I.A.S.R.I. / P.R. –03/2020
आई. ए. एस. आर. आई. / पी. आर. 03/2020
Institute Project Code:
AGENIASRISIL201701200098

परियोजना रिपोर्ट

PROJECT REPORT

मशीन लर्निंग तकनीकों के प्रयोग से सूखे सूचकांक का मॉडलिंग एवं पूर्वानुमान

Modelling and forecasting of drought index using machine learning techniques

हर कदम, हर डगर
किसानों का हमसफर
भारतीय कृषि अनुसंधान परिषद

Agrisearch with a human touch

के. एन. सिंह	K. N. Singh (PI, From 12th Sept. 2018 onwards
राजीव रंजन कुमार	Rajeev Ranjan Kumar (PI, From 22nd May 2017 to 11th Sept. 2018)
रविन्द्र सिंह शेखावत	Ravindra Singh Shekhawat (Co-PI)
संजीव पनवार	Sanjeev Panwar (Co-PI)

पूर्वानुमान एवं कृषि प्राणाली मॉडलिंग प्रभाग
भा.कृ.अनु.प.-भारतीय कृषि सांख्यिकी अनुसंधान संस्थान
लाइब्रेरी एवेन्यू, पूसा, नई दिल्ली –110012

Division of Forecasting and Agricultural System Modeling
ICAR-Indian Agricultural Statistics Research Institute
Library Avenue, Pusa, New Delhi – 110012

2020

एक विश्वसनीय और कम्प्यूटेशनल रूप से कुशल सूखा मॉडल जल संसाधन प्रबंधन के लिए एक उपयोगी उपकरण है क्योंकि सूखे-जोखिम का पूर्व ज्ञान इसकी भविष्य की घटना का सटीक अनुमान लगाने की हमारी क्षमता पर निर्भर करता है। चरम घटनाओं पर जलवायु परिवर्तन रिपोर्ट पर अंतर-सरकारी पैनल ने सूखे को एक चरम जलवायु घटना के रूप में मान्यता दी है जिसे इसके नकारात्मक प्रभावों को कम करने के लिए इस से निपटने की आवश्यकता है। इसलिए, सूखे प्रबंधन के लिए तेज, सटीक और विश्वसनीय सूखा पूर्वानुमान मॉडल जो भविष्य के सूखे-जोखिम पर लीड-टाइम जानकारी प्रदान करते हैं, एक उपयोगी उपकरण है। सूखा आर्थिक क्षेत्रों (जैसे, कृषि) के लिए जोखिम पैदा करता है, और इसलिए इसे वर्षा की कमी के बजाय जल संसाधनों की कमी से परिभाषित किया जाना चाहिए। इसलिए, जल संसाधनों की अवधारणा के आधार पर DIs का पूर्वानुमान (उदाहरण के लिए, प्रभावी DI, EDI; बयून और विल्हाइट 1999), सूखे जल विज्ञान और जल संसाधन प्रबंधन के क्षेत्र में निर्णय लेने के लिए उपयोगी है।

सूखा-जोखिम के बारे में जानकारी निकालने के लिए, वैश्विक जलवायु मॉडल (GCMs) से वर्षा की गिरावट सांख्यिकीय या गतिशील तरीकों का उपयोग करके की जाती है। डायनामिक तरीके से छोटे-पैमाने के सूखे की जानकारी प्राप्त करने के लिए GCM से सीमा स्थितियों का उपयोग करते हुए एक सीमित क्षेत्र के उच्च-रिज़ॉल्यूशन मॉडल (जैसे, क्षेत्रीय जलवायु मॉडल) का उपयोग करते हैं, जबकि सांख्यिकीय मॉडल भविष्य की जलवायु और बड़े पैमाने पर भविष्यवक्ताओं के बीच संबंध की खोज करते हैं। मशीन लर्निंग (एमएल) एल्गोरिदम पर आधारित सांख्यिकीय मॉडल में, अनुकरण किए गए डेटा से अनुभवजन्य रूप से कैलिब्रेट किए जाते हैं। एमएल को स्थानीय स्तर पर सूखा पूर्वानुमान के विकल्प के रूप में मान्यता प्राप्त है। एमएल कम जटिल है क्योंकि यह जलवायु रुझानों से डेटासेट को 'सीखने' के लिए आत्मसात करता है। अन्य लाभों में जटिल भौतिक समीकरणों की आवश्यकता के बिना जलवायु का पूर्वानुमान शामिल है, आसान मॉडल विकास, कम कम्प्यूटेशनल लागत, तेजी से प्रशिक्षण और परीक्षण, बिंदु-आधारित पूर्वानुमान डेटा और उनके प्रतिस्पर्धी प्रदर्शन के सापेक्ष स्थानीय अनुप्रयोगों (जैसे, खेतों) की संभावना आदि।

