*Clayton* [1978]
Frank	$-\frac{1}{\theta}\ln\left[1+\frac{(\exp(-\theta u)-1)(\exp(-\theta v)-1)}{\exp(-\theta)-1}\right]$	$\theta\in\mathbb{R}\backslash 0$	*Li et al.* [2013]
Gumbel	$\exp\left\{-\left[(-\ln(u))^{\theta}+(-\ln(v))^{\theta}\right]^{1/\theta}\right\}$	$\theta\in[1,\infty)$	*Li et al.* [2013]
Independence	uv		*Nelsen* [2003]
Ali-Mikhail-Haq (AMH)	$\frac{uv}{1-\theta(1-u)(1-v)}$	$\theta\in[-1,1)$	*Ali et al.* [1978]
Joe	$1-\left[(1-u)^{\theta}+(1-v)^{\theta}-(1-u)^{\theta}(1-v)^{\theta}\right]^{1/\theta}$	$\theta\in[1,\infty)$	*Li et al.* [2013]
Farlie-Gumbel-Morgenstern (FGM)	$uv[1+\theta(1-u)(1-v)]$	$\theta\in[-1,1]$	*Nelsen* [2007]
Gumbel-Barnett	$u+v-1+(1-u)(1-v)\exp\left[-\theta\ln(1-u)\ln(1-v)\right]$	$\theta\in[0,1]$	*Gumbel* [1960] and *Barnett* [1980]
Plackett	$\frac{1+(\theta-1)(u+v)-\sqrt{[1+(\theta-1)(u+v)]^2-4\theta(\theta-1)uv}}{2(\theta-1)}$	$\theta\in(0,\infty)$	*Plackett* [1965]
Cuadras-Auge	$[\min(u,v)]^{\theta}(uv)^{(1-\theta)}$	$\theta\in[0,1]$	*Cuadras and Augé* [1981]
Raftery	$\begin{cases}u-\frac{1-\theta}{1+\theta}u^{\frac{1}{1-\theta}}\left(v^{\frac{-\theta}{1-\theta}}-v^{\frac{1}{1-\theta}}\right), \text{if } u\le v\\ v-\frac{1-\theta}{1+\theta}v^{\frac{1}{1-\theta}}\left(u^{\frac{-\theta}{1-\theta}}-u^{\frac{1}{1-\theta}}\right), \text{if } v\le u\end{cases}$	$\theta\in[0,1)$	*Nelsen* [2007]
Shih-Louis	$\begin{cases}(1-\theta)uv+\theta\min(u,v), & \text{if } \theta\in(0,\infty)\\ (1+\theta)uv+\theta(u+v-1)\Psi(u+v-1), & \text{if } \theta\in(-\infty,0]\end{cases}$ $\Psi(a)=1$ if $a\ge 0$ and $\Psi(a)=0$ if $a<0$		*Shih and Louis* [1995]
Linear-Spearman	$\begin{cases}[u+\theta(1-u)]v, & \text{if } v\le u \text{ and } \theta\in[0,1]\\ [v+\theta(1-v)]u, & \text{if } u<v \text{ and } \theta\in[0,1]\\ (1+\theta)uv, & \text{if } u+v<1 \text{ and } \theta\in[-1,0]\\ uv+\theta(1-u)(1-v), & \text{if } u+v\ge 1 \text{ and } \theta\in[-1,0]\end{cases}$	$\theta\in[-1,1]$	*Joe* [2014]
Cubic	$uv[1+\theta(u-1)(v-1)(2u-1)(2v-1)]$	$\theta\in[-1,2]$	*Durrleman et al.* [2000]
Burr	$u+v-1+\left[(1-u)^{-1/\theta}+(1-v)^{-1/\theta}-1\right]^{-\theta}$	$\theta\in(0,\infty)$	*Frees and Valdez* [1998]
Nelsen	$\frac{-1}{\theta}\log\left\{1+\frac{[\exp(-\theta u)-1][\exp(-\theta v)-1]}{\exp(-\theta)-1}\right\}$	$\theta\in(0,\infty)$	*Nelsen* [2007]
Galambos	$uv\exp\left\{(-\ln(u))^{-\theta}+(-\ln(v))^{-\theta}\right\}^{-1/\theta}$	$\theta\in[0,\infty)$	*Huynh et al.* [2014]
Marshall-Olkin	$\min\left[u^{(1-\theta_1)}v, uv^{(1-\theta_2)}\right]$	$\theta_1,\theta_2\in[0,\infty)$	*Huynh et al.* [2014]
Fischer-Hinzmann	$\left\{\theta_1[\min(u,v)]^{\theta_2}+(1-\theta_1)[uv]^{\theta_2}\right\}^{1/\theta_2}$	$\theta_1\in[0,1]$, $\theta_2\in\mathbb{R}$	*Fischer and Hinzmann* [2007]
Roch-Alegre	$\exp\left\{1-\left[\left(\left((1-\ln(u))^{\theta_1}-1\right)^{\theta_2}+\left((1-\ln(v))^{\theta_1}-1\right)^{\theta_2}\right)^{1/\theta_2}+1\right]^{1/\theta_1}\right\}$	$\theta_1\in(0,\infty)$, $\theta_2\in[1,\infty)$	*Roch and Alegre* [2006]
Fischer-Kock	$uv\left[1+\theta_2(1-u^{1/\theta_1})(1-v^{1/\theta_1})\right]^{\theta_1}$	$\theta_1\in[1,\infty)$, $\theta_2\in[-1,1]$	
BB1	$\left\{1+\left[(u^{-\theta_1}-1)^{\theta_2}+(v^{-\theta_1}-1)^{\theta_2}\right]^{1/\theta_2}\right\}^{-1/\theta_1}$	$\theta_1\in(0,\infty)$, $\theta_2\in(1,\infty)$	*Genest and Favre* [2007]

Table 1. (continued)

Name	Mathematical Description[a]	Parameter Range	Reference
BB5	$\exp\left\{-\left[(-\ln(u))^{\theta_1}+(-\ln(v))^{\theta_1}-\left((-\ln(u))^{-\theta_1\theta_2}+(-\ln(v))^{-\theta_1\theta_2}\right)^{-1/\theta_2}\right]^{1/\theta_1}\right\}$	$\theta_1 \in [1,\infty)$, $\theta_2 \in (0,\infty)$	*Genest and Favre* [2007]
Tawn	$\exp\left\{\ln\left(u^{(1-\theta_1)}\right)+\ln\left(v^{(1-\theta_2)}\right)-\left[(-\theta_1\ln(u))^{\theta_3}+(-\theta_2\ln(v))^{\theta_3}\right]^{1/\theta_3}\right\}$	$\theta_1,\theta_2 \in [0,1]$, $\theta_3 \in [1,\infty)$	*Huynh et al.* [2014]

[a]These formulations might not be unique. Please refer to the associated reference.
[b]ϕ represents a standard Gaussian distribution.
[c]t_{θ_2} represents a Student's t distribution with θ_2 degrees of freedom.

joint probability values of $\tilde{\mathbf{Y}}$ are derived from the observations using Gringorten plotting position, and inputs to the copula models, the uniform marginals u and v, are also estimated in a similar manner (with a difference of using univariate analysis).

2.3. Bayesian Analysis

Bayesian analysis has been successfully employed in different fields [*Geweke*, 1989; *Jarrell and Gubernatis*, 1996; *Huelsenbeck et al.*, 2001; *Huelsenbeck and Ronquist*, 2001; *Ellison*, 2004; *Box and Tiao*, 2011], including hydrology [*Wood and Rodríguez-Iturbe*, 1975; *Kuczera*, 1999; *Thiemann et al.*, 2001; *Kavetski et al.*, 2006; *Cheng et al.*, 2014], for model inference and uncertainty quantification purposes. Bayes' theorem updates the prior probability (belief) of a certain hypothesis when new information is acquired. Bayes' law conveniently attributes all modeling uncertainties to the parameters and estimates the posterior distribution of model parameters through

$$p(\boldsymbol{\theta}|\tilde{\mathbf{Y}})=\frac{p(\boldsymbol{\theta})p(\tilde{\mathbf{Y}}|\boldsymbol{\theta})}{p(\tilde{\mathbf{Y}})}, \tag{4}$$

in which $p(\boldsymbol{\theta})$ and $p(\boldsymbol{\theta}|\tilde{\mathbf{Y}})$ signify prior and posterior distribution of parameters, respectively. $p(\tilde{\mathbf{Y}}|\boldsymbol{\theta}) \cong \mathcal{L}(\boldsymbol{\theta}|\tilde{\mathbf{Y}})$ denotes likelihood function, and $p(\tilde{\mathbf{Y}})=\int_{\boldsymbol{\theta}} p(\tilde{\mathbf{Y}}|\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}$ is coined evidence. Evidence, being a constant value in each modeling practice, can be simply removed from the analysis, if the main goal is to estimate the posterior distribution of parameters, and posterior parameter distributions can be estimated through

$$p(\boldsymbol{\theta}|\tilde{\mathbf{Y}}) \propto p(\boldsymbol{\theta})p(\tilde{\mathbf{Y}}|\boldsymbol{\theta}). \tag{5}$$

In the absence of useful information regarding the prior distribution of parameters, one may employ a flat uniform prior [*Thiemann et al.*, 2001]. Assuming error residuals are uncorrelated, homoscedastic, and Gaussian-distributed with mean zero, the likelihood function can be formulated as [*Sorooshian and Dracup*, 1980]

$$\mathcal{L}(\boldsymbol{\theta}|\tilde{\mathbf{Y}})=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi\tilde{\sigma}^2}}\exp\left\{-\frac{1}{2}\tilde{\sigma}^{-2}[\tilde{y}_i-y_i(\boldsymbol{\theta})]^2\right\}, \tag{6}$$

where $\tilde{\sigma}$ is an estimate of standard deviation of measurement error. As *Thyer et al.* [2009] state, "it is perhaps the most widely used calibration criterion in hydrology." For simplicity and numerical stability, this equation is usually logarithmically transformed to

$$\ell(\boldsymbol{\theta}|\tilde{\mathbf{Y}})=-\frac{n}{2}\ln(2\pi)-\frac{n}{2}\ln\tilde{\sigma}^2-\frac{1}{2}\tilde{\sigma}^{-2}\sum_{i=1}^{n}[\tilde{y}_i-y_i(\boldsymbol{\theta})]^2. \tag{7}$$

