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, govind@purdue.edu

Source: http://thecalloftheland.files.wordpress.com/2009/02/drought.jpg

Copula-based Approaches to Characterization of Droughts

Xin Lu Tan[1], Sergey Kirshner[1], Rao S. Govindaraju[2]

[1]Department of Statistics, Purdue University, West Lafayette, IN 47907, USA, tanx@purdue.edu, skirshne@purdue.edu

[2]School of Civil Engineering, Purdue University, West Lafayette, IN 47907, USA, govind@purdue.edu

PURDUE UNIVERSITY

PURDUE ENGINEERING

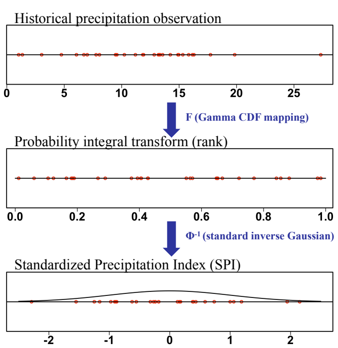
Background:

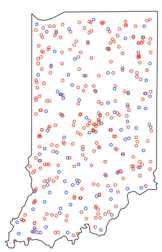
- Droughts are used to reflect water shortages. A precise mathematical definition of droughts does not exist, and a drought is loosely worded as “a prolonged absence or marked deficiency of precipitation”.
- Numerous drought indices have been developed using different hydrologic variables.
- In this 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.

Given the high dimensional nature of the data, we combine tree-structured graphical models with copulas to design a new index for drought characterization.

What is SPI?

- Among the various drought indices, the Standardized Index (SI) (McKee et al., 1993) has gained wide recognition due to its (1) computational simplicity, (2) versatility in comparing different hydrologic variables at different time scales.
- When applied to precipitation data (from one station), it is called Standardized Precipitation Index (SPI).



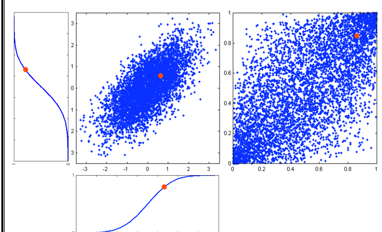


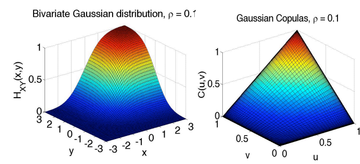
Stations over Indiana
Blue: ≥ 50 years
Red: ≤ 50 years

- 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.
- We aim to build on SPI to capture drought for a region, i.e., to construct a joint index for multiple locations.

Copula:

- Sklar (1959) showed that for a *d*-dimensional continuous random variables $\{x_1, \ldots, x_d\}$ with joint-CDF H and marginal CDFs $u_j = F_j(x_j)$, $j = 1, \ldots, d$, there exists one unique *d*-copula C such that $H(x_1, \ldots, x_d) = C(u_1, \ldots, u_d)$.
- 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.

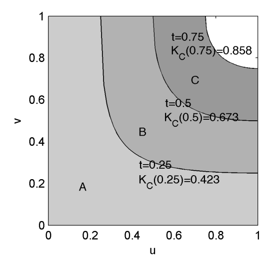




The joint cdf of Bivariate Gaussian distribution with ρ = 0.1 and its corresponding copula.

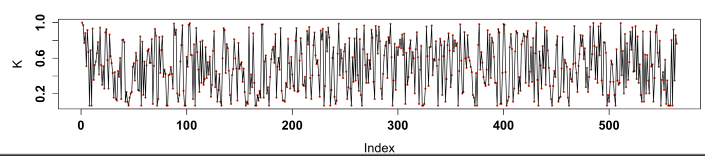
Proposed Index: Kendall Distribution Function

- 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 $K_C(t) = P(C(U, V) \le t)$.
- Kendall distribution function serves as a candidate drought index. This index computes the proportion of the events for which the cumulative distribution function (cdf) does not exceed the cdf for the given event.
- A similar index was proposed by Kao and Govindaraju 2010 for modeling a joint index for different temporal scales of drought.



Estimation

- We consider two settings for the index estimate $K_C(t)$ from historical precipitation observations
  - Parametric: assume that historical observations come from an Archimedean copula - a class of parametric copulas characterized by a (parametrized) generator function φ. In the bivariate case, $C(u, v) = \varphi^{[-1]}(\varphi(u) + \varphi(v))$. We first estimate the generator function φ and then evaluate C.
  - Non-parametric: C can be estimated using empirical copula. $K_C(t)$ can be estimated as the proportion of observations with C ≤ t.



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 1884 to year 2009 with missing data (less than 121 year × 12 month monthly observations).
- ***Curse of dimensionality***: the number of samples required to estimate a *d*-dimensional distribution grows exponentially in *d*.

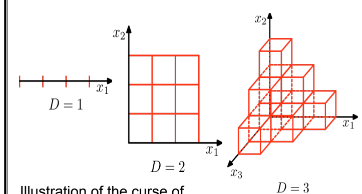
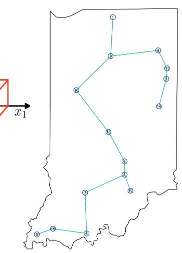


Illustration of the curse of dimensionality, showing how the number of regions in a regular grid grow exponentially w.r.t. dimensionality D (Bishop, 2008).



- To address the challenge, we (1) concatenate neighboring stations exhibiting similar characteristics, and (2) introduce conditional independence assumptions using trees, a graphical model to reduce dimension of the joint distribution.
- In graphical model, separation of nodes in the graph corresponds to conditional independence of the separated variables conditioned on the separating variables. (Markov property)
- Introducing a tree-structure into the distribution enables the joint density $c(u_1, \ldots, u_d)$ to factorize as a product of $c(u_i, u_j)$ for each (u_i, u_j) belonging to the set of edges in the tree.
- Then, estimating a tree-structured copula reduces to estimation of bivariate copulas for its edge pairs of variables.
- Empirical tree-structured copula can be computed using the sum-product algorithm.

Future Work:

- A future challenge is to convert $K_C(t)$ to drought classes (i.e. normal, moderate, severe and so forth).
- However, the estimation of $K_C(t)$ 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.

This work has been supported by the award AGS-1025430 by the National Science Foundation, and by the DURI program at Purdue University.