इस अध्ययन में प्रभावी सूखे सूचकांक के पूर्वानुमान के लिए एक विश्वसनीय मशीन लर्निंग मॉडल विकसित करने का प्रयास किया गया है।

PREFACE

A reliable and computationally efficient drought model is a useful tool for water resources management since a prior knowledge of drought-risk depends on our ability to accurately forecast its future occurrence. The Inter-governmental Panel on Climate Change Report on extreme events has recognized drought as an extreme climatic event that needs to be mitigated to reduce its negative effects. Therefore, fast, accurate and reliable drought forecasting models that provide lead-time information on the future drought-risk is a useful tool for drought management. Droughts pose risks to economic areas (e.g., agriculture), and so must be defined by the deficiency of water resources rather than a rainfall deficiency. Therefore, forecasting of the DIs based on the concept of water resources (e.g., the effective DI, EDI; Byun and Wilhite 1999) are useful for decision-making in the field of drought hydrology and water resource management.

To extract information on drought-risk, the downscaling of rainfall from global climate models (GCMs) is performed using statistical or dynamical methods. Dynamical methods utilize a limited-area high-resolution model (e.g., regional climate model) using boundary conditions from GCMs to derive small-scale drought information while statistical models discover links between future climate and large-scale predictors. In statistical models using machine learning (ML) algorithms, the simulations are empirically calibrated from the observed data. ML is recognised as an alternative for local-scale drought forecasting. ML is less complex as it assimilates datasets to 'learn' from climatic trends. Other advantages include the forecasting of climate without the need for complex physical equations, easy model development, low computational cost, fast training and testing, the possibility of local applications (e.g., farms) using point-based forecasting data and their competitive performance relative to fully-dynamic models.

In this study an attempt has been made to develop a reliable machine learning model for the forecasting of effective drought index.

Table of Contents

Chapter-1

Introduction and Review of Literature

1.1 Introduction

Drought is a complex hydrologic feature of arid and semiarid regions with strong implications on the sustainability of water resources, agriculture and environmental management. The Inter-governmental Panel on Climate Change Report on extreme events has recognized drought as an extreme climatic event that needs to be mitigated to reduce its negative effects (Field 2012). Drought forecasting is a critical part of addressing drought and is important in risk management and mitigation (Mishra and Singh 2011; Belayneh et al. 2016). Forecasts are performed using drought indices (DIs) that are standardized metrics of rainfall, temperature or evapotranspiration. Drought models are used to forecast drought indices (DIs) that quantify drought by its onset, termination, and subsequent properties such as the severity, duration and peak intensity in order to monitor and evaluate the impacts of future drought. Numerically expressed DIs are more functional than raw rainfall data as such metrics can be used as better triggers for detecting the initiation and termination of drought levels necessary for recovery planning, mitigation and decision-making. Therefore, fast, accurate and reliable drought forecasting models that provide lead-time information on the future drought-risk is a useful tool for drought management. For forecasting of drought, basically two types of models are considered in literature:

- Physical models which predict coupled effects of the ocean and the atmosphere, known as Global Circulation Model (GCM) and
- Statistical models that assimilate observed values of hydro-meteorological properties (e.g. temperature or rainfall) to forecast future drought events.

Machine learning (ML) (or statistical model) is now being experimented in a wide variety of climate applications. The ML models utilize, assimilate and ‘learn’ from the evidence of past climate trends using observational dataset to predict the future. The practical advantages of the ML algorithm over the GCM are the explanation of the externally driven climate without the need for complex physical models, easiness of experimentation, validation and evaluation, low computational burden, much more simple and fast in the training and the testing phases, the

applicability to the data for a specific point of measurement (a specific area) and the competitive performance compared to physical models (Ortiz-Garcia et al., 2014).