Given the underlying assumptions about error residuals, $\tilde{\sigma}$ can be estimated as

$$\tilde{\sigma}^2=\frac{\sum_{i=1}^{n}[\tilde{y}_i-y_i(\boldsymbol{\theta})]^2}{n}. \tag{8}$$

We can, then, further simplify equation (7) to

$$\ell(\boldsymbol{\theta}|\tilde{\mathbf{Y}}) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2} - \frac{n}{2}\ln\frac{\sum_{i=1}^{n}[\tilde{y}_i - y_i(\boldsymbol{\theta})]^2}{n}. \quad (9)$$

For comparison purposes, we can conveniently remove the constants and present log-likelihood function as,

$$\ell(\boldsymbol{\theta}|\tilde{\mathbf{Y}}) \simeq -\frac{n}{2}\ln\left\{\frac{\sum_{i=1}^{n}[\tilde{y}_i - y_i(\boldsymbol{\theta})]^2}{n}\right\}. \quad (10)$$

Bayes' equation (5) is usually difficult, if not impossible, to solve analytically and numerical methods, such as MCMC simulation, are adopted to sample from the posterior distribution.

Further clarification is warranted here that we impose no assumption, whatsoever, on the posterior distribution of parameters. The aforementioned assumptions of "homoscedasticity, no correlation, and Gaussian distribution with mean zero" only apply to the distribution of error residuals which is used to construct the likelihood function that summarizes the distance between observations (empirical bivariate probability values) and model simulations (copula predicted bivariate probability values) into a single scalar.

2.4. Markov Chain Monte Carlo Simulation

Markov Chain Monte Carlo (MCMC) algorithms are a class of statistical methods to sample from high-dimensional complex distributions [*Andrieu and Thoms*, 2008]. The equilibrium state of MCMC, if the transition kernel warrants ergodicity, represents the target distribution. We propose a newly developed hybrid-evolution MCMC approach that employs adaptive proposal distributions to delineate the posterior parameter region in a Bayesian context. The hybrid-evolution MCMC benefits from an intelligent starting point selection [*Duan et al*., 1993] and employs Adaptive Metropolis (AM) [*Roberts and Sahu*, 1997; *Haario et al*., 1999, 2001; *Roberts and Rosenthal*, 2009], differential evolution (DE) [*Storn and Price*, 1995, 1997; *Ter Braak*, 2006; *ter Braak and Vrugt*, 2008; *Vrugt et al*., 2009], and snooker update [*Gilks et al*., 1994; *ter Braak and Vrugt*, 2008] algorithms to search the feasible space. This hybrid-evolution MCMC sampling is described in algorithm 1.

In algorithm 1, *LN* is number of samples drawn from the prior distribution, [$p(\theta)$], using Latin Hypercube Sampling (LHS), and *N* is number of Markov chains (*CH*). *D* notifies the dimension of the entire parameter space, whereas *d* represents the dimension of the subspace of the parameters randomly selected for update (Metropolis within Gibbs sampling). *T* is the total number of iterations, and N_{AM} signifies the number of chains selected for AM algorithm. γ_{1-4} are jump factors with γ_1 randomly selected from $[1.2, 2.2]$ [*ter Braak and Vrugt*, 2008], $\gamma_2 = 2.38/\sqrt{d}$, $\gamma_3 = 0.1/\sqrt{d}$ [*Roberts and Rosenthal*, 2009], and $\gamma_4 = 2.38/\sqrt{2d}$ [*Ter Braak*, 2006]. γ_4 takes a value of 1, with a probability of 10%, to allow for full exploration of multimodal distributions [*Ter Braak*, 2006].

This algorithm starts as a random search of the entire prior space using the LHS algorithm. Drawing insight from the Shuffled Complex Evolution (SCE) algorithm of *Duan et al*. [1993], these samples are then randomly assigned to *N* complexes, and the sample with highest likelihood value, $\mathcal{L}(.|.)$, is selected as the starting point for a Markov chain. We call this step "intelligent prior sampling" since it starts chains from points with highest chance of success and improves the overall acceptance rate. This approach, unlike SCE, randomly assigns samples to different complexes to enable starting points from different regions of attraction and avoid degeneration. To diversify the jumping possibilities, the hybrid-evolution MCMC algorithm then employs snooker direction update with 10% probability and parallel direction update with 90% probability. For the parallel direction update, we even further diversify the jumping algorithm by allowing N_{AM} chains to use the AM kernel, and the remainder of the *N* chains to use DE for sampling. We select $N = 2D$ following *Ter Braak* [2006]. It is particularly useful to employ only a few chains (we used one) with AM sampling and let the rest of the chains follow DE update. AM approach diversifies the jump direction, and therefore enhances the search in the early stages of MCMC. However, DE shows a higher potential in converging to the target distribution. Since AM uses an adaptive covariance matrix, Σ_d, based on the last 50% samples of the Markov chains, the proposed hybrid-evolution MCMC benefits from a more powerful search in the early stages of the algorithm. As the algorithm progresses, samples drawn from DE will dominate the scale and orientation of the covariance matrix, Σ_d, and AM also shows similar behavior to DE. The Metropolis ratio is used to

Algorithm 1: Hybrid-evolution MCMC

Intelligent prior sampling: do:

Draw LN ($\geq N$) samples from prior ($p(\theta)$) using Latin Hypercube Sampling (LHS)

Randomly assign the LHS samples to N complexes

Select the best sample in each complex as starting point of a Markov chain (CH)

end do

for $t=2:T$ **do**

  for $i=1:N$ **do**

    Snooker update: With a 10% probability **do**:

    Draw 3 samples, r_{1-3}, from parameter space $\{1:D\}\backslash\{i\}$

    Find the update direction $Z=CH_i-CH_{r_1}$

    Project CH_{r_2} and CH_{r_3} onto Z to get Z_{p_1} and Z_{p_2}

    Create proposal $CH^*=CH_i+\gamma_1(Z_{p_2}-Z_{p_1})$

    Compute Metropolis ratio MR$=\frac{\mathcal{L}(CH^*)||CH^*-CH_{r_1}||^{D-1}}{\mathcal{L}(CH_i)||CH_i-CH_{r_1}||^{D-1}}$

    end do

    Adaptive Metropolis and differential evolution update: With a 90% probability **do**:

    Randomly select d dimensions from D-dimensional parameter space to update (within Gibbs sampling)

    if $i \leq N_{AM}$ **then**

      Create proposal $CH^*(\boldsymbol{d})=CH_i(\boldsymbol{d})+(1-\beta)\,N\left(\boldsymbol{0_d},\gamma_2^2\,\Sigma_{\boldsymbol{d}}\right)+\beta\,N\left(\boldsymbol{0_d},\gamma_3^2\,\boldsymbol{I_d}\right)$

    else

      Create proposal $CH^*(\boldsymbol{d})=CH_i(\boldsymbol{d})+\gamma_4\left(CH_{r_2}(\boldsymbol{d})-CH_{r_1}(\boldsymbol{d})\right)+\boldsymbol{e}$

    end if

    Compute Metropolis ratio MR$=\frac{\mathcal{L}(CH^*)}{\mathcal{L}(CH_i)}$

    end do

    Accept proposal, CH^*, with probability max $(\mathrm{MR},1)$, and update current chain, CH_i

  end for

end for

Check for Gelman-Rubin $\hat{R}$ convergence diagnostic

accept/reject proposal samples, and a Gelman-Rubin $\hat{R}$ diagnostic is then used to monitor the convergence of the chains [*Gelman and Rubin*, 1992].

2.5. Goodness of Fit Measures

In this study, we use several goodness of fit measures to evaluate the performance of different copula models, including likelihood value, AIC, BIC, RMSE, and NSE. Likelihood value is calculated through equation (6). A parameter set that provides the maximum likelihood minimizes the residuals between model simulations and observations. It therefore provides, in this sense, the best fit to the observed data.

Higher model complexity (more degrees of freedom) provides the advantage of greater model flexibility and hence usually results in a better fit to the observed data. However, this might stimulate overconditioning of the model. AIC, in contrast to the ad hoc likelihood value, takes into account both complexity of the

model and minimization of error residuals and provides a more robust measure of quality of model predictions. AIC avoids the problem of overconditioning by adding a penalty term based on the number of parameters. AIC is formulated as [*Akaike*, 1974, 1998; *Aho et al*., 2014]

$$\mathrm{AIC}=2D-2\ell, \tag{11}$$

in which D is the number of parameters of the statistical model and ℓ is the log-likelihood value of the best parameter set (equation (7)). This equation can be simplified to

$$\mathrm{AIC}=2D+n\ln\left\{\frac{\sum_{i=1}^{n}\left[\tilde{y}_i-y_i(\boldsymbol{\theta})\right]^2}{n}\right\}-2CS, \tag{12}$$

given the Gaussian assumption of error residuals, $\left(\tilde{\sigma}^2=\frac{\sum_{i=1}^{n}\left[\tilde{y}_i-y_i(\boldsymbol{\theta})\right]^2}{n}\right)$, and a constant CS. A lower AIC value associates with a better model fit.

Similar to AIC, BIC is presented as [*Schwarz et al*., 1978]

$$\mathrm{BIC}=D\ln n-2\ell, \tag{13}$$

which similarly simplifies to

$$\mathrm{BIC}=D\ln n+n\ln\left\{\frac{\sum_{i=1}^{n}\left[\tilde{y}_i-y_i(\boldsymbol{\theta})\right]^2}{n}\right\}-2CS, \tag{14}$$

if residuals are independent and identically distributed following a Gaussian distribution centered around zero. Similar to AIC, a lower BIC value associates with a better model fit.

