JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 113, D14105, doi:10.1029/2008JD009856, 2008

Probabilistic prediction of hydroclimatic variables with nonparametric quantification of uncertainty

Rajib Maity[1] and D. Nagesh Kumar[2]

Received 23 January 2008; revised 6 April 2008; accepted 8 May 2008; published 16 July 2008.

[1] A semiparametric, copula-based approach is proposed to capture the dependence between teleconnected hydroclimatic variables for the prediction of response variable using the information of climate precursors. The copulas have an excellent property to study the scale-free dependence structure while preserving such dependence during simulation. This property is utilized in the proposed approach. The usefulness of the proposed method can be recognized in three distinct aspects: (1) It captures the dependence pattern preserving scale-free or rank-based "measure of association" between the variables. (2) The proposed method is able to quantify the uncertainty associated with the relationship between teleconnected variables due to various factors; thus, the probabilistic predictions are available along with information of uncertainty. (3) Instead of parametric probability distribution, nonparametrically estimated probability densities for data sets can be handled by the proposed approach. Thus, the proposed method can be applied to capture the relationship between teleconnected hydroclimatic variables having some linear and/or nonlinear cause-effect relationship. The proposed method is illustrated by an example of the most discussed problem of Indian summer monsoon rainfall (ISMR) and two different large-scale climate precursors, namely, El Niño–Southern Oscillation (ENSO) and Equatorial Indian Ocean Oscillation (EQUINOO). The dependence between them is captured and investigated for its potential use to predict the monthly variation of ISMR using the proposed method. Predicted rainfall is shown to correspond well with the observed rainfall with a correlation coefficient of 0.81 for the summer monsoon months, i.e., June through September. Moreover, the uncertainty associated with the predicted values is also made available through boxplots. The method, being general, can be applied to similar analysis to assess the dependence between teleconnected hydroclimatic variables for other regions of the world and for different temporal scales such as seasonal.

Citation: Maity, R., and D. Nagesh Kumar (2008), Probabilistic prediction of hydroclimatic variables with nonparametric quantification of uncertainty, *J. Geophys. Res.*, *113*, D14105, doi:10.1029/2008JD009856.

1. Introduction

[2] The hydroclimatic teleconnection between hydrologic variables and large-scale atmospheric circulation is being investigated across the world [*Kahya and Dracup*, 1993; *Dracup and Kahya*, 1994; *Moss et al.*, 1994; *Eltahir*, 1996; *McKerchar et al.*, 1998; *Jain and Lall*, 2001; *Zubair*, 2003; *Chowdhury and Ward*, 2004; *Verdon et al.*, 2004]. It is recently inferred that temporal structure of hydrologic time series is significantly forced by large-scale atmospheric circulation patterns through hydroclimatic teleconnection [*Jain and Lall*, 2001; *Maity and Nagesh Kumar*, 2006b]. The historical relationship between the teleconnected hydroclimatic variables is the key issue of empirical modeling approach. However, having established the physical mechanism of their association, an advanced statistical technique is essential to suitably capture the dependence between the teleconnected hydroclimatic variables. The existing techniques of statistical modeling to capture such dependence are not adequate in various aspects. For instance, preservation of the linear association through Pearson product moment correlation coefficient may not be a suitable "measure of association" while capturing the dependence. This is due to the fact that a nonlinear, complex relationship is expected in most of the cases. So, preservation of a scale-free "measure of association" is required. It may be noted that the term "scale-free" or "scale-invariant" indicates that the measure of association will not change under nonlinear transformations of the random variable. Second, uncertainty associated with the dependence pattern between teleconnected hydroclimatic variables needs to be quantified as it is essential and more useful. Finally, assumption of any parametric

[1]Department of Civil Engineering, Indian Institute of Technology Bombay, Mumbai, India.
[2]Department of Civil Engineering, Indian Institute of Science, Bangalore, India.

0148-0227/08/2008JD009856$09.00

probabilistic distribution may not be able to represent the data set properly in many cases and therefore, a nonparametric estimate of probability density is preferable. The challenge lies in capturing the dependence using copula for nonparametrically estimated probability density of the data sets. An important advantage (and flexibility) of the copula is that the distribution of data sets can be simulated separately, and also they need not have to follow any particular distribution.

[3] The usefulness of the proposed approach can be recognized in three specific ways as follows:

[4] 1. Pearson's product moment correlation (ρ), known as "measure of linear association," considers only the linear association between random variables, whereas, scale-free "measure of association," such as Kendall's tau (τ) and Spearman's rho (ρ_s) are more general and robust [*Wilks*, 2006]. The proposed method preserves the scale-free "measure of association" while capturing the dependence pattern between the teleconnected hydroclimatic variables.

[5] 2. The proposed method can provide the information about uncertainty associated with the conditional distribution between the teleconnected hydroclimatic variables. This is an important aspect from hydroclimatic modeling perspective as the probabilistic predictions are made available to the user.

[6] 3. In this method, it is not necessary to consider any parametric probabilistic distributional assumptions for the variables as it can deal with nonparametrically estimated probability density for the data set of the teleconnected hydroclimatic variables. This aspect renders the method general and hence can be applied to wide range of similar problems, even if the data set does not follow any particular probabilistic distributional form, which is quite possible in many cases.

[7] The proposed method is applied in the context of nonparametric prediction of monthly Indian summer monsoon rainfall (ISMR) by capturing the dependence between monthly variation of ISMR and the large-scale climate precursors.

[8] The variation of Indian summer monsoon rainfall (ISMR) is crucial for the agroeconomic aspects of India. Long-range forecast of Indian summer monsoon rainfall has a long history since the work of *Blanford* [1884]. In early twentieth century, the link between summer monsoon rainfall over India with the Southern Oscillation was established by *Walker* [1923, 1924]. Later, several researchers had contributed toward a reliable long-range forecast of ISMR [*Banerjee et al.*, 1978; *Shukla and Paolino*, 1983; *Mooley et al.*, 1986; *Krishna Kumar et al.*, 1992; *Shukla and Mooley*, 1987]. Significant influence of the large-scale circulation patterns over tropical Pacific Ocean region and Indian Ocean region, on the variation of ISMR, is established in earlier studies [*Pant and Parthasarathy*, 1981; *Rasmusson and Carpenter*, 1983; *Khandekar and Neralla*, 1984; *Mooley and Paolino*, 1989; *Kane*, 1998; *Gadgil et al.*, 2004]. Two such large-scale circulation patterns are El Niño Southern Oscillation (ENSO) and Equatorial Indian Ocean Oscillation (EQUINOO). ENSO is the coupled Ocean-atmosphere mode of tropical Pacific Ocean [*Cane*, 1992] and EQUINOO is the atmospheric part of Indian Ocean Dipole (IOD) mode [*Saji et al.*, 1999; *Gadgil et al.*, 2004].

[9] Still the quality of the forecasts is not fully satisfactory from most of the existing models [*Rajeevan et al.*, 2004]. This is due to the fact that the relationship between ISMR and these large-scale circulation patterns is very complex [*Gadgil et al.*, 2004]. The linear association between them is not very strong particularly at monthly timescale, with Pearson product moment correlation coefficients of 0.33 and 0.19 with ENSO index and EQUINOO index respectively [*Maity and Nagesh Kumar*, 2006a]. Thus, the traditional approaches to capture the variation of ISMR, even at seasonal timescale, are not successful [*Gadgil et al.*, 2005].

[10] The possibility of combined influence of ENSO and EQUINOO on seasonal and monthly variation of ISMR is recently established [*Gadgil et al.*, 2004; *Maity and Nagesh Kumar*, 2006b]. At seasonal timescale, all the extremes in ISMR (greater than ± one standard deviation) from 1958 to 2003 are shown to be statistically associated with favorable (unfavorable) phases of ENSO or EQUINOO or both [*Gadgil et al.*, 2004]. In monthly scale, the combined influence of ENSO and EQUINOO on ISMR was established and captured to predict the variation of ISMR using Bayesian dynamic linear model (BDLM) [*Maity and Nagesh Kumar*, 2006b]. However, Gaussian probability distribution (parametric) of the data sets is the basic assumption in the BDLM approach. In this paper, as an illustration of the proposed method, the scale-free dependence pattern between monthly variation of ISMR and combined index of ENSO and EQUINOO is investigated and the monthly variation of ISMR is nonparametrically predicted, without the assumptions of any parametric distribution for the data sets.

[11] The rest of the paper is organized broadly as follows. The proposed method is detailed in section 2. In section 3, application of the proposed method to capture the dependence of ISMR on the combined index of ENSO and EQUINOO is elucidated along with the results of nonparametric prediction of monthly ISMR. Summary and conclusions are presented in section 4.

