Computers and Electronics in Agriculture 152 (2018) 149–165

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Original papers

Multi-stage committee based extreme learning machine model incorporating the influence of climate parameters and seasonality on drought forecasting

Mumtaz Ali, Ravinesh C. Deo*, Nathan J. Downs, Tek Maraseni

School of Agricultural, Computational and Environmental Sciences, Institute of Agriculture and Environment, University of Southern Queensland, Springfield, QLD 4300, Australia

ARTICLE INFO

Keywords:
Standardized precipitation index
Drought forecasting
Committee model
Extreme learning machine
Particle swarm optimization based adaptive neuro fuzzy inference system
Multi-linear regression

ABSTRACT

Drought forewarning is an important decisive task since drought is perceived a recurrent feature of climate variability and climate change leading to catastrophic consequences for agriculture, ecosystem sustainability, and food and water scarcity. This study designs and evaluates a soft-computing drought modelling framework in context of Pakistan, a drought-stricken nation, by means of a committee extreme learning machine (Comm-ELM) model in respect to a committee particle swarm optimization-adaptive neuro fuzzy inference system (Comm-PSO-ANFIS) and committee multiple linear regression (Comm-MLR) model applied to forecast monthly standardized precipitation index (SPI). The proposed Comm-ELM model incorporates historical monthly rainfall, temperature, humidity, Southern Oscillation Index (SOI) at monthly lag ($t-1$) and the respective month (*i.e.*, periodicity factor) as the explanatory variable for the drought's behaviour defined by SPI. The model accuracy is assessed by root mean squared error, mean absolute error, correlation coefficient, Willmott's index, Nash-Sutcliffe efficiency and Legates McCabe's index in the independent test dataset. With the incorporation of periodicity as an input factor, the performance of the Comm-ELM model for Islamabad, Multan and Dera Ismail Khan (D. I. Khan) as the test stations, was remarkably improved in respect to the Comm-PSO-ANFIS and Comm-MLR model. Other than the superiority of Comm-ELM over the alternative models tested for monthly SPI forecasting, we also highlight the importance of the periodicity cycle as a pertinent predictor variable in a drought forecasting model. The results ascertain that the model accuracy scales with geographic factors, due to the complexity of drought phenomenon and its relationship with the different inputs and data attributes that can affect the overall evolution of a drought event. The findings of this study has important implications for agricultural decision-making where future knowledge of drought can be used to develop climate risk mitigation strategies for better crop management.

1. Introduction

Drought is a socio-economic hazard posing severe threats to groundwater reservoirs, leading to a scarcity of water resources, crop failure and socio-economic challenges (Mpelasoka et al., 2008; Riebsame et al., 1991; Wilhite et al., 2000; Deo et al., 2009). A general scarcity of sufficient palatable water due to rapid growth of human populations and an expansion of agricultural, energy and industrial sectors is also a growing concern (IPCC, 2012; McAlpine et al., 2007). Climate change further escalates the potential severity and frequency of drought events (Deo et al., 2009; McAlpine et al., 2009). Large-scale climate mode indices including the Southern Oscillation Index (SOI) is significantly correlated with fluctuations in rainfall and onset of drought (Mishra and Singh, 2010; Morid et al., 2007; Nguyen-Huy et al., 2017; Özger et al., 2012). Hence drought forecasts that incorporate climate factors can assist hydrologists, agriculturalists and resource planners in strategic decisions to address socio-economic challenges posed by a drought, particularly in cases of prolonged events (Bates et al., 2008; Deo et al., 2016a; Mishra and Singh, 2010, 2011; Wilhite and Hayes, 1998).

