

Identification of hydrologic drought triggers from hydroclimatic predictor variables

Rajib Maity,¹ Meenu Ramadas,² and Rao S. Govindaraju²

Received 13 August 2012; revised 30 May 2013; accepted 2 June 2013; published 31 July 2013.

[1] Drought triggers are patterns in hydroclimatic variables that herald upcoming droughts and form the basis for mitigation plans. This study develops a new method for identification of triggers for hydrologic droughts by examining the association between the various hydroclimatic variables and streamflows. Since numerous variables influence streamflows to varying degrees, principal component analysis (PCA) is utilized for dimensionality reduction in predictor hydroclimatic variables. The joint dependence between the first two principal components, that explain over 98% of the variability in the predictor set, and streamflows is computed by a scale-free measure of association using asymmetric Archimedean copulas over two study watersheds in Indiana, USA, with unregulated streamflows. The M6 copula model is found to be suitable for the data and is utilized to find expected values and ranges of predictor hydroclimatic variables for different streamflow quantiles. This information is utilized to develop drought triggers for 1 month lead time over the study areas. For the two study watersheds, soil moisture, precipitation, and runoff are found to provide the fidelity to resolve amongst different drought classes. Combining the strengths of PCA for dimensionality reduction and copulas for building joint dependence allows the development of hydrologic drought triggers in an efficient manner.

Citation: Maity, R., M. Ramadas, and R. S. Govindaraju (2013), Identification of hydrologic drought triggers from hydroclimatic predictor variables, *Water Resour. Res.*, 49, 4476–4492, doi:10.1002/wrcr.20346.

1. Introduction

[2] Drought, as a prolonged status of water deficit, is perceived as one of the most expensive and the least understood natural disasters. In monetary terms alone, a typical drought costs American farmers and businesses \$6–8 billion dollars each year, more than damages incurred from floods and hurricanes [*Federal Emergency Management Agency*, 1995]. The consequences tend to be more severe in areas, where agriculture is a major economic driver. *Dra-cup et al.* [1980] stated that proper definition of drought depends on the nature of water deficit relevant to the study area. As water moves through the hydrologic cycle, precipitation deficits (meteorological droughts) lead to low soil moisture levels (agricultural droughts) that translate into low streamflows, reservoir, and/or groundwater levels (hydrologic droughts).

[3] The occurrence and magnitude of hydrologic droughts are heralded by triggers that may be manifested in specific patterns of hydroclimatic variables. Identification of these triggers at appropriate lead times is necessary for devising effective drought mitigation plans. Estimating water deficits

and drought categories at weekly, monthly, seasonal, and annual lead times are needed for scheduling irrigation events and managing water resources of a region. Drought characterization is currently accomplished by indices such as standardized precipitation index (SPI), palmer drought severity index, crop moisture index, surface water supply index, and reclamation drought index ([http://drought.unl.edu/Planning/Monitoring/ ComparisonofIndicesIntro.aspx](http://drought.unl.edu/Planning/Monitoring/ComparisonofIndicesIntro.aspx)). Drought indices are typically designed for assessing current conditions and have little predictive capability. Large-scale oceanic and atmospheric indicators such as the El Niño–Southern oscillation phases, North Atlantic oscillations, Pacific North American index, Atlantic multidecadal oscillations, and Pacific decadal oscillations are used as long-term precursors to annual/seasonal forecasts of precipitation [*Ropelewski and Halpert*, 1996; *McHugh and Rogers*, 2001; *Maity and Nagesh Kumar*, 2008a]. However, for many parts of the world, including Indiana, USA, these indicators have been found to have little to no influence [*Charusombat and Niyogi*, 2011]. Further, their incapability to provide short-term predictions (several weeks, to 6 month range) render them unsuitable as drought triggers for such time scales. We hypothesize that hydrological droughts, reflected in unregulated streamflows, would have precursors in local hydrometeorologic variables related to rainfall and soil moisture over the corresponding watersheds. *McKay et al.* [1989] suggested that accurate drought predictions will need models that link between climate and weather factors to streamflows and river stage data.

[4] Several considerations come into play for the development of drought triggers including drought types, data availability, choice of hydrologic variables (precipitation,

¹Department of Civil Engineering, Indian Institute of Technology, Kharagpur, India.

²School of Civil Engineering, Purdue University, West Lafayette, Indiana, USA.

Corresponding author: R. Maity, Department of Civil Engineering, Indian Institute of Technology, Kharagpur 721302, India. (rajib@civil.iitkgp.ernet.in)

temperature, streamflows, storage levels, etc.), temporal scales, and validity of the trigger. Over the past two decades, drought triggers have been developed by several states and utilities [Steinemann, 2003]. However, these have met with limited success because of (i) anomalies between results from different drought indicators and (ii) lack of a strong record length for proper model development and validation exercises. Moreover, these triggers are often defined as some preset thresholds to be crossed by various drought indices at the same instance of time for which drought status is being analyzed. Thus, they may not recognize early warning signals that may be present in the record.

[5] Though droughts are fundamentally triggered by insufficient precipitation, the evolution of water deficits from precipitation to soil moisture and to streamflows is not instantaneous and is controlled by complex physical mechanisms. As hydrologic droughts are based on abnormally low flows, estimation of streamflows is, therefore, a necessary prerequisite to drought analysis. Since a drought trigger governs the level of future response, it is important that the trigger be based on methods that convey predictive uncertainty. There are many methods available for estimation of streamflows, classified mainly into physics-based, conceptual, and data-driven approaches. Several watershed models have been developed that rely upon the physical knowledge of the watershed and the hydrological cycle, often resulting in complex representations that require intensive computer effort for model calibration and corroboration. Data-driven techniques do not require detailed understanding of the inherent physical mechanisms, but have shown comparable accuracy for streamflow prediction as physics-based models [Wu *et al.*, 2009]. The time scale of 1 month lead forecasts is particularly challenging because physics-based models (HEC-HMS, MIKE-SHE, etc.) are not able to project using input data beyond several hours to days without a disaggregation procedure. Process-based models such as SWAT perform simulations at a daily time step [Srinivasan and Arnold, 1994], and model outputs have to be aggregated to obtain monthly values. However, the strength of such models lies in examining long-term consequences of management practices rather than monthly forecasts. There are many conceptual lumped-parameter models developed in the last four decades, mainly for flood forecasting, with one day or shorter time resolutions [Xu and Singh, 2004], but their predictive capabilities are very limited if the time horizon exceeds several days.

[6] Statistical approaches have been utilized to model the complex relationships between streamflows and the large-scale atmospheric circulation phenomena [Anmala *et al.*, 2000; Maity and Nagesh Kumar, 2008b]. The predictors used in majority of these data-driven approaches were hydroclimatic variables such as mean temperature, mean sea-level pressure, soil moisture, precipitation, runoff, and wind speed. While these studies have stressed the importance of hydroclimatic variables for enhancing streamflow prediction, they were primarily targeted toward long-range forecasting [Salas *et al.*, 2011]. Even with the predictor set identified, new approaches are needed for achieving short-term (few weeks to months) forecasts. The use of advanced statistical models based on Markov properties [e.g., Mallya

et al., 2013] have helped in probabilistic classification of drought states and alleviated the need for user-specified thresholds for drought categorization. Thus, though robust models exist for forecasting streamflows and upcoming hydrologic droughts, these models are not suitable for development of triggers that require identification of the ranges of predictor variables that herald a particular drought.

[7] The joint probability density function between streamflows and hydroclimatic predictor variables is needed to identify and develop drought triggers. Copulas are a natural choice for this task [Nelsen, 2006]. They allow the dependence structure to be modeled without any restriction on the distributions of the marginals [Genest and Favre, 2007] and have been gaining popularity with hydrologic applications. Favre *et al.* [2004] used Frank and Clayton 2-copulas to model the dependence between streamflow peaks and volumes. Salvadori and De Michele [2004] adopted copulas in their study of the return period of hydrological events. Zhang and Singh [2006] used copulas to determine bivariate distributions between flood peaks, volumes and durations, and employed them to define joint and conditional return periods needed for hydrologic design calculations. The joint distribution of intensity, duration, and severity of droughts was modeled using copulas by Shiau *et al.* [2007], Wong *et al.* [2010], and Madadgar and Moradkhani [2013]. Maity and Nagesh Kumar [2008a] analyzed the dependencies among the teleconnected hydroclimatic variables using copulas for the prediction of response variables using large-scale oceanic and atmospheric indicators. Kao and Govindaraju [2010a] utilized copulas to construct an intervariable drought index, where the dependence structure of precipitation and streamflow marginals was preserved. The review by Mishra and Singh [2010] highlights the expanding role of copulas in drought assessment studies.

[8] Given the large number of potential hydroclimatic variables in the predictor set, the direct use of copulas to model their joint dependence with streamflows is impractical because of the mathematical complexity in constructing higher-dimensional copulas. If the dependence between all the interacting variables cannot be represented by multivariate Gaussian (or meta-elliptical) copulas, then models at even the trivariate level can be very challenging [Kao and Govindaraju, 2008, 2010b]. Moreover, with multiple interacting variables, the curse of dimensionality adds further challenges to estimation of model parameters from limited record lengths. While many options exist for modeling bivariate dependence between variables, models for higher dimensions are not easily available.

[9] Principal component analysis (PCA) provides an elegant way of projecting the precursor hydroclimatic variables onto a feature space, and representing the original data through a reduced number of effective features called principal components (PCs) [Jolliffe, 1986; Preisendorfer, 1988]. If the first few (two in this case) features are able to explain most of the variability (>90%) in the original data set, then substantial dimensionality reduction may be achieved through unsupervised learning. PCA is recognized as the most widely used tool for dimensionality reduction for multivariate data problems. Lins [1985] utilized PCA to

construct parsimonious models for multisite streamflows. Maurer *et al.* [2004] showed the effectiveness of PCA for both reducing the dimensionality of large data sets and better graphical representation of the modes of variability in streamflows. Tripathi and Govindaraju [2008] developed algorithms for data compression using PCA for data sets with noise. PCA was adopted by Keyantash and Dracup [2004] to achieve dimensionality reduction for developing an aggregate drought index.

[10] The goal of this paper is two-fold. The first goal is to model the joint distribution of streamflows and the important principal components of precursor hydroclimate variables using an appropriate copula family for two study watersheds in Indiana, USA. This copula model is tested for its capability to forecast low streamflows that are of concern for hydrologic droughts. The second goal is to utilize the PCA-copula framework to develop drought trigger information. While copulas and PCA have been widely used individually, to the best of our knowledge, no prior studies exist for identifying drought triggers in this fashion. The details of study watersheds are provided in section 2. The methodology adopted in the study with details of principal components analysis, copula models, and drought trigger analysis are explained in section 3. These are followed by results and discussion in section 4, and the summary and conclusions of the study in section 5.

2. Study Area and Data Used

2.1. Study Area

[11] The study was carried out over two watersheds in the state of Indiana, USA. Both the watersheds form a part of the Ohio River Basin. The first watershed (WS I) extending from 38°34'N to 39°49'N and 85°24'W to 86°31'W spreads over 6259 square kilometers. The second watershed (WS II) lies between 40°47'N to 41°24'N and 85°08'W to 86°20'W and extends over an area of 1657 square kilometers. The two watersheds are shown in Figure 1. The land use in these watersheds consists of mainly agricultural and forest lands, followed by public and urban built-up lands. Agriculture being the major economic activity prevalent in WS I and WS II, high irrigation water demands exist during the growing season. The choice of the watersheds was governed by the need to conduct drought analyses for locations, where streamflows were not influenced by human activities.