A frequently used ML algorithm in climate sciences is the artificial neural network (ANN). ANN is a powerful and versatile data-driven algorithm for capturing and representing complex input and output relationships. However a major challenge encountered by the ANN is the requirement of iterative tuning of model parameters, slow response of the gradient based learning algorithm used and the relatively low prediction accuracy compared to the more advanced ML algorithms (e.g. (Acharya et al., 2013; Şahin et al., 2014)). The extreme learning machine (ELM) model has gained popularity in hydrological forecasting as an improved artificial intelligence approach that requires significantly less computational time than the classical artificial neural network (ANN) model for training datasets. The concept of extreme learning machine (ELM) was proposed by Huang (2006) for single-hidden layer feedforward neural networks (SLFNs) which randomly chooses hidden nodes and analytically determines the output weights of SLFNs. This algorithm tends to provide good generalization performance at extremely fast learning speed. The ELM algorithm has a fast three step method designed using a single layer feed forward neural network with hidden neurons and randomly generated weights. The hidden layer parameters (hidden neurons and biases) are first randomized for a particular network architecture (governed by the number of input variables and the number of hidden neurons chosen optimally by the ELM network performance). Afterwards, the input variables are propagated through the hidden layer and finally the output weights are solved as a linear system of equations.

1.2 Review of literature

Huang (2006) proposed a new learning algorithm called extreme learning machine (ELM) for single-hidden layer feedforward neural networks (SLFNs) which randomly chooses hidden nodes and analytically determines the output weights of SLFNs. This algorithm tends to provide good generalization performance at extremely fast learning speed.

Kim and Byun (2009) investigated the effect of global warming on drought patterns over Asia at the end of the twenty-first century by a multi-model ensemble method based on daily precipitation data of 15 coupled climate models. The projected precipitation climatology was translated into the change in drought climatology using the effective drought index. The results of the models were

consistent in that they project an increase in the mean and the standard deviation of precipitation over most of Asia, and the increase was considerably greater in higher latitude areas.

Kim et al. (2011) investigated drought in South Korea were using daily precipitation data for 1777–2008. The climatological characteristics of the effective drought index was used to quantify the drought intensity. They developed a spatiotemporal EDI map using wavelet transformation.

Belayneh and Adamowski (2012) compared the effectiveness of three data-driven models i.e. artificial neural networks (ANNs), support vector regression (SVR), and wavelet neural networks (WN) for forecasting drought conditions in the Awash River Basin of Ethiopia. The forecast results indicate that the coupled wavelet neural network (WN) models were the best models for forecasting SPI values over multiple lead times in the Awash River Basin in Ethiopia.

Rathinasamy and Khosa (2012) presented a comparative evaluation of different wavelet forms when employed for forecasting future states of various kinds of time series. The results suggest that those wavelet forms that have a compact support, for example the Haar wavelet, have a better time localization property and show improved performance in the case of time series that have a short memory with short duration transient features.

Huang (2012) showed that both LS-SVM and PSVM can be simplified further and a unified learning framework of LS-SVM, PSVM, and other regularization algorithms referred to extreme learning machine (ELM) can be built. ELM provides a unified learning platform with a widespread type of feature mappings and can be applied in regression and multiclass classification applications directly.

Li and Cheng (2014) proposed a conjunction model of wavelet neural networks with ELM (WNN-ELM) for 1-month ahead streamflow discharge forecasting. The results indicate that the SLFNs-ELM performs slightly better than the SVM for peak discharge estimation, and the proposed model WNN-ELM provides more accurate forecast precision than SLFNs-ELM and SVM.

Abdullah et al. (2015) investigated the efficiency of Extreme Learning Machines (ELM) algorithm at predicting evapotranspiration in southern part of Iraq. They showed that ELM is efficient, simple in application, of high speed, and has very good generalization performance. They recommended this algorithm for locations similar to the geographical and meteorological conditions of Iraq that consists of both arid and semiarid regions.

Deo et al. (2015a) employed the EDI for analysis of historical drought events in Australia and demonstrated its skill for quantifying dry spells as well as continuing drought events.

Belayneh et al. (2016) explored the ability of coupled machine learning models and ensemble techniques to predict drought conditions in the Awash River Basin of Ethiopia. They explored the potential of wavelet transforms coupled with the bootstrap and boosting ensemble techniques to develop reliable artificial neural network (ANN) and support vector regression (SVR) models for drought prediction.

Deo et al. (2016) developed a wavelet-based drought model using the extreme learning machine (W-ELM) algorithm where the input data are first screened through the wavelet pre-processing technique for better accuracy to forecast the monthly effective DI (EDI). They demonstrate enhanced forecast skill of the drought models that use wavelet pre-processing of the predictor dataset.