2. Proposed Approach

[12] The method, proposed in this study, is based on the theory of copula. The theory was successfully applied in many fields of application [*Bouyé et al.*, 2000; *Bagdonavicius et al.*, 1999; *Frees and Valdez*, 1998]. Application of copula in the field of water resources is still in its nascent stage and, to our knowledge, mostly limited to frequency analysis [*Favre et al.*, 2004; *Salvadori and De Michele*, 2004; *Grimaldi and Serinaldi*, 2006; *Zhang and Singh*, 2006] and few others [*Wang*, 2001; *Salvadori and De Michele*, 2006]. A brief description of the theory of copula helps to understand its basic advantages.

2.1. Theory of Copula

[13] A copula, C, is a function that joins or couples multiple distribution functions to their one-dimensional marginal distribution functions [*Nelsen*, 2006]. Application of copula to probability and statistics is achieved through Sklar Theorem [*Sklar*, 1959], which states that if $H_{X,Y}(x, y)$ is a joint distribution function, then there exists a copula $C(u, v)$ such that for all x, y in $\bar{R} \in (-\infty, \infty)$, $H_{X,Y}(x, y) =$

Table 1. Functional Forms of Different Copulas Along With Their φ (•) Functions

Copula	C_θ (u, v)	φ_θ (t)	$\theta\in$	$\tau=$
Frank	$-\frac{1}{\theta}\ln\left(1+\frac{\left(e^{-\theta u}-1\right)\left(e^{-\theta v}-1\right)}{e^{-\theta}-1}\right)$	$-\ln\left(\frac{e^{-\theta t}-1}{e^{-\theta}-1}\right)$	$(-\infty,\infty)$, excluding 0	$1-\frac{4}{\theta}[D_1(-\theta)-1]$[a]
Clayton	$[\max(u^{-\theta}+v^{-\theta}-1,0)]^{-1/\theta}$	$\frac{1}{\theta}(t^{-\theta}-1)$	$[1,\infty)$, excluding 0	$\frac{\theta}{\theta+2}$
Ali-Mikhail-Haq (AMH)	$\frac{uv}{1-\theta(1-u)(1-v)}$	$\ln\frac{1-\theta(1-t)}{t}$	$[-1, 1)$	$\left(\frac{3\theta-2}{\theta}\right)-\frac{2}{3}\left(1-\frac{1}{\theta}\right)^2\ln(1-\theta)$
Gumble-Hougaard (GH)	$\exp\left(-[(-\ln u)^\theta+(-\ln v)^\theta]^{1/\theta}\right)$	$(-\ln t)^\theta$	$[-1,\infty)$	$\frac{\theta-1}{\theta}$

[a] D_1, first-order Debye function; $D_1(\theta)=\frac{1}{x}\int_0^\theta\frac{t}{\exp(t)-1}dt$ for $\theta>0$; $D_1(-\theta)=D_1(\theta)+\frac{\theta}{2}$ [*Genest*, 1987; *Zhang and Singh*, 2006].

$C(F_X(x), F_Y(y))$, where, $F_X(x)$ and $F_Y(y)$ are the marginal distributions of X and Y respectively.

[14] The conceptual definition of copula can be expressed in another way also. Let X and Y be a pair of random variables with cumulative distribution functions (CDF) of $F_X(x)$ and $F_Y(y)$ respectively. Also let their joint CDF be $H_{X,Y}(x, y)$. Each pair, (x, y), of real numbers leads to a point, $(F_X(x), F_Y(y))$, in the unit square $[0, 1] \times [0, 1]$ and this ordered pair, in turn, corresponds to a number, $H_{X,Y}(x, y)$, in [0, 1]. This correspondence is indeed a function, which is known as Copula [*Nelsen*, 2006].

[15] It is worthwhile to note here that such correspondence is independent of the marginal distributions of the random variables. In other words, any form of marginal distribution can be coupled to get their joint distribution, which is the main reason for the popularity of copula theory in many areas of research.

[16] The theory of copula is used in this study because of its excellent property to study the scale-free dependence structure while preserving such dependence during simulation. Kendall's tau (τ) and Spearman's rho (ρ_s) are the most widely used scale-free ''measures of association'' for dependence structure between random variables. Kendall's tau (τ) is used in this study, which is defined as the difference between probability of concordance and discordance. The definition of concordant pair, discordant pair and sample estimate of Kendall's tau is explained later. It is important to note here that Pearson's product moment correlation (ρ), is a ''measure of linear association'' between random variables. It is obvious that the estimate of ρ changes under the nonlinear transformation of random variables. However, τ and ρ_s are scale-invariant and copulas are able to capture those scale-invariant properties of joint distribution which are invariant under strictly increasing transformation [*Schweizer and Wolff*, 1981].

[17] The Kendall's tau (τ) can be expressed mathematically in terms of copula as

$$\tau = 4\int C(u,v)dC(u,v) - 1 \tag{1}$$

where C is the corresponding copula [*Nelsen*, 2006]. This mathematical form can be expressed in a much simpler way for the class of Archimedean Copula. Among the various classes (Elliptical, Farlie-Gumbel-Morgenstern) the class of Archimedean Copula is most popular among researchers because of its unique mathematical properties.

[18] A copula is classified as an ''Archimedean copula'' if it can be expressed in terms of $C(u, v) = \varphi^{[-1]}(\varphi(u) + \varphi(v))$, where, $\varphi(\bullet)$ is known as generator of the copula and $\varphi^{[-1]}(\bullet)$ is the ''pseudo inverse'' of $\varphi(\bullet)$. $\varphi^{[-1]}(\bullet)$ is defined as,

$$\varphi^{[-1]}(t) = \begin{cases} \varphi^{-1}(t), & 0 \le t \le \varphi(0) \\ 0, & \varphi(0) \le t \le \infty \end{cases} \tag{2}$$

The basic properties of this class of copulas make them suitable for most of the research applications and hence used in this study also. *Frank* [1979], *Clayton* [1978], Ali-Mikhail-Haq (AMH) and Gumble-Hougaard (GH) [*Gumbel*, 1960; *Hougaard*, 1986] are a few examples of copulas belonging to the Archimedean class. The functional forms of these copulas are provided in Table 1. It can be shown that if X and Y are two random variables having an Archimedean copula, C, generated by φ, equation (1) for Kendall's tau (τ) can be reduced to [*Nelsen*, 2006]:

$$\tau = 1 + 4\int_0^1 \frac{\varphi_\theta(t)}{\varphi'_\theta(t)}dt \tag{3}$$

[19] Dependence parameter θ is estimated by replacing the population version of Kendall's tau with its sample estimate ($\hat{\tau}$), in equation (3). For some cases a close form equation is obtained. For example, let us take Clayton copula, for which, $\varphi_\theta(t) = \frac{1}{\theta}(t^{-\theta} - 1)$. Thus, $\frac{\varphi_\theta(t)}{\varphi'_\theta(t)} = \frac{t^{\theta+1}-t}{\theta}$ when $\theta \neq 0$ and $\frac{\varphi_0(t)}{\varphi'_0(t)} = t\ln t$ for $\theta = 0$. Finally, using equation (3), $\tau = 1 + 4\int_0^1 \frac{t^{\theta+1}-t}{\theta}dt \Rightarrow \tau = 1 + \frac{4}{\theta}\left(\frac{1}{\theta+2} - \frac{1}{2}\right) \Rightarrow \tau = \frac{\theta}{\theta+2}$. However, for some copulas such a close form does not exist. A numerical integration or special functions are to be used. These calculations are well established and can be found in standard text books [*Nelsen*, 2006; *Zhang and Singh*, 2006]. A list of such relationships is also provided in the Table 1.

[20] The theory explained so far is sufficient for this paper. Further details of copula and related proofs can be found elsewhere [*Genest and MacKay*, 1986a; *Genest and Rivest*, 1993; *Nelsen*, 2006].