2.2. Data Used

[12] The 30 m resolution DEMs obtained from USGS National Elevation Data set was used to delineate the watersheds. Though the choice of coarser resolution affects the identification of drainage features in low relief landscapes, there is substantial reduction in computational efforts involved in the processing of the 30 m digital elevation model (DEM) over a high-resolution DEM. Modeling the dependencies and analysis of drought triggers require a long record of historic observations. Therefore, monthly data with a minimum record length of 50 years were adopted in the present study. The various hydroclimatic variables used in the study are listed in Table 1. The 0.5° × 0.5° climate prediction center (CPC) global monthly data sets [Huang *et al.*, 1996; Fan and van den Dool, 2004],

available from 1948 onwards, were used. The land model was treated as a one-layer “bucket” water balance model, when generating the CPC data sets. The data used in our study include modeled monthly soil moisture values, modeled monthly runoff values, observed monthly precipitation values, observed monthly temperature values, and modeled monthly evaporation values. The location of CPC stations is marked by circles in Figure 1. Given the small watershed sizes determined by the need for unregulated streamflows, the number of CPC grid points directly over the study areas is quite small. The variables: Sea-level pressure, u-wind, and v-wind were obtained from the NCEP/NCAR Reanalysis-1 project data, at a spatial resolution of 2.5° × 2.5° [Kalnay *et al.*, 1996]. The resultant of the u-wind and v-wind components was adopted as the wind speed variable in the present study. Given the monthly time scale chosen for this study, the time of concentration for these watersheds is in the order of days. Thus, variables were multiplied by the Thiessen weights at different grid points to obtain their spatially averaged values over the study watersheds. The US Geological Survey (USGS) monthly streamflow data from 1958 to 2010 recorded at the USGS 03371500 (East Fork White River near Bedford, Indiana) were used for WS I, while the data at USGS streamflow gage 03328500 (Eel River near Logansport, Indiana) from 1948 to 2010 were used for WS II.

3. Methodology

3.1. Dimensionality Reduction Using Principal Components Analysis

[13] The formulation of a dependence model between the seven predictor variables in Table 1 and streamflows is impractical even when using copulas. PCA was performed to transform the set of correlated n -dimensional ($n = 7$ here) predictor set into another set of n -dimensional uncorrelated vectors (called principal components). The PCs are arranged in order of their ability to explain the variability in the data. The conventional or standard PCA, which is formulated as an eigenvalue problem, was used for unsupervised dimensionality reduction [Jolliffe, 1986]. Prior to extracting the principal components, the mean value was subtracted from each of the predictors to obtain a series of predictor anomalies. The covariance matrix was obtained for the anomaly data sets, and the eigenvalues and eigenvectors of this covariance matrix were computed. The degree of dimensionality reduction achieved in the predictor set was determined by variance explained by the first two principal components.

3.2. Asymmetric Archimedean Class of Copulas

[14] A copula is a function that models the dependence between multiple random variables, regardless of their marginals. A d -dimensional copula is a multivariate cumulative density function (CDF) C defined in the unit d -dimensional space $[0, 1]^d$ with uniform margins $[0, 1]$ and with the following properties: (i) $\forall u \in [0, 1]^d$, $C(u) = 0$ if at least one coordinate of u is equal to 0, and $C(u) = u_k$ if all the coordinates of u are equal to 1 except u_k ; (ii) $\forall a$ and $b \in [0, 1]^d$ such that $a \leq b$, $V_c([a, b]) \geq 0$, where V is the C -volume [Nelsen, 2006]. The copula approach to dependence modeling has its roots in the theorem by Sklar [1959],

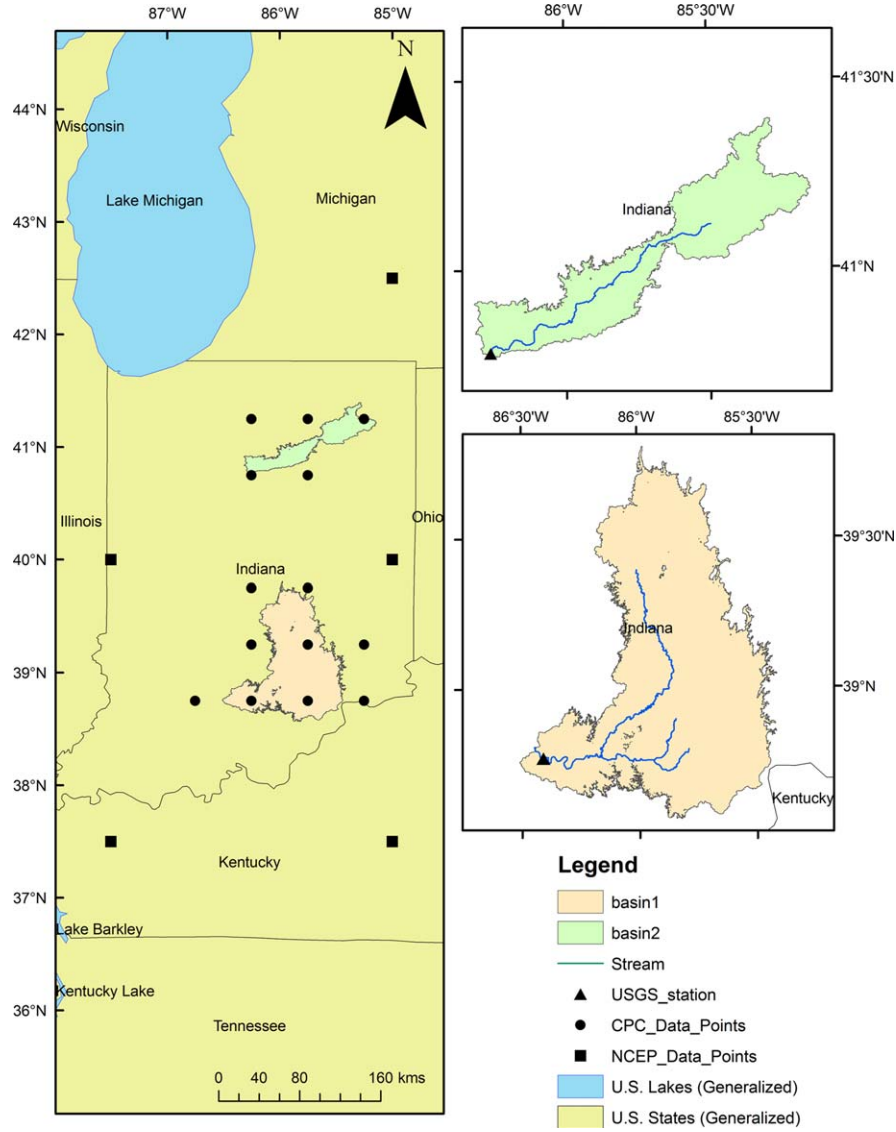


Figure 1. Map of the study watersheds WS I and WS II.

according to which a d -dimensional CDF with univariate margins F_1, F_2, \dots, F_d is defined by:

$$H(x_1, x_2, \dots, x_d) = C(F_1(x_1), F_2(x_2), \dots, F_d(x_d)) = C(u_1, u_2, \dots, u_d) \quad (1)$$

where $F_k(x_k) = u_k$ for $k = 1, 2, \dots, d$ with $U_k \in U(0, 1)$ if F_k is continuous.

[15] Archimedean copulas are very popular, with both symmetric and asymmetric forms available in the literature [Joe, 1997; Nelsen, 2006]. They possess closed form expressions and allow modeling of a variety of different dependence structures. An Archimedean symmetric d -copula is of the form:

$$C(u) = \varphi^{-1} \left(\sum_{k=1}^d \varphi_k(u_k) \right) \quad (2)$$

where the function φ (called the generator of the copula) is a continuous strictly decreasing function from $[0, 1]$ to $[0, \infty)$, such that $\varphi(0) = \infty$ and $\varphi(1) = 0$, and its inverse φ^{-1} is completely monotone on $[0, \infty)$, that is, φ^{-1} has derivatives of all orders which alternate in sign [Nelsen, 2006]:

$$(-1)^k \cdot \frac{d^k \varphi^{-1}(t)}{dt^k} \geq 0 \quad (3)$$

for all t in $[0, \infty)$ and $k = 1, 2, \dots, d$.

[16] In equation (2), if a certain u_k is assigned the value 1, then the joint distribution of $(u_1, u_2, \dots, u_d | u_k)$ is obtained. Since $\varphi(u_k) = 0$ when $k = 1$, the $(d-1)$ -dimensional marginal of the symmetric Archimedean copula is also an Archimedean copula. The expressions for these $(d-1)$ -dimensional copulas are identical regardless of the choice of k . As a result, only one Archimedean 2-copula is required to model all mutual dependencies among the

Table 1. List of Variables Used in the Study

Sl. No.	Variable	Unit	Period of Data
1	Soil moisture	mm	1948–2010
2	Runoff	mm	1948–2010
3	Temperature	°C	1948–2010
4	Precipitation	mm	1948–2010
5	Evaporation	mm	1948–2010
6	Sea-level pressure	mbar	1948–2010
7	Wind speed	m/s	1948–2010

variables. This exchangeability property that can be modeled by symmetric copulas limits the nature of the dependence structures. Since the study took into account correlated variables such as streamflows and principal components that possess different bivariate dependence structures, a more general multivariate extension of the Archimedean 2-copula, namely the fully nested or asymmetric copula as described in *Whelan* [2004], was adopted here. This copula is given by $d - 1$ distinct generating functions as:

$$C(u_1, u_2, \dots, u_d) = C_1(u_d, C_2(u_{d-1}, \dots, C_{d-1}(u_2, u_1) \dots)) \quad (4)$$

[17] For example, in a fully nested 3-copula, two variables u_1 and u_2 are coupled using copula C_2 and the copula of u_1 and u_2 , is coupled with u_3 by copula C_1 . In general, there are $d(d - 1)/2$ ways of coupling d variables. When the bivariate joint probability of two variables conditioned on the third variable is computed, different dependence structures are obtained based on the conditioning variable. *Grimaldi and Serinaldi* [2006] used asymmetric Archimedean copulas to model trivariate joint distribution of flood peaks, volumes, and durations. A nested 3-copula was adopted in the present study to model the dependence between the monthly streamflow anomaly and the first two principal components of a set of predictor variables. There are two parameters for the nested 3-copula model: θ_1 and θ_2 such that $\theta_1 \leq \theta_2$ implying a higher degree of dependence for the inner nested variables. It has been found that only two dependence structures can be reproduced for three possible pairs [Grimaldi and Serinaldi, 2006]. When two variables u_1 and u_2 are likely correlated with the third one u_3 , and the degree of dependence between u_1, u_2 is stronger than that of either u_1 and u_2 with u_3 , the asymmetric three-

dimensional model may be applied. The dependence between the variables is expressed in terms of the Kendall's correlation coefficient, τ . Kendall's τ for a random vector $(X, Y)^T$ is simply the probability of concordance minus the probability of discordance [Embrechts et al., 2003]:

$$\tau_{XY} = \text{Prob} \{ (X - \tilde{X})(Y - \tilde{Y}) > 0 \} - \text{Prob} \{ (X - \tilde{X})(Y - \tilde{Y}) < 0 \} \quad (5)$$

[18] The various asymmetric Archimedean copula families selected for the study, their permissible θ values, and dependence ranges are listed in Table 2.