Deo et al. (2017) studied the drought modelling using multivariate adaptive regression splines (MARS), least square support vector machine (LSSVM), and M5Tree models by forecasting SPI in eastern Australia. They highlighted the importance of periodicity in drought forecasting and also ascertains that model accuracy scales with geographic/seasonal factors due to complexity of drought and its relationship with inputs and data attributes that can affect the evolution of drought events.

Dayal et al. (2017) investigated the feasibility of the Artificial Neural Network (ANN) algorithms for prediction of a drought index: Standardized Precipitation-Evapotranspiration Index (SPEI). They developed an ANN model to predict the index in two selected regions in Queensland, Australia.

1.3 Motivation

- Machine learning techniques is less complex as it assimilates datasets to ‘learn’ from climatic trends and it is recognised as an alternative for local-scale drought forecasting.
- By the use of machine learning technique, we can forecast the drought without the need for complex physical equations.
- Forecasting of drought index through machine learning has an easy model development, low computational cost, fast training and testing, the possibility of local applications (e.g.,

farms) using point-based forecasting data and their competitive performance relative to fully-dynamic models.

- The modelling of drought index using machine learning (ML) algorithms for drought forecasting is still lacking, especially in India.
- The extreme learning machine (ELM) model has gained popularity in hydrological forecasting as an improved artificial intelligence approach that requires significantly less computational time than the classical artificial neural network (ANN) model for training datasets.
- ELM satisfies the universal approximation condition with good generalization performance (Huang et al. 2015), and therefore, is a suitable ML approach for drought forecasting.
- Though so many advantages of ELM, existing ELM algorithms pay little attention to optimizing the choice of kernels, which is indeed crucial to the performance of ELM for applications in drought forecasting.
- Wavelet transformation of signals can allow a ML model to better analyze the time evolution of hydrological processes at different scales. Hence, wavelet decomposition of inputs is considered a multi-resolution tool for pre-processing non-stationary signals.
- Wavelet coupled multiple kernel ELM, to model the drought index has not been explored anywhere.

1.4 Objectives

1) To develop methodology for modelling and forecasting of drought index using different machine learning techniques
2) To develop and evaluate forecasting model for drought index using multiple kernel extreme learning machine
3) To develop and evaluate wavelet based multiple kernel extreme learning machine for forecasting drought index

1.5 Plan of Report

This report deals with various aspects of development of multiple kernel extreme learning machine (MK-ELM) and wavelet based multiple kernel extreme learning machine (W-MK-ELM) for the

forecasting of drought index. This report is divided into five chapters. Chapter-I of this report i.e. the present chapter gives a brief introduction to the problem and objective of the study and review of literature on earlier work done in the direction of forecasting of drought index. Chapter-II deals with machine learning techniques for modelling and forecasting of drought index. Chapter-III discusses multiple kernel extreme learning machine model for drought index. Chapter–IV deals with wavelet based multiple kernel extreme learning machine model for forecasting drought index.

Chapter-2

Machine learning techniques for modelling and forecasting of drought index

2.1 Introduction

For the forecasting of future drought, basically two types of models are considered in literature: physical models and statistical models. Physical models predict coupled effects of the ocean and the atmosphere, known as Global Circulation Model (GCM). An alternative to the physical model is the machine learning (ML) (or statistical model) that is now being experimented in a wide variety of climate applications. Statistical models assimilate observed values of hydro-meteorological properties (e.g. temperature or rainfall) to forecast future drought events. The practical advantages of the ML algorithm over the GCM are the explanation of the externally driven climate without the need for complex physical models, easiness of experimentation, validation and evaluation, low computational burden etc. The ML models utilise, assimilate and 'learn' from the evidence of past climate trends using observational dataset to predict the future. Many types of ML algorithms have recently been proposed in literature, including the co-integration methods that analyse relationships between stationary and non-stationary data, regression approaches for evaluating time-series properties of air temperature, neural networks for predicting rainfall (Abbot and Marohasy, 2014) etc. Consequently many studies are using different ML algorithms to demonstrate nearly coincident or in some cases, even better prediction yields than the GCM models.

A frequently used ML algorithm used in climate sciences is the artificial neural network (ANN). ANN is a powerful and versatile data-driven algorithm for capturing and representing complex input and output relationships. However a major challenge encountered by the ANN is the requirement of iterative tuning of model parameters, slow response of the gradient based learning algorithm used and the relatively low prediction accuracy compared to the more advanced ML algorithms (e.g. (Acharya et al., 2013; Şahin et al., 2014)).