2.2. Methodology

[21] Before the methodology is presented, it is worthwhile to mention that assumption of some parametric probabilistic distribution of data set is very common in many modeling approaches because of its convenient computational properties. However, a nonparametric estimate of probability density is preferable for many cases as explained

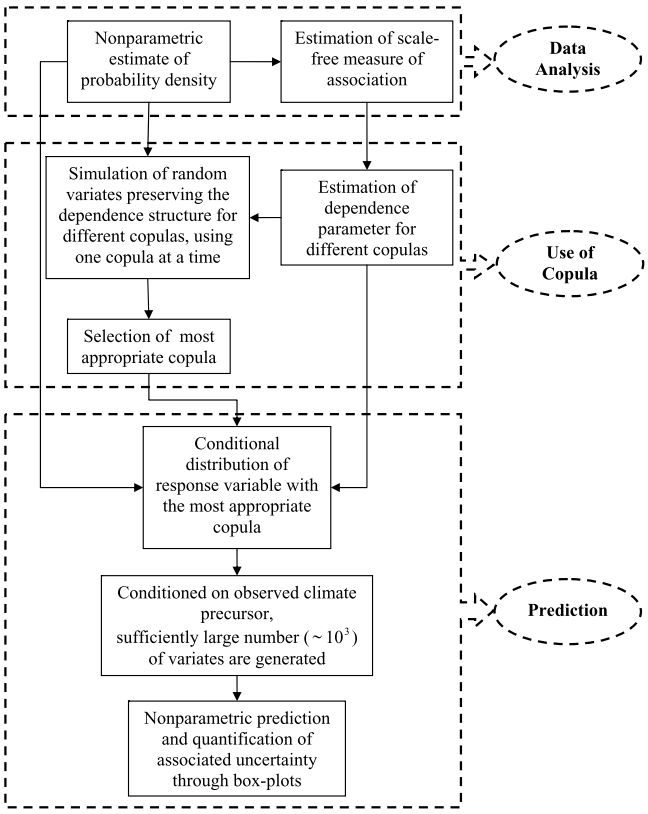

Figure 1. Flowchart summarizing the proposed method. The major steps of the methodology are explained in section 2.2 for a sequential understanding of the method.

earlier. So, in this study, nonparametric estimates of probability density of data sets are used in the proposed copula-based methodology, to capture the scale-free dependence pattern between the teleconnected hydroclimatic variables. However, the methodology is also applicable for parametric probabilistic distribution of the data set as well. As it cannot always be expected to have a parametric distributional form of the hydroclimatic time series, the methodology is explained for the nonparametrically estimated distribution of the data set to make the methodology more robust and applicable for any hydroclimatic data set.

[22] A flowchart, summarizing the proposed methodology, is presented in Figure 1. The major steps of the methodology are presented in the flowchart for a sequential understanding of the method. As shown in Figure 1, there are three major steps: (1) data analysis, (2) use of copula theory, and (3) prediction. These steps are explained below in detail.

2.2.1. Data Analysis

2.2.1.1. Estimation of Nonparametric Probability Density

[23] Kernel density estimator is most popular for estimation of nonparametric density [*Bosq*, 1998]. The kernel estimate of probability density, for a real-valued time series X_i, $i = 1, \ldots, n$ can be expressed as,

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - X_i) \tag{4}$$

where, $K_h(z) = \frac{1}{h} Kr(\frac{z}{h})$, in which, h is the smoothing parameter and Kr is the kernel function. Different types of kernel functions are naïve, normal and Epanechnikov, which are used in the study. Mathematical formulations of these kernel functions are shown below [*Bosq*, 1998].

$$\text{Naïve Kernel}: \quad Kr(u) = 1 \qquad -\frac{1}{2} \le u \le \frac{1}{2}$$

$$\text{Normal Kernel}: \quad Kr(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) \qquad -\infty \le u \le \infty$$

$$\text{Epanechnikov Kernel}: \quad Kr(u) = \frac{3}{4\sqrt{5}}\left(1 - \frac{u^2}{5}\right) \qquad -\sqrt{5} \le u \le \sqrt{5}$$

[24] The cumulative probability density is obtained from the corresponding nonparametrically estimated probability density.

2.2.1.2. Estimation of Scale-Free Measure of Association

[25] Let (x_1, y_1), (x_2, y_2), ... , (x_n, y_n) be the paired samples of two random variables. Two pairs (x_i, y_i) and (x_j, y_j) are known to be concordant if $(x_i - x_j)(y_i - y_j) > 0$ and discordant if $(x_i - x_j)(y_i - y_j) < 0$. Sample estimate of Kendall's tau is obtained as the difference between the probability of concordance and probability of discordance. Out of n paired samples, there are ${}^nC_2$ different ways to select two pairs. If there are c number of concordant pairs and d number of discordant pairs, sample estimate of Kendall's tau is expressed as,

$$\hat{\tau} = P\left[(X_i - X_j)(Y_i - Y_j) > 0\right] - P\left[(X_i - X_j)(Y_i - Y_j) < 0\right] = \frac{c}{{}^nC_2} - \frac{d}{{}^nC_2} = \frac{c-d}{{}^nC_2} \tag{5}$$

2.2.2. Use of Copula

2.2.2.1. Estimation of Dependence Parameter

[26] After obtaining the sample estimate of Kendall's tau ($\hat{\tau}$), the dependence parameter θ of an Archimedean copula is estimated by replacing the population version of Kendall's tau with its sample estimate ($\hat{\tau}$), in equation (3). Thus, paired random variates can be simulated through this Archimedean copula preserving their dependence structure.

2.2.2.2. Simulation of Random Variates Preserving the Dependence Structure

[27] Such simulation can be achieved by the algorithm [*Genest and MacKay*, 1986b] as explained below.

[28] 1. For an Archimedean copula, functional forms of $\varphi^{[-1]}(\bullet)$, $\varphi'(\bullet)$ and $\varphi'^{(-1)}(\bullet)$ are obtained where $\varphi_\theta(\bullet)$ is the generator function with parameter θ. Equation (2) is used to obtain $\varphi^{[-1]}(\bullet)$. Same can be used for $\varphi'^{(-1)}(\bullet)$ also after obtaining $\varphi'(\bullet)$, which is the derivative of $\varphi(\bullet)$ with respect to $\bullet$.

[29] 2. Two independent uniformly distributed ($\sim U(0,1)$) random variates, u and r, are generated.

[30] 3. Two new variables, s and w, are obtained as $s = \varphi'(u)/r$ and $w = \varphi'^{(-1)}(s)$.

[31] 4. Another variable, v, is obtained as $v = \varphi^{[-1]}(\varphi(w) - \varphi(u))$ [*Genest and MacKay*, 1986b]. The pair u and v are the simulated pair, preserving the dependence structure.

[32] 5. Both these u and v are in the range [0 1]. These simulated pairs of u and v are then back transformed by their corresponding inverse cumulative distribution functions to

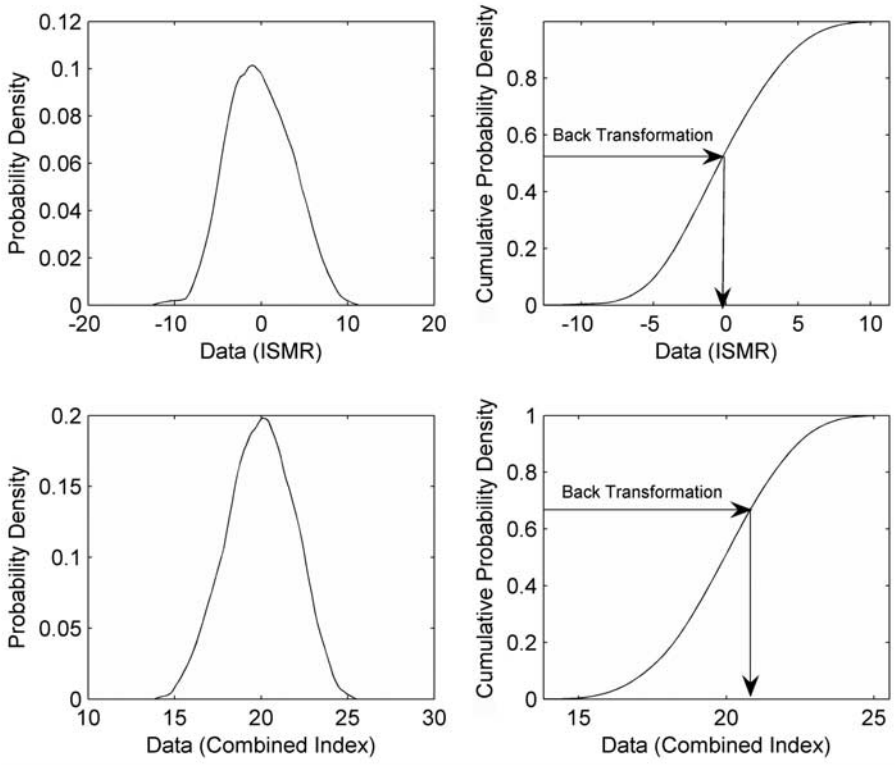

Figure 2. Nonparametric estimates of probability density for monthly ISMR and combined index using Epanechnikov kernel as estimator.

map them in the original scale of the random variable. In other words, these simulated pairs of u and v are replaced by cumulative probability density of the corresponding random variable and the value of simulated random variable (original scale) is obtained. Figure 2 (right) may be referred to for a clearer visual understanding. Details of Figure 2 are explained later. At this moment, only the principle of back transformation may be considered.