3.3. Parameter Estimation

[19] Several copula parameter estimation methods are available in the literature namely, the method of moments, canonical maximum likelihood (CML) method, and inference from margins method. When one-parameter bivariate copulas are adopted, the popular approach is the simple method of moments based on inversion of Spearman's or Kendall's rank correlation [Genest and Favre, 2007]. In the multivariate-multiparameter case, this method becomes less elegant and may lead to inconsistencies. In such instances, a more natural estimation technique is the CML method [Genest et al., 1995; Kojadinovic and Yan, 2011]. The parameters of the five nested 3-copula families used in this study were estimated using the CML method. This method performs a nonparametric estimation of the marginals by using the respective scaled ranks. The dependence parameters θ_1 and θ_2 are obtained by maximizing the log-likelihood function $l(\theta)$ given by:

$$l(\theta) = \sum_{i=1}^n \log [c_{\theta} \{ \hat{F}_1(x_{i1}), \hat{F}_2(x_{i2}), \dots, \hat{F}_d(x_{id}) \}] \quad (6)$$

where c_{θ} denotes the density of the copula C_{θ} , and $\hat{F}_k(x_{ik})$ (also denoted as u_k) is the rank-based nonparametric marginal probability of k th variable given by:

$$\hat{F}_k(x_{ik}) = \frac{1}{n+1} \sum_{i=1}^n I(X_{ik} \leq x_{ik}) \quad k = 1, 2, \dots, d \quad (7)$$

where $I(\bullet)$ is indicator function returning 1 if the argument is true and 0 otherwise.

Table 2. Asymmetric Archimedean Copula Families Used in the Study

Family	Nested Copula $C_{\theta_1}(u_3, C_{\theta_2}(u_1, u_2))$	$\theta_2 \geq \theta_1 \in$	$\tau_{12}, \tau_{23}, \tau_{13} \in$	Reference
M3	$-\theta_1^{-1} \log \left\{ 1 - (1 - e^{-\theta_1})^{-1} (1 - [1 - (1 - e^{-\theta_2})^{-1} (1 - e^{-\theta_2 u_1}) (1 - e^{-\theta_2 u_2})]^{(\theta_1/\theta_2)}) (1 - e^{-\theta_1 u_3}) \right\}$	$[0, \infty)$	$[0, 1]$	Joe [1997]
M4	$[(u_1^{-\theta_2} + u_2^{-\theta_2} - 1)^{(\theta_1/\theta_2)} + u_3^{-\theta_1} - 1]^{(-1/\theta_1)}$	$[0, \infty)$	$[0, 1]$	Joe [1997]
M5	$1 - \left[\left\{ (1 - u_1)^{\theta_2} (1 - (1 - u_2)^{\theta_2}) + (1 - u_2)^{\theta_2} \right\}^{(\theta_1/\theta_2)} (1 - (1 - u_3)^{\theta_1}) + (1 - u_3)^{\theta_1} \right]^{(1/\theta_1)}$	$[1, \infty)$	$[0, 1]$	Joe [1997]
M6	$e^{\{ -[(\log u_1)^{\theta_2} + (\log u_2)^{\theta_2}]^{(\theta_1/\theta_2)} + (-\log u_3)^{\theta_1} \}^{(1/\theta_1)}}$	$[1, \infty)$	$[0, 1]$	Joe [1997] and Embrechts et al. [2003]
M12	$\left\{ \left([(u_1^{-1} - 1)^{\theta_2} + (u_2^{-1} - 1)^{\theta_2}]^{(\theta_1/\theta_2)} + (u_3^{-1} - 1)^{\theta_1} \right)^{(1/\theta_1)} + 1 \right\}^{-1}$	$[1, \infty)$	$[0.333, 1]$	Embrechts et al. [2003]

3.4. Goodness-of-Fit Tests for Asymmetric Copulas

[20] When there exist more than one feasible copula families that satisfy the dependence range for the given data, the final selection of a suitable copula is based on the best fit to observations. This fit can be assessed graphically by comparing the scatter plots of observed and simulated data in the case of bivariate distributions, but becomes difficult for higher dimensions. Goodness-of-fit tests examine the null hypothesis $H_0: C \in C_0$ for a copula class C_0 against $H_1: C \notin C_0$. These tests compare the distance between the empirical distribution of copula, C_n and an estimation of C_{θ_n} of C obtained under H_0 [Genest et al., 2009]. Formally, the goodness-of-fit tests are based on the statistic:

$$\Omega = \sqrt{n}\{C_n(u) - C_{\theta_n}(u)\} \quad u \in [0, 1]^d \quad (8)$$

where the empirical copula of the data X_1, X_2, \dots, X_d is defined by Deheuvels [1981] as:

$$C_n(u) = \frac{1}{n} \sum_{i=1}^n I(U_i \leq u), u \in [0, 1]^d \quad (9)$$

[21] In this study, the rank-based versions of Cramér-von Mises and Kolmogorov-Smirnov statistics were used for testing the goodness-of-fit of the nested copulas. The Cramér-von Mises statistic S_n has been a popular goodness-of-fit test procedure for copula models [Genest et al., 2009]. The statistic S_n was determined using equation (10), substituting the value of C_θ evaluated from the copula expression.

$$S_n = \sum_{i=1}^n \{C_n(\hat{U}_i) - C_{\theta_n}(\hat{U}_i)\}^2 \quad (10)$$

where C_n is the empirical copula computed as per equation (9).

[22] The Kolmogorov-Smirnov statistic T_n utilizes the absolute maximum distance between the empirical copula probability distribution and that simulated using the estimated parameters to measure the fit of the copulas as shown below [Genest et al., 2009].

$$T_n = \max_{u \in [0, 1]^d} |\sqrt{n}\{C_n(u) - C_{\theta_n}(u)\}| \quad (11)$$

[23] Additionally, the probability plots of the empirical distribution and the nested copula families were compared to assess the performance of copulas. The family providing the best fit based on the above criteria was selected for subsequent analysis.

3.5. Streamflow Forecasting and Drought Analysis

[24] The joint dependence modeled using the best copula was employed to estimate 1 month ahead streamflows. The probabilistic predictions of streamflows at different quantiles were made using the copula function. The expected values of monthly streamflows during the model development and model testing periods were computed. The range of forecasts was quantified by estimating predictions at 2.5% and 97.5% probabilities, i.e., 95% confidence interval

for the prediction. The forecasts of streamflow were analyzed to identify the occurrence of extremes, particularly for droughts in the study area. Given the focus on streamflows in this study, hydrological droughts were characterized by the standardized streamflow index that is similar to the SPI introduced by McKee et al. [1993] for meteorological drought analysis. The long-term streamflows record was fitted to a gamma probability distribution and then transformed to a standard normal distribution through the quantiles so that the mean standardized index for a certain location and particular period (1 month) is zero [Edwards and McKee, 1997]. A positive value of the index shows the degree of wetness, while a negative value indicates the severity of streamflow deficit. The ranges of this drought index for different hydrological conditions, labeled exceptionally dry (D4) to exceptionally wet (W4), are presented in Table 3. This drought severity classification based on SPI values was adopted from <http://droughtmonitor.unl.edu/classify.htm>. The streamflows estimated using copula were used for the prediction of droughts in the study areas.

3.6. Analysis for Drought Triggers

[25] The occurrence of hydrological extremes in the study areas was highly correlated with the local hydroclimatic variables at 1 month lead times, and as such short-term predictions of droughts could be achieved. The joint dependence information contained in the copula was exploited to obtain the expected values of the climate precursor anomalies conditioned on a streamflow anomaly. This allowed for identification of patterns in the precursors that could trigger hydrological droughts of different categories.

4. Results and Discussion

4.1. Principal Components Analysis

[26] The anomalies of hydroclimatic predictors and streamflows at monthly scale were obtained by subtracting their respective monthly means. The dependence between the first two principal components of the anomalies of these variables was represented by a joint asymmetric copula in the present study and was used to predict streamflows. The data from January 1958 to December 1993 were used for developing the statistical model for WS I, whereas model development period for WS II was from January 1948 to December 1990. Thus, two thirds of the data were used for model training and the remainder used for evaluating model performance.

[27] Starting from the large suite of potential predictors, PCA was used for dimensionality reduction. The results of principal components analysis performed on the predictor variables for the two watersheds are given in Table 4. As the first two components (PCs) were found to explain more than 98% of the variance, only these were selected for modeling streamflows. Next, the correlation values of different pairs (streamflow anomaly and two PCs) for different lags (1–3 months) were computed. PCs from predictor variables lagged by only 1 month were adopted for streamflow forecasting, as significant correlations were observed at this lag for both WS I and WS II.

4.2. Analysis of Asymmetric Archimedean Copula

[28] The joint dependence between the streamflow anomaly, PC-1 and PC-2 require that the nature of

Table 3. Range of Drought Index for Different Hydrological States

State	Description	Drought Index
D4	Exceptional drought	−2 or less
D3	Extreme drought	−1.6 to −1.9
D2	Severe drought	−1.3 to −1.5
D1	Moderate drought	−0.8 to −1.2
D0	Abnormally dry	−0.5 to −0.7
Normal	Normal condition	−0.4 to 0.4
W0	Abnormally wet	0.5–0.7
W1	Moderately wet	0.8–1.2
W2	Severely wet	1.3–1.5
W3	Extremely wet	1.6–1.9
W4	Exceptionally wet	2 or more

association between them be identified. The scatter plots of the pairs of predict and predictor variables indicated a higher degree of dependence between the streamflow anomaly and PC-1 with a correlation of 0.43 and 0.37 for WS I and WS II, respectively. The correlation between streamflow anomaly and PC-2 is 0.08 and 0.02, respectively, for WS I and WS II, whereas the first two PCs are uncorrelated by nature. Correlations between higher order PCs are very close to zero.

[29] The scatter plots show that the pairs of variables have different bivariate dependence structures that cannot be modeled by the symmetric copulas. They are not included in the paper for the sake of brevity. The Kendall's τ values of the various pairs of these variables are listed in Table 5. Given this nature of dependence, a class of asymmetric Archimedean copulas were adopted wherein the streamflow anomaly and PC-1 was coupled by a copula C_2 , and this structure was then associated with PC-2 by another copula C_1 .

[30] From the streamflow anomaly values and the two PCs, their rank-based nonparametric marginal probabilities u_1 , u_2 and u_3 , respectively, were calculated for modeling the copula function. The properties of asymmetric Archimedean copulas are mentioned in section 3.2. However, as the study data set did not conform to the requirement of the M12 nested 3-copula family that $\tau_{12}, \tau_{23}, \tau_{13} \in [0.333, 1]$ (Table 2), this copula family was rejected for both study watersheds.