NSE and RMSE are also two widely used measures of goodness of fit, which only focus on minimization of residuals,

$$\mathrm{RMSE}=\sqrt{\frac{\sum_{i=1}^{n}\left[\tilde{y}_i-y_i(\boldsymbol{\theta})\right]^2}{n}}, \tag{15}$$

$$\mathrm{NSE}=1-\frac{\sum_{i=1}^{n}\left[\tilde{y}_i-y_i(\boldsymbol{\theta})\right]^2}{\sum_{i=1}^{n}\left[\tilde{y}_i-\bar{\tilde{y}}_i\right]^2}. \tag{16}$$

A perfect model fit is associated with RMSE = 0, RMSE $\in[0,\infty)$, and NSE = 1, NSE $\in(-\infty,1]$. All these metrics evaluate, in different ways, the performance of copulas in terms of how close modeled bivariate probabilities ($\mathbf{Y}$) are to their empirical observed counterparts ($\tilde{\mathbf{Y}}$). Although number of copula parameters (model complexity) impacts some of the evaluation metrics, parameter ranges have zero impact on them.

2.6. Multivariate Copula Analysis Toolbox

We now present the Multivariate Copula Analysis Toolbox (MvCAT) that includes 26 different copula families (Table 1) and optimization methods. Figure 1 shows the Graphical User Interface (GUI) of MvCAT, as well as its inputs, and sample outputs. Users can conveniently browse their input data and select any set of desired copula families for the analysis.

Next step is to select either local optimization or MCMC. The local optimization algorithm, included in this package, uses a gradient-based "interior-point" optimization algorithm [*Byrd et al*., 2000; *Waltz et al*., 2006], which estimates Hessian by a dense quasi-Newton approximation. This algorithm finds the best solution by searching the interior of feasible space. Interior-point method is selected in this study since our parameter space is conditioned (due to feasible parameter ranges), and faster methods such as simplex are not particularly amenable to conditional problems. For each copula estimation, we repeat the search 30 times from different random starting points to minimize the probability of getting trapped in a local minimum. This,

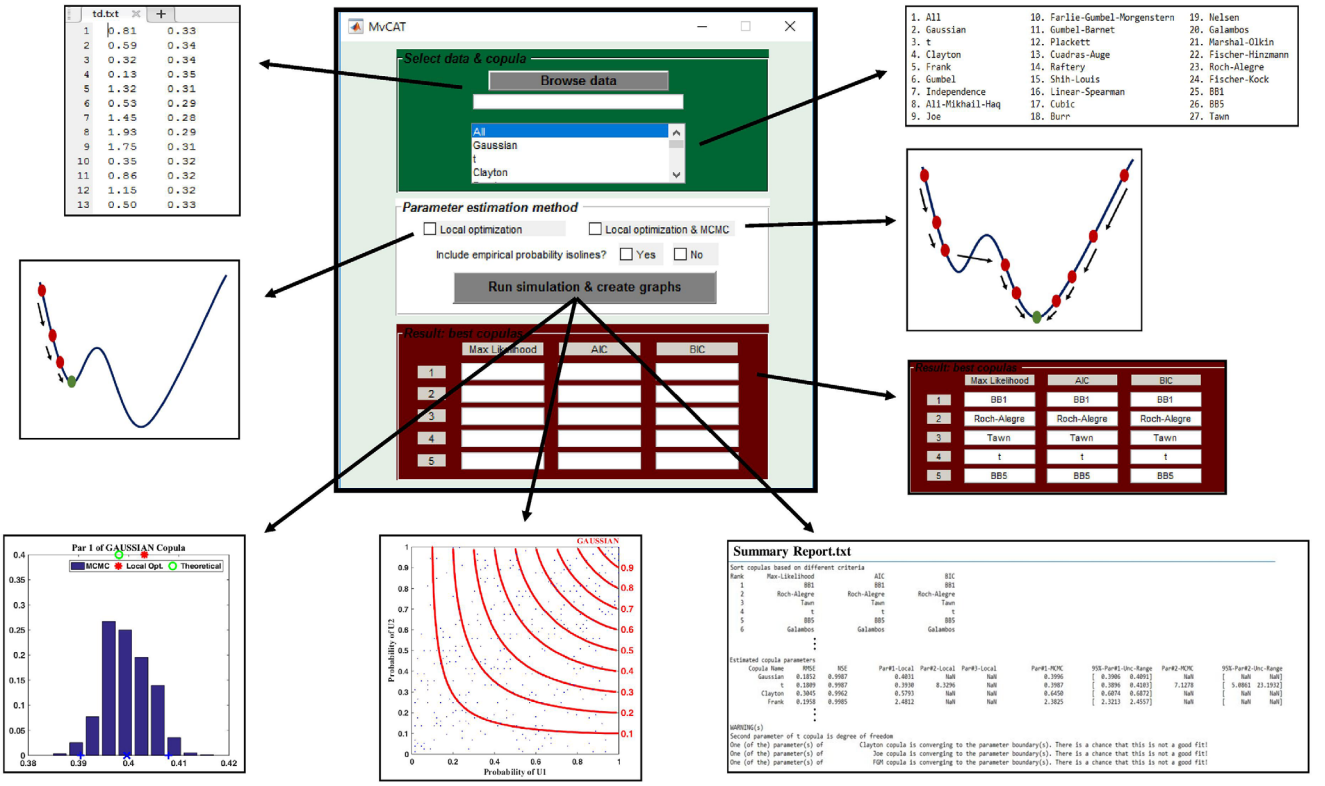

Figure 1. MvCAT: Multivariate Copula Analysis Toolbox.

however, cannot guarantee finding the global optimum. Conversely, the MCMC algorithm benefit from multiple starting points in a single run. Multiple parallel chains communicate on the fly to adapt the scale and orientation of the search. The proposed algorithm is capable of searching multiple regions of attraction, and not only finds an estimate of the global optimum but also approximates the posterior distribution of parameters. Local optimization methods benefit from an efficient and swift search, with the trade-off of susceptibility to finding a local optimum. MCMC algorithm is computationally extensive compared to the local optimization algorithms, but is superior in that it guarantees finding an estimate of the global optimum and characterizes the underlying uncertainty.

MvCAT automatically saves all the results in a folder entitled "Results," which includes a summary report, posterior parameter distribution plots (if MCMC is selected), modeled probability isoline plots, as well as a file, with ".mat" extension, that contains all the variables from MATLAB workspace. A summary report details the ranking of different copula families based on the three criteria of goodness of fit, namely, (1) likelihood, (2) AIC, and (3) BIC. It will also report the parameter values and their 95% uncertainty range, for each selected copula family, as well as associated NSE and RMSE values. Finally, a warning section will draw user's attention to potential concerns. The MvCAT's GUI will also report, on the screen, the five best copula families. If user is interested in the uncertainty analysis of the copulas, a further postprocessing on the MCMC samples is required.

3. Illustrative Case Studies and Results

3.1. Example 1: Drought Analysis

In this section, we first focus our attention on drought analysis of Del Norte county in northern California with an emphasis on the uncertainty quantification of the employed copula framework. Droughts are generally categorized into four major classes: (1) meteorological (lack of precipitation), (2) agricultural (lack of soil moisture), (3) hydrological (lack of surface or ground water resources), and (4) socioeconomical (lack of

commodities) [*Hao and AghaKouchak*, 2013; *Heim*, 2002]. One of the most commonly used indices to characterize drought is the Standardized Precipitation Index (SPI) [*McKee et al*., 1993; *Hayes et al*., 2011; *Mo*, 2011; *Shukla and Wood*, 2008; *Svoboda et al*., 2002]. SPI is simple to compute and interpret and only considers precipitation in the analysis. However, droughts usually involve multiple hydrological variables, such as precipitation, soil moisture, water deficit, evaporative demand, etc. A more robust approach to analyze droughts necessitates application of joint distribution of multiple hydrological variables [*Dracup et al*., 1980]. This stimulated the development and application of several joint drought indices [*Keyantash and Dracup*, 2004; *Kao and Govindaraju*, 2010; *Vicente-Serrano et al*., 2010].

One such approach is to employ a multivariate, multi-index drought analysis framework by combining the drought information from precipitation and soil moisture through their joint distribution [*Hao and AghaKouchak*, 2013]. We follow this framework, and use MvCAT to construct the joint probability distribution of precipitation (mm/d) and soil moisture (mm) anomalies in Del Norte county in northern California. We empirically estimate the marginal distribution of each hydrological variable and construct the joint distribution of precipitation and soil moisture anomalies using the 26 copulas of Table 1. Precipitation and soil moisture data are obtained from the Climate Prediction Center of the National Weather Service available at http://www.cpc.ncep.noaa.gov/. These variables are available from 1948 to 2015, and we analyzed them at monthly and yearly scales. To scrutinize the impact of length of data on the dependence structure and the underlying modeling uncertainty, we used a subset of recent 34 years (1982–2015) of annual data, in addition to the original 68 years (1948–2015) of record. In other words, for this study we used 34 (annual), 68 (annual), and 816 (monthly) observation pairs of data to constrain the parameters of different copula families using local optimization and MCMC.

