Journal of Hydrology 526 (2015) 116–126

Contents lists available at ScienceDirect

Journal of Hydrology

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

Probabilistic drought classification using gamma mixture models

Ganeshchandra Mallya [a,*], Shivam Tripathi [b], Rao S. Govindaraju [a]

[a] School of Civil Engineering, Purdue University, West Lafayette, IN 47907, USA
[b] Department of Civil Engineering, Indian Institute of Technology, Kanpur, UP 208016, India

ARTICLE INFO

Article history:
Available online 11 November 2014

Keywords:
Droughts
Standardized Precipitation Index
Probabilistic SPI
Probabilistic drought classification
Gamma mixture models
Bayesian inference

SUMMARY

Drought severity is commonly reported using drought classes obtained by assigning pre-defined thresholds on drought indices. Current drought classification methods ignore modeling uncertainties and provide discrete drought classification. However, the users of drought classification are often interested in knowing inherent uncertainties in classification so that they can make informed decisions. Recent studies have used hidden Markov models (HMM) for quantifying uncertainties in drought classification. The HMM method conceptualizes drought classes as distinct hydrological states that are not observed (hidden) but affect observed hydrological variables. The number of drought classes or hidden states in the model is pre-specified, which can sometimes result in model over-specification problem. This study proposes an alternate method for probabilistic drought classification where the number of states in the model is determined by the data. The proposed method adapts Standard Precipitation Index (SPI) methodology of drought classification by employing gamma mixture model (Gamma-MM) in a Bayesian framework. The method alleviates the problem of choosing a suitable distribution for fitting data in SPI analysis, quantifies modeling uncertainties, and propagates them for probabilistic drought classification. The method is tested on rainfall data over India. Comparison of the results with standard SPI show important differences particularly when SPI assumptions on data distribution are violated. Further, the new method is simpler and more parsimonious than HMM based drought classification method and can be a viable alternative for probabilistic drought classification.

1. Introduction

Drought classification schemes classify a drought based on its severity or intensity. Water resources planners rely on drought classification to decide drought mitigation strategies and hence weather agencies throughout the world routinely issue drought classification bulletins. For example, the US Drought Monitor releases a weekly update of drought status in U.S.A. by classifying droughts into five classes – D0–D4 with the latter representing exceptional drought. India Meteorological Department (IMD) issues drought bulletins classifying droughts into three categories, namely, mild, moderate, and severe.

The most common quantitative drought classification schemes work in two steps – first, by defining a drought index using hydro-meteorological observations and next, by categorizing droughts based on pre-defined thresholds on the index value. Examples include IMD classification that uses departure of rainfall from its long period average as a drought index, and US Drought Monitor classification that, along with other indices, uses Standardized Precipitation Index (SPI) as a drought index. Mallya et al. (2012) proposed an alternative method that does not require pre-specification of thresholds. Their method provides a probabilistic drought classification by learning thresholds from the data. Both the approaches have drawbacks arising either from the limitations of the drought index or shortcomings in the procedure for defining thresholds. The following paragraphs briefly describe some of those limitations that we have attempted to address in this work.

Drought classification schemes employ drought indices that measure degree of departure of hydro-meteorological variables, such as precipitation and streamflow, from their long-term averages. Drought indices have been used for identifying droughts and their triggers (Steinemann, 2003), assessing drought status (Kao and Govindaraju, 2010), forecasting droughts (AghaKouchak, 2014), performing drought risk analysis (Hayes et al., 2004) and studying relationship of droughts with local-scale regional hydrological variables like water quality (Sprague, 2005) and large-scale climate patterns like El Niño-Southern Oscillation (Cole and Cook, 1998; Liu and Juárez, 2001; Ryu et al., 2010). Among several drought indices proposed in the literature (Dai, 2011; Heim, 2002; Mishra and Singh, 2010), the Standardized Precipitation Index (SPI; McKee et al., 1993) is very popular because of its computational simplicity and versatility in comparing different

* Corresponding author.

http://dx.doi.org/10.1016/j.jhydrol.2014.11.008
0022-1694/

hydro-meteorological variables at different time scales. In SPI, historical observations are used to compute the probability distribution of the monthly and seasonal (4-months, 6-months, and 12-months) precipitation totals. The fitted probability distributions are then normalized using the standard inverse Gaussian function to calculate SPI values. A negative value of SPI indicates precipitation less than the median rainfall, and the magnitude of departure from zero represents the severity of a drought based on which drought classes are defined. As many drought classification schemes in the literature use SPI, they inherit its weaknesses.

Standard SPI based drought classification schemes ignore uncertainties arising from data errors, model assumptions, and parameter estimations providing discrete classification. Thus, the users are not aware of inherent uncertainties in drought classification often required for making informed decisions. Further, in the context of SPI there is an ongoing debate on the selection of the parametric distribution for fitting data. McKee et al. (1995) in their original paper on SPI recommends gamma distribution. Lloyd-Huges and Saunders (2002) found gamma distribution to be an appropriate model for Europe. Guttman (1999) suggested Pearson-III distribution as the best universal model for SPI because it provides more flexibility than the gamma distribution. Rossi and Cancelliere (2003) found normal, lognormal, and gamma distributions to be suitable for different datasets in their study. Loukas and Vasiliades (2004) investigated different theoretical distributions using Kolmogorov–Smirnov (K–S) test and Chi-squared test and found Extreme Value-I distribution to be most suitable for studying drought over Thessaly, Greece. Mishra et al. (2007) argues that different distributions may be appropriate for different drought durations (window size), and recommends K–S test for choosing an appropriate distribution. Bonaccorso et al. (2013) used Lilliefors test to choose among normal, lognormal, and gamma distributions while Russo et al. (2013) used the three parameter generalized extreme value (GEV) distribution for SPI analysis. Thus there is no consensus on the choice of distribution for SPI analysis.