4.3. Parameter Estimation

[31] The parameters of the nested copula were estimated using the CML method [Genest et al., 1995; Kojadinovic and Yan, 2011]. The parameter values must conform to the range specified for each class of copula. The condition that

Table 4. Principal Components and the Explained Variance

Principal Component	Eigenvalues		Explained Variance (%)	
	WS I	WS II	WS I	WS II
1	4158.98	3535.89	80.52	81.17
2	943.94	773.00	18.27	17.75
3	33.95	29.50	0.66	0.68
4	22.83	11.54	0.44	0.26
5	2.96	3.13	0.06	0.07
6	2.19	2.59	0.04	0.06
7	0.51	0.57	0.01	0.01

the more nested variables have a stronger degree of dependence among them, that is, $\theta_2 \geq \theta_1 \in [0, \infty)$ was satisfied by the M3 and M4 families, and the condition $\theta_2 \geq \theta_1 \in [1, \infty)$ was satisfied by the M5 and M6 families of copula. The estimated values of the copula parameters and the maximum likelihood value obtained for each of the copula families are listed in Table 5.

4.4. Goodness-of-Fit Tests

[32] From the copula families evaluated in the study, the best copula was selected using popular goodness-of-fit measures. The probability distribution function of different copula families and the empirical copula are plotted in Figure 2. The performance statistics computed for the probability distribution function between the empirical and estimated copulas are given in Table 6. The M6 copula family was found to have lowest value of S_n and T_n statistics calculated for WS I. The goodness-of-fit for this copula family is also evident from Figure 2a. The lowest value of S_n and T_n was obtained for M6 copula in the case of WS II. It also provided the best distribution fit among all copula models in Figure 2b. Plots in Figure 3 show the performance of only the M6 copula for different months, suggesting that the dependence structure of the first two principal components of anomalies of the hydroclimatic variables and streamflow anomalies could be modeled by the same M6 copula family for all months in both study watersheds.

4.5. Streamflow Prediction Using Copula

[33] Given u_2 and u_3 (the rank-based values of PCs extracted from the predictors), the probability distribution of u_1 (derived from streamflow anomalies) was generated using the M6 copula model (Table 2). The streamflow anomalies corresponding to different quantiles were calculated from this CDF. The rank-based nonparametric marginal probabilities at 0.025, 0.5, and 0.975 quantiles were

Table 5. Kendall's τ and Parameter θ for Different Copulas

						Maximum Likelihood Estimate						
τ_{12}		τ_{13}		τ_{23}		Nested Copula Family	θ_1		θ_2		Maximum Likelihood Value	
WS I	WS II	WS I	WS II	WS I	WS II		WS I	WS II	WS I	WS II	WS I	WS II
0.34	0.28	0	0.02	0	0	M3	0.005	0.185	3.35	2.71	55.71	47.12
						M4	0.005	0.001	0.69	0.63	45.34	48.93
						M5	1.08	1.10	1.57	1.35	44.73	28.33
						M6	1.04	1.05	1.45	1.31	56.17	41.01

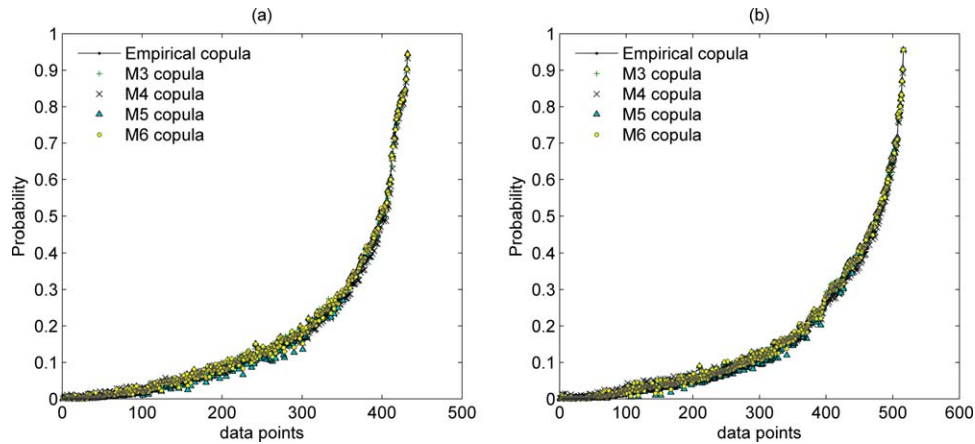


Figure 2. Comparison plots of probability distributions of different copula families used in (a) WS I and (b) WS II.

calculated and transformed into the streamflow anomaly values; subsequently, the estimates of streamflows for the next month were obtained. Streamflows simulated for the model development period were compared with the observed flows for evaluating model performance.

[34] The model developed for WS I was tested for the period January 1994 to December 2010, while model testing was carried out for the period 1991–2010 for WS II. The PCA coefficients obtained for predictors during model development period were used to obtain the PCs for the testing period as well. The predicted streamflow values for the model development and testing periods are compared with corresponding observed flows in Figures 4a, 4b, 5a,

and 5b for the two watersheds. The uncertainty in the predictions is quantified by the plot of interquantile range of predicted streamflows. Most of the observed flows lie within the predicted range during the model development periods in WS I. Typically, low flows in the late 1960s and 1970s are in close agreement with the expected values of streamflows obtained from the model (Figure 4a). The low flows during the testing period, especially in the 1990s, match well with the expected values in Figure 4b. However, this is not the case with high flows in WS I during both training and testing periods, where 1 month lead forecasts underestimate the observed peaks. In WS II, the recorded flows fall within the range of probabilistic predictions offered by the developed model. In Figure 5a, the predicted low flows in the 1950s, 1960s, and 1980s conform to observations. During the testing period also, the model performed well with low flow predictions (Figure 5b). The peak flows for both training and testing periods were typically underestimated perhaps because of the small numbers of training samples in this range. Additionally, the box plots for model development and testing periods in WS I and WS II in Figures 4c and 5c, respectively, indicate that though the model performance is not satisfactory in the

Table 6. Goodness-of-Fit Test Statistics for Different Copulas

Nested Copula Family	S_n		T_n	
	WS I	WS II	WS I	WS II
M3	0.064	0.061	0.038	0.038
M4	0.105	0.116	0.051	0.044
M5	0.046	0.053	0.040	0.044
M6	0.043	0.043	0.038	0.041

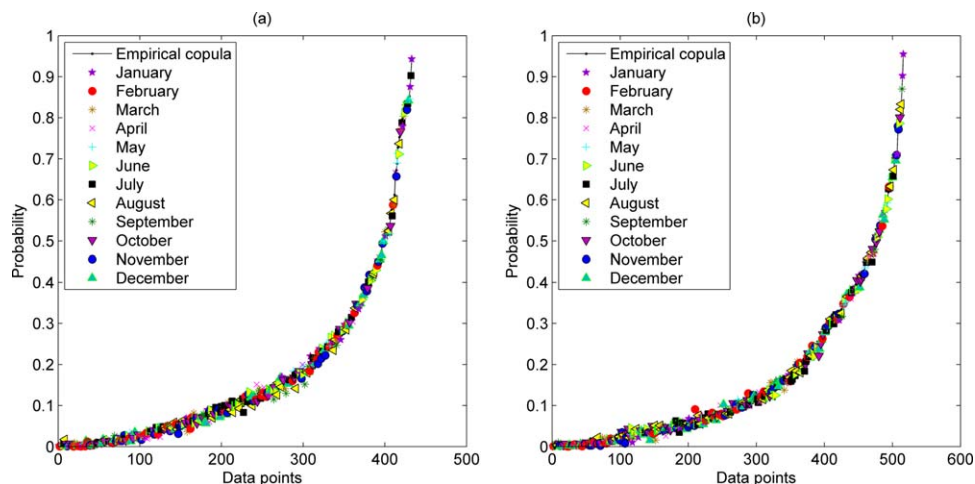


Figure 3. Plots showing M6 copula fit for each month in (a) WS I and (b) WS II.

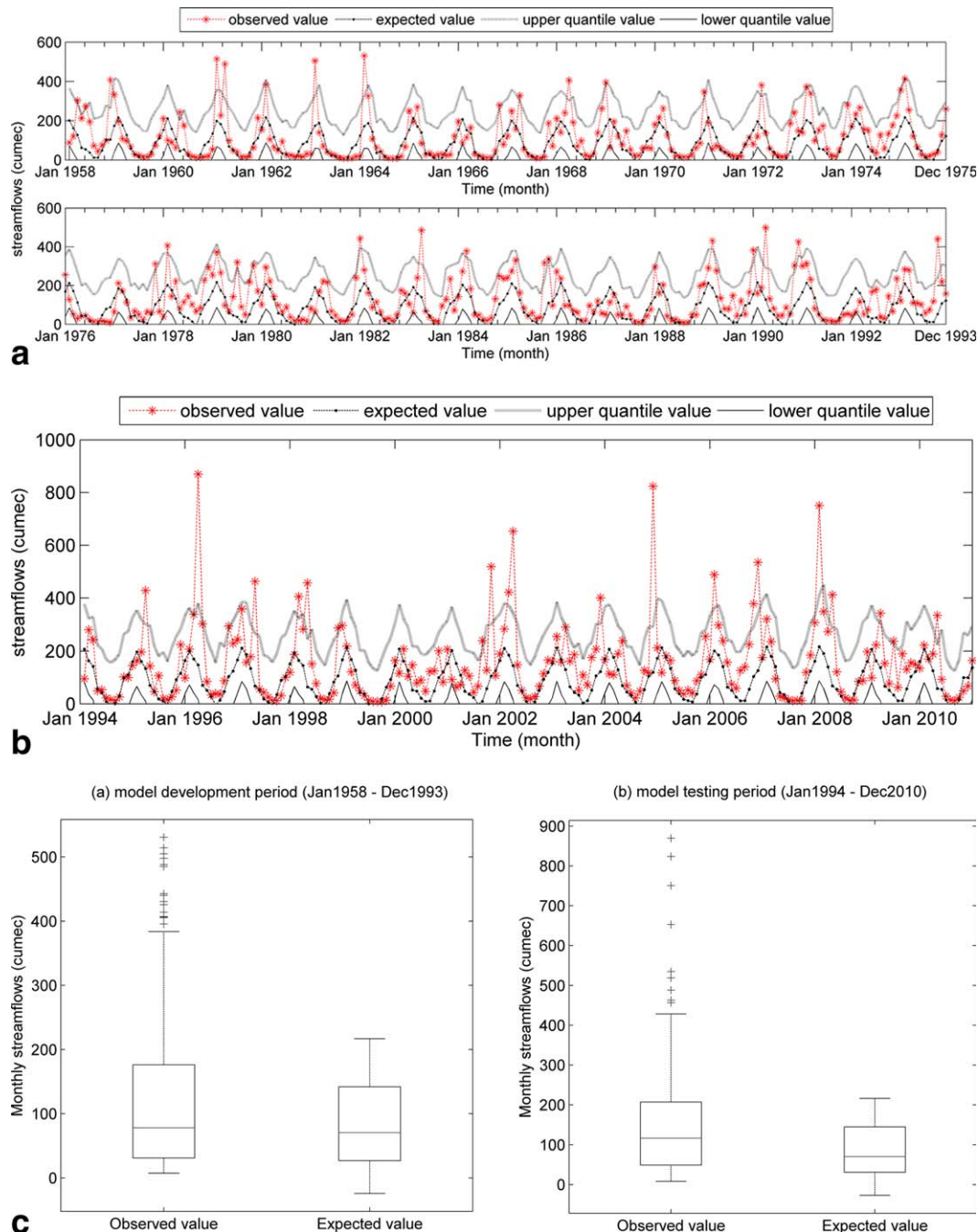


Figure 4. (a, b) Comparison plots of observed and predicted streamflows in WS I during (a) model development period and (b) model testing period (lower and upper quantile curves correspond to 0.025 and 0.975 quantiles respectively), and (c) box plots for observed and predicted (expected) values of monthly streamflows during model development and testing periods in WS I. On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually with a plus symbol.

case of high flows, low flows are estimated well. Overall, the predictive capability of the model was found to favor low flow conditions, prompting us to explore the development of droughts over the two study watersheds. The coefficients of determination (R^2) values obtained were 0.64 and 0.53, respectively, for the model development and testing periods in WS I, and 0.58 and 0.50, respectively, for WS II. Comparisons with state-of-the-art statistical models

[Tripathi and Govindaraju, 2008] using the same set of predictors for streamflow showed similar performance, but the results are not reported here for brevity.