Traditionally, drought indices are used to measure, monitor and forecast drought. Several different indices have been developed according to their appropriateness for a given geographic region (Mishra and Singh, 2010, 2011). The Palmer (1965) Drought Severity Index (PDSI) measures the overall dryness based on precipitation and temperature datasets. PDSI is particularly useful in pointing out long-term drought events and it is not appropriate for a region with generally high surface run-off (Mishra and Singh, 2010, 2011). To tackle the

* Corresponding author.
E-mail addresses: Mumtaz.Alli@usq.edu.au (M. Ali), ravinesh.deo@usq.edu.au (R.C. Deo).

https://doi.org/10.1016/j.compag.2018.07.013
Received 21 April 2018; Received in revised form 4 July 2018; Accepted 5 July 2018

complexities associated with PDSI, researchers have developed the Crop Moisture Index (CMI), particularly for assessing agricultural drought events (Palmer, 1968). CMI is based on the rank of the precipitation records and aims to compute both positive and negative precipitation anomalies. However, CMI in the short-term is insufficient to offset long-term issues. Byun and Wilhite (1999) developed the Effective Drought Index (EDI) based on daily precipitation data. EDI is a good index for operational monitoring of meteorological and agricultural drought, although the consideration of precipitation alone does not take into account other environmental parameters that also cause drought (*e.g.*, the impact of temperature). Several drought indices have been developed built on PDSI to take into account additional data on precipitation and crop moisture, however the standardized precipitation index (SPI) is used universally as a standard metric (Deo et al., 2017a).

The importance of modelling SPI is derived from the notion that: (1) SPI is able to assess the water shortage situations built on a statistical distribution of rainfall that can enable both short-term (i.e. monthly) and long-term (i.e., seasonal and annual) drought assessments (ranging from 1 up to 48 months). (2) SPI is presented as a normalized standard metric of rainfall surpluses/deficits in relation to a benchmark climatological period (Hayes et al., 1999; McKee et al., 1993; Yuan and Zhou, 2004), and therefore, it can enable a comparison of the drought behaviour in geographically and climatologically diverse regions. (3) SPI has been explored and validated for drought mitigation studies in diverse climatic regions (Almedeij, 2016; Choubin et al., 2016; Svoboda et al., 2012). (4) SPI has the potential to represent both short (1 and 3 months) and long-term (6–12 months) drought in a probabilistic fashion, largely on multiple timescales. In the case of agricultural drought, the SPI makes it possible to examine the soil moisture status with respect to precipitation anomalies on a comparatively short timescale using hydrological reservoirs to replicate long-term climate anomalies (Svoboda et al., 2012). Due to the advantageous features, the SPI is an ideal metric for the management of not only hydrological, but also for agricultural drought events (Deo et al., 2017a; Guttman, 1999). In this study, SPI based on 1 month is utilized for drought forecasting as there is no study on a monthly drought in the selected study regions, although the study of Ali et al. (2018a) designed a study using ensemble-ANFIS model for long term (3–12 months) drought forecasting and Ali et al. (2018b) developed a multi-stage hybridized online sequential extreme learning machine integrated with Markov Chain Monte Carlo copula-Bat algorithm for rainfall forecasting, both focussed in Pakistan.

In the existing literature, data-driven models have been used to model SPI for drought forecasting. For example, an SPI-based methodology was designed by Cancelliere et al. (2007) to forecast probabilistic drought alterations in Sicily, Italy. Jalalkamali et al. (2015) in Yazd, Iran conducted a study to forecast the SPI using several model variants, including a multilayer perceptron artificial neural network (MLP ANN), an adaptive neuro-fuzzy inference system (ANFIS), support vector machines (SVM), and an autoregressive integrated moving average (ARIMA) multivariate model. In another study, models based on ANFIS and ANN Wavelet tools were adopted by Shirmohammadi et al. (2013) to forecast SPI in Azerbaijan. An SPI-based forecasting study was also performed by Santos et al. (2009) using an ANN model for San Francisco, USA. Drought forecasts using SPI have an extensive history in the current literature (Adamowski et al., 2012; Bonaccorso et al., 2003; Cancelliere et al., 2006; Choubin et al., 2016; Deo et al., 2017b; Guttman, 1999; Hayes et al., 1999; Jalalkamali et al., 2015; Moreira et al., 2015; Moreira et al., 2008; Paulo and Pereira, 2007; Sönmez et al., 2005). However, SPI based drought forecasts are yet to be explored for agricultural regions in Pakistan where the influence of drought is a major impeding factor for crop productivity and farmers' livelihoods.