The purpose of this investigation is then three fold: A first objective consists of evaluating the capability of the ELM model considered in the problem of predicting monthly Effective Drought Index using meteorological datasets and climate mode indices as input variables. A second objective consists of a deep statistical analysis of the results disseminated by the ELM model in

relation to the predictions by ANN model. A third objective is to deduce the optimum network architecture of the MK-ELM, W-MK-ELM and the ANN models, computational speeds and performance parameters. The conclusion is that the wavelet based MK-ELM model performs significantly better than the ANN model and therefore, is potentially useful for research in areas of effective management of water environments, agriculture and infrastructure. In this study we have adopted a much improved class of ML algorithm, known as extreme learning machine (ELM) as a statistical model in a problem of predicting the monthly Effective Drought Index (EDI) (Byun and Wilhite, 1999).

2.2 Methodology

2.2.1 Extreme learning machine (ELM)

Extreme learning machine (ELM) developed by Huang et al., (2006) is the state-of-art novel machine learning algorithm for Single Layer Feedforward Neural Network (SLFNs). Consequently the ELM model has been widely used for the solution of estimation problems in many different fields and is now gaining attention within the climate research and applied engineering community (Acharya et al., 2013; Belayneh and Adamowski, 2012; Şahin et al., 2014). These investigations and others have demonstrated important advantages of the ELM model over the traditional neural network or vector classification schemes. The ELM model is easy to use and no parameters need to be tuned except the predefined network architecture, thus avoiding many complications faced by the gradient-based algorithms such as learning rate, learning epochs, and local minima. Importantly the ELM model has also been proven to be a faster algorithm compared with other conventional learning algorithms such as backpropagation (BP) or support vector machines (SVM). In the ELM approach most of the training is accomplished in time span of seconds or at least in minutes in large complex applications which are not easily achieved by using the traditional neural network models (Acharya et al., 2013). The ELM model possesses similar generalization performance to the backpropagation, the SVM and the singular value decomposition (SVD) algorithms in data classification and prediction problems. Therefore, the ELM model has been considered as an ideal computational algorithm for forecasting atmospheric and meteorological variables including solar energy, air temperature and rainfall that generally have large and complex datasets.

ELM was proposed for ‘‘generalized’’ single-hidden layer feedforward networks where the hidden layer need not be neuron alike (Huang & Chen, 2006; Huang, Zhou, et al., 2012). The output function of ELM for generalized SLFNs is

$$f_L(\boldsymbol{X}) = \sum_{i=1}^{L} \boldsymbol{\beta}_i h_i(\boldsymbol{X}) = \boldsymbol{h}(\boldsymbol{X})\boldsymbol{\beta}$$

where

$\boldsymbol{\beta} = [\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_L]^T$ is the output weight vector between the hidden layer of L nodes to the m≥ 1 output nodes, and $\boldsymbol{h}(\boldsymbol{X}) = [h_1(\boldsymbol{X}), \ldots, h_L(\boldsymbol{X})]$ is ELM nonlinear feature mapping, e.g., the output (row) vector of the hidden layer with respect to the input $\boldsymbol{X}$.

$h_i(\boldsymbol{X})$ is the output of the *ith* hidden node output. The output functions of hidden nodes may not be unique. Different output functions may be used in different hidden neurons.