Figure 2 shows the dependence structure between precipitation and soil moisture anomalies using 68 years of monthly (Figure 2a), 68 years of annual (Figure 2b), and 34 years of annual (Figure 2c) observed data. What is most notable in this figure is the asymmetric and skewed dependence structure of the monthly data (Figure 2a). The probability isolines derived with the Tawn copula are visibly skewed to the top left corner (low probability of precipitation and high probability of soil moisture). This behavior, however, is not replicated in the annual data. It is worth noting that a visual inspection of the 26 copulas fitted to the monthly data shows that only the Tawn and Marshall-Olkin copulas are capable of characterizing the asymmetric dependence structure of this data set. Neither is among the commonly used copulas in the hydrological literature. The Tawn copula provides a very good fit to the data with a NSE = 0.9988 (NSE = 1 is associated with a perfect fit) and is selected as the best copula according to AIC, BIC, maximum likelihood, and other residual-based metrics. Indeed, Tawn copula is a specific version of a class of copula, namely, Khoudraji's device copula [*Khoudraji*, 1996], designed to generate asymmetric copulas [*Frees and Valdez*, 1998],

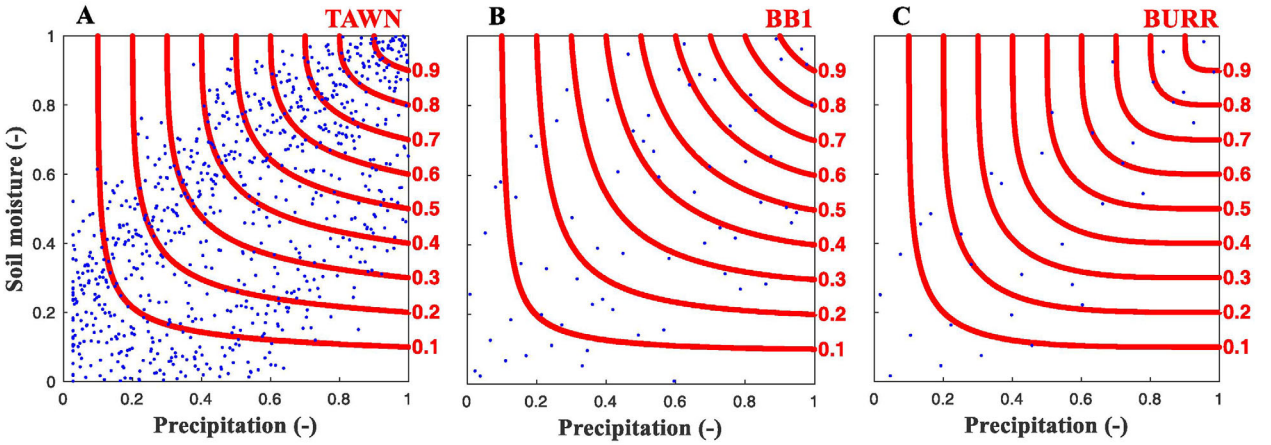

Figure 2. Dependence structure of precipitation and soil moisture anomaly for Del Norte county in northern California, USA. Both precipitation (*x* axis) and soil moisture (*y* axis) are presented in probability space. Red lines present the copula isolines and blue dots show observed data. Tawn copula is used to model the dependence of monthly precipitation and soil moisture anomalies between 1948 and 2015 (plot: a), whereas BB1 copula was selected for the annual scale of this data in the period of 1948–2015 (plot: b) and Burr copula was used for annual data in the period of 1982–2015 (plot: c).

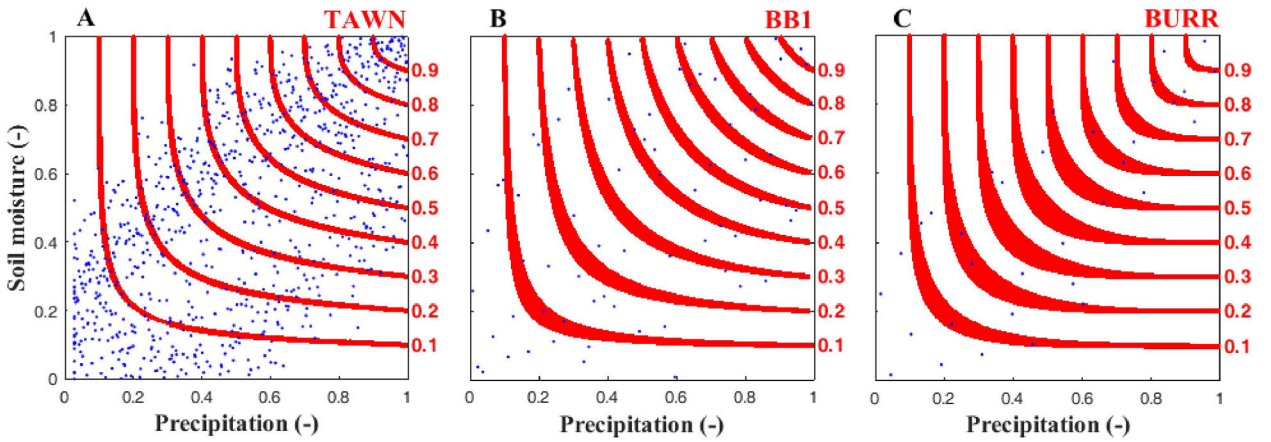

Figure 3. Underlying uncertainty in the dependence structure of precipitation and soil moisture anomaly for Del Norte county in northern California, USA. Both precipitation (*x* axis) and soil moisture (*y* axis) are presented in probability space. Uncertainty ranges of copula isolines are shown with red and observed data are presented with blue dots. The uncertainty ranges are only due to the parameter uncertainty impacts derived in a Bayesian analysis. Tawn copula is used to model the dependence of monthly precipitation and soil moisture anomalies between 1948 and 2015 (plot: a), whereas BB1 copula was selected for the annual scale of this data in the period of 1948–2015 (plot: b) and Burr copula was used for annual data in the period of 1982–2015 (plot: c).

$$C_{\kappa,\lambda}(u,v)=u^{1-\kappa}v^{1-\lambda}C(u^{\kappa},v^{\lambda}),\quad 0\le\kappa,\lambda\le 1, \tag{17}$$

in which C is an exchangeable bivariate copula and form the limiting case of the nonexchangeable bivariate copula $C_{\kappa,\lambda}$. If C is selected to be Gumbel, equation (17) simplifies to the closed-form formula of Tawn copula [*Nikoloulopoulos and Karlis*, 2008].

It is also notable that three different copula families are selected for the three scenarios of this study (68 years monthly, 68 years annual, and 34 years annual data), with different model complexities (Tawn: three parameters, BB1: two parameters, and Burr: one parameter). Indeed, a higher length of data is more desirable for constraining copula parameters. Bear in mind that this does not imply that one should necessarily use a more complex copula family (with higher number of parameters) when longer data is available.

The estimation uncertainties, however, are not yet discussed. We use Bayesian analysis with MCMC simulation to find the posterior parameter distributions, and translate them to probability isoline uncertainties. Figure 3a presents uncertainty ranges of probability isolines associated with the posterior parameters of the Tawn copula for the monthly precipitation and soil moisture anomalies. It is visible that the uncertainty ranges are tightly constrained for this scenario. This is due to the presence of enough information in the relatively long record of precipitation and soil moisture observation to constrain the copula parameters. However, if we use a shorter set of 86 data points (annual), the uncertainty ranges become wider (Figure 3b). Such a large range of uncertainty in the probability isolines is then translated to the drought indices such as Multivariate Standardized Drought Index (MSDI), rendering the analysis (e.g., drought classification) more uncertain. Figure 3c shows the uncertainty ranges of the probability isolines due to parameter uncertainty for the third scenario with 34 years of annual data (1982–2015). This scenario shows a considerably wider uncertainty ranges compared to the others. Such behavior points to the insufficiency of the data set to constrain the parameter(s) of the copula model. Note that Tawn, BB1, and Burr copulas are selected based on their performances for scenarios A, B, and C, respectively. If other copula families are to be employed, the uncertainty ranges will only widen.

It is worth noting that the uncertainty range of probability isolines in scenario C (34 years of annual data) is so wide that it can encapsulate the best prediction of most of the copula families studied herein. This highlights the importance of uncertainty analysis of copula applications, given most studies in hydrology and climatology only include limited lengths of data. MvCAT allows users to efficiently evaluate the uncertainty of the parameter estimates and the best choices of copulas relative to their length of record.

3.2. Example 2: Frequency Analysis

In this example, we use MvCAT to perform multivariate frequency analysis and characterize the uncertainties in flood return periods. We use flood peak [Q (m^3/s)] and volume [V (m^3)] data from the Saguenay River

