A Spatio-Temporal Drought Analysis for the Midwestern US

Shih-Chieh Kao[1], Rao S. Govindaraju[2], and Dev Niyogi[3]

[1] Computational Sciences and Engineering Division, Oak Ridge National Laboratory, Oak Ridge 37831-6017, USA; email: kaos@ornl.gov
[2] Civil Engineering, Purdue University, West Lafayette, IN 47907, USA; email: govind@purdue.edu
[3] Agronomy and Earth and Atmospheric Sciences, Indiana State Climate Office, Purdue University, West Lafayette, IN 47907, USA; email: climate@purdue.edu

ABSTRACT

Droughts are prolonged abnormalities of moisture deficits that vary widely across temporal and spatial scales. Many hydrometeorologic variables are used to monitor the status of a drought. However, because of the dependence structure between all affecting variables under various temporal windows, an integrated spatio-temporal analysis of droughts cannot be easily achieved. In this study, a copula-based drought analysis was performed by using long-term monthly precipitation dataset for the upper Midwest United States. The spatio-temporal dependence relationships between various drought variables were investigated, and their joint probability distribution was constructed by combining drought marginals and the dependence structure. A copula-based joint deficit index (JDI) was adopted for an objective (probability-based) description of the overall drought status and compared to the Palmer drought severity index results. Results from the copula-based JDI provide information for drought identification, and further allow a month-by-month assessment for future drought recovery.

KEY WORDS: Copulas, Drought, Standardized Precipitation Index, Palmer Drought Severity Index

INTRODUCTION

Drought has been a challenging topic in water resources management. It is perceived as one of the most expensive and least understood natural disasters. Drought impacts tend to be more severe in areas such as the mid-western United States, where agriculture is the major economic driver. Droughts are categorized based on their impacts and duration. For example, meteorological droughts consider deficits in precipitation, agricultural droughts primarily consider deficits in soil moisture, and hydrologic droughts respond to streamflow deficits (Dracup *et al.*, 1980). Though these different types of deficits are generally positively correlated and are likely responding to the same trigger, they exhibit diverse temporal and spatial scales. An integrated drought indicator, that covers multiple types of deficits over different temporal scales, is therefore difficult to develop owing to the complicated dependencies in the variables that are used to characterize droughts.

Given their somewhat nebulous nature, the status of droughts is often assessed by various indices that are derived from hydrometorologic variables. Palmer (1965) proposed a moisture index (Palmer Drought Severity Index, PDSI) based on water budget accounting using precipitation and temperature data. PDSI soon became a popular choice for drought assessment and is widely used even today (Dalezios *et al.*, 2000; Kim *et al.*, 2003). Another popular index - Standardized Precipitation Index (SPI) was introduced by McKee *et al.* (1993). Based on a given window size, the rainfall depth is transformed to its corresponding cumulative probability, and then mapped onto the standard normal scale. The probabilistic nature of SPI allows it to be comparable among various locations and variables, and it can be further interpreted in terms of recurrence intervals (return periods). Nevertheless, to judge an overall drought status, different SPIs with multiple temporal scales (e.g. 3-, 6-, 9-, 12-month) need to be examined.

Despite the availability of different drought indices, no single index is ideal for characterizing droughts. The drought status that is assessed from one indicator often does not correspond well with that obtained from another. Therefore, to successfully assess the various drought characteristics, information from various sources need to be examined simultaneously. This is currently successfully managed by the US Drought Monitor where the severity of a drought (D0 ~ D4) is determined based on various indicators (PDSI, CPC Soil Moisture, USGS weekly, Percent of normal, SPI, and VCI), and from human input as described by Svoboda *et al.* (2002). Nevertheless, though an Objective Blend of Drought Indicators (OBDI, a linear weighted average of several indicators) is adopted as a measure of the overall severity, the decision of final drought status relies heavily on the subjective judgment of many people, instead of invariant objective standards. Because of the subjective intervention in specification of drought status, a rigorous analysis of the physical and probabilistic characterization of droughts is challenging.

There is a need to develop objective standards for analyzing records and specifying drought status based on multiple variables. So far, one of the major stumbling blocks to such an approach has been our inability to describe the complicated dependent relationships between various drought-related variables. To accomplish this goal, this study employed copulas to identify and construct the dependence structure of droughts. Building on the analysis, we propose a new copula-based drought indicator that enables the computation of a probability-based overall water deficit index from multiple drought-related quantities (or indices). By invoking copulas, we construct joint distributions of droughts so that marginal distributions of relevant variables and their dependence structures can be fully preserved. The joint distribution affords an objective description of the overall deficit, and assists the computation of probabilistic quantities such as return period and associated risk of a drought.