4.6. Drought Analysis

[35] The results of the drought analysis carried out for the model development period (January 1948 to December 1993) for WS I are shown in Figure 6a. There were few

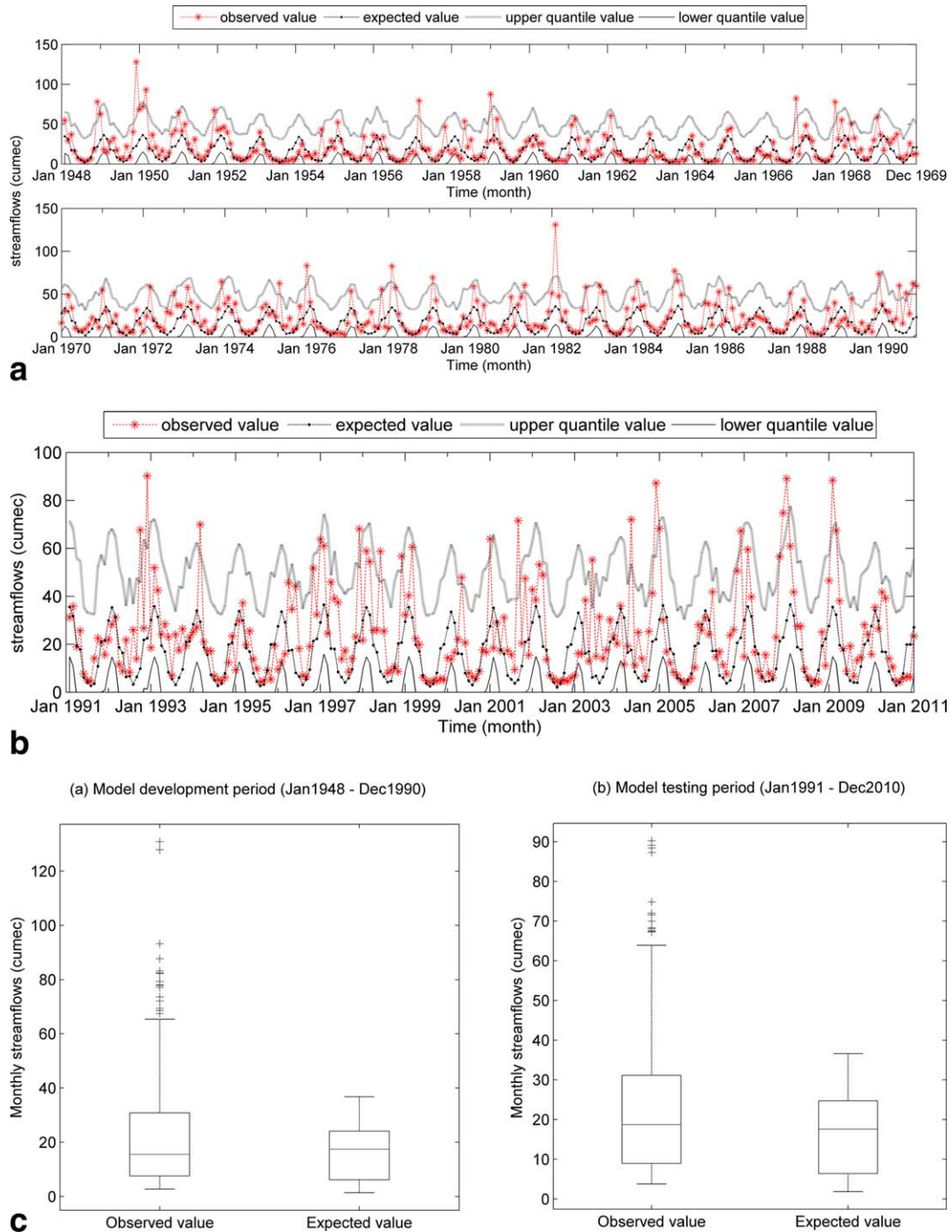


Figure 5. (a, b) Comparison plots of observed and predicted streamflows in WS II during (a) model development period and (b) model testing period (lower and upper quantile curves correspond to 0.025 and 0.975 quantiles respectively), and (c) box plots for observed and predicted (expected) values of monthly streamflows during model development and testing periods in WS II. On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually with a plus symbol.

occurrences of D3 and D4 classes of droughts during the model development periods, and mild (D0) and moderate (D1) droughts prevailed in most of the drought months. The drought index values obtained from the expected streamflows provided good forecasts of dry as well as wet

conditions. The drought analysis was then carried out for the testing period and compared with the observed conditions. Few occurrences of D2 and D1 classes of droughts marked the testing period. Wet conditions dominated during this period, with most of them being underestimated by

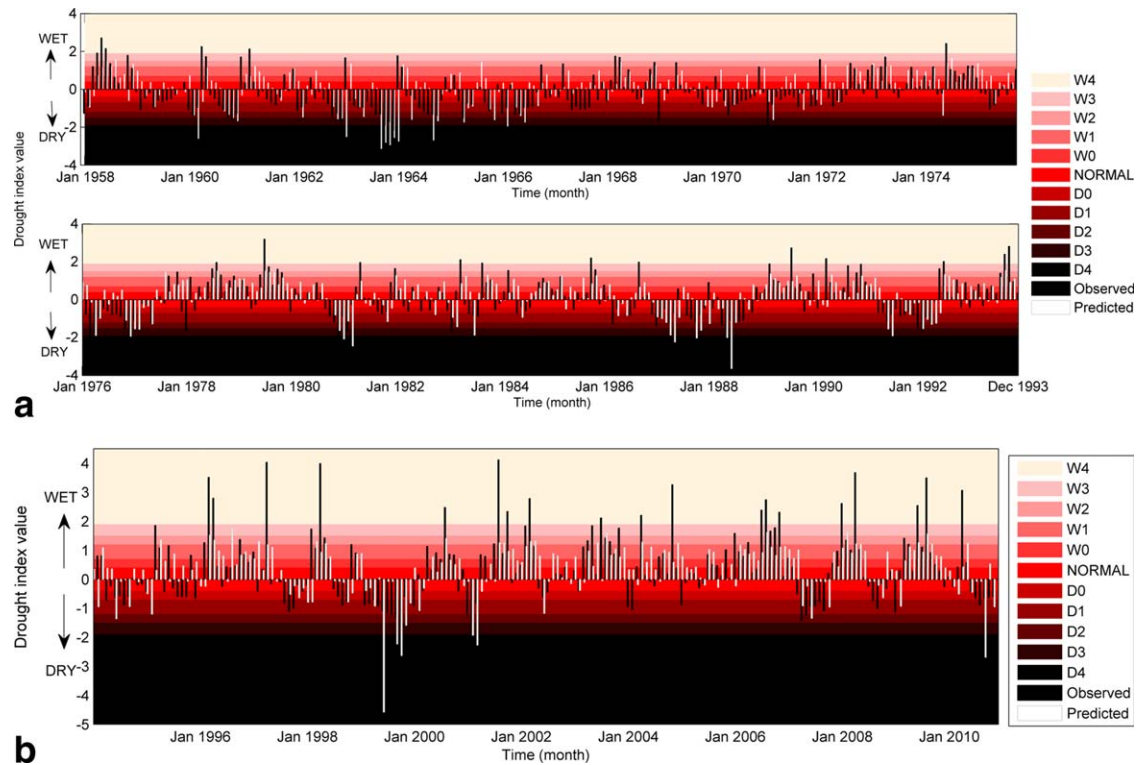


Figure 6. (a) Drought index values during the model development period in WS I and (b) drought index values during the model testing period in WS I.

the model (Figure 6b). The plots for drought indices calculated for WS II in Figures 7a and 7b also indicate that different drought categories were better predicted than the wet

categories. The sequences of drought months in different subperiods during the entire model development and testing periods were also well predicted.

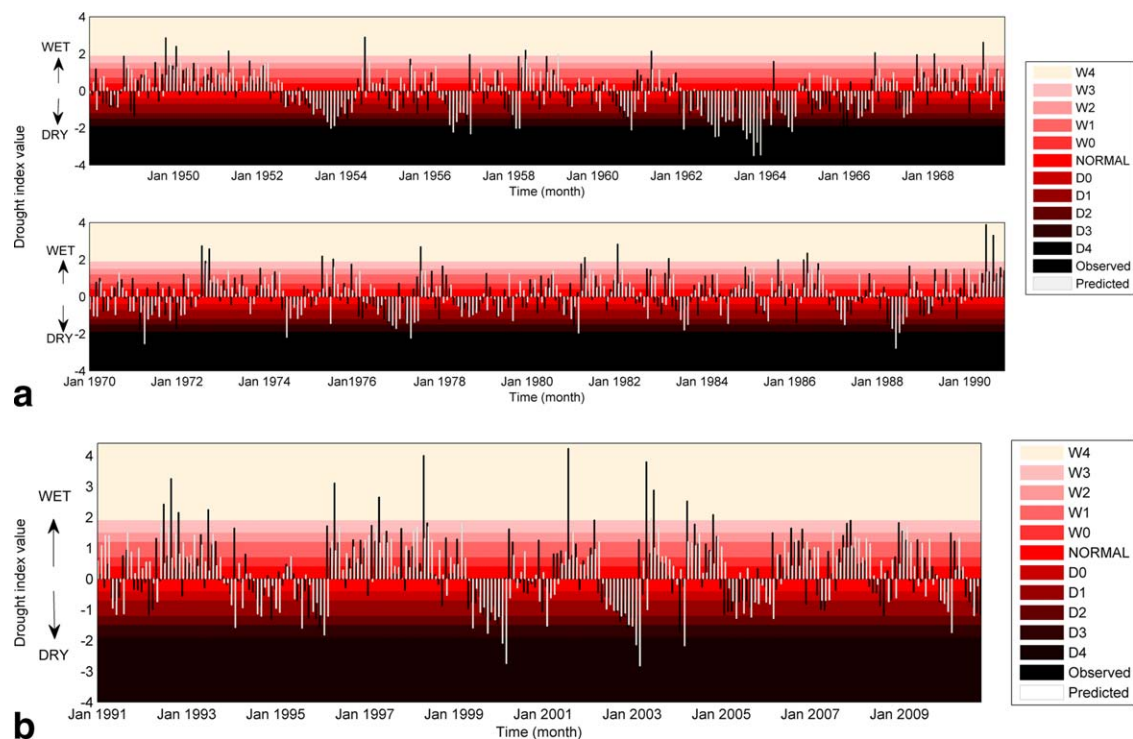


Figure 7. (a) Drought index values during the model development period in WS II and (b) drought index values during the model testing period in WS II.