[33] The above steps are repeated for different Archimedean copulas considered in the analysis.

2.2.2.3. Selection of the Most Appropriate Copula

[34] Steps involved to select the most appropriate copula are as follows [*Genest and Rivest*, 1993; *Zhang and Singh*, 2006].

[35] 1. For a particular Archimedean copula C with generator function φ, a parametric function $K(z)$ is defined as $K(z) = z - \frac{\varphi(z)}{\varphi'(z^+)}$. $K(z)$ is the distribution function of random variable, $C(U, V)$ where u and v are the uniformly (0,1) distributed random variables [*Nelsen*, 2006].

[36] 2. A nonparametric estimate of the above function, $K_n(z)$ is obtained as the proportion of $z_i < z$, where z_i is

$$z_i = \frac{\text{Number of } (x_j, y_j) \text{ such that } x_j < x_i \text{ and } y_j < y_i}{(N-1)}, \quad \text{where } i = 1, \mathrm{K}\ , N$$

[37] 3. A scatterplot between $K(z)$ and $K_n(z)$ is prepared.

[38] 4. Steps 1–3 are repeated for all the copulas considered.

[39] 5. The better the fit, the closer the corresponding scatter to a 45° line through origin. Sum of square errors (SSE) from the 45° line through origin are obtained. The copula with the least SSE is selected as the most appropriate.

2.2.3. Prediction

2.2.3.1. Conditional Distribution of Response Variable

[40] After selecting the best copula, conditional distribution of response variable, conditioned on the observed climate precursor, is obtained as follows [*Zhang and Singh*, 2006]

$$C_{X/Y=y}(x) = \frac{\partial}{\partial y} C_{X,Y}(x,y) \Big/_{Y=y} \tag{6}$$

where X is the response variable and Y is the observed climate precursor.

[41] Once this distribution is obtained, a sufficiently large number ($\sim 10^3$) of variates are generated. These values are back transformed by using the nonparametrically estimated cumulative probability density of the response variable, obtained in step 1.

2.2.3.2. Nonparametric Prediction and Quantification of Associated Uncertainty

[42] Statistical properties of the generated values of the response variable, conditioned on the observed climate precursor, are investigated through box plot. The median of these values is used as a prediction, corresponding to the observed value of the climate precursor. The interquartile range (75th percentile to 25th percentile) of these values indicates the associated uncertainty.

[43] It may be noted here that the association between the teleconnected variables is simulated preserving their dependence through scale-free measure of association between them. Thus, the dependence pattern between the variables is implicitly captured. This will be made clear in the next section while describing the application of the proposed methodology. Another point necessary to mention here is that the approach needs a well representative and sufficiently large data sets, so that a proper estimate of density functions can be obtained.

[44] It is also necessary to mention that the estimation of probability density is carried out nonparametrically using kernel density estimator as stated before. It is also mentioned that the prediction and quantification of associated uncertainty are also carried out nonparametrically. However, while simulating the random variates preserving the mutual dependence, the dependence structure is preserved through a parameter θ in copula. Thus the proposed methodology is semiparametric.

3. Application of the Proposed Methodology

[45] The methodology presented in the previous section is applied to capture the dependence between monthly variation of ISMR and composite index of ENSO and EQUINOO. The data for the period 1958–1985 are used to develop the model and the model performance is tested with the data for the period 1986–2003. The variation of rainfall across the months is considered through their anomaly values for individual monsoon months. The dependence of these anomaly values on the composite index is analyzed.

3.1. Data

3.1.1. Monthly ISMR Data

[46] Monthly rainfall data [*Parthasarathy et al.*, 1994] over entire India are obtained for the period, January 1958 to December 2003, from the Website of Indian Institute of Tropical Meteorology (IITM), Pune, India (http://www.tropmet.res.in/data.html). The monthly rainfall indicates the aggregated rainfall over a month. The variation of rainfall across the month is not considered. Monthly anomaly values for monsoon months (June through September), are used in the present study.

3.1.2. Monthly ENSO and EQUINOO Data

[47] Sea surface temperature anomaly from Niño 3.4 region (5°S–5°N, 120–170°W) is used as "ENSO index" in this study. Monthly sea surface temperature data from Niño 3.4 region for the period, January 1958 to December 2003, are obtained from the Website of National Weather Service, Climate Prediction Centre of NOAA (http://www.cpc.noaa.gov/data/indices/). EQWIN, the negative of zonal wind anomaly over equatorial Indian Ocean region (60–90°E, 2.5°S–2.5°N), is used as "EQUINOO index" [*Gadgil et al.*, 2004]. Monthly surface wind data for the period, January 1958 to December 2003, are obtained from National Center for Environmental Prediction (http://www.cdc.noaa.gov/Datasets).

3.1.3. Combined Index of ENSO and EQUINOO

[48] The combined index of ENSO and EQUINOO indices is used in the analysis using the results obtained in an earlier study [*Maity and Nagesh Kumar*, 2006b]. A brief description is presented here. Both the indices are transformed using equations: $X'(i) = X(i)^* \frac{\sigma_R}{\sigma_X}$ and $X''(i) = X'(i) + (\mu_R - \mu_{X'}) + \gamma$, where the variable X is the observed values of an index; σ_X and σ_R are the standard deviations of the index and the observed monthly rainfall anomaly respectively; μ_R is the mean monthly rainfall anomaly; $\mu_{X'}$ is the mean of the transformed index X'; γ is any constant to shift the entire time series by some desirable amount. γ is selected as 20 on the basis of our earlier studies. The transformed indices are combined as $CI_{i,j} = c_1\, EN_{i,j-\kappa} + c_2\, EQ_{i,j-\lambda}$, where, CI is the combined index; EN and EQ are the ENSO and EQUINOO indices respectively; c_1 and c_2 are the relative weightage factors; first and second subscripts are the year and month respectively. On the basis of our earlier studies, the best values for c_1, c_2, κ and λ were found to be 0.61, 0.39, 2 months and 1 month respectively [*Maity and Nagesh Kumar*, 2006b]. These parameters are obtained using the data for the period 1958–1985. Once these parameters are obtained, the combined index (CI) is constructed for each month during the testing period too. For example, CI for July 1998 is constructed using ENSO from May 1998 (2 months lag) and EQUINOO from June 1998 (1 month lag). The physical meaning of combined index can be viewed as that both the ENSO and EQUINOO have their influences on the variation of ISMR. The effect of circulation pattern over tropical Pacific Ocean on ISMR is modified by the circulation pattern over tropical Indian Ocean. Thus, a joint influence is established in earlier studies [*Gadgil et al.*, 2004; *Maity and Nagesh Kumar*, 2006b]. This joint influence is being considered by the composite index. The lags of both ENSO and EQUINOO (2 months and 1 month respectively) indicate that the EQUINOO is more immediate factor than ENSO, which is due to the fact that the origin of EQUINOO is closer to India than that of ENSO. It can be noted that the transformation of the raw climate indices is performed to extract the climate signal more effectively. It is observed that the performance of the model has enhanced on using the transformed indices [*Maity and Nagesh Kumar*, 2006b]. So, the indices are thus transformed before combining them.

[49] It may be noted that the monthly rainfall anomaly values are uncorrelated over successive time steps, which is ensured from the autocorrelation diagram. Thus, pairwise anomaly values are considered to be independent over successive time steps, which are used to capture the scale-free dependence.

3.2. Results and Discussions

[50] Kernel estimate of nonparametric probability densities of the monthly rainfall anomaly and the combined index is obtained by considering the Epanechnikov kernel (Figure 2, left) based on the model development period (1958–1985). However, it was noticed that the probability densities do not differ much, even if "normal" kernel function is opted (not shown). Corresponding cumulative probability densities are numerically estimated through trapezoidal rule (Figure 2, right). It may be noted that the smoothing parameter is chosen as $\sigma(\frac{4}{3n})^{\frac{1}{5}}$, where σ is the standard deviation and n is length of data set.