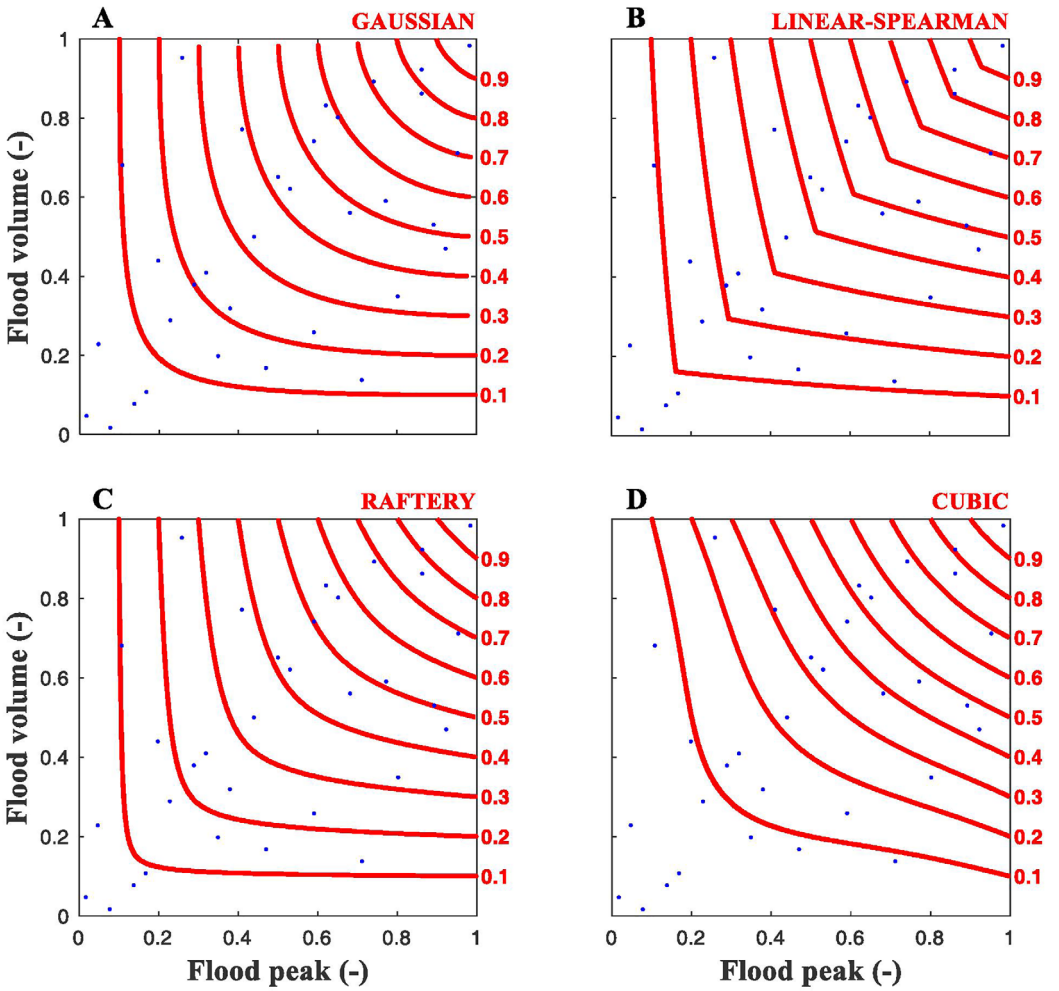

Figure 4. Dependence structure of flood peak and volume of the Saguenay River in Quebec, Canada. Both flood peak (*x* axis) and flood volume (*y* axis) are presented in probability space. Red lines present the copula isolines and blue dots show observed data. (a) Gaussian, (b) Linear-Spearman, (c) Raftery, and (d) Cubic copula families are used to explain the bivariate dependence of this data.

in Quebec, Canada, for the period of 1963–1995 [*Yue et al.*, 1999; *Chebana and Ouarda*, 2011; *Zhang and Singh*, 2007; *AghaKouchak*, 2014]. This data set, including 33 pairs of annual flood peak and volume data, is extracted from daily streamflow data. Marginal distribution of each variable is estimated empirically, and copula families of Table 1 are calibrated against the empirical joint probabilities to construct the joint density functions, and model the dependence structure between flood peaks and volumes. The joint return period (RP) is then estimated as [*AghaKouchak*, 2014]

$$\mathrm{RP} = \frac{1}{P(Q \geq q_p\,,\, V \geq v_p)} = \frac{1}{1 - C(u,v)}, \tag{18}$$

in which the flood peak (Q), or volume (V), or both exceed a prespecified threshold value (q_p, v_p). As mentioned in the previous example, there are several copula families that can be employed, and researchers need to find the copula family that best fits the purpose. Figure 4 shows different copula families can return different dependence structures, keeping in mind that all these four copula families (Gaussian, Linear-Spearman, Raftery, and Cubic) model a common dependence structure in the flood peak and volume of the Saguenay River in Quebec, Canada. They are dissimilar in form, while being quite similar in the performance metric of RMSE. Indeed, Gaussian [RMSE = 0.1000], Raftery [RMSE = 0.0970], and Linear-Spearman copulas [RMSE = 0.1079] can be considered identically good in terms of RMSE, whereas Cubic copula [RMSE = 0.3621] is inferior to the others. This highlights the importance of choice of copula, and more

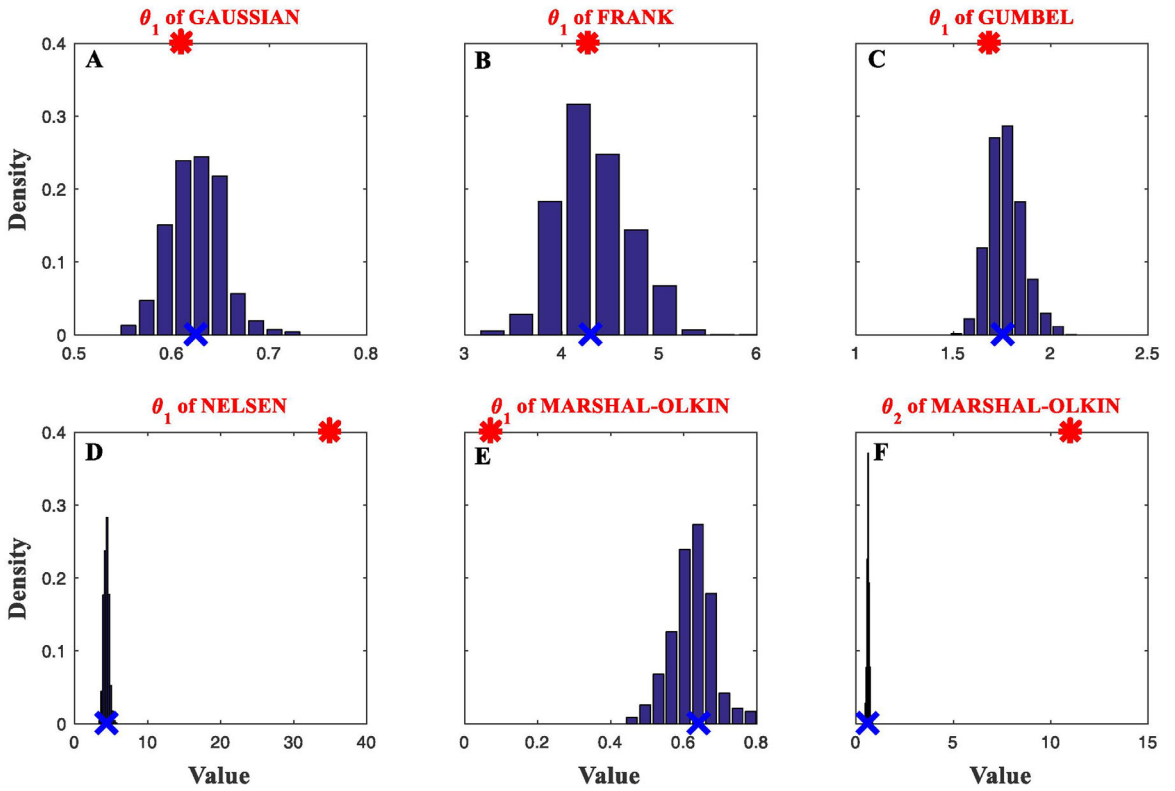

Figure 5. Posterior distribution of (a) Gaussian, (b) Frank, (c) Gumbel, (d) Nelsen, and (e–f) Marshall-Olkin copulas derived by the MCMC simulation within a Bayesian framework. Red asterisks on top of each plot show the copula parameter value derived by the local optimization approach, whereas the blue bins are the MCMC-derived parameters and blue cross shows the maximum likelihood parameter of the MCMC.

importantly quantifying the underlying copula modeling uncertainties. Uncertainty quantification is specifically more imperative in cases that lack the sufficient constraining information in their observation data.

We now turn our attention back to the uncertainties of the copula modeling and plot the posterior parameter distributions of a set of five representative copula families (Figure 5). Posterior parameters of Gaussian, Frank, and Gumbel copulas (Figures 5a–5c) are well constrained. More importantly, the copula parameters derived by the local optimization algorithm (red asterisks on top of each plot) coincide with the mode of the distribution (most likely parameter, shown with blue cross on the bottom of each plot) derived from the MCMC simulation. However, this does not hold for all copula families (Figures 5d–5f). The inferred parameters of Nelsen and Marshall-Olkin copulas from the local optimization algorithm diverge significantly from their counterparts from the MCMC simulation. A closer look at the results of the two approaches reveals that MCMC yields in a reasonable retrieval of the target values. The best parameter value of MCMC simulation returns an RMSE value of 0.1138 and 0.1239 for the Nelsen and Marshall-Olkin copulas, respectively, whereas the local optimization algorithm yields RMSE values of 0.3154 and 0.3459 for these copulas. In more details, six (Gaussian, t, Clayton, Gumbel, Nelsen, and Marshall-Olkin) of the 26 copulas studied herein show an inferior fit (according to RMSE) with local optimization compared to the MCMC results. Keeping in mind that we repeat local optimization with 30 different starting points for each copula family, this behavior is rather disappointing. This raises the question of whether the local optimization algorithms, widely used in the literature, provide a robust and reliable solution to the dependence analysis problem or not. We argue that it is important to employ global optimization algorithms and/or MCMC framework that benefit from multiple start points, such as the one used in MvCAT, for copula inference.