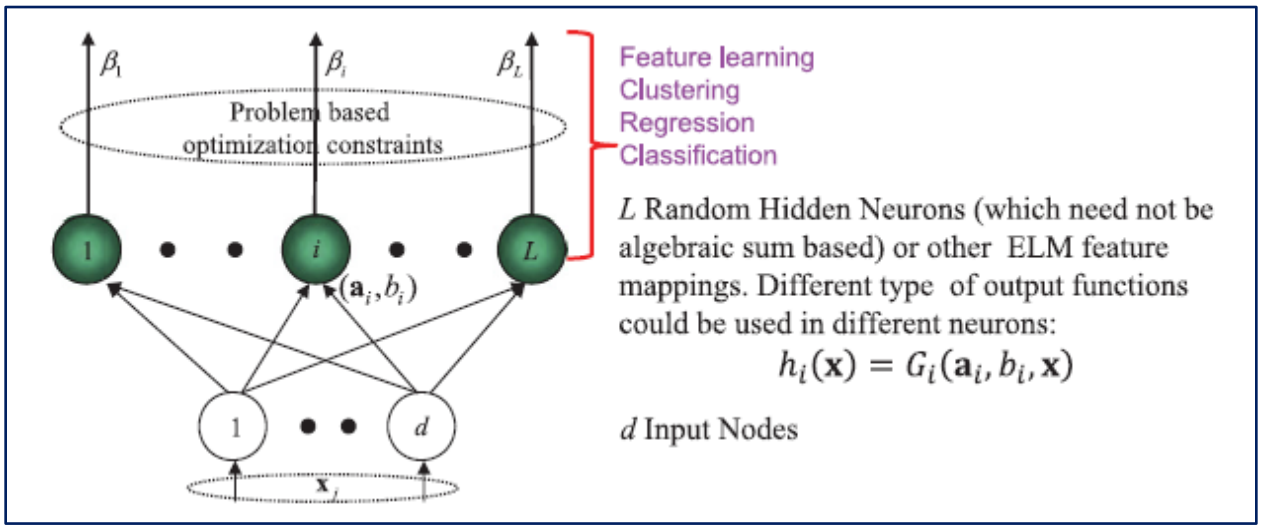

Figure: Schematic view of ELM

In particular, in real applications $h_i(\boldsymbol{X})$ can be

$$h_i(\boldsymbol{X}) = G(\boldsymbol{a}_i,\ b_i, \boldsymbol{X}), \qquad \boldsymbol{a}_i \in \boldsymbol{R}^d,\ b_i \in \mathrm{R}$$

where

G(**a**, b, **x**) (with hidden node parameters (a, b)) is a nonlinear piecewise continuous function satisfying ELM universal approximation capability theorems (Huang, et al., 2006).

Basically, ELM trains an SLFN in two main stages:

(1) random feature mapping and

(2) linear parameters solving.

In the first stage, ELM randomly initializes the hidden layer to map the input data into a feature space by some nonlinear mapping functions.

The nonlinear mapping functions in ELM can be any nonlinear piecewise continuous functions. In ELM, the hidden node parameters (a, b) are randomly generated (independent of the training data) according to any continuous probability distribution instead of being explicitly trained, leading to remarkable efficiency compared to traditional BP neural networks. In the second stage of ELM learning, the weights connecting the hidden layer and the output layer, denoted by β, are solved by minimizing the approximation error in the squared error sense:

$$\underset{\beta \in R^{L\times m}}{min} \ \|\boldsymbol{H}\beta - \boldsymbol{T}\|^2$$

where

H is the hidden layer output matrix (randomized matrix):

$$\boldsymbol{H} = \begin{bmatrix} \boldsymbol{h}(\boldsymbol{X}_1) \\ \vdots \\ \boldsymbol{h}(\boldsymbol{X}_N) \end{bmatrix} = \begin{bmatrix} h_1(\boldsymbol{X}_1) & \dots & h_L(\boldsymbol{X}_1) \\ \vdots & \vdots & \vdots \\ h_1(\boldsymbol{X}_N) & \cdots & h_L(\boldsymbol{X}_N) \end{bmatrix}$$

and **T** is the training data target matrix:

$$T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix} = \begin{bmatrix} t_{11} & \dots & t_{1m} \\ \vdots & \vdots & \vdots \\ t_{N1} & \cdots & t_{Nm} \end{bmatrix}$$

where

$\|\cdot\|$ denotes the Frobenius norm.

The optimal solution is given by:

$$\beta^* = \boldsymbol{H}^{\intercal}\boldsymbol{T}$$

where

$\boldsymbol{H^{\tau}}$ denotes the Moore–Penrose generalized inverse of matrix H.

Generalization performance:

Most algorithms proposed for feedforward neural networks do not consider the generalization performance when they are proposed first time. ELM aims to reach better generalization performance by reaching both the smallest training error and the smallest norm of output weights:

$$minimize: \|\beta\|_p^{\sigma_1} + C\|H\beta - T\|_p^{\sigma_2}$$

where

$$\sigma_1 > 0,\ \sigma_2 > 0,\ \ p, q = 0, \frac{1}{2}, 1, 2, \ldots, +\infty$$

The first term in the objective function is a regularization term which controls the complexity of the learned model.