[51] The scale-free measure of association between monthly rainfall anomaly and combined index is obtained as 0.24, in terms of Kendall's tau, using the data for the

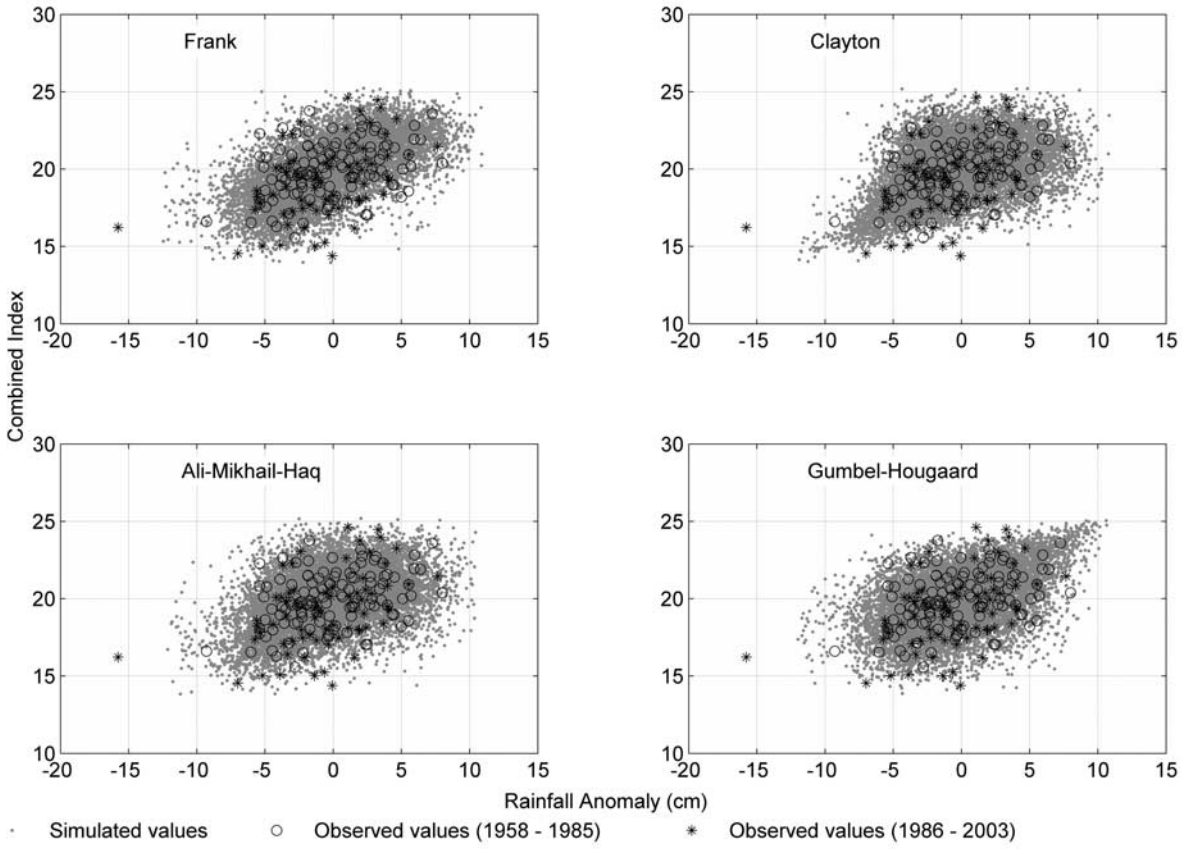

Figure 3. Simulated values of monthly rainfall anomaly and combined index using different copulas.

period 1958 to 1985. It may be noted that the dependence between climate precursors and hydrologic response variables is not expected to be very high, because of the associated uncertainty and complex nature of the relationship. However, the Kendall's tau obtained between the combined index and ISMR is statistically significant as tested by a standard t test. It may be noted that that a t test is only an approximate test for τ. However, the data set of climate indices (ENSO and EQUINOO) and monthly rainfall anomaly follows an approximate Normal distribution [*Maity and Nagesh Kumar*, 2006b]. As the composite index is a linear combination of ENSO and EQUINOO indices, the composite index may also be assumed to approximately follow a Normal distribution. So the t test is adopted here. Using a standard t test, it is observed that the null hypothesis, $\tau = 0$, is comfortably rejected with p value as 1.76×10^{-4}.

[52] Four different copulas from Archimedean family, namely, (1) Frank, (2) Clayton, (3) Ali-Mikhail-Haq (AMH), and (4) Gumbel-Hougaard (GH) copulas, are selected to simulate the association between monthly rainfall anomaly and combined index. A large number of simulated values (30,000) are shown in Figure 3 for each copula. Observed values of monthly rainfall anomaly and combined index, for the period 1958 to 1985, are shown by circles in Figure 3. It is noticed that in each case, simulated values cover the entire range of the observed data for the period 1958–1985, based on which the simulation was performed.

[53] To check visually whether the simulation covers the possible range of unforeseen data, the observed values of monthly ISMR and combined index, for the period 1986 to 2003, are also plotted (as shown by asterisks in Figure 3) over the "cloud" of simulated values. It is found that during this period, one data point remains as a distinct outlier in all cases. This point corresponds to July 2002, for which the observed rainfall was entirely unprecedented (49% below normal). Being the historical outlier, this point correspondingly remains as distinct outlier in all cases. Except for this point, all other unforeseen data points are covered well by the simulated cloud. However, to select the best fitted copula, the procedure explained in step 5 of methodology (section 2.2) is followed. Scatterplots between $K(z)$ and $K_n(z)$ are prepared for all the copulas and the corresponding sum of square errors (SSE) values are obtained. SSE values for Frank, Clayton, Gumbel-Hougaard (GH) and Ali-Mikhail-Haq (AMH) are 0.024, 0.057, 0.026 and 0.046 respectively. Thus, the Frank copula (SSE = 0.024) performs best among all copulas, with GH copula (SSE = 0.026) as its close competitor. So, joint probability density between rainfall anomaly and combined index is obtained using Frank copula (Figure 4).

[54] Next, the nature of the rainfall anomaly is examined for different ranges of combined index. Simulated rainfall anomalies, corresponding to extremes of simulated combined index, are compared. Rainfall anomalies are identified, corresponding to combined index lying (1) below 5th percentile, (2) in between 5th and 95th percentile, and

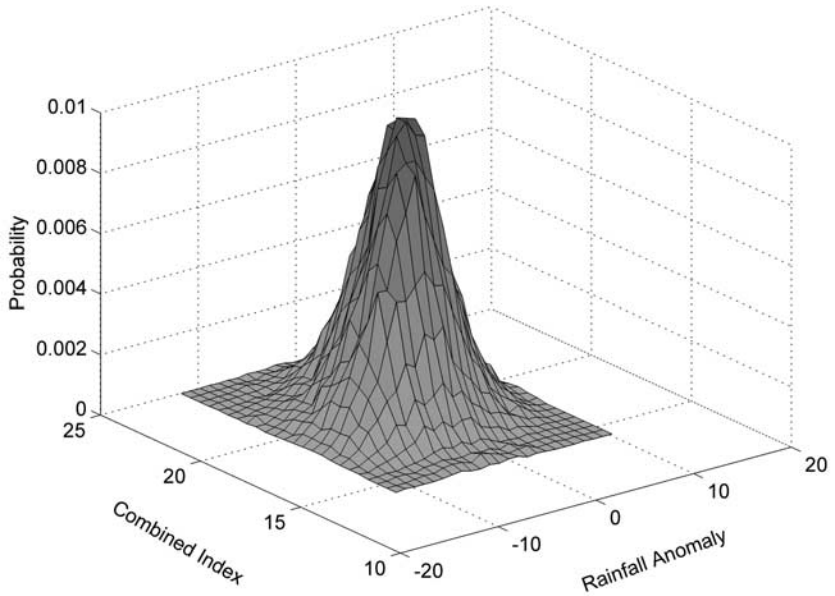

Figure 4. Approximate joint probability density of rainfall anomaly and combined index.

(3) above 95th percentile. Rainfall anomalies in each category are presented by three different boxplots in Figure 5. It may be noted that the number of data points in extreme cases (∼1500) as well as those in between 5th and 95th percentile (∼27000) are simulated in sufficiently large number. On the basis of these simulated data, it is noticed that rainfall anomalies, corresponding to combined index below 5th percentile, are mostly below normal and those, corresponding to combined index above 95th percentile, are above normal. This conclusion corroborates with earlier studies [*Gadgil et al.*, 2004; *Maity and Nagesh Kumar*, 2006b]. It is analytically cross-checked by the proposed copula-based methodology to show that this methodology is able to preserve the dependence.