Mallya et al. (2012) uses hidden Markov model (HMM) for drought classification by conceptualizing hidden states in the model to represent drought states. Their model avoided the need for specifying thresholds for drought classification and provided probabilistic drought classification by accounting model uncertainties; however, the number of hidden states (drought classes) is pre-specified. To facilitate comparison of HMM drought classification with standard methods they specified 11 hidden states. Since the number of states is imposed on the model, it is possible that for datasets with short record length the model suffers from *over-specification problem*, i.e. the model structure is more complicated than supported by the dataset. Specifically, in the HMM context, over-specification means that the number of specified hidden states are more than that needed to model the data. Over-specification can result in *parameter identification problem* leading to unreliable results.

The main objective of this paper is to propose an alternate method for probabilistic drought classification. The proposed method adapts SPI drought classification methodology by

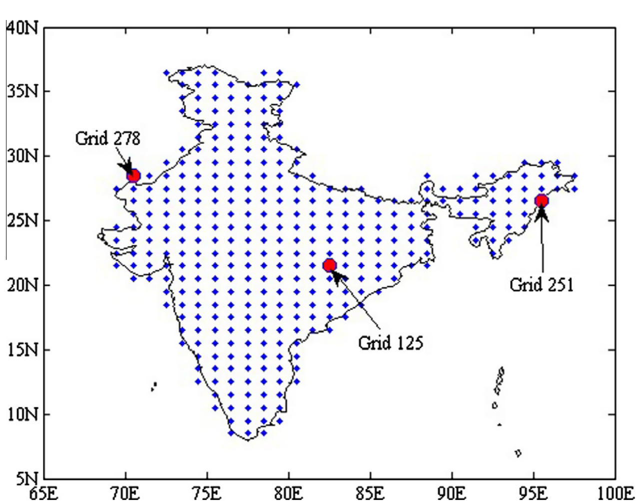

Fig. 1. Map showing the study area along with the location of grids for which rainfall data were provided by IMD.

Table 1
US Drought Monitor classification scheme. SPI ranges are prescribed for the inverse of the normal distribution. Corresponding thresholds on CDF are given in the last column.

Category	Description	SPI range	Threshold on CDF
D0	Abnormally dry	−0.5 to −0.8	0.212–0.309
D1	Moderate drought	−0.8 to −1.3	0.097–0.212
D2	Severe drought	−1.3 to −1.6	0.055–0.097
D3	Extreme drought	−1.6 to −1.9	0.023–0.055
D4	Exceptional drought	−2.0 or less	0.023 or less

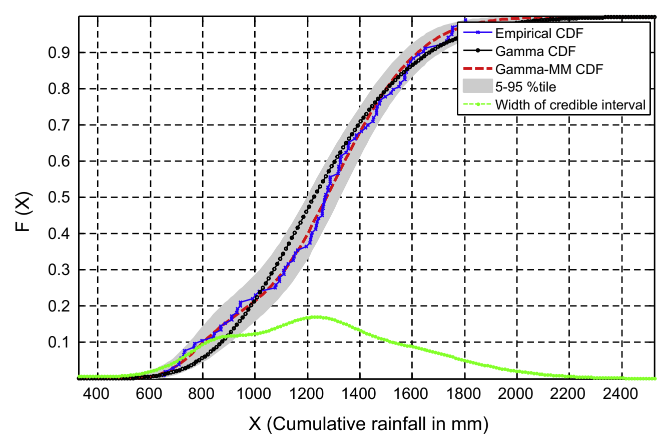

Fig. 2. Empirical CDF along with CDFs obtained by fitting gamma distribution (Gamma CDF) and gamma mixture model (Gamma-MM CDF) to the cumulative rainfall in a water-year at Grid 125. The grey band shows 5th and 95th percentile of the Gamma-MM CDF and the green dotted line shows width of its credible interval. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

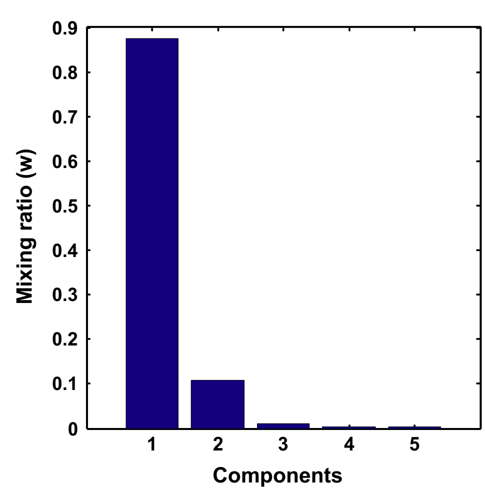

Fig. 3. Mixing ratios of the components of a Bayesian Gamma-MM. Two components are identified to be significant for characterizing water-year drought at Grid 125.