Huang, *et al.* (2012) especially studied the stability and generalization performance of ELM with σ1 = σ2 = p = q = 2:

$$\underset{\beta \in R^{L\times m}}{min} \quad \frac{1}{2}\|\boldsymbol{\beta}\|^2 + \frac{C}{2}\sum_{i=1}^{N}\|e_i\|^2 \qquad s.t.\ \ \boldsymbol{h}(\boldsymbol{X}_i)\beta = \boldsymbol{t}_i^T - \boldsymbol{e}_i^T,\ \ i = 1, \ldots, N.$$

By substituting the constraints of above equation into its objective function, we obtain the following equivalent unconstrained optimization problem:

$$\underset{\beta \in R^{L\times m}}{min}\ L_{ELM} = \frac{1}{2}\|\boldsymbol{\beta}\|^2 + \frac{C}{2}\|\boldsymbol{T} - \boldsymbol{H\beta}\|^2$$

If **H** has more rows than columns (N > L), which is usually the case where the number of training patterns is larger than the number of the hidden neurons, we have the following closed form solution for β:

$$\boldsymbol{\beta}^* = \left(\boldsymbol{H}^T\boldsymbol{H} + \frac{\boldsymbol{I}}{C}\right)^{-1}\boldsymbol{H}^T\boldsymbol{T}$$

where

I is an identity matrix of dimension L.

If the number of training patterns is less than the number of hidden neurons (N < L), then **H** will have more columns than rows, so

$$\boldsymbol{\beta}^* = \boldsymbol{H}^T \left(\boldsymbol{H}\boldsymbol{H}^T + \frac{I}{C} \right)^{-1} \boldsymbol{T}$$

where, **I** is an identity matrix of dimension N.

2.2.2 Artificial Neural Network (ANN)

ANNs have been applied for modelling purposes for many years in mathematics, engineering, medicine, economics, hydrology, meteorology, psychology, neurology and other subjects. Their popularity has grown since their first inception in 1943 (McCulloch and Pitts, 1943) mainly to solve prediction problems with variables of stochastic nature, nonlinear or unknown variations or those that must be determined from less controlled environments (Moustris et al., 2011). An ANN model weaves through mathematical components derived from stochastic time-series datasets to tackle the prediction of very complex systems (e.g. rainfall or drought prediction). As they are flexible and less assumption-dependent, there is no need to define the underlying physical process between the inputs and outputs (Morid, et al., 2007). This makes the ANN very suitable for drought forecasting where variables that trigger a drought may not be fully understood. Basically, the ANN model learns from previous history of how the input signal has varied over the time. It constructs logically an input–output mapping system to perform the future predictions. In order to train and test an ANN model or predicting a variable, the input data and its corresponding output values are necessary (Şahin et al., 2013)

ANN is a computational paradigm composed of non-linear elements (neurons) operating in parallel and massively connected by networks characterized by different weights. A single neuron computes the sum of its inputs, adds a bias term, and drives the result through a generally nonlinear activation function to produce a single output termed the activation level of the neuron. ANN models are specified by network topology, neuron characteristics, and training or learning rules (Lippman, 1987) with inputs, output(s) and hidden layers with interconnections. The fundamentals processing unit is a neuron, which computes a weighted sum of its input signals, y_i, for i = 0,1,2,. . . , n, hidden layers, w_{ij} and then applies a nonlinear activation function to produce an output signals u_j.

A neuronal model consists of an externally applied bias, b_k which has the effect of increasing or decreasing the net input of the activation functions depending on whether it is positive or negative. Mathematically, a neuron *k* may be described by

$$u_k = \sum_{j=1}^{m} w_k x_j$$

$$y_k = \Phi(u_k + b_k)$$

where $x_1, x_2, \ldots, x_m$ are the inputs signals; $w_{k1}, w_{k2}, \ldots, w_{km}$ are the synaptic weights of neuron *k*; u_k is the linear combiner output due to input signals; b_k is the bias; Φ(.) is the activation function and y_k is the output signal of the neuron. Bias b_k has the effect of applying an affine transformation to the output u_k of the linear combiner in the model (Fig. 2b) as shown by

$$v_k = u_k + b_k$$

In particular, depending on whether b_k is positive or negative, the relationship between the induced local field or activation potential v_k of neuron *k* and linear combiner output u_k can be modified. Note that as a result of this affine transformation, the graph of v_k versus u_k no longer passes through the origin. The bias b_k is an external parameter of artificial neuron *k*.