[55] It is interesting to note from Figure 5 that the median of rainfall anomaly, given the "low" combined index, is −2.0935 cm, where as the median of rainfall anomaly, given the "high" combined index, is +2.3422 cm. Moreover, interquartile range of rainfall anomaly, given the "low" combined index, is −4.6806 to −0.4835 cm, whereas the same, given the "high" combined index, is 0.2522 to +4.6178 cm. On the other hand, unconditional median of rainfall anomaly is −0.4596 cm and unconditional interquartile range of rainfall anomaly is −3.0167 to +2.3838 cm (Figure 5). Thus, it is clearly observed that the conditional distribution of rainfall anomaly, given the combined index, is much more informative compared to the unconditional distribution. It may be noted that the rainfall anomaly mostly varies within the range of −5 to +5 cm and better information about the positive and negative anomaly can be obtained by considering the combined index.

[56] Another important point is that significant overlap of interquartile ranges, as reflected from the corresponding boxplots, indicates that the dependence is associated with uncertainty. Thus the relationship is associated with uncertainty and quantification of the uncertainty is necessary while the dependence is used for prediction of monthly ISMR.

[57] The prediction performance of the proposed method is investigated using the data for the period 1986–2003. A plot between the observed and predicted monthly rainfall is shown in Figure 6. The predicted rainfall values are shown by bar plots along with box plots to show the information regarding uncertainty associated with the predicted values. The magnitudes of observed rainfall are shown as asterisks with a connecting line to visually compare the predicted rainfall with the observed one. The climatological mean values are also shown by circles. It is obvious that asterisks lying above circles indicate positive anomaly and vice versa.

[58] It can be observed that during the testing period (1986–2003), out of 72 cases, the anomalies were correctly indicated for 50 cases and wrongly indicated for 20 cases. For the remaining two cases predictions are almost same as

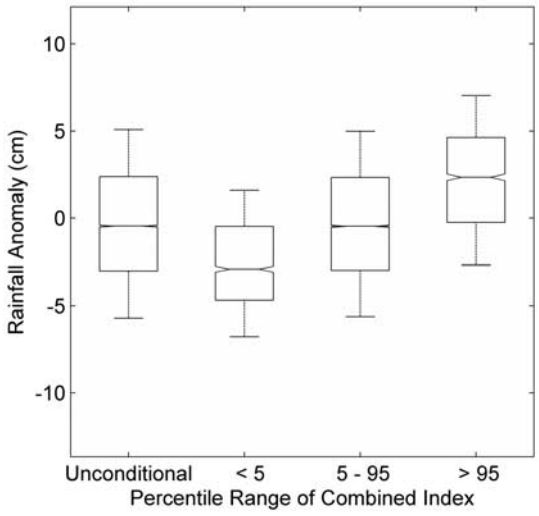

Figure 5. Box plots of simulated rainfall anomalies for different percentile ranges of combined index.

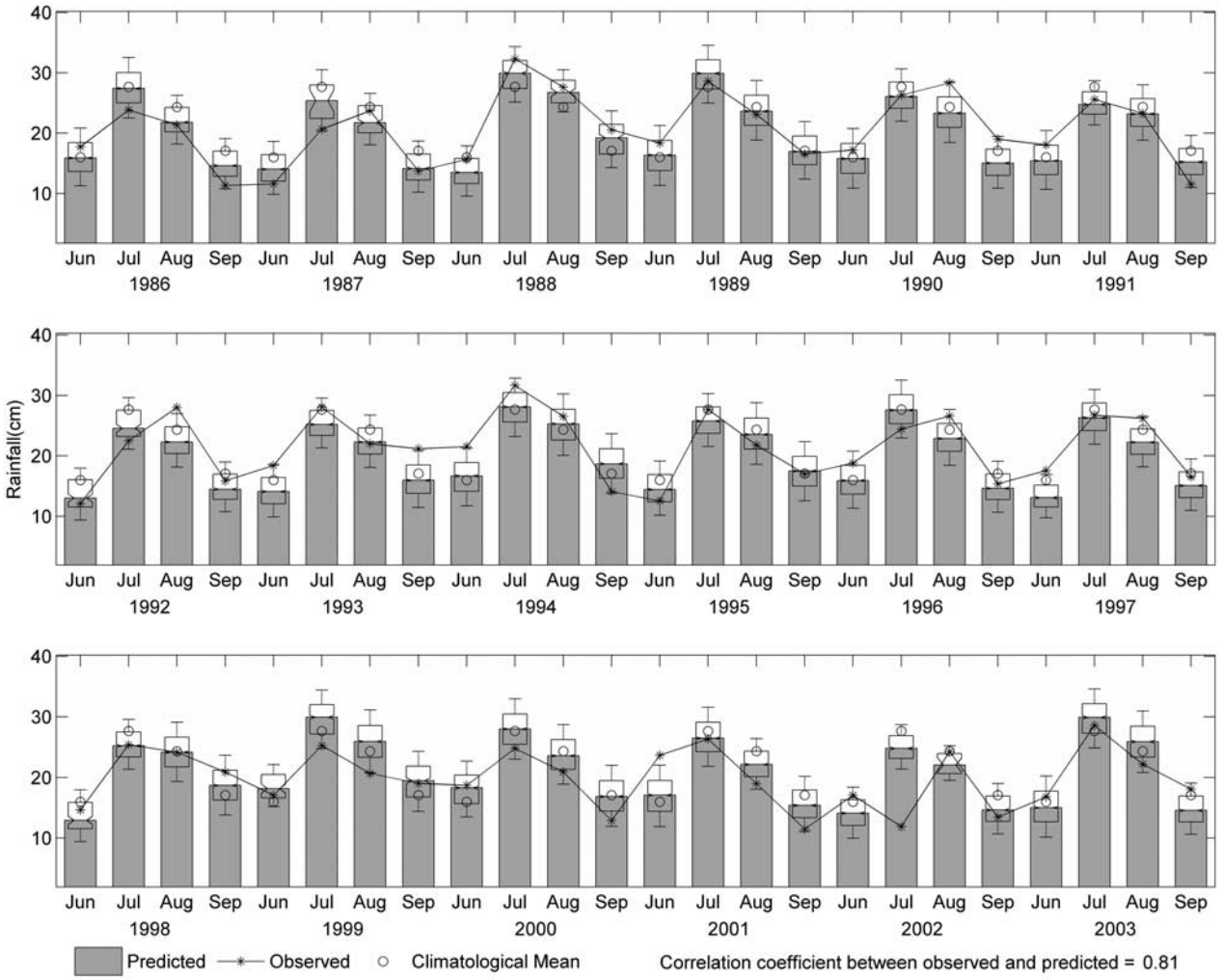

Figure 6. Comparison between observed and predicted rainfall along with box plots, showing the information regarding uncertainty associated with the predicted values

the climatological means. Thus, predictions are successfully made for most of the years except for July 2002. More importantly, lower than normal rainfall in the year 1987 and higher than normal rainfall in the year 1988, are both successfully predicted. For other years, predictions are also made with reasonable accuracy. In fact, predicted values highly correspond to the observed values with a Pearson product moment correlation coefficient of 0.81 (significant correlation coefficient being 0.21 at 95% statistical confidence level) between them. Another important point is that the predicted values are available along with the information of uncertainty as shown by the interquartile range of the associated boxplots. However, the approaches based on linear association can also provide the uncertainty estimates. But, if the relationship is nonlinear, the approaches based on the linear association may not be adequate and would obviously be inferior. In such cases, the uncertainty captured by the proposed method will be superior to that by the approaches based on linear association. The performance of a simple linear regression model, using the information of ENSO and EQUINOO indices, is very poor. The Pearson product moment correlation coefficient between observed and predicted monthly rainfall is 0.27 [*Maity and Nagesh Kumar*, 2006b] as against 0.81 by the present method. This is due to the complex relationship between ISMR and large-scale climate precursors, which is not within the capability of simple linear regression.

