Journal of Hydrology 526 (2015) 164–182

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Multivariate drought index: An information theory based approach for integrated drought assessment

Deepthi Rajsekhar [a,*], Vijay. P. Singh [b,a], Ashok. K. Mishra [c]

[a] Biological and Agricultural Engineering (BAEN) Department, Texas A & M University, College Station, TX 77840, United States
[b] Civil Engineering Department, Texas A & M University, College Station, TX 77840, United States
[c] Glenn Department of Civil Engineering, Clemson University, Clemson, SC 29634, United States

ARTICLE INFO

Article history:
Available online 26 November 2014

Keywords:
Entropy
Kernel ECA
Kernel PCA
Multivariate drought index

SUMMARY

Most of the existing drought indices are based on a single variable (e.g. precipitation) or a combination of two variables (e.g., precipitation and streamflow). This may not be sufficient for reliable quantification of the existing drought condition. It is possible that a region might be experiencing only a single type of drought at times, but multiple drought types affecting a region is quite common too. To have a comprehensive representation, it is better to consider all the variables that lead to different physical forms of drought, such as meteorological, hydrological, and agricultural droughts. Therefore, we propose to develop a multivariate drought index (MDI) that will utilize information from hydroclimatic variables, including precipitation, runoff, evapotranspiration and soil moisture as indicator variables, thus accounting for all the physical forms of drought. The entropy theory was utilized to develop this proposed index, that led to the smallest set of features maximally preserving the information of the input data set. MDI was then compared with the Palmer drought severity index (PDSI) for all climate regions within Texas for the time period 1950–2012, with particular attention to the two major drought occurrences in Texas, viz. the droughts which occurred in 1950–1957, and 2010–2011. The proposed MDI was found to represent drought conditions well, due to its multivariate, multi scalar, and nonlinear properties. To help the user choose the right time scale for further analysis, entropy maps of MDI at different time scales were used as a guideline. The MDI time scale that has the highest entropy value may be chosen, since a higher entropy indicates a higher information content.

1. Introduction

Droughts are the costliest of all natural disasters with an estimated annual loss of $6–8 billion in the United States (Wilhite, 2000) and collectively affects more people than any other natural disaster. Thus, there is a need for developing a system to quantify, monitor and predict droughts (Mishra and Singh, 2011). However, given the wide variety of sectors affected by drought and its diverse geographical and temporal distribution, it is difficult to develop a single, precise definition for drought.

Droughts are classified into four categories: meteorological or climatological, agricultural, hydrological, and socioeconomic (The American Meteorological Society, 2004; Mishra and Singh, 2010). A prolonged deficit in precipitation leads to meteorological drought. A dryness in the surface layers (root zone), which occurs at a critical time during the growing season, can result in an agricultural drought that severely reduces crop yield, even though deeper soil levels may be saturated. The onset of an agricultural drought may follow a meteorological drought, depending on the prior moisture status of the surface soil layers. Precipitation deficits over a prolonged period that affect surface or subsurface water supply, thus reducing streamflow, groundwater, reservoir and lake levels, may lead to a hydrological and ground water drought, which will persist long after a meteorological drought has ended (Heim, 2002). The ground water drought, can be different from hydrological drought due to the involvement of complex hydrological processes (Mishra and Singh, 2010). Socioeconomic drought associates the supply and demand of some economic goods with certain elements of meteorological, agricultural, and hydrological droughts. The relationship between hydroclimatic variables and different types of droughts is complex and hence it is difficult to develop an accurate index to quantify and compare droughts.

Currently, there exist a number of drought indices that are used to represent different types of droughts. Some of the earlier drought indices include: Munger's Index (Munger, 1916),

* Corresponding author.
E-mail address: deepthir86@gmail.com (D. Rajsekhar).

http://dx.doi.org/10.1016/j.jhydrol.2014.11.031

Blumenstock's Index (Blumenstock, 1942), and Antecedent Precipitation Index(McQuigg, 1954) which are all basically precipitation based indices. In 1965, Palmer (1965) introduced the widely popular Palmer Drought Severity Index (PDSI) which is based on precipitation and temperature as input variables in a water budget model. Despite its wide usage, it has several limitations like lack of physical meaning, slowness in detecting the onset of drought events, unclear temporal scale and problems with Thornthwaite's method used for calculation of PDSI. McKee et al. (1993) introduced another popular drought index named as Standardised Precipitation Index (SPI). SPI has several advantages like comparability among various locations, and wide range of time scales ranging from 1-month to 24-months. However, multiple SPIs with various time scales may also lead to confusion in assessment of drought condition. Similar to SPI, there are other indices like Standardized Runoff Index (SRI; Shukla and Wood, 2008) and Standardized Streamflow Index (SSFI; Modarres, 2007) which use runoff and streamflow as drought indicator variables. Other commonly used indices include Crop Moisture Index (CMI; Palmer, 1968) for agricultural drought, Vegetation Condition Index (VCI; Kogan, 1995), Climate prediction center (CPC) Soil Moisture Index (SMI; Huang et al., 1996), and Standardized Precipitation Evapotranspiration Index (SPEI; Vicente-Serrano et al., 2010).