Drought hazard continues to severely affect the agriculturally dependent nation of Pakistan (Report, 1950-2015). The severe drought event of 1998 led to a significant reduction in Pakistan's national agricultural productivity by 2.6% over 2000–2001 (Ahmad et al., 2004). In spite of such pressing issues, drought models for local agricultural zones in Pakistan have been very limited: (1) Khan and Gadiwala (2013) aimed to investigate drought behaviour using SPI at multi timescales for the province of Sindh: (2) Xie et al. (2013) applied a spatiotemporal variability analysis based on the SPI data to also forecast drought behavior, and most recently, (3) the study of Ali et al. (2017) forecasted drought based on Standardized Precipitation-Evapotranspiration Index (SPEI) where a multilayer perceptron-based artificial neural network model was employed. (4) The study of Ahmed et al. (2016) has utilized antecedent SPI data for the characterization of future seasonal drought events in Balochistan, Pakistan and (5) Ali et al. (2018a) implemented an ensemble strategy based on the ANFIS model to forecast the SPI using the historical SPI to forecast future SPI. While it was not specifically on drought forecasting, a recent study of Ali et al. (2018b) has forecasted rainfall in Pakistan using a multi-stage online sequential extreme learning machine integrated with Markov Chain Monte Carlo copula-Bat algorithm. However, there has been no study in Pakistan specifically on future drought models or drought indices utilising different climatological parameters.

Considering a lack of drought models for developing nations like Pakistan, the aim of this research is to develop and evaluate Comm-ELM (a data-intelligent model) with universal approximation capabilities. The specific objectives are:

(i) To develop a committee based extreme learning machine (Comm-ELM) model following its successful application elsewhere, *e.g.*, (Barzegar et al., 2018; Prasad et al., 2018) and evaluate its preciseness for forecasting future SPI incorporating the relevant hydro-meteorological dataset (*i.e.*, temperature, rainfall, humidity, Southern Oscillation Index) and the seasonality metric (i.e., periodicity) as predictors for the period of 1081–2015. Here, a multi-stage model strategy is adopted incorporating antecedent climate-based variables at $(t-1)$ and corresponding periodicity in the first stage, and the forecasted SPI.
(ii) To elucidate the importance of periodicity as a pertinent factor in drought forecast models and seasonal influences on drought progression using data monitoring for Pakistan, a nation vulnerable to significant agricultural and water management issues.
(iii) To compare the performances of Comm-ELM in respect to Comm-PSO-ANFIS and Comm-MLR models for SPI-forecasting.

2. Theoretical background

2.1. Standardized precipitation index (SPI)

The SPI drought forecasting metric, centred on normalized probabilities of dryness relative to a base climatology, provides a depiction of irregular wetness and dryness situation. Before developing the forecast model for the designated regions in Pakistan, the monthly SPI index was calculated using precipitation (PCN) data (Deo et al., 2017a; McKee et al., 1993). Pearson Type III distribution/gamma distribution function is given by the following mathematical expression:

$$g(PCN) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)}(PCN)^{\alpha-1}e^{-x/\beta} \tag{1}$$

where α and β represents the estimated parameters using the maximum likelihood. The cumulative probability can be given by

$$G(PCN) = \int_0^P g(PCN)dPCN = \frac{1}{\beta^{\alpha}\Gamma(\alpha)}\int_0^P x^{\alpha-1}e^{-P/\beta}dPCN \tag{2}$$

Suppose that $m = PCN/\beta$, this reduces Eq. (2) to an incomplete gamma function:

$$G(PCN) = \frac{1}{\Gamma(\alpha)}\int_0^m m^{\alpha-1}e^{-m}dm \tag{3}$$