[59] The results based on the linear regression for different cases are also worked out and presented in Table 2. The regression analyses are carried out between ISMR and (1) ENSO (with 2 months lag), (2) EQUINOO (with 1 month lag), (3) ENSO and EQUINOO (with 2 months lag for ENSO and 1 month lag for EQUINOO) and (4) composite index (CI). Both mean square error (MSE) and Pearson product moment correlation coefficient (CC) between observed and predicted ISMR are obtained for each of these models. MSE and CC for the present copula-based approach are also presented in the table for comparison. It can be observed that the performance of proposed copula-based approach is the best. Another point needs to be discussed here. It is obvious that among the first three models (1–3), model 3 is performing best. It may strike to the mind as to why model 4 is not performing at least to be comparable to that of model 3. The comparable results

Table 2. Performance of Different Linear Regression Models and Copula-Based Approach

Model Number	Model for Prediction of ISMR	Performance Statistics	
		MSE	CC
1	regression using ENSO (with 2 months lag)	27.13	0.222
2	regression using EQUINOO (with 1 month lag)	27.82	0.177
3	regression using ENSO and EQUINOO (with 2 months lag for ENSO and 1 month lag for EQUINOO)	26.76	0.268
4	regression using CI	28.92	0.043
5	proposed copula-based approach	10.40	0.814

would have been obtained if the composite index was developed by simple least square technique. It should be kept in mind that CI is not developed by simple least square technique rather it is developed considering the dynamic relationship between ISMR and large-scale circulation indices (ENSO and EQUINOO) [*Maity and Nagesh Kumar*, 2006b]. If some index (better not to call it CI to avoid confusion) would have been developed by least square technique (let us denote it as CI_LST), the performance of the regression model between ISMR and CI_LST would probably be comparable to that of model 3. In fact, the MSE and CC values by the regression model using CI_LST were found to be 26.69 and 0.272 respectively (not shown in Table 2). Still the results from the proposed copula-based approach are seen to be much better than all these models.

[60] While investigating the prediction performance for individual months, it is found that variances for prediction error in June, July, August and September are 5.79 cm^2, 14.61 cm^2, 9.31 cm^2 and 8.50 cm^2 respectively. On the other hand, variances for rainfall in June, July, August and September are 9.25 cm^2, 19.98 cm^2, 7.79 cm^2 and 10.75 cm^2 respectively. Thus, except for the month of August, the variance is reduced by 37.5% for June, 26.9% for July and 21.0% for September. However, for the month of August the variance is not getting reduced. This is due to the fact that in 1990 and 1992 the predicted values are very low compared to the observed ones resulting in high errors. If we exclude these two cases, the variance of error for the month of August becomes 6.81, which is smaller than that of observed rainfall for this month.

[61] Most important point of the proposed methodology is that the rainfall values are successfully predicted for those years, in which the long established negative correlation between ENSO and ISMR was not perceived. For example, normal rainfall in the year of 1997 is predicted successfully even though a drought was expected because of the largest El Niño of the nineteenth century. However, so far as physical mechanism is concerned, this is due to the consideration of EQUINOO along with ENSO [*Gadgil et al.*, 2004; *Maity and Nagesh Kumar*, 2006b]. Thus, the dependence is successfully captured by the proposed method. In the year 2002, seasonal rainfall (total rainfall during June through September) was significantly lower than normal. However, at monthly scale, observed rainfall values for June and August were normal and those for July and September were significantly lower than normal. In fact, July rainfall was completely unprecedented being 49% below normal. Significantly lower than normal rainfall in September and normal rainfalls for June and August are successfully predicted by this approach as shown in Figure 6. However, July rainfall, being completely unprecedented, is not predicted with reasonable accuracy. Still the predicted rainfall is below normal, thus, the direction of the anomaly is indicated. Sudden failure of July rainfall in 2002 may be attributed to some other factors, which are beyond the scope of the present study. Nonetheless, the proposed method is shown to be able to capture the overall dependence and, in general, able to predict the monthly variation of rainfall with reasonable accuracy, along with the information of uncertainty associated with the predicted values.

[62] At a smaller spatial scale, i.e., within the homogeneous rainfall regions (identified by Indian Meteorological Department), similar approach may be followed. Smaller temporal scale (lower than monthly) may not be of much use at this spatial scale. However, weekly or fortnightly temporal analysis at further smaller spatial scale, for example, river basin scale, may be of importance. In such small spatiotemporal scale, influence of some local meteorological variables may be significant apart from the large-scale atmospheric circulations (i.e., ENSO and EQUINOO). Such meteorological variables may have to include air temperature, local pressure, relative humidity, wind speed etc. Such further analysis may be considered as a future scope.

4. Summary and Conclusions

[63] In this paper, a method is proposed to capture the dependence between teleconnected hydroclimatic variables for successful prediction of the response variable. The method is based on the theory of copula, which is a function that couples the marginal distribution of variables to their joint distribution. The theory of copula is having very strong theoretical background. The basic motivation to use the copula is that it helps to study the scale-free dependence structure and preserves such dependence during simulation. Moreover, the distribution of data sets can be simulated separately, and they need not have to follow any particular distribution.

[64] The copula-based method, proposed in this paper, is demonstrated in the context of Indian summer monsoon rainfall and its dependence on combined index of ENSO and EQUINOO. The performance of the model is seen to be alluring. The proposed method is general in many aspects and can therefore be applied to similar studies of teleconnected hydroclimatic variables to capture the dependence of any form between them. First of all, any relationship between the teleconnected hydroclimatic variables is subject to variation under different temporal scales. For instance, interannual, interdecadal variability of ISMR with respect to ENSO is well documented in literature. The proposed method is shown to be able to capture the overall relationship. Thus, the method is able to capture the scale-free dependence pattern between the teleconnected hydroclimatic variables. Scale-free dependence pattern provides the information of relationship irrespective of its nature.

[65] Second, while capturing the relationship, uncertainty associated with the relationship is also captured. This

enables to provide the information of uncertainty associated with the prediction using the proposed method.

[66] Finally, kernel-based nonparametric estimates of probability density of the data sets (Figure 2) are used in the proposed method. This is an important aspect of the method as it avoids any parametric probabilistic distributional assumptions for the data sets. Parametric assumptions may not be able to suitably represent the data sets for many cases. Thus, the method is robust with respect to the distributional properties of the data set. In brief, while the relationship between response and causal variables is complex and uncertainty quantification in a nonparametric way is necessary, it is recommended to use the proposed method.

[67] However, having presented the convenience of the approach, one point, related to the theory of copula, must be kept in mind. In some cases, bivariate data may not follow any copula derived from one of the classical copula families. Even though the Sklar's theorem shows that any multivariate distribution can be expressed in a copula form, this does not imply that one of the "classical" copulas will be adequate for this purpose. Thus, it is required to investigate whether the assumed copula is able to acceptably describe the data at hand or not as explained in this paper.

[68] **Acknowledgments.** This work is partially supported by the SERC division, Department of Science and Technology, Government of India, through project SR/FTP/ETA-26/07 and INCOH, Ministry of Water Resources, Government of India, through project 23/52/2006-R&D. The financial support from the seed grant from IIT Bombay to the first author through project 07IR020 is also acknowledged.

References

Bagdonavicius, V., S. Malov, and M. Nikulin (1999), Characterizations and parametric regression estimation in Archimedian copulas, *J. Appl. Stat. Sci.*, *89*, 137–154.

Banerjee, A. K., P. N. Sen, and C. R. V. Raman (1978), On foreshadowing southwest monsoon rainfall over India with midtropospheric circulation anomaly of April, *Indian J. Meteorol. Hydrol. Geophys.*, *29*, 425–431.

Blanford, H. F. (1884), On the connexion of the Himalaya snowfall with dry winds and seasons of drought in India, *Proc. R. Soc. London*, *37*, 3–22, doi:10.1098/rspl.1884.0003.

Bosq, D. (1998), *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction, Lecture Notes in Statistics*, vol. 110, 210 pp., Springer, New York.

Bouyé, E., A. Durrleman, A. Nikeghbali, G. Riboulet, and T. Roncalli (2000), Copulas for finance–A reading guide and some applications, technical report, Groupe de Rech. Oper., Credit Lyonnais, Paris.

Cane, M. A. (1992), Tropical Pacific ENSO models: ENSO as a mode of the coupled system, in *Climate System Modeling*, edited by K. E. Trenberth, pp. 583–689, Cambridge Univ. Press, New York.

Chowdhury, M. R., and N. Ward (2004), Hydro-meteorological variability in the greater Ganges-Brahmaputra-Meghna basins, *Int. J. Climatol.*, *24*, 1495–1508, doi:10.1002/joc.1076.

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*, 141–151, doi:10.1093/biomet/65.1.141.

Dracup, J. A., and E. Kahya (1994), The relationship between U. S. streamflow and La Niña events, *Water Resour. Res.*, *30*(7), 2133–2141, doi:10.1029/94WR00751.

Eltahir, E. A. B. (1996), El Niño and the natural variability in the flow of the Nile River, *Water Resour. Res.*, *32*(1), 131–137, doi:10.1029/95WR02968.

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.

Frank, M. J. (1979), On the simultaneous associativity of $F(x, y)$ and $x + y - F(x, y)$, *Aequationes Math.*, *19*, 194–226, doi:10.1007/BF02189866.