Table 7. Contingency Table and Degree of Association Between Observed and Predicted Drought Categories for WS I (Top) and WS II (Bottom)

Predicted Category	Model Development Period (1958–1993)			Model Testing Period (1994–2010)		
	Observed Category			Observed Category		
	Dry	Normal	Wet	Dry	Normal	Wet
Dry	71	18	11	18	9	5
Normal	78	38	39	30	23	16
Wet	31	62	84	8	30	65
Q		85.75			52.92	
DOF		4			4	
p value		<0.0001			<0.0001	
C		0.407			0.454	
C_{\max}		0.817			0.817	
C/C_{\max}		0.498			0.556	

Predicted Category	Model Development Period (1948–1990)			Model Testing Period (1991–2010)		
	Observed Category			Observed Category		
	Dry	Normal	Wet	Dry	Normal	Wet
Dry	107	24	23	40	14	15
Normal	63	50	45	24	19	13
Wet	33	80	91	12	39	64
Q		105.54			54	
DOF		4			4	
p value		<0.0001			<0.0001	
C		0.412			0.429	
C_{\max}		0.817			0.817	
C/C_{\max}		0.505			0.525	

[36] Apart from visual inspection, the model performance for multiple category classification of streamflows was assessed by computing the contingency coefficient C , proposed by *Pearson* [1904]. This coefficient is a measure of degree of association between multiple categories in a contingency table classifying N samples [*Gibbons and Chakraborti*, 2011] and mathematically expressed as:

$$C = \sqrt{\left(\frac{Q}{Q+N}\right)} \quad (12)$$

where Q is a statistic that tests the null hypothesis that there is no association between observed and predicted categories. It is expressed as:

$$Q = \sum_{i=1}^r \sum_{j=1}^k \frac{(NX_{ij} - X_i Y_j)^2}{NX_i Y_j} \quad (13)$$

where r and k are the number of categories, X_{ij} is the number of cases falling in i th observed and j th predicted category, $X_i = \sum_{j=1}^k X_{ij}$ and $Y_j = \sum_{i=1}^r X_{ij}$.

[37] The statistic Q approximately follows chi-square distribution with degrees of freedom (dof) equal to $(r-1)(k-1)$. Thus, the null hypothesis (no association) can be rejected if the p value is very low. Higher values of C correspond to better association. The value of C cannot exceed 1 theoretically and has an upper bound of $C_{\max} (= \sqrt{(t-1)/t})$ where $t = \min(r, k)$ [*Gibbons*

and *Chakraborti*, 2011]. The ratio C/C_{\max} is often used as a measure of degree of association.

[38] In order to ensure sufficient data for robust statistics, a contingency table with three different categories, dry, normal, and wet, was prepared. The extreme categories were merged to ensure that the observations and predictions are available sufficiently in all categories. These contingency tables are shown in Table 7 for WS I and WS II, respectively. Thus, both r and k are 3, and dof is 4. The statistic Q , contingency coefficient C , and the measure of degree of association C/C_{\max} are shown at the end of Table 7. The low p values for the statistic Q indicate that the null hypothesis of no association between observed and predicted categories should be rejected. The degree of association was found to be reasonable for both the watersheds during model development as well as testing periods.

4.7. Analysing the Drought Triggers

[39] Using the modeled asymmetric copula dependence function, the conditions that trigger hydrological droughts or extremes in the watershed were examined. The triggers for various streamflow conditions were generated using the conditional copula. The procedure is illustrated as follows. Given a certain streamflow anomaly quantile α , let y_1^α and y_2^α correspond to the first and second PCs conditioned on the streamflow anomaly value. The quantities y_1^α and y_2^α are obtained from the M6 copula for the particular watershed. Since these two PCs explain over 98% of the total variation, the other principal components remain unaffected by the choice of the streamflow quantile. Our goal is to find the expected values of the precursor variables $x_i^\alpha, i =$

Table 8. Expected Principal Component Values for Various Quantiles of Streamflow

Streamflow Anomaly Quantile	Streamflow Anomaly (cumecs)		Expected PC-1 Value		Expected PC-2 Value	
	WS I	WS II	WS I	WS II	WS I	WS II
0.01	-172.84	-27.65	-49.10	-34.41	-3.79	-5.33
0.1	-99.47	-17.27	-34.31	-23.69	-3.25	-4.86
0.2	-63.19	-10.62	-26.54	-17.36	-2.44	-4.03
0.4	-27.13	-4.86	-15.73	-6.11	-1.97	-3.26
0.5	-16.94	-3.26	-8.79	-1.98	-1.64	-3.10
0.6	-5.71	-1.20	-1.63	2.98	-0.96	-1.79
0.8	58.38	8.79	25.20	18.02	0.57	-0.86
0.9	123.30	20.38	43.28	31.62	1.08	0.54
0.99	310.66	53.54	121.66	80.46	4.20	5.56

1, 2, ..., 7 that would correspond to this particular streamflow quantile. If a_{ij} are the PCA coefficients for the data set, then the following equation provides the conditional expectation of the precursor variables:

$$[A]\{x^\alpha\} = \{y^\alpha\} \quad (14)$$

where a_{ij} is the ij th element of the matrix $[A]$, y_1^α and y_2^α are computed from the M6 copula, and y_j , $j = 3, 4, \dots, 7$ are simply the expected values of the principal components (≈ 0).

[40] The expected values of PC-1 and PC-2 conditioned on various streamflow anomaly quantiles (corresponding to different α values) are shown in Table 8 for both watersheds.

[41] The expected anomaly values of all the predictor variables for different values corresponding to different streamflow anomalies are shown in Table 9. Low flows correspond to smaller values of soil moisture, temperature, precipitation, evaporation, and runoff of the previous month in both watersheds. Sea level pressure anomaly varied inversely with the streamflow anomaly for WS I and

WS II suggesting that increase in sea-level pressure from the long-term mean can enhance the chances of droughts in the regions. Increase in wind speed was found to trigger droughts in WS I, in contrast to the trend observed in the case of WS II. The dissimilar trends in some variables suggest that drought triggers are likely to be specific to each watershed.

[42] The conditional expectations of anomalies of different precursors corresponding to different streamflow quantiles (Table 9) were utilized to develop potential triggers for each drought category. The long-term monthly means of hydroclimatic variables were added to their expected anomaly values to carry out this analysis. The resulting precursor values were then associated with the 1 month lead drought index values. From the expected streamflow anomaly, streamflows for each month were computed and corresponding drought indices were calculated. The trigger analysis is limited to low flow conditions corresponding to droughts reflecting the better model performance for flows in this range. The plots in Figures 8a and 8b show the expected precursor range in each month obtained for different drought classes for WS I and WS II, respectively. If the values of the hydroclimatic variables fall within the suggested range for any class of drought, then that drought would likely occur in the succeeding month. For WS I, soil moisture, precipitation, and runoff are able to offer a range of predictor values for different drought categories as shown in Figure 8a. Some months (May to July) do not show any range of potential predictor values for certain drought classes, implying the likelihood of such droughts being very low in those periods in WS I. While soil moisture, precipitation, and runoff show some variability with drought classes in WS II, the other variables stay within a very tight band for any given month (Figure 8b). Thus, only these three variables are capable of resolving amongst different drought classes for the study watersheds. Low variability is manifested in the expected anomaly values of temperature, evaporation, sea-level pressure, and wind speed in Table 9.

Table 9. Conditional Expectations (in Terms of Anomalies of Hydro-Climatic Variables) Associated With Streamflow Anomaly Values

Hydro-Climatic Triggers in Terms of Expected Values of Anomalies							
Expected Streamflow Anomaly (cumecs)	Soil Moisture Anomaly (mm)	Temperature Anomaly ($^{\circ}\text{C}$)	Precipitation Anomaly (mm)	Evaporation Anomaly (mm)	Sea-Level Pressure Anomaly (mbar)	Wind Speed Anomaly (m/s)	Runoff Anomaly (mm)
<i>WS I</i>							
-172.84	-37.33	-0.0032	-31.28	-1.26	0.21	0.0017	-7.18
-99.47	-25.74	-0.0072	-22.35	-0.86	0.15	0.0005	-5.02
-63.19	-19.94	-0.0050	-17.23	-0.67	0.12	0.0005	-3.88
-27.13	-11.52	-0.0072	-10.63	-0.38	0.07	-0.0003	-2.31
-16.94	-6.12	-0.0085	-6.38	-0.19	0.05	-0.0008	-1.30
-5.71	-0.74	-0.0070	-1.72	-0.01	0.01	-0.0008	-0.25
58.38	19.96	-0.0096	14.94	0.69	-0.10	-0.0024	3.66
123.30	34.22	-0.0157	25.74	1.18	-0.17	-0.0040	6.29
310.66	95.51	-0.0345	73.30	3.29	-0.49	-0.0100	17.70
<i>WS II</i>							
-27.65	-27.61	-0.127	-20.92	-0.94	0.126	-0.046	-3.42
-17.27	-18.44	-0.102	-15.45	-0.62	0.097	-0.036	-2.37
-10.62	-13.29	-0.080	-11.73	-0.44	0.075	-0.028	-1.74
-4.86	-3.80	-0.050	-5.75	-0.17	0.042	-0.016	-0.64
-3.26	-0.27	-0.041	-3.66	0.01	0.031	-0.012	-0.24
-1.20	3.47	-0.016	-0.17	0.13	0.008	-0.004	0.26
8.79	16.20	0.023	7.75	0.57	-0.036	0.012	1.74
20.38	27.44	0.065	15.39	0.95	-0.079	0.027	3.08
53.54	67.84	0.215	42.84	2.34	-0.238	0.082	7.89