COPULAS

Over the last decade, copulas have emerged as a powerful approach in simplifying multivariate stochastic analysis. Sklar (1959) showed that for d-dimensional continuous random variables $\{X_1,..., X_d\}$ with marginal cumulative distribution functions (CDFs) $u_j = F_{X_j}(x_j)$, $j = 1,..., d$, there exists one unique *d*-copula $C_{U_1,...,U_d}$ such that:

$$H_{X_1,...,X_d}(x_1,...,x_d) = C_{U_1,...,U_d}(u_1,...,u_d) \tag{1}$$

where u_j is the j^{th} marginal and $H_{X_1,...,X_d}$ is the joint-CDF of $\{X_1,..., X_d\}$. Copulas $C_{U_1,...,U_d}$ can be regarded as a transformation of $H_{X_1,...,X_d}$ from $[-\infty,\infty]^d$ to $[0,1]^d$, in which the marginal distributions are segregated from $H_{X_1,...,X_d}$. Hence, $C_{U_1,...,U_d}$ becomes only relevant to the association between variables, and it gives a complete description of the entire dependence structure. This approach is available for all existing joint distributions, and it further provides a general method for constructing suitable joint distributions. Kao and Govindaraju (2007) showed that copulas allow for easier computation of probabilistic quantities such as means and standard deviations of rainfall excess. The detailed theoretical background and descriptions for the use of copulas can be found in Nelsen (2006). Specific properties that are central to the implementation of study are described below.

Choices of Copulas for Higher Dimensional Joint Distributions

Previous studies have indicated that copulas perform well for bivariate problems, and in particular Archimedean copulas have been a popular choice. However, the direct extension of Archimedean copulas to higher orders (>2) is very limited because they impose severe restrictions on the pair-wise mutual dependencies that can be accommodated, and compatibility conditions are far more difficult to satisfy in higher dimensional problems (Kao and Govindaraju, 2008). Only few parametric copula families with general applicability such as meta-elliptical copulas are available at higher dimensions (Genest *et al.*, 2007). Nevertheless, the computational burden associated with meta-elliptical copulas increases rapidly with increasing dimensions since explicit expressions exist only for the copula densities and not for the copulas themselves. The dimension of drought dependence structure is expected to be quite large to capture the temporal behavior, and consequently nonparametric empirical copulas are chosen in this study.

Similar to the concept of plotting position formula used in univariate statistical analysis (e.g. Weibull formula), empirical copulas are rank-based empirical joint cumulative probability measures (Nelsen, 2006). For sample size *n*, the *d*-dimensional empirical copula C_n is:

$$C_n\left(k_1/n, k_2/n,..., k_d/n\right) = a/n \tag{2}$$

where *a* is the number of samples $\{x_1,..., x_d\}$ with $x_1 \le x_{1(k_1)}$, …, $x_d \le x_{d(k_d)}$, and $x_{1(k_1)}$, …, $x_{d(k_d)}$ with $1 \le k_1,...,k_d \le n$ are the order statistics from the sample. Empirical copulas are mostly used for model verification and are treated as the observed (real) dependence structure. When a sufficiently large sample size is available, empirical copulas can be used to construct non-parametric joint empirical probability distributions, which tend to be more computationally efficient. This is a desirable feature for drought analysis.

Distribution Function of Copulas - K_C

For given d-variate sample marginals $\{u_{1x},...,u_{dx}\}$, a copula $C_{U_1,...,U_d}(u_{1x},...,u_{dx})$ is the cumulative probability measure $P[U_1 \le u_{1x},...,U_d \le u_{dx}] = q$. One can expect that there exist other sample marginals $\{u_{1y},...,u_{dy}\}$ with the same value of cumulative probability $C_{U_1,...,U_d}(u_{1y},...,u_{dy}) = q$. If this cumulative probability q is treated as an indicator, i.e. events with same value q are assumed to cause similar impact (e.g. in drought analysis it can be defined as joint deficit status, a smaller q implying overall drier conditions), then it will be of interest to know what the probability of random event $\{U_1,...,U_d\}$ with $C_{U_1,...,U_d}(U_1,...,U_d) = q$, or $C_{U_1,...,U_d}(U_1,...,U_d) \le q$. In this context, the distribution function K_C of copulas can be defined as the probability measure of the set $\{(U_1,...,U_d) \in [0,1]^d \mid C_{U_1,...,U_d}(U_1, ...,U_d) \le q\}$:

$$K_C(q) = P[C_{U_1,...,U_d}(U_1,...,U_d) \le q] \tag{3}$$

Salvadori and De Michele (2004) adopted K_C for defining a secondary return period for bivariate Archimedean copulas. The most appealing feature of K_C is that it can help project multivariate information onto a single axis. Though an analytical expression of K_C might not exist for non-Archimedean copulas, it can be numerically constructed from Monte Carlo simulations, and then the empirical distribution function K_C may be constructed as (example shown in Kao and Govindaraju, 2008):

$$K_{C_n}(l/n) = b/n \tag{4}$$

where b is the number of samples $\{x_1,..., x_d\}$ with $C_n(k_1/n,...,k_d/n) \le l/n$. K_C allows us to compute the probabilistic measure of the joint deficit status, which can be further translated to a joint drought index.

DATA USED IN THIS STUDY

To construct reliable multivariate statistical models, sufficiently long observations are desirable. In the proposed study, the Time Bias Corrected Divisional Dataset (TD-9640, http://www7.ncdc.noaa.gov/CDO/CDODivisionalSelect.jsp#) is adopted for the mid-western US (including states of IL, IN, IA, KS, MI, MN, MO, NE, ND, OH, SD, and WI). The data consists of monthly average temperature, precipitation, SPI and PDSI of each climate division within US from January 1895 to present (see Karl (1986) and Karl *et al.* (1986) for more details), and also provides an areal averaged value of the drought status. To investigate the droughts on a finer scale, the precipitation records over Indiana were obtained from the daily surface dataset (TD 3200) of cooperative stations (COOP) from National Climate Data Center (NCDC). After data processing (such as quality assuring with neighboring stations), a total number of 73 stations with record lengths greater than 80 years were used in the analysis discussed ahead. Monthly precipitation was computed based on aggregated daily values. In cases where data was missing for the entire month (around 1.22% of the entire precipitation dataset), it was replaced by the historic mean of that specific month (assuming normal moisture status for that unknown month). We adopted copulas for conducting at-site analyses to capture temporal evolution of droughts, and used spatial interpolation for assessing the behavior over the regional scale.

COPULA-BASED JOINT DROUGHT DEFICIT INDEX

In this study, the standardized precipitation index (SPI) introduced by McKee *et al.* (1993) was adopted for developing a copula-based joint drought deficit index. There are several reasons for adopting SPI: (1) it can be applied to precipitation, streamflow, and other variables as well, (2) it does not incur model assumptions that are typical for other indices such as PDSI, and (3) it is a probability measure of cumulative precipitation by definition. Therefore, drought severity (dryness) is normalized in terms of probabilities that can be compared between various temporal scales, locations and variables. Since the current SPI approach cannot account for seasonal variability, an amendment proposed by Kao and Govindaraju (2009) was adopted in order to obtain a more statistically sound SPI. For a *w*-month SPI, the corresponding *w*-month marginal u_w can be expressed as

$$u_w = \phi(\mathrm{SPI}_w) \tag{5}$$

in which ϕ is the standard Gaussian CDF. Referring to the definition of drought categories used in US Drought Monitor (Svoboda *et al.*, 2002), Table 1 shows the range of SPI values along with their probabilities of occurrence and corresponding drought conditions.

Table 1. The categories used in US Drought Monitor along with the corresponding SPI values.

SPI Values	Prob. of Occurrence (%)	Drought Condition	Drought Monitor Category
-0.84 ~ -0.52	20 ~ 30	Abnormally dry	D0
-1.28 ~ -0.84	10 ~ 20	Drought - moderate	D1
-1.64 ~ -1.28	5 ~ 10	Drought - severe	D2
-2.05 ~ -1.64	2 ~ 5	Drought - extreme	D3
< -2.05	< 2	Drought - exceptional	D4