We now proceed with plotting the histogram of the posterior parameter distributions of four other copulas, namely, AMH, FGM, Gumbel-Barnett, and Tawn (Figure 6). An interesting observation in the AMH, FGM, and Gumbel-Barnett copulas (Figures 6a–6c) is that their parameter distributions merge to the parameter bounds, with the maximum likelihood (best) parameter placed on the boundary. This implies that the

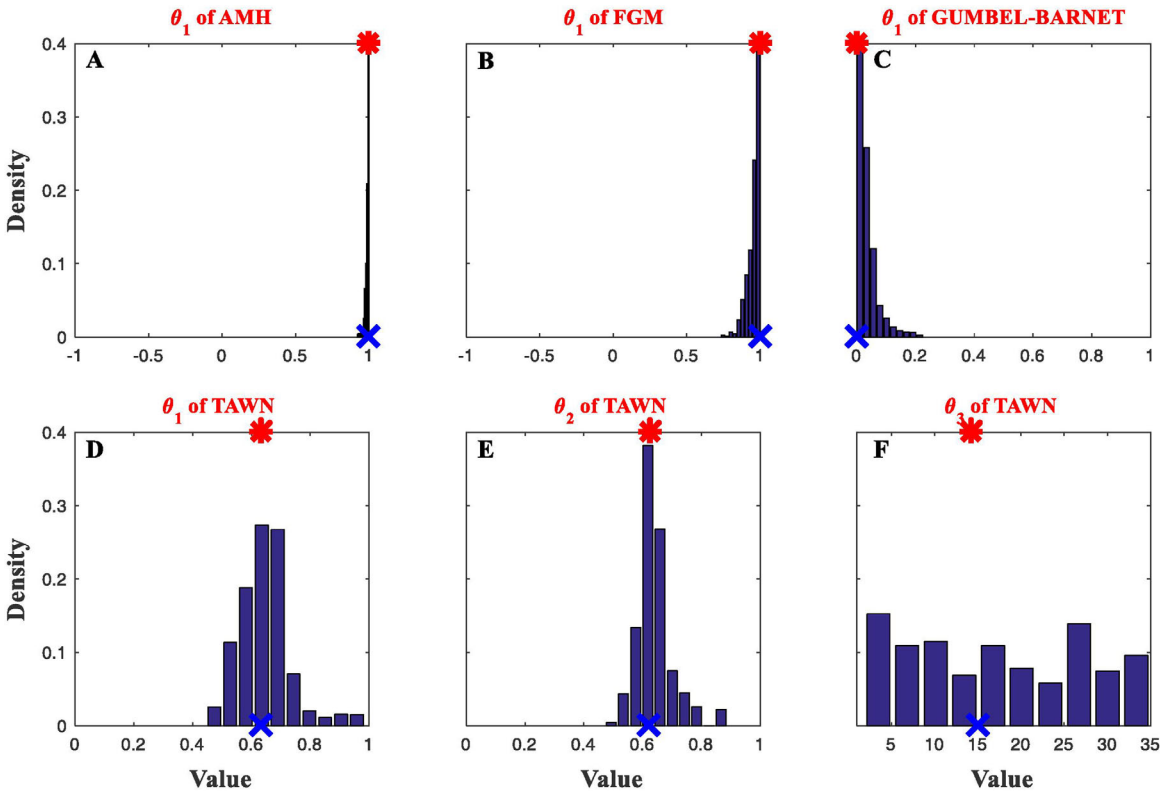

Figure 6. Posterior distribution of (a) AMH, (b) FGM, (c) Gumbel-Barnett, and (d–f) Tawn copula parameters derived by the MCMC simulation within a Bayesian framework. Red asterisks on top of each plot show the copula parameter value derived by the local optimization approach, whereas the blue bins are the MCMC-derived parameters and blue cross shows the maximum likelihood parameter of the MCMC.

optimization algorithm is trying to improve the fit by going outside the bound, which is forcefully not permitted. One implication might be that the choice of such copulas for this data is not proper. A closer look reveals the Gumbel-Barnett parameter merges to zero, which corresponds to independence. Indeed, the independence and Gumbel-Barnett copulas return a similar fit to the data with an RMSE of 0.3845. A rather interesting behavior is observed in the posterior distributions of the Tawn copula parameters (Figures 6d–6f). It is noticeable that the third parameter of the Tawn copula has almost a uniform marginal distribution, which suggests the information in the observed data (34 data points) is not enough to constrain three parameters of the Tawn copula. As for the first and second parameters of the Tawn copula (θ_1 and θ_2), the MCMC-derived posterior distributions are nicely constrained, and their mode coincide with the parameter value inferred by the local optimization (red asterisks).

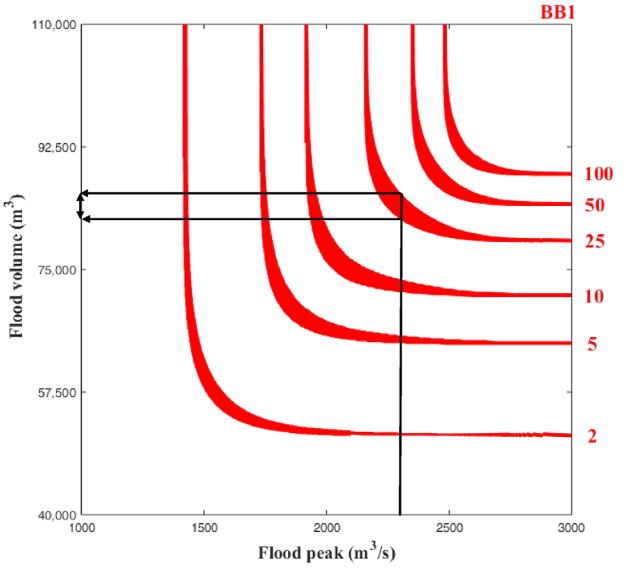

Figure 7. Underlying uncertainty in the dependence structure of flood peak and volume in the Saguenay River in Quebec, Canada. The uncertainty ranges (red) are only due to the BB1 copula parameter uncertainty impacts derived in a Bayesian analysis.

Finally, we will explore the uncertainty ranges of the return periods based on flood peak and volume information (Figure 7). To

estimate these return period levels, we use equation (18) which in turn depends on the copula analysis of flood peak and volume. MvCAT returns BB1 as the best copula family (in terms of all performance criteria including AIC, BIC, and maximum likelihood), and we use that to model the dependence structure between flood peak and volume. The modeled dependence structure and equation (18) is then used to estimate the uncertainties of different return period levels in Figure 7. This figure shows a relatively large uncertainty in each return period level. For example, for a flood event with a peak value of 2300 (m^3/s) and a return period of 25 years, the flood volume can lie in the range of approximately 82,000–85,500 (m^3). An uncertainty of approximately 3500 (m^3) for a return period of 25 years is indeed significant and cannot be neglected. MvCAT offers the opportunity to provide not only the dependence structure but also a measure of parameter uncertainty.

4. Summary and Discussion

Hydrological variables are interconnected, and understanding their dependencies is often fundamental to reliable modeling, prediction, and risk assessment. Copulas have been widely used for modeling and understanding the relationship between two or more variables. Copulas have been used in a wide range of applications including multivariate frequency analysis, multi-index drought assessment, extreme value analysis, and coastal flood risk assessment.

While there are a large number of copulas in the literature, only few have been widely used in the hydrological and climatological literature. Furthermore, little attention has been paid to the underlying uncertainties associated with fitting copulas to data sets with limited length (limited information). In this paper, we present the Multivariate Copula Analysis Toolbox (MvCAT). This toolbox has 26 copulas including a number of copulas that have not been widely used in the field. MvCAT includes not only the commonly used local optimization method, but also an advanced MCMC framework. The latter employs MCMC simulation within a Bayesian framework to estimate copula parameters and their underlying uncertainties. The MCMC component improves the parameter estimation and offers information on the posterior distribution of parameter values. The posterior distribution can then provide uncertainty information on the fitted copula probability isolines.

MvCAT ranks the performance of selected copulas based on maximum likelihood, Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC). The summary report provides detailed information on the best parameter set, and 95% uncertainty ranges of parameters of each copula, as well as its best performance in terms of root mean square error (RMSE) and Nash-Sutcliff efficiency (NSE) criteria. We used MvCAT in two example applications: (a) multi-index drought assessment using joint precipitation-soil moisture anomalies in Del Norte county, California, USA and (b) flood peak and volume frequency analysis in the Saguenay River in Quebec, Canada. Most important conclusions of this study are:

1. Bayesian analysis can be used to estimate the posterior distribution of copula parameters and the underlying uncertainties in copula modeling. Our analysis has shown there is a large uncertainty in the simulation of probability isolines, when limited information is available in constraining data (see Figures 3 and 7). Simulation of probability isolines refer to creating them from the copula model, given the derived posterior parameter sets.
2. In hydrological and climatological applications the length of record is typically short (e.g., a 30 year climatology). This study shows that length of record significantly affects the uncertainty of results (see Figure 3). Uncertainty quantification of copula applications has not received the attention it deserves. MvCAT offers uncertainty bounds for the copula probability isolines. This information is particularly useful in multivariate frequency analysis studies.
3. Local optimization algorithms, widely used in the literature to estimate the copula parameter values, may get trapped in local optima and lead to biased results (see Figure 5). MCMC simulation, on the contrary, starts its search for the posterior region of interest from multiple random points and is designed to efficiently explore the entire feasible space. The proposed hybrid-evolution MCMC algorithm will not only find a good estimate of the global optimum but also provide an approximation of the underlying uncertainties in a Bayesian framework.
4. There are several copula families that can describe different probabilistic properties (see Table 1), but only a limited number are used in the hydrological literature (mainly, Gaussian, t-, Clayton, Frank, and Gumbel). Some variables exhibit asymmetric skewed dependence structures that cannot be described

by the commonly used copulas (see Figure 3). MvCAT includes a wide range of families with different degrees of freedom that can capture both symmetric and asymmetric dependencies.

This paper presents the theoretical background of MvCAT. The graphical user interface (GUI) of this program enables users to conveniently browse the input data and select the desired copula family (all or a subset of the models). In this study, we have focused our attention to mainly model parameter uncertainties. Efforts are underway for a more comprehensive analysis of different uncertainty sources (forcing data, model structural and calibration data) which could be beneficial to the community. MvCAT is freely available to public, and the interested users can download the source codes from http://amir.eng.uci.edu/software.php and http://coen.boisestate.edu/hydroclimate/softwares/.

Acknowledgments
This work was supported by the National Science Foundation award CMMI-1635797 and National Oceanic and Atmospheric Administration award NA14OAR4310222. Data for the drought analysis can be freely downloaded from the National Weather Service Climate Prediction Center's website available at http://www.cpc.ncep.noaa.gov/. We would like to appreciate the comments of four anonymous reviewers, and the Associate Editor which substantially improved the quality of this paper.