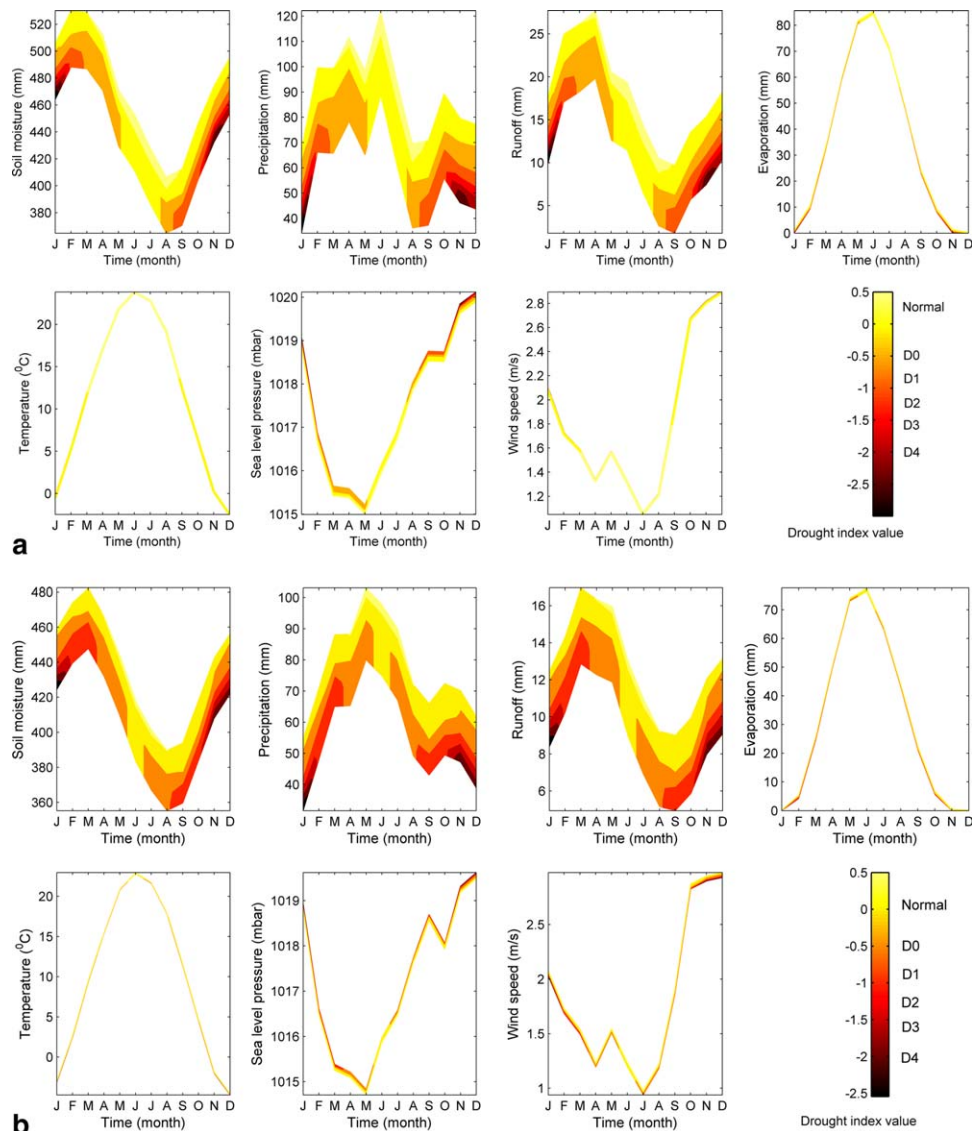


Figure 8. Contour plots showing expected ranges of different hydroclimatic variables as precursors to droughts in (a) WS I and (b) WS II.

[43] The precursor ranges developed in this manner were validated by means of scatter plots between the observed and modeled values of variables over the model development and testing periods (Figures 9a and 9b) for all classes of droughts. These scatter plots demonstrate good agreement between the observed and modeled triggers in both watersheds. The scatter is less in the case of soil moisture, precipitation, runoff, evaporation, and temperature in both watersheds. Among the predictors, wind speed shows the most scatter making it the least reliable precursor for both watersheds. The modeled triggers for soil moisture, precipitation, and runoff values are underpredicted compared to observations during calibration as well as validation. Additionally, correlation values for all the trigger variables were calculated and tabulated in Table 10. High correlations in some predictors (for example, temperature and evaporation in WS I and WS II), however, were not useful as they were found incapable of resolving among the different drought categories.

[44] The results indicate that drought trigger information retrieved in this manner has potential for applications in

hydrologic drought preparedness. Even though individual variables show scatter, if multiple variables fall close to their trigger values, the confidence in their effectiveness as hydrologic drought triggers will improve. Hence, the combined behavior of predictor variables needs to be considered when estimating potential drought triggers.

5. Summary and Conclusions

[45] This study provides a novel method for developing drought triggers by combining the strengths of PCA for dimensionality reduction and copulas for modeling the joint dependence between variables. The first two PCs were found capable of explaining the variability in the anomaly set of predictor variables for both study watersheds. The joint dependence of the streamflow anomaly and the two principal components was modeled by a scale-free association using a suitable asymmetric 3-copula selected based on goodness-of-fit statistics. The developed model was first tested for forecasting streamflows in two study watersheds.

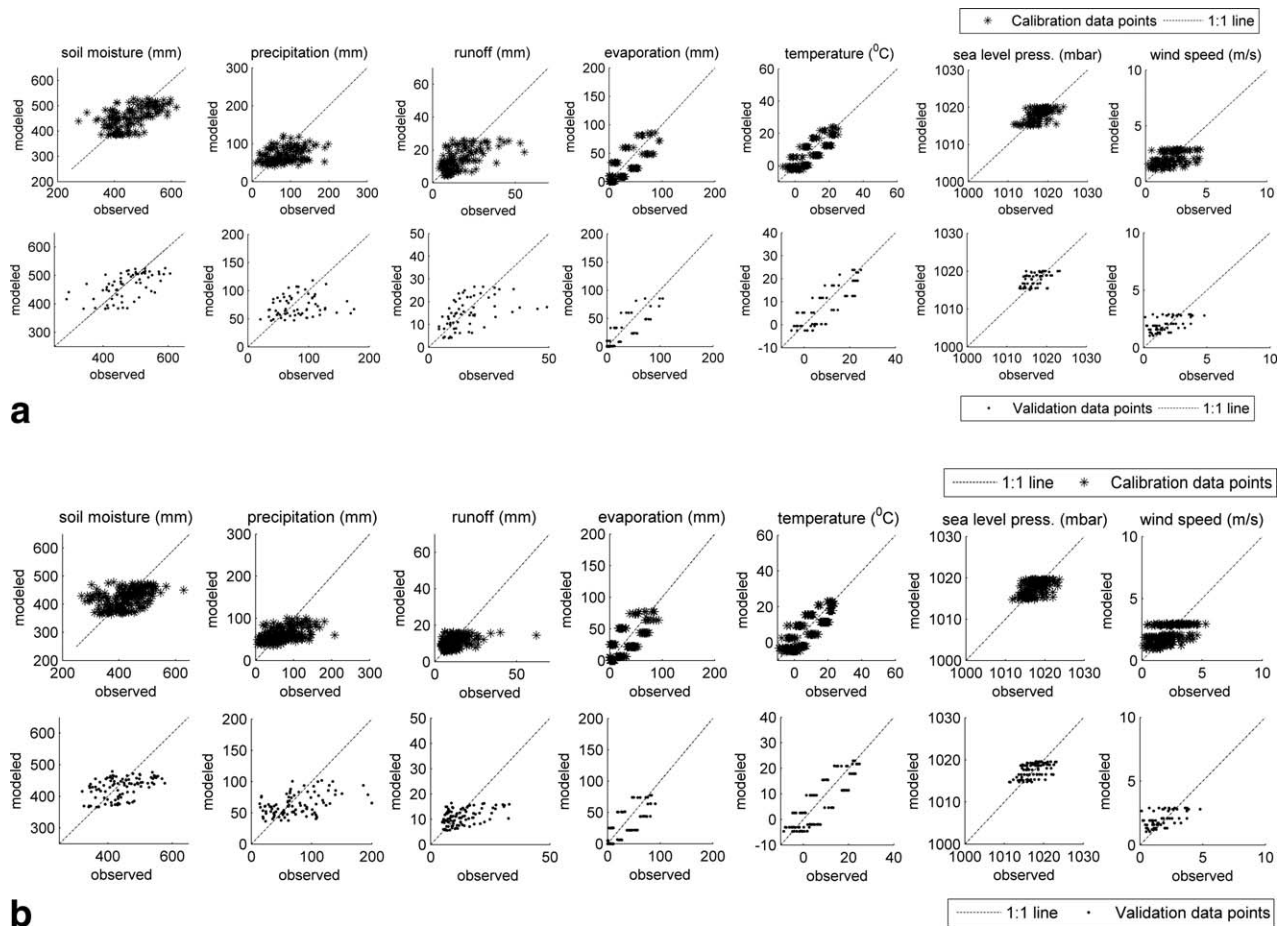


Figure 9. Scatter plots of different hydroclimatic precursors (modeled versus observed) for model development and testing periods in (a) WS I and (b) WS II.

The study focused on 1 month lead predictions because correlations between the principal components and streamflow anomaly diminished rapidly beyond a lag of 1 month. Underprediction of peak flows was observed in the results of both watersheds, but low streamflows were reasonably predicted allowing hydrologic drought studies. Drought index values based on standardized flows were computed to identify the occurrences of droughts during the model development and testing periods in the two study regions

[46] The conditional dependence of the principal components PC-1 and PC-2 on streamflow anomaly was used to determine the drought triggers in the two watersheds. The

precursors to droughts were expressed in terms of the anomaly values of the climatic variables. Negative anomalies of soil moisture, precipitation, evaporation, temperature, and runoff, and increased sea-level pressure and wind speeds were obtained as potential drought triggers for WS I. Similarly, increased sea level pressure conditions and reduced soil moisture, precipitation, evaporation, temperature, runoff, and wind speeds from their respective long-term means led to drought conditions in WS II.

[47] Further, the patterns of various hydroclimatic variables as potential precursors to different categories of droughts were examined for the two watersheds. The ranges of predictor values that led to different drought conditions were estimated from the expected precursor values for low streamflow quantiles. The trigger analysis results were validated by comparing the observed hydroclimatic variables with their expected trigger values for the model development and testing periods. The correlation values computed indicated that the analysis could yield reliable information on the pattern of drought triggers for both the watersheds.

[48] The following conclusions are derived from this study:

[49] 1. Drought triggers are likely to be specific to watersheds. Even though the two study watersheds are located in the same part of the world and have similar land use distribution, local conditions influence streamflows especially at monthly time scales.

Table 10. Correlation Values Between Observed and Modeled Drought Precursors

HydroClimatic Precursor	WS I		WS II	
	Calibration	Validation	Calibration	Validation
Soil moisture	0.57	0.58	0.41	0.44
Precipitation	0.35	0.29	0.48	0.44
Runoff	0.59	0.47	0.41	0.45
Evaporation	0.80	0.83	0.80	0.82
Temperature	0.81	0.82	0.82	0.85
Sea level pressure	0.58	0.43	0.50	0.52
Wind speed	0.45	0.56	0.48	0.52

[50] 2. Using copulas, conditional expectations of first two PCs based on different quantiles of streamflow anomalies provide a method for estimating drought triggers. Among all the precursors, soil moisture, precipitation, and runoff showed the greatest potential for assessing different classes of droughts for both watersheds. The other variables, despite showing strong seasonal trends, demonstrated little capability for resolving the different drought classes.

[51] 3. Validation results for triggers over all drought classes show results with different degrees of variability. Even with the scatter present for single (individual) variables, if triggers for multiple variables fall within expected ranges, the confidence in the trigger would improve. Hence, it is recommended that precursors for droughts be examined in combination by using multiple input variables.