Frees, E. W., and E. A. Valdez (1998), Understanding relationships using copulas, *N. Am. Actuarial J.*, *2*(1), 1–25.

Gadgil, S., P. N. Vinayachandran, P. A. Francis, and S. Gadgil (2004), Extremes of the Indian summer monsoon rainfall, ENSO and equatorial Indian Ocean oscillation, *Geophys. Res. Lett.*, *31*, L12213, doi:10.1029/2004GL019733.

Gadgil, S., M. Rajeevan, and R. Nanjundiah (2005), Monsoon prediction—Why yet another failure?, *Curr. Sci.*, *88*(9), 1389–1400.

Genest, C. (1987), Frank's family of bivariate distributions, *Biometrica*, *74*, 549–555, doi:10.1093/biomet/74.3.549.

Genest, C., and J. MacKay (1986a), The joy of copulas: Bivariate distributions with uniform marginals, *Am. Stat.*, *40*(4), 280–283, doi:10.2307/2684602.

Genest, C., and J. MacKay (1986b), Copules archimédiennes et familles de lois bidimensionnelles dont les marges sont données Copules, *Can. J. Stat.*, *14*(2), 145–159, doi:10.2307/3314660.

Genest, C., and L.-P. Rivest (1993), Statistical inference procedures for bivariate Archimedean copulas, *J. Am. Stat. Assoc.*, *88*(423), 1034–1043, doi:10.2307/2290796.

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.

Gumbel, E. J. (1960), Bivariate exponential distributions, *J. Am. Stat. Assoc.*, *55*, 698–707, doi:10.2307/2281591.

Hougaard, P. (1986), A class of multivariate failure time distributions, *Biometrika*, *73*, 671–678.

Jain, S., and U. Lall (2001), Floods in a changing climate: Does the past represent the future?, *Water Resour. Res.*, *37*(12), 3193–3205, doi:10.1029/2001WR000495.

Kahya, E., and J. A. Dracup (1993), U. S. streamflow patterns in relation to the El Niño/Southern Oscillation, *Water Resour. Res.*, *29*(8), 2491–2503, doi:10.1029/93WR00744.

Kane, R. P. (1998), Extremes of the ENSO phenomenon and Indian summer monsoon rainfall, *Int. J. Climatol.*, *18*, 775–791, doi:10.1002/(SICI)1097-0088(19980615)18:7<775::AID-JOC254>3.0.CO;2-D.

Khandekar, M. L., and V. R. Neralla (1984), On the relationship between the sea surface temperatures in the equatorial Pacific and the Indian monsoon rainfall, *Geophys. Res. Lett.*, *11*, 1137–1140, doi:10.1029/GL011i011p01137.

Krishna Kumar, K., K. Rupa Kumar, and G. B. Pant (1992), Premonsoon ridge location over India and its relation to monsoon rainfall, *J. Clim.*, *5*, 979–986, doi:10.1175/1520-0442(1992)005<0979:PRLOIA>2.0.CO;2.

Maity, R., and D. Nagesh Kumar (2006a), Hydroclimatic association of the monthly summer monsoon rainfall over India with large-scale atmospheric circulations from tropical Pacific Ocean and the Indian Ocean region, *Atmos. Sci. Lett.*, *7*, 101–107, doi:10.1002/asl.141(RMetS).

Maity, R., and D. Nagesh Kumar (2006b), Bayesian dynamic modeling for monthly Indian summer monsoon rainfall using El Niño–Southern Oscillation (ENSO) and Equatorial Indian Ocean Oscillation (EQUINOO), *J. Geophys. Res.*, *111*, D07104, doi:10.1029/2005JD006539.

McKerchar, A. I., C. P. Pearson, and B. B. Ftzharris (1998), Dependency of summer lake inflows and precipitation on spring SOI, *J. Hydrol. Amsterdam*, *205*(1–2), 66–80, doi:10.1016/S0022-1694(97)00144-3.

Mooley, D. A., and D. A. Paolino (1989), The response of the Indian monsoon associated with the change in sea surface temperature over the eastern south equatorial Pacific, *Mausam*, *40*, 369–380.

Mooley, D. A., B. Parthasarathy, and G. B. Pant (1986), Relationship between all-India summer monsoon rainfall and location of ridge at 500-mb level along 75°E, *J. Clim. Appl. Meteorol.*, *25*, 633–640, doi:10.1175/1520-0450(1986)025<0633:RBISMR>2.0.CO;2.

Moss, M. E., C. P. Pearson, and A. I. McKerchar (1994), The Southern Oscillation index as a predictor of the probability of low flows in New Zealand, *Water Resour. Res.*, *30*(10), 2717–2723, doi:10.1029/94WR01308.

Nelsen, R. B. (2006), *An Introduction to Copulas*, 2nd ed., 269 pp., Springer, New York.

Pant, G. B., and B. Parthasarathy (1981), Some aspect of an association between the Southern Oscillation and Indian summer monsoon, *Arch. Meteorol. Geophys. Bioklimetol.*, *89*, 179–195.

Parthasarathy, B., A. A. Munot, and D. R. Kothawale (1994), All India monthly and seasonal rainfall series: 1871–1993, *Theor. Appl. Climatol.*, *49*, 217–224, doi:10.1007/BF00867461.

Rajeevan, M., D. S. Pai, S. K. Dikshit, and R. R. Kelkar (2004), IMD's new operational models for long-range forecast of southwest monsoon rainfall over India and their verification for 2003, *Curr. Sci.*, *86*(3), 422–431.

Rasmusson, E. M., and T. H. Carpenter (1983), The relationship between eastern equatorial Pacific sea surface temperature and rainfall over India and Sri Lanka, *Mon. Weather Rev.*, *111*, 517–528, doi:10.1175/1520-0493(1983)111<0517:TRBEEP>2.0.CO;2.

Saji, N. H., B. N. Goswami, P. N. Vinayachandran, and T. Yamagata (1999), A dipole mode in the tropical Indian Ocean, *Nature*, *401*, 360–363.

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.

Salvadori, G., and C. De Michele (2006), Statistical characterization of temporal structure of storms, *Adv. Water Resour.*, *29*, 827–842, doi:10.1016/j.advwatres.2005.07.013.

Schweizer, B., and E. E. Wolff (1981), On nonparametric measures of dependence for random variables, *Ann. Stat.*, *9*(4), 879–885, doi:10.1214/aos/1176345528.

Shukla, J., and D. A. Mooley (1987), Empirical prediction of the summer monsoon rainfall over India, *Mon. Weather Rev.*, *115*, 695–703, doi:10.1175/1520-0493(1987)115<0695:EPOTSM>2.0.CO;2.

Shukla, J., and D. A. Paolino (1983), The Southern Oscillation and long-range forecasting of the summer monsoon rainfall over India, *Mon. Weather Rev.*, *111*, 1830–1837, doi:10.1175/1520-0493(1983)111<1830:TSOALR>2.0.CO;2.

Sklar, A. (1959), Fonctions de répartition à *n* dimensions et leurs marges, *Publ. Inst. Stat. Univ. Paris*, *8*, 229–231.

Verdon, D. C., A. M. Wyatt, A. S. Kiem, and S. W. Franks (2004), Multi-decadal variability of rainfall and streamflow: Eastern Australia, *Water Resour. Res.*, *40*, W10201, doi:10.1029/2004WR003234.

Walker, G. T. (1923), Correlation in seasonal variations of weather. III: A preliminary study of world weather, *Mem. Indian Meteorol. Dep.*, *24*, 75–131.

Walker, G. T. (1924), Correlation in seasonal variations of weather. IV: A further study of world weather, *Mem. Indian Meteorol. Dep.*, *24*, 275–332.

Wang, Q. J. (2001), A Bayesian joint probability approach for flood record augmentation, *Water Resour. Res.*, *37*(6), 1707–1712, doi:10.1029/2000WR900401.

Wilks, D. S. (2006), *Statistical Methods in the Atmospheric Sciences*, 627 pp., Elsevier, New York.

Zhang, L., and V. P. 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).

Zubair, L. (2003), El Niño–Southern Oscillation influences on the Mahaweli Streamflow in Sri Lanka, *Int. J. Climatol.*, *23*, 91–102, doi:10.1002/joc.865.

R. Maity, Department of Civil Engineering, Indian Institute of Technology Bombay, Mumbai 400076, India. (rajib@civil.iitb.ac.in)

D. Nagesh Kumar, Department of Civil Engineering, Indian Institute of Science, Bangalore 560012, India. (nagesh@civil.iisc.ernet.in)