References

AghaKouchak, A. (2014), Entropy–copula in hydrology and climatology, *J. Hydrometeorol.*, *15*(6), 2176–2189.
AghaKouchak, A. (2015), A multivariate approach for persistence-based drought prediction: Application to the 2010–2011 East Africa drought, *J. Hydrol.*, *526*, 127–135.
AghaKouchak, A., L. Cheng, O. Mazdiyasni, and A. Farahmand (2014), Global warming and changes in risk of concurrent climate extremes: Insights from the 2014 California drought, *Geophys. Res. Lett.*, *41*, 8847–8852, doi:10.1002/2014GL062308.
Aho, K., D. Derryberry, and T. Peterson (2014), Model selection for ecologists: The worldviews of AIC and BIC, *Ecology*, *95*(3), 631–636.
Akaike, H. (1974), A new look at the statistical model identification, *IEEE Trans. Autom. Control*, *19*(6), 716–723.
Akaike, H. (1998), Information theory and an extension of the maximum likelihood principle, in *Selected Papers of Hirotugu Akaike*, pp. 199–213, Springer, New York.
Ali, M. M., N. Mikhail, and M. S. Haq (1978), A class of bivariate distributions including the bivariate logistic, *J. Multivar. Anal.*, *8*(3), 405–412.
Andrieu, C., and J. Thoms (2008), A tutorial on adaptive MCMC, *Stat. Comput.*, *18*(4), 343–373.
Balakrishna, N., and C. D. Lai (2009), Distributions expressed as copulas, in *Continuous Bivariate Distributions*, pp. 67–103, Springer, New York.
Bárdossy, A. (2006), Copula-based geostatistical models for groundwater quality parameters, *Water Resour. Res.*, *42*, W11416, doi:10.1029/2005WR004754.
Barnett, V. (1980), Some bivariate uniform distributions, *Commun. Stat. Theory Methods*, *9*(4), 453–461.
Box, G. E., and G. C. Tiao (2011), *Bayesian Inference in Statistical Analysis*, vol. 40, John Wiley, Hoboken, N. J.
Brahimi, B., F. Chebana, and A. Necir (2015), Copula representation of bivariate L-moments: A new estimation method for multiparameter two-dimensional copula models, *Statistics*, *49*(3), 497–521.
Byrd, R. H., J. C. Gilbert, and J. Nocedal (2000), A trust region method based on interior point techniques for nonlinear programming, *Math. Program.*, *89*(1), 149–185.
Charpentier, A., and J. Segers (2007), Lower tail dependence for Archimedean copulas: Characterizations and pitfalls, *Insur. Math. Econ.*, *40*(3), 525–532.
Chebana, F., and T. B. Ouarda (2011), Multivariate quantiles in hydrological frequency analysis, *Environmetrics*, *22*(1), 63–78.
Cheng, L., A. AghaKouchak, E. Gilleland, and R. W. Katz (2014), Non-stationary extreme value analysis in a changing climate, *Clim. Change*, *127*(2), 353–369.
Clayton, D. G. (1978), A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence, *Biometrika*, *65*(1), 141–151.
Cuadras, C. M., and J. Augé (1981), A continuous general multivariate distribution and its properties, *Commun. Stat. Theory Methods*, *10*(4), 339–353.
De Michele, C., and G. Salvadori (2003), A generalized Pareto intensity-duration model of storm rainfall exploiting 2-copulas, *J. Geophys. Res.*, *108*(D2), 4067, doi:10.1029/2002JD002534.
De Michele, C., G. Salvadori, M. Canossi, A. Petaccia, and R. Rosso (2005), Bivariate statistical approach to check adequacy of dam spillway, *J. Hydrol. Eng.*, *10*(1), 50–57.
Dracup, J. A., K. S. Lee, and E. G. Paulson (1980), On the definition of droughts, *Water Resour. Res.*, *16*(2), 297–302.
Duan, Q., S. Sorooshian, and V. Gupta (1992), Effective and efficient global optimization for conceptual rainfall-runoff models, *Water Resour. Res.*, *28*(4), 1015–1031.
Duan, Q., V. K. Gupta, and S. Sorooshian (1993), Shuffled complex evolution approach for effective and efficient global minimization, *J. Optim. Theory Appl.*, *76*(3), 501–521.
Durrleman, V., A. Nikeghbali, and T. Roncalli (2000), A note about the conjecture on spearman's rho and kendall's tau. [Available at http://dx.doi.org/10.2139/ssrn.1032558.]
Ellison, A. M. (2004), Bayesian inference in ecology, *Ecol. Lett.*, *7*(6), 509–520.
Embrechts, P., F. Lindskog, and A. McNeil (2001), Modelling dependence with copulas, rapport technique, Dép. de Math., Inst. Féd. de Technol. de Zurich, Zurich, Switzerland.
Favre, A.-C., S. El Adlouni, L. Perreault, N. Thiémonge, and B. Bobée (2004), Multivariate hydrological frequency analysis using copulas, *Water Resour. Res.*, *40*, W01101, doi:10.1029/2003WR002456.
Fischer, M., and G. Hinzmann (2007), A new class of copulas with tail dependence and a generalized tail dependence estimator, Citeseer, Friedrich-Alexander University Erlangen-Nuremberg, Germany.
Fisher, N. I. (1997), Copulas, in *Encyclopedia of Statistical Sciences*, John Wiley, Hoboken, N. J.
Frees, E. W., and E. A. Valdez (1998), Understanding relationships using copulas, *North Am. Actuarial J.*, *2*(1), 1–25.
Gelman, A., and D. B. Rubin (1992), Inference from iterative simulation using multiple sequences, *Stat. Sci.*, *7*, 457–472.
Genest, C., and A.-C. Favre (2007), Everything you always wanted to know about copula modeling but were afraid to ask, *J. Hydrol. Eng.*, *12*(4), 347–368.
Geweke, J. (1989), Bayesian inference in econometric models using Monte Carlo integration, *Econometrica*, *57*, 1317–1339.
Gilks, W. R., G. O. Roberts, and E. I. George (1994), Adaptive direction sampling, *Statistician*, *43*, 179–189.
Gräler, B., M. van den Berg, S. Vandenberghe, A. Petroselli, S. Grimaldi, B. De Baets, and N. Verhoest (2013), Multivariate return periods in hydrology: A critical and practical review focusing on synthetic design hydrograph estimation, *Hydrol. Earth Syst. Sci.*, *17*(4), 1281–1296.