Droughts have varying durations and a single SPI cannot capture the overall status. By invoking copulas, the dependence structure of selected SPIs was constructed, and the overall deficit status was expressed as joint cumulative probability (a lower probability measure corresponds to overall dry conditions) thus allowing for a more comprehensive assessment of droughts. To capture the short- as well a the long-term droughts, window sizes from 1-, 2- to 12-month ($w = 1, 2, \ldots, 12$) were selected. Empirical copulas (Eq. 2) were adopted to construct the dependence structure for precipitation marginals $\{u_1, u_2, ..., u_{12}\}$. Since the record lengths are quite large (>1000 data points for most of stations and climate divisions), empirical copulas were assumed to provide reliable results. The choice of $\{u_1, u_2, ..., u_{12}\}$ in forming high dimensional copulas increases the complexity of the dependence model. Nevertheless, it was needed because drought durations exhibit wide temporal variations, and only by encompassing various durations (e.g., from 1- to 12-month) the drought characteristics can be analyzed. An annual cycle also accounts for seasonal effects naturally. Moreover, this construct allows for a month-by-month assessment for future conditions, as will be shown later.

After a copula has been constructed, it yields the cumulative probability $P[U_1 \le u_1,...,U_{12} \le u_{12}] = q$. Clearly, when a drought occurs, most of the marginals u_w will be small and result in a small value of *q*. Therefore, the cumulative probability *q* can be viewed as the *joint deficit status* with a smaller *q* implying overall drought conditions while a higher *q* would imply overall wet conditions. We note that the joint deficit status *q* is linked to the given set of marginals and is only comparable to other *q* with the same set of marginals. For instance, q_1 for $\{u_1, u_2,..., u_{12}\}$ and q_2 for $\{u_1, u_2,..., u_{11}\}$ are not comparable. Therefore, a more general index derived from *q* is desirable. By assuming that events with the same value of *q* will have similar joint drought severity, it will be of interest to know the cumulative probability for events with joint deficit status less than or equal to a given threshold *q* (i.e., $P[C_{U_1,...,U_{12}}(U_1,...,U_{12}) \le q]$). The distribution function of copulas K_C (Eq. 3) provides this probability. Therefore, a *joint deficit index* (JDI) can be defined analogously to SPI:

$$\mathrm{JDI} = \phi^{-1}(K_C(q)) = \phi^{-1}(P[C_{U_1,...,U_{12}}(U_1,...,U_{12}) \le q]) \qquad (6)$$

where positive JDI ($0.5 < K_C < 1$) implies overall wet conditions, negative JDI ($0 < K_C < 0.5$) indicates overall dry conditions, and JDI zero ($K_C = 0$) indicates normal conditions. In other words, JDI is based on the cumulative probability of joint deficit status *q*. An extreme drought will result in a small *q*, and the JDI will correspondingly yield a low probability. Since JDI is on an inverse normal scale (same as SPI), the classifications listed in Table 1 can be adopted for JDI as well.

An illustration of the JDI using precipitation marginals of station Alpine 2 NE in Indiana (COOPID: 120132) is shown in Figure 1. JDI, SPI_w, $w = 1$, 2, …, 12 and the corresponding 12-month precipitation values are presented for four select cases. In Fig. 1(a), SPI_w values observed in all window sizes for June 1988 report large precipitation deficits indicating a severe drought, which is also noted in the JDI. Nevertheless, only by invoking JDI can the overall deficit status be expressed as a probability-based index. Fig. 1(b) shows an opposite case, in which JDI and SPI_w observed in all window sizes for August 1985 report sufficient precipitation. One can notice that JDI is slightly higher than all other SPI_w in this instance. Since JDI is based on the joint probability of all SPI_w, it suggests that the joint behavior of a drought cannot be assessed by a simple weighted average of SPI_w due to the effect of dependence structure. Fig. 1(c) shows a case of emerging drought in February 1978, in which SPI_w is quite small due to the large precipitation deficit in February while other SPI_w are above normal. This case is difficult to interpret since most of the indices do not capture droughts in a timely manner. The JDI reflects a drought condition based on the entire dependence structure. Fig. 1(d) shows a prolonged drought in Oct. 1988, in which SPI_1 and SPI_2 report sufficient precipitation in September and October while other SPI_w report precipitation deficit due to the preceding drought. Therefore, an important feature of JDI is that the overall deficit status is based on the dependence structure of deficit indices with various temporal windows. When the deficit indices are found to be uniformly low, the resulting joint drought index will be extremized due to its rareness.