[52] Even though the results and conclusions are specific to study watersheds, the method shows promise for application to different watersheds. An important limitation is that the level of dimensionality reduction that can be achieved in different watersheds cannot be known a priori. If multiple predictors were to be important, the model for constructing the joint distribution would be too complex for practical purposes except in limited cases modeled using Gaussian copulas. Data limitations also continue to be a serious challenge for many hydrologic studies. Large amount of data need to be used for capturing the trigger behaviors in drought studies. The model development and testing periods were short in this study, and the methodology performs reasonably well even for the small record lengths available here. Future efforts employing more hydroclimatic variables and different watersheds will help develop better understanding of trigger mechanisms for droughts.

[53] **Acknowledgments.** The first author, as a BOYSCAST Fellow, acknowledges the support of Department of Science and Technology (DST), Govt. of India, for offering the fellowship. He also wishes to acknowledge the support of School of Civil Engineering, Purdue University, USA where this study was undertaken. Studies of the second and third authors were supported in part by the National Science Foundation under Grants ACI 0753116 and AGS 1025430, and by USDA NIFA award number 2011-67019-21122. This support is gratefully acknowledged. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation or the USDA.

References

- Anmala, J., B. Zhang, and R. S. Govindaraju (2000), Comparison of ANNs and empirical approaches for predicting watershed runoff, *J. Water Resour. Plann. Manage.*, 126(3), 156–166, doi:10.1061/(ASCE)0733-9496(2000)126:3(156).
- Charusombat, U., and D. Niyogi (2011), A hydroclimatological assessment of regional drought vulnerability: A case study of Indiana droughts, *Earth Interact.*, 15(26), 1–65, doi:10.1175/2011EI343.1.
- Deheuvels, P. (1981), *A Non Parametric Test for Independence*, Publ. de l'Inst. de Stat. de l'Univ. de Paris, vol. 26, pp. 29–50, Inst. de Stat. Univ. de Paris, Paris, France.
- Dracup, J. A., K. S. Lee, and E. G. Paulson (1980), On the definition of droughts, *Water Resour. Res.*, 16(2), 297–302, doi:10.1029/WR016i002p00297.
- Edwards, D. C., and T. B. McKee (1997), Characteristics of 20th century drought in the United States at multiple time scales, *Climatol. Rep.*, 97-2, Dep. of Atmos. Sci., Colo. State Univ., Fort Collins.
- Embrechts, P., F. Lindskog, and A. McNeil (2003), Modelling dependence with copulas and applications to risk management, in *Handbook of Heavy Tailed Distributions in Finance*, pp. 329–384, Elsevier, New York.
- Fan, Y., and H. van den Dool (2004), Climate Prediction Center global monthly soil moisture data set at 0.5° resolution for 1948 to present, *J. Geophys. Res.*, 109, D10102, doi:10.1029/2003JD004345.
- Favre, A.-C., S. El Adlouni, L. Perreault, N. Thiémond, and B. Bobée (2004), Multivariate hydrological frequency analysis using copulas, *Water Resour. Res.*, 40, W01101, doi:10.1029/2003WR002456.
- Federal Emergency Management Agency (1995), *National Mitigation Strategy; Partnerships for Building Safer Communities*, 45 pp., Mitigation Dir., Washington, D. C.
- Genest, C., and A. Favre (2007), Everything you always wanted to know about copula modeling but were afraid to ask, *J. Hydrol. Eng.*, 12(4), 347–368, doi:10.1061/(ASCE)1084-0699(2007)12:4(347).
- Genest, C., K. Ghoudi, and L.-P. Rivest (1995), A semiparametric estimation procedure of dependence parameters in multivariate families of distribution, *Biometrika*, 82(3), 543–552.
- Genest, C., B. Remillard, and D. Beaudoin (2009), Goodness-of-fit tests for copulas: A review and a power study, *Insur. Math. Econ.*, 44(2), 199–213, doi:10.1016/j.insmatheco.2007.10.005.
- Gibbons J. D., and S. Chakraborti (2011), *Nonparametric Statistical Inference*, 5th ed., Chapman and Hall, Boca Raton, Fla.
- Grimaldi, S., and F. Serinaldi (2006), Asymmetric copula in multivariate flood frequency analysis, *Adv. Water Resour.*, 29(8), 1155–1167, doi:10.1016/j.advwatres.2005.09.005.
- Huang, J., H. van den Dool, and K. P. Georgakakos (1996), Analysis of model-calculated soil moisture over the United States (1931–93) and application to long-range temperature forecasts, *J. Clim.*, 9(6), 1350–1362.
- Joe, H. (1997), *Multivariate Models and Dependence Concepts*, Chapman and Hall, London.
- Jolliffe, I. T. (1986), *Principal Component Analysis*, Springer, New York.
- Kalnay, E., et al. (1996), The NCEP/NCAR 40-year reanalysis project, *Bull. Am. Meteorol. Soc.*, 77(3), 437–471, doi:10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.
- Kao, S.-C., and R. S. Govindaraju (2008), Trivariate statistical analysis of extreme rainfall events via the Plackett family of copulas, *Water Resour. Res.*, 44, W02415, doi:10.1029/2007WR006261.
- Kao, S.-C., and R. S. Govindaraju (2010a), A copula-based joint deficit index for droughts, *J. Hydrol.*, 380(1–2), 121–134, doi:10.1016/j.jhydrol.2009.10.029.
- Kao, S.-C., and R. S. Govindaraju (2010b), Reply to Comment by T. P. Hutchinson on “Trivariate statistical analysis of extreme rainfall events via Plackett family of copulas,” *Water Resour. Res.*, 46, W04802, doi:10.1029/2009WR008774.
- 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.
- Kojadinovic, I., and J. Yan (2011), A goodness-of-fit test for multivariate multiparameter copulas based on multiplier central limit theorems, *Stat. Comput.*, 21(1), 17–30, doi:10.1007/s11222-009-9142-y.
- Lins, H. F. (1985), Interannual streamflow variability in the United States based on principal components, *Water Resour. Res.*, 21(5), 691–701, doi:10.1029/WR021i005p00691.
- Madadgar, S., and H. Moradkhani (2013), Drought analysis under climate change using copula, *J. Hydrol. Eng.*, 746–759, doi:10.1061/(ASCE)HE.1943-5584.0000532.
- Maity, R., and D. Nagesh Kumar (2008a), Probabilistic prediction of hydroclimatic variables with nonparametric quantification of uncertainty, *J. Geophys. Res.*, 113, D14105, doi:10.1029/2008JD009856.
- Maity, R., and D. Nagesh Kumar (2008b), Basin-scale stream-flow forecasting using the information of large-scale atmospheric circulation phenomena, *Hydrol. Processes*, 22(5), 643–650, doi:10.1002/hyp.6630.
- Mallya, G., S. Tripathi, S. Kirshner, and R. S. Govindaraju (2013), Probabilistic assessment of drought characteristics using a hidden Markov model, *J. Hydrol. Eng.*, 18(7), 834–845, doi:10.1061/(ASCE)HE.1943-5584.256.0000699.
- Maurer, E. P., D. P. Lettenmaier, and N. J. Mantua (2004), Variability and potential sources of predictability of North American runoff, *Water Resour. Res.*, 40, W09306, doi:10.1029/2003WR002789.
- McHugh, M. J., and J. C. Rogers (2001), North Atlantic oscillation influence on precipitation variability around the Southeast African convergence zone, *J. Clim.*, 14(17), 3631–3642, doi:10.1175/1520-0442(2001)014<3631:NAOIOP>2.0.CO;2.
- McKay, G. A., R. B. Godwin, and J. Maybank (1989), Drought and hydrological drought research in Canada: An evaluation of the state of the art, *Can. Water Resour. J.*, 14(3), 71–84, doi:10.4296/cwrj1403071.
- 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, pp. 179–184, Am. Meteorol. Soc., Boston, Mass.
- Mishra, A. K., and V. P. Singh (2010), A review of drought concepts, *J. Hydrol.*, 39(1–2), 202–216.
- Nelsen, R. B. (2006), *An Introduction to Copulas*, Springer, New York.
- Pearson, K. (1904), *On the Theory of Contingency and its Relation to Association and Normal Correlation*, *Draper's Comp. Res. Mem. Biometric Ser. I*, Dulau and Co., London, U. K.
- Preisendorfer, R. W. (1988), *Principal Component Analysis in Meteorology and Oceanography*, Elsevier, New York.
- Ropelewski, C. F., and M. S. Halpert (1996), Quantifying Southern oscillation-precipitation relationships, *J. Clim.*, 9(5), 1043–1059, doi:10.1175/1520-0442(1996)009<1043:QSOPR>2.0.CO;2.
- Salas, J. D., C. Fu, and B. Rajagopalan (2011), Long-range forecasting of Colorado streamflows based on hydrologic, atmospheric, and oceanic data, *J. Hydrol. Eng.*, 16, 508–520, doi:10.1061/(ASCE)HE.1943-5584.0000343.
- Salvadori, G., and C. De Michele (2004), Frequency analysis via copulas: Theoretical aspects and applications to hydrological events, *Water Resour. Res.*, 40, W12511, doi:10.1029/2004WR003133.
- Shiau, J.-T., S. Feng, and S. Nadarajah (2007), Assessment of hydrological droughts for the Yellow River, China, using copulas, *Hydrol. Processes*, 21(16), 2157–2163, doi:10.1002/hyp.6400.
- Sklar, A. (1959), *Fonctions de répartition à n dimensions et leurs marges*, *Publ. de l'Inst. de Stat. de l'Univ. de Paris*, vol. 8, pp. 229–231, Inst. de Stat. Univ. de Paris, Paris, France.
- Srinivasan, R., and J. G. Arnold (1994), Integration of a basin-scale water quality model with GIS, *J. Am. Water Resour. Assoc.*, 30(3), 453–462, doi:10.1111/j.1752-1688.1994.tb03304.x.
- Steinemann, A. (2003), Drought indicators and triggers: A stochastic approach to evaluation, *J. Am. Water Resour. Assoc.*, 39(5), 1217–1233, doi:10.1111/j.1752-1688.2003.tb03704.x.
- Tripathi, S., and R. S. Govindaraju (2008), Engaging uncertainty in hydrologic data sets using principal component analysis: BaNPCA algorithm, *Water Resour. Res.*, 44, W10409, doi:10.1029/2007WR006692.
- Whelan, N. (2004), Sampling from Archimedean copulas, *Quant. Finan.*, 4(3), 339–352, doi:10.1088/1469-7688/4/3/009.
- Wong, G., M. Lambert, M. Leonard, and A. Metcalfe (2010), Drought analysis using trivariate copulas conditional on climatic states, *J. Hydrol. Eng.*, 15(2), 129–141, doi:10.1061/(ASCE)HE.1943-5584.0000169.
- Wu, C. L., K. W. Chau, and Y. S. Li (2009), Predicting monthly streamflow using data-driven models coupled with data-preprocessing techniques, *Water Resour. Res.*, 45, W08432, doi:10.1029/2007WR006737.
- Xu, C.-Y., and V. P. Singh (2004), Review on regional water resources assessment models under stationary and changing climate, *Water Resour. Manage.*, 18(6), 591–612, doi:10.1007/s11269-004-9130-0.
- Zhang, L., and V. Singh (2006), Bivariate flood frequency analysis using the copula method, *J. Hydrol. Eng.*, 11(2), 150–164, doi:10.1061/(ASCE)1084-0699(2006)11:2(150).