Grimaldi, S., and F. Serinaldi (2006), Asymmetric copula in multivariate flood frequency analysis, *Adv. Water Resour.*, *29*(8), 1155–1167.
Grimaldi, S., A. Petroselli, G. Salvadori, and C. De Michele (2016), Catchment compatibility via copulas: A non-parametric study of the dependence structures of hydrological responses, *Adv. Water Resour.*, *90*, 116–133.
Gumbel, E. J. (1960), Bivariate exponential distributions, *J. Am. Stat. Assoc.*, *55*(292), 698–707.
Haario, H., E. Saksman, and J. Tamminen (1999), Adaptive proposal distribution for random walk metropolis algorithm, *Comput. Stat.*, *14*(3), 375–396.
Haario, H., E. Saksman, and J. Tamminen (2001), An adaptive metropolis algorithm, *Bernoulli*, *7*, 223–242.
Hao, Z., and A. AghaKouchak (2013), Multivariate standardized drought index: A parametric multi-index model, *Adv. Water Resour.*, *57*, 12–18.
Hayes, M., M. Svoboda, N. Wall, and M. Widhalm (2011), The Lincoln declaration on drought indices: Universal meteorological drought index recommended, *Bull. Am. Meteorol. Soc.*, *92*(4), 485–488.
Heim, R. R., Jr. (2002), A review of twentieth-century drought indices used in the United States, *Bull. Am. Meteorol. Soc.*, *83*(8), 1149–1165.
Huelsenbeck, J. P., and F. Ronquist (2001), Mrbayes: Bayesian inference of phylogenetic trees, *Bioinformatics*, *17*(8), 754–755.
Huelsenbeck, J. P., F. Ronquist, R. Nielsen, and J. P. Bollback (2001), Bayesian inference of phylogeny and its impact on evolutionary biology, *Science*, *294*(5550), 2310–2314.
Huynh, V.-N., V. Kreinovich, and S. Sriboonchitta (2014), *Modeling Dependence in Econometrics*, Springer, New York.
Jarrell, M., and J. E. Gubernatis (1996), Bayesian inference and the analytic continuation of imaginary-time quantum Monte Carlo data, *Phys. Rep.*, *269*(3), 133–195.
Joe, H. (2014), *Dependence Modeling With Copulas*, CRC Press, Boca Raton, Fla.
Jongman, B., et al. (2014), Increasing stress on disaster-risk finance due to large floods, *Nat. Clim. Change*, *4*(4), 264–268.
Kao, S.-C., and R. S. Govindaraju (2010), A copula-based joint deficit index for droughts, *J. Hydrol.*, *380*(1), 121–134.
Kavetski, D., G. Kuczera, and S. W. Franks (2006), Bayesian analysis of input uncertainty in hydrological modeling: 2. Application, *Water Resour. Res.*, *42*, W03408, doi:10.1029/2005WR004376.
Keyantash, J. A., and J. A. Dracup (2004), An aggregate drought index: Assessing drought severity based on fluctuations in the hydrologic cycle and surface water storage, *Water Resour. Res.*, *40*, W09304, doi:10.1029/2003WR002610.
Khoudraji, A. (1996), Contributions a l'etude des copules et a la modelisation de valeurs extremes bivariees, PhD thesis, Université Laval, Quebec City, Que., Canada.
Kuczera, G. (1999), Comprehensive at-site flood frequency analysis using Monte Carlo Bayesian inference, *Water Resour. Res.*, *35*(5), 1551–1557.
Kwon, H.-H., and U. Lall (2016), A copula-based nonstationary frequency analysis for the 2012–2015 drought in California, *Water Resour. Res.*, *52*, 5662–5675, doi:10.1002/2016WR018959.
Lee, T., R. Modarres, and T. Ouarda (2013), Data-based analysis of bivariate copula tail dependence for drought duration and severity, *Hydrol. Processes*, *27*(10), 1454–1463.
Li, C., V. P. Singh, and A. K. Mishra (2013), A bivariate mixed distribution with a heavy-tailed component and its application to single-site daily rainfall simulation, *Water Resour. Res.*, *49*, 767–789, doi:10.1002/wrcr.20063.
McKee, T. B., N. J. Doesken, and J. Kleist (1993), The relationship of drought frequency and duration to time scales, in Proceedings of the 8th Conference on Applied Climatology, vol. *17*, pp. 179–183, Am. Meteorol. Soc., Boston, Mass.
Min, A., and C. Czado (2010), Bayesian inference for multivariate copulas using pair-copula constructions, *J. Financ. Econ.*, *8*(4), 511–546.
Mishra, A., and V. P. Singh (2009), Analysis of drought severity-area-frequency curves using a general circulation model and scenario uncertainty, *J. Geophys. Res.*, *114*, D06120, doi:10.1029/2008JD010986.
Mo, K. C. (2011), Drought onset and recovery over the United States, *J. Geophys. Res.*, *116*, D20106, doi:10.1029/2011JD016168.
Nelsen, R. B. (2003), Properties and applications of copulas: A brief survey, in *Proceedings of the First Brazilian Conference on Statistical Modeling in Insurance and Finance*, edited by J. Dhaene, N. Kolev, and P. A. Morettin, pp. 10–28, Univ. of Sao Paulo, Sao Paulo, Brazil.
Nelsen, R. B. (2007), *An Introduction to Copulas*, Springer, New York.
Nikoloulopoulos, A. K., and D. Karlis (2008), Fitting copulas to bivariate earthquake data: The seismic gap hypothesis revisited, *Environmetrics*, *19*(3), 251–269.
Parent, E., A.-C. Favre, J. Bernier, and L. Perreault (2014), Copula models for frequency analysis what can be learned from a Bayesian perspective?, *Adv. Water Resour.*, *63*, 91–103.
Pitt, M., D. Chan, and R. Kohn (2006), Efficient Bayesian inference for Gaussian copula regression models, *Biometrika*, *93*(3), 537–554.
Plackett, R. L. (1965), A class of bivariate distributions, *J. Am. Stat. Assoc.*, *60*(310), 516–522.
Rachev, S. T. (2003), *Handbook of Heavy Tailed Distributions in Finance: Handbooks in Finance*, vol. 1, Elsevier, North Holland, Netherlands.
Renard, B., and M. Lang (2007), Use of a Gaussian copula for multivariate extreme value analysis: Some case studies in hydrology, *Adv. Water Resour.*, *30*(4), 897–912.
Ribatet, M., and M. Sedki (2013), Extreme value copulas and max-stable processes, *J. Soc. Fr. Stat.*, *154*(1), 138–150.
Roberts, G. O., and J. S. Rosenthal (2009), Examples of adaptive MCMC, *J. Comput. Graphical Stat.*, *18*(2), 349–367.
Roberts, G. O., and S. K. Sahu (1997), Updating schemes, correlation structure, blocking and parameterization for the Gibbs sampler, *J. R. Stat. Soc., Ser. B*, *59*(2), 291–317.
Roch, O., and A. Alegre (2006), Testing the bivariate distribution of daily equity returns using copulas. An application to the Spanish stock market, *Comput. Stat. Data Anal.*, *51*(2), 1312–1329.
Sadegh, M., and J. Vrugt (2013), Bridging the gap between glue and formal statistical approaches: Approximate Bayesian computation, *Hydrol. Earth Syst. Sci.*, *17*, 4831–4850.
Sadegh, M., and J. A. Vrugt (2014), Approximate Bayesian computation using Markov chain Monte Carlo simulation: $\text{DREAM}_{(\text{ABC})}$, *Water Resour. Res.*, *50*, 6767–6787, doi:10.1002/2014WR015386.
Sadegh, M., J. A. Vrugt, C. Xu, and E. Volpi (2015), The stationarity paradigm revisited: Hypothesis testing using diagnostics, summary metrics, and $\text{DREAM}_{(\text{ABC})}$, *Water Resour. Res.*, *51*, 9207–9231, doi:10.1002/2014WR016805.
Salvadori, G., and C. De Michele (2004a), Analytical calculation of storm volume statistics involving Pareto-like intensity-duration marginals, *Geophys. Res. Lett.*, *31*, L04502, doi:10.1029/2003GL018767.
Salvadori, G., and C. De Michele (2004b), Frequency analysis via copulas: Theoretical aspects and applications to hydrological events, *Water Resour. Res.*, *40*, W12511, doi:10.1029/2004WR003133.
Salvadori, G., F. Durante, C. De Michele, M. Bernardi, and L. Petrella (2016), A multivariate copula-based framework for dealing with hazard scenarios and failure probabilities, *Water Resour. Res.*, *52*, 3701–3721, doi:10.1002/2015WR017225.
Schwarz, G. (1978), Estimating the dimension of a model, *Ann. Stat.*, *6*(2), 461–464.

Serinaldi, F. (2015), Dismissing return periods!, *Stochastic Environ. Res. Risk Assess.*, *29*(4), 1179–1189.
Serinaldi, F., and C. G. Kilsby (2013), The intrinsic dependence structure of peak, volume, duration, and average intensity of hyetographs and hydrographs, *Water Resour. Res.*, *49*, 3423–3442, doi:10.1002/wrcr.20221.
Shih, J. H., and T. A. Louis (1995), Inferences on the association parameter in copula models for bivariate survival data, *Biometrics*, *51*, 1384–1399.
Shukla, S., and A. W. Wood (2008), Use of a standardized runoff index for characterizing hydrologic drought, *Geophys. Res. Lett.*, *35*, L02405, doi:10.1029/2007GL032487.
Sklar, M. (1959), Fonctions de Répartition À N Dimensions Et Leurs Marges, Univ. Paris 8, Saint-Denis, France.
Smith, M. S., Q. Gan, and R. J. Kohn (2012), Modelling dependence using skew t copulas: Bayesian inference and applications, *J. Appl. Econ.*, *27*(3), 500–522.
Sorooshian, S., and J. A. Dracup (1980), Stochastic parameter estimation procedures for hydrologie rainfall-runoff models: Correlated and heteroscedastic error cases, *Water Resour. Res.*, *16*(2), 430–442.
Storn, R., and K. Price (1995), *Differential Evolution—A Simple and Efficient Adaptive Scheme for Global Optimization Over Continuous Spaces*, vol. *3*, Int. Comput. Sci. Inst., Berkeley, Calif.
Storn, R., and K. Price (1997), Differential evolution—A simple and efficient heuristic for global optimization over continuous spaces, *J. Global Optim.*, *11*(4), 341–359.
Svoboda, M., et al. (2002), The drought monitor, *Bull. Am. Meteorol. Soc.*, *83*(8), 1181–1190.
Szolgay, J., L. Gaál, T. Bacigál, S. Kohnová, K. Hlavčová, R. Vỳleta, J. Parajka, and G. Blöschl (2016), A regional comparative analysis of empirical and theoretical flood peak-volume relationships, *J. Hydrol. Hydromech.*, *64*(4), 367–381.
Ter Braak, C. J. (2006), A Markov chain Monte Carlo version of the genetic algorithm differential evolution: Easy Bayesian computing for real parameter spaces, *Stat. Comput.*, *16*(3), 239–249.
ter Braak, C. J., and J. A. Vrugt (2008), Differential evolution Markov chain with snooker updater and fewer chains, *Stat. Comput.*, *18*(4), 435–446.
Thiemann, M., M. Trosset, H. Gupta, and S. Sorooshian (2001), Bayesian recursive parameter estimation for hydrologic models, *Water Resour. Res.*, *37*(10), 2521–2535.
Thyer, M., B. Renard, D. Kavetski, G. Kuczera, S. W. Franks, and S. Srikanthan (2009), Critical evaluation of parameter consistency and predictive uncertainty in hydrological modeling: A case study using Bayesian total error analysis, *Water Resour. Res.*, *45*, W00B14, doi:10.1029/2008WR006825.
Vernieuwe, H., S. Vandenberghe, B. De Baets, and N. Verhoest (2015), A continuous rainfall model based on vine copulas, *Hydrol. Earth Syst. Sci.*, *19*(6), 2685–2699.
Vicente-Serrano, S. M., S. Beguería, and J. I. López-Moreno (2010), A multiscalar drought index sensitive to global warming: The standardized precipitation evapotranspiration index, *J. Clim.*, *23*(7), 1696–1718.
Vrugt, J. A., and M. Sadegh (2013), Toward diagnostic model calibration and evaluation: Approximate Bayesian computation, *Water Resour. Res.*, *49*, 4335–4345, doi:10.1002/wrcr.20354.
Vrugt, J. A., C. Ter Braak, C. Diks, B. A. Robinson, J. M. Hyman, and D. Higdon (2009), Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling, *Int. J. Nonlinear Sci. Numer. Simul.*, *10*(3), 273–290.
Wahl, T., S. Jain, J. Bender, S. D. Meyers, and M. E. Luther (2015), Increasing risk of compound flooding from storm surge and rainfall for major us cities, *Nat. Clim. Change*, *5*(12), 1093–1097.
Waltz, R. A., J. L. Morales, J. Nocedal, and D. Orban (2006), An interior algorithm for nonlinear optimization that combines line search and trust region steps, *Math. Program.*, *107*(3), 391–408.
Wood, E. F., and I. Rodríguez-Iturbe (1975), Bayesian inference and decision making for extreme hydrologic events, *Water Resour. Res.*, *11*(4), 533–542.
Yue, S., T. Ouarda, B. Bobée, P. Legendre, and P. Bruneau (1999), The Gumbel mixed model for flood frequency analysis, *J. Hydrol.*, *226*(1), 88–100.
Zhang, L., and V. P. Singh (2007), Bivariate rainfall frequency distributions using Archimedean copulas, *J. Hydrol.*, *332*(1), 93–109.