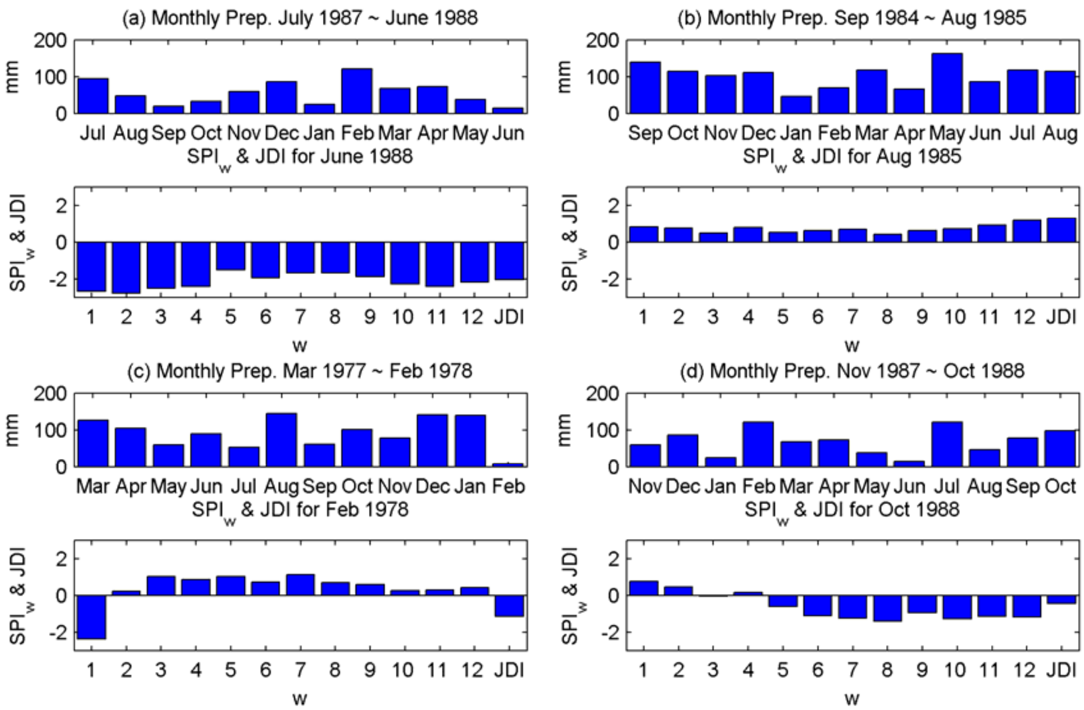

Figure 1. Illustration of JDI and SPI_w, w = 1, 2,…, 12 of precipitation station Alpine 2 NE, Indiana (COOPID: 120132) for four selected cases.

COMPARISON BETWEEN VARIOUS DROUGHT INDICES

To gain more insights into the proposed drought index, a comparison between JDI, SPI_1, SPI_3, SPI_6, SPI_9, SPI_{12} and PDSI was performed using the TD-9640 dataset for the Midwest. Since PDSI has been included in TD-9640, only JDI and modified SPIs were computed for each climate division. Spearman's rank correlation coefficients between different indices were constructed and are shown in Table 2.

Table 2. Average Spearman's rank correlation coefficient between various drought indices of the entire Midwest

	SPI_1	SPI_3	SPI_6	SPI_9	SPI_{12}	JDI
JDI	0.72	0.80	0.79	0.76	0.70	
PDSI	0.45	0.66	0.75	0.77	0.76	0.79

The comparison indicates that the correlation between JDI and SPIs are generally high (0.7 ~ 0.8). This is anticipated since JDI is built on SPIs with various window sizes. In other words, JDI represents the overall drought status with respect to different temporal scales. We then compare both JDI and SPI to the commonly used PDSI, which is based on a somewhat different water accounting approach. While the correlation between PDSI and SPIs increases with window sizes, it is of interest to note that the JDI shows good similarity to PDSI, especially for the eastern part of Midwestern US as illustrated in Figure 2. Though a more in-depth analysis should be conducted to see why these two different approaches can independently produce similar results, it indicates that JDI can effectively combine SPIs to form an overall drought index. The JDI approach can also be applied to other mixtures of indices, and is worthy for further investigation.

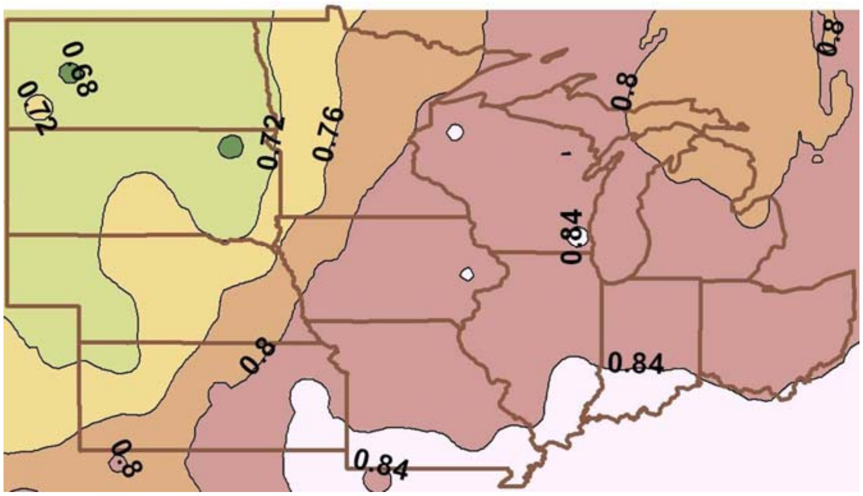

Figure 2. Spearman's rank correlation between JDI and PDSI of the Midwest

USING JDI TO ASSESS THE POTENTIAL OF FUTURE DROUGHTS

The adoption of high dimensional marginals facilitates a month-by-month future drought potential assessment since JDI is based on temporal dependence structure of deficits. Therefore, it will be of interest to know under given current conditions, what amount of precipitation is required in the following months to bring the joint deficit status to normal (JDI = 0 or $K_C = 0.5$). In other words, how much precipitation is required in a future time horizon to recover from an existing drought? This can be achieved via an iterative procedure described in Kao and Govindaraju (2009). A regional illustration of the required precipitation for July 1988 to achieve normality based on the observations made from August 1987 to June 1988 and the corresponding probability of exceedance is shown in Figure 3. As indicated in Fig. 3, a majority of Indiana would have needed over 150-mm of rain in July 1988. Based on historical July precipitation, this information can be further transformed into exceedance probability, and it suggests that the probability for recovering to normal conditions is small (less than 0.1 for most of the state). Such drought maps can be an effective way of relaying drought information, as most of the current indices are not amenable to statistical interpretation and are artificially converted into drought severity levels.

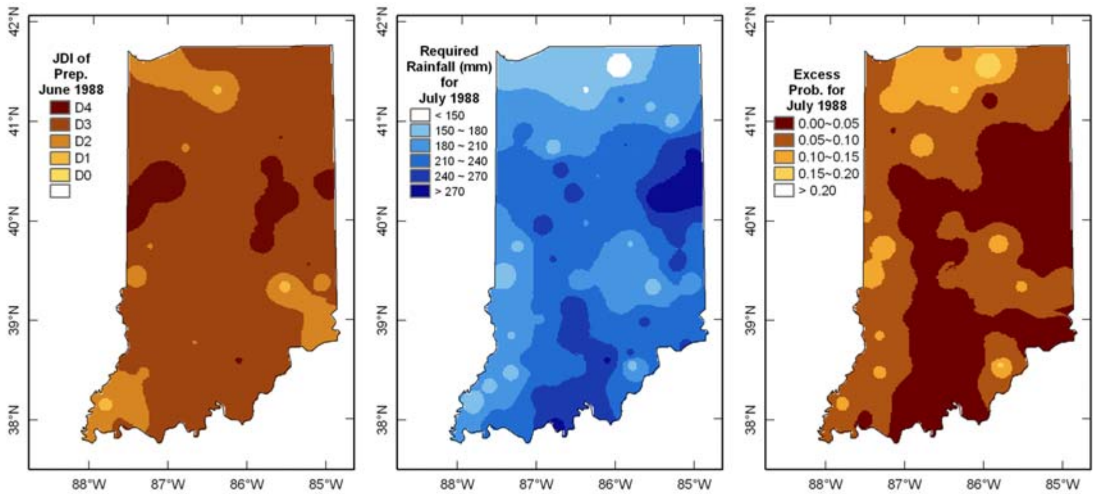

Figure 3. Illustration of (a) left: JDI for June 1988, (b) center: required precipitation for July 1988 to achieve normal status ($K_C = 0.5$ or JDI = 0), and (c) right: the corresponding excess probability of Indiana based on the observations made from Aug 1987 to June 1988