$$v_k = \sum_{j=0}^{m} w_{kj} x_j$$

$$y_k = \Phi(v_k)$$

The tangent sigmoid, $\phi(x)$ logarithmic sigmoid, $\Psi(x)$ and linear, $\chi(x)$ transfer function are described as follows

$$\phi(x) = \frac{2}{1 + e^{-2x}} - 1$$

$$\Psi(x) = \frac{1}{1 + e^{-x}}$$

$$\chi(x) = linear(x) = x$$

Computationally efficient ANN networks employ second-order training methods, primarily the Levenberg-Marquardt (LM) or the Broyden-Fletcher-Goldfarb-Shanno (BFGS) quasi-Newton backpropagation learning algorithms that minimize the mean squared error between the predicted and observed variable (Tiwari and Adamowski, 2013). An LM algorithm uses an approximation to the Hessian matrix as follows

$$x_{k+1} = x_k - [J^T J + \mu I]^{-1} J^T e$$

where *J* is the Jacobian matrix calculated using standard backpropagation techniques and is less complex than computing the Hessian matrix. The *J* contains first derivatives of network errors with respect to the weights and biases and **e** is a vector of errors. The BFGS quasi-Newton is an alternative to the conjugate gradient methods for fast optimization, which uses the following equation:

$$x_{k+1} = x_k - A_k^{-1} g_k$$

where A_k^{-1} is the Hessian matrix (second derivatives) of the performance index at the current values of the weights and biases.

2.2.3 Least-squares support vector regression (LSSVR)

In this study we incorporated the SVR class of models as a comparison with the neural network models. SVR-based models adhere to the structural risk minimization principle in contrast to neural network models which are developed to reduce empirical risk. SVR models are designed to improve generalization properties of the forecast model by placing a regularization constraint on the size of the model weights while neural network models aim to reduce the empirical error without regularization—in short, SVR models seek to attain good performance with small weight values while neural network models simply seek to attain good performance, with no mechanism to control the size of their weights. In general, the LSSVR algorithm (a specific model from the SVR class) is utilized to solve a linear programming problem where the training (input) vectors, taken as support vectors, are used to nonlinearly map inputs into forecasts. A linear formulation of LSSVR helps alleviate the convex quadratic programming problem associated with the original SVR model. In the standard SVR, the level of ‘‘slackness’’ (i.e., the property of the SVR which determines which training data vectors are chosen as support vectors) is set by selecting an

inequality constraint. This forces the inclusion of another parameter in the model that must be optimized (i.e., via the quadratic programming problem mentioned above). However, the LSSVR model avoids this problem (and solves the regression problem as a set of linear equations) since each input vector is chosen as a support vector and a ‘‘slackness’’ term is not required (this aspect of LSSVR is often seen as an advantage over SVR due to its faster training speed, higher stability, and better control. Consequently, an LSSVR model provides global solutions to the error function which is an advantage over gradient based algorithms (e.g., FFBP-ANN) that suffer from difficulties posed by the presence of local minima. A kernel function (K) and its relevant parameters are chosen such that a bound on Vapnik–Chervonenkis (VC) dimension is minimized, yielding stable solutions.

Given an input–output data set, X and Y comprised of N training data samples, i.e., $(x_t,\ y_t)$ for *t = 1, 2,…,N*, where $x_t \in \mathbb{R}^d$ and $y_t \in \mathbb{R}^d$; the LSSVR model is minimized with respect to the loss function L defined by

$$L(W\ ,\ e) = \frac{1}{2}W^T W + C\frac{1}{2}\sum_{t=1}^{N} e_t^2,$$

where

e_t^2 is the quadratic loss term,

W is the weight vector and

C is the regularization parameter

To solve for the LSSVR parameters, the Lagrangian multipliers method is used as follows (Suykens *et al.* 2002):

$$L(W\ ,k,\ e,\ \alpha) = L(W\ ,\ e) - \sum_{t=1}^{N} \alpha_t\{W^T\phi(x_t) + k + e_t - y_t\}$$

where

ϕ(x) is a nonlinear mapping function

$\alpha \epsilon \mathbb{R}^N$ is the set of Lagrange multipliers and

$k \epsilon \mathbb{R}$ is the bias term.

The conditions which prove to be optimal in solving the LSSVR parameters are determined by taking partial derivatives of the extended loss function [i.e., L($W,\ k, e\ \alpha$)] with respect to each term ($W,\ k, e\ \alpha$) as follows:

$