6-21-2011

Copula-based Approaches to Characterization of Droughts

Xin Lu Tan
Purdue University, tanx@purdue.edu

Sergey Kirshner
Department of Statistics, Purdue University, skirshne@purdue.edu

Rao S. Govindaraju
Purdue University, govin@purdue.edu

Follow this and additional works at: http://docs.lib.purdue.edu/ddad2011

http://docs.lib.purdue.edu/ddad2011/22
Copula-based Approaches to Characterization of Droughts

Xin Lu Tan¹, Sergey Kirshner¹, Rao S. Govindaraju²

¹Department of Statistics, Purdue University, West Lafayette, IN 47907, USA, tanx@purdue.edu, skirshne@purdue.edu
²School of Civil Engineering, Purdue University, West Lafayette, IN 47907, USA, govind@purdue.edu

Background:
- Droughts are used to reflect water shortages. A precise mathematical definition of droughts does not exist, and a drought is loosely worded as “a prolonged absence or marked deficiency of precipitation.”
- Numerous drought indices have been developed using different hydrologic variables.
- In this study, we explore the potential of copulas in describing the joint water deficit over multiple stations in a region.
- Focus on precipitation data for stations within Indiana.

Goal: To develop a joint water deficit index for multiple locations to enable regional quantification of droughts.

What is SPI?
- Among the various drought indices, the Standardized Index (SI) (McKee et al., 1993) has gained wide recognition due to its computational simplicity, (2) versatility in comparing different hydrologic variables.
- When applied to precipitation data (from one station), it is called the Standardized Precipitation Index (SPI).
- SPI is a statistical index designed to provide a means of comparing the current precipitation situation with historical precipitation at the same location.
- SPI can be calculated for any given period of time (e.g., monthly, quarterly, or annually) and can be used to assess the severity of drought conditions relative to historical conditions.

The SPI is calculated using the following formula:

\[ SPI = \frac{P - \mu}{\sigma} \]

where:
- \( P \) is the precipitation for a given period of time
- \( \mu \) is the mean precipitation for the same period
- \( \sigma \) is the standard deviation of precipitation for the same period

Standardized Precipitation Index (SPI)

The SPI is a dimensionless index that ranges from negative to positive values. Positive values indicate periods of wetter-than-normal conditions, while negative values indicate periods of drier-than-normal conditions.

A negative value of SPI indicates precipitation less than median rainfall (0 < rank < 0.5), and the magnitude of departure from 0 indicates the severity of drought.

Copula:

- Sklar (1959) showed that for a d-dimensional continuous random variables \( (x_1, …, x_d) \) with joint-CDF \( F \) and marginal CDFs \( u_j = F(x_j) \), \( j = 1, …, d \), there exists one unique d-copula \( C \) such that \( H(x_1, …, x_d) = C(u_1, …, u_d) \).
- Copulas can be viewed as a joint distribution over ranks of the individual variables. Since SPI is based on the probability integral transform of precipitation, a copula for observations at multiple precipitation locations can be viewed as a joint distribution of functions of SPIs at these locations.
- A statistic of such precipitation-based copula can serve as a joint water deficit index.

Proposed Index: Kendall Distribution Function

- If \( X \) and \( Y \) are continuous random variables with joint-CDF \( H \), then the Kendall distribution function of \( (X, Y) \) is the distribution function of the random variable \( H(X, Y) = C(U, V) \) (Nelsen et al., 2003), i.e., the Kendall distribution function \( K(t) = P(C(U, V) \leq t) \).
- Kendall distribution function serves as a candidate drought index.
- A statistic of such precipitation-based copula can serve as a joint water deficit index.

Challenges:
- We are using the NCDC precipitation stations within Indiana. [ftp://ftp.ncdc.noaa.gov/pub/data/inventories/COOP.TXT]
- The data set contains daily precipitation data for 372 stations (variables), spanning from year 1894 to year 2009 with missing data (less than 121 year month monthly observations).
- Curse of dimensionality: the number of samples required to estimate a d-dimensional distribution grows exponentially in d.

Future Work:
- A future challenge is to convert \( K(t) \) to drought classes (i.e. normal, moderate, severe, and so forth).
- However, the estimation of \( K(t) \) is difficult (numerical) unless the copula falls within the class of Archimedean copulas, but on the other hand, Archimedean copulas generally do not decompose along a tree-structured graphical model.