CONCLUSIONS

The complex relationships between drought-related variables have hindered characterization of overall drought status in the past. However, copulas provide a promising method for characterizing the complicated dependence structure by using a modified SI based on observed rainfall and streamflow data. The dependence structures of precipitation marginals with window sizes varying from 1- to 12-months were constructed via empirical copulas. The reliability of empirical copulas was founded on the large sample sizes adopted in this study. Study results can be summarized as follows:

(1) A joint deficit index (JDI) was proposed in this study using the distribution function K_C of copulas. The JDI offers a probability-based drought index from a set of SPIs with various temporal window sizes. This is an improvement over conventional drought status indices that could be estimated through subjective judgments or by linear weighted SPIs. Besides providing objective description of the overall drought status, the JDI was shown to be capable of capturing both emerging and prolonged droughts in a timely manner.

(2) Comparing to SPIs with various window sizes, the JDI is found to resemble behavior of PDSI especially for the eastern part of Midwestern US. Thus, JDI can effectively combine drought status with various temporal scales to form an overall deficit index. The similarity in the performance of two independently derived indices (PDSI and JDI) deserves further investigations.

(3) Drought severity can be assessed through the proposed JDI approach. For instance, the required precipitation for achieving normal conditions (JDI = 0) can be estimated. The required rainfall depth along with its exceedance probability provides good interpretation of drought status and mitigation needs.

The utilization of copulas in drought characterization was demonstrated in this study. Copulas can play an important role in drought analysis and perhaps in other hydrometeorological studies as they enable multidimensional stochastic analysis.

ACKNOWLEDGEMENTS

Support for this study was provided by National Science Foundation (grant numbers OCI:0815811 and DBI:0619086). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

Dalezios, N. R., A. Loukas, L. Vasiliades, and E. Liakopoulos (2000), Severity-duration- frequency analysis of droughts and wet periods in Greece, *Hydrol. Sci. J.*, 45(5), 751-770.

Dracup, J. A., K. S. Lee, and E. G. Paulson, Jr. (1980), On the definition of droughts, *Water Resour. Res.*, *16*(2), 297-302.

Genest, C., A.-C. Favre, J. Béliveau, and C. Jacques (2007), Metaelliptical copulas and their use in frequency analysis of multivariate hydrological data, *Water Resour. Res.*, *43*, W09401.

Kao, S.-C., and R. S. Govindaraju (2007), Probabilistic structure of storm surface runoff considering the dependence between average intensity and storm duration, *Water Resour. Res.*, *43*, W06410.

Kao, S.-C. and R. S. Govindaraju (2008), Trivariate statistical analysis of extreme rainfall events via Plackett family of copulas, *Water Resour. Res.*, *44*, W02415.

Kao, S.-C. and R. S. Govindaraju (2009), A copula-based joint deficit index for droughts, submitted to *J. Hydrol.*, in revision.

Karl, T. R. (1986), The sensitivity of the Palmer drought severity index and Palmer's Z-index to their calibration coefficients including potential evapotranspiration, *J. Appl. Meteorol.*, *25*(1), 77-78.

Karl, T. R., C. N. Williams Jr., P. J. Young, and W. M. Wendland (1986), A model to estimate the time of observation bias associated with monthly mean maximum, minimum and mean temperatures for the United States, *J. Appl. Meteorol.*, *25*(2), 145-160.

Kim, T.-W., J. B. Valdés, and C. Yoo (2003), Nonparametric approach for estimating return periods of droughts in arid regions, *J. Hydrol. Eng.*, *8*(5), 237-246.

McKee, T. B., N. J. Doeskin, and J. Kleist (1993), The relationship of drought frequency and duration to time scales, in Proceedings of the 8th Conference on Applied Climatology, 179-184, January 17-22, Anaheim, California.

Nelsen, R. B. (2006), *An Introduction to Copulas*, Springer, New York.

Palmer, W. C. (1965), Meteorological drought, Res. Paper No. 45, Weather Bureau, Washington, D. C.

Salvadori, G., and C. De Michele (2004), Frequency analysis via copulas: Theoretical aspects and applications to hydrological events, *Water Resour. Res.*, 40, W12511.

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

Svoboda, M., D. LeComte, M. Hayes, R. Heim, K. Gleason, J. Angel, B. Rippey, R. Tinker, M. Palecki, D. Stooksbury, D. Miskus, and S. Stephens (2002), The drought monitor, *B. Am. Meteorol. Soc.*, *83*(8), 1181-1190.