All of these indices consider one specific physical form of drought: hydrological, meteorological, or agricultural. This might not be adequate to get a comprehensive idea of the drought condition since it is dependent on multiple variables. Hence, in general it can be concluded that the drought status indicated by one drought index might not be consistent with the findings obtained while using a different drought index.

To overcome these limitations, a group of indices that consider multiple variables to represent drought were developed. The drought monitor developed by Svoboda et al. (2002) considers an Objective Blend of Drought Indicators (OBDI) which is the linear weighted average of several drought indices. Aggregated Drought Index (ADI; Keyantash and Dracup, 2004) comprehensively considers all physical forms of drought through variables like precipitation, streamflow, evapotranspiration, reservoir storage, soil moisture content and snow water content. ADI aggregates all these variables into a single time series through principal component analysis (PCA). However, the use of PCA has several limitations like linearity assumption in data transformation, and the assumption that most information is contained in those directions where input data variance is maximum. These assumptions however need not be always met in reality. Recently, bivariate drought indices have been derived using copulas to quantify the joint behavior of drought types. Kao and Govindaraju (2010) introduced a Joint Drought Index (JDI) using copula for obtaining the joint probabilities while considering precipitation and streamflow. Hao and AghaKouchak (2013b) introduced Multivariate Standardized Drought Index (MSDI) which uses copula to form joint probabilities of precipitation and soil moisture content. The use of copula for multivariate analysis is, no doubt, highly effective. However, for higher dimensional cases (i.e., more than three variables), this method will not be a feasible choice due to the lack of flexibility in modeling the dependence structure.

Feature extraction technique is an effective approach to aggregate the various drought types into a single index. The PCA, which has been commonly used in hydrology and water resources, is a popular technique that falls under the class of linear feature extraction models. Over time, other techniques were developed, which tackled the non-linearity problem through local approaches (Roweis and Saul, 2000), neural networks (Kramer, 1991), or kernel algorithms (Scholkopf et al., 1999). The kernel based methods, like the kernel principal component analysis (KPCA) and kernel partial least squares (KPLS), have attracted a lot of attention, particularly in the last decade as an effective non-linear approach for dimensionality reduction. These methods target at finding projections that maximize the variance of input data in the feature space. However, the method assumes that the maximum information that can be obtained from the input data is oriented along the direction of maximum variance. It has been proved that entropy is a much better measure of information than variance (Dionisio et al., 2007). Entropy is related to the higher order moments of a distribution, and thus, unlike the variance, it can offer a better characterization of the input data, since it uses more information from the probability distribution (Ebrahimi et al., 1999).

The objective of this study, therefore, is to make use of a kernel entropy component analysis (KECA) for extracting a drought index named as multivariate drought index (MDI) from the set of input variables representing the various physical forms of drought. We consider the variables: precipitation (P), runoff (R), evapotranspiration (ET), and soil moisture (SM), thus accounting for all the major elements in the water balance model. The method is essentially a novel feature extraction technique that combines the concept of entropy and KPCA. The KECA or entropy PCA performs dimensionality reduction by projecting the data onto those kernel principal component axes that maximally contribute to the entropy estimate of the input dataset. These axes will not necessarily correspond to the top eigenvalues or eigenvectors of the kernel matrix, as in the case of KPCA (Jenssen, 2010). The KECA thus overcomes the disadvantages of PCA and KPCA. The advantages of KECA include: (1) It does not make the linearity assumption; (2) final multivariate index is obtained in such a way that it preserves the entropy of the input data, which means it tries to preserve the maximum amount of information of the input data; and (3) unlike KPCA, it does not make the assumption that the maximum information from the input data is oriented along the direction of maximum variance. KPCA essentially preserves only the second order statistics of data set, whereas KECA preserves the higher order statistics also through the use of entropy. Additionally, this study also explored the multiscalar nature of MDI by comparing the entropy values of different temporal scales. This would guide the user to choose the most suitable time scale required for further analysis or decision making.

The paper is organized as follows. The second section deals with the study area. Section three discusses data, its sources and the description of the model used for simulating the input variables. The methodology is described in section four, followed by results in the fifth section. The sixth section discuss the results and the conclusions drawn from the study.

2. Study area

The study area considered is the state of Texas in the USA. It is the second largest state in United States with a total land area of 261,914 square miles. Because of its size and geographical location, the state has a diverse climate ranging from arid to subtropical humid. There are five distinct climate zones in Texas, namely arid, semi-arid, continental steppe, sub-tropical semi-humid and subtropical humid zones. The basic climatic pattern within Texas is fairly simple: annual mean temperature increases from north to south and annual mean precipitation increases from west to east. Hot spots are found in Rio Grande and Red River Basin, whereas the mountains in west Texas experience the coolest summertime temperatures (Nielson-Gammon, 1995). The varied physiography in Texas from forests in the east and coastal plains in the south to the elevated plateaus and basins in the north and west results in a wide variety of weather throughout the year (Benke and Cushing, 2005). The land surface elevation follows a decreasing trend from west to east, with arid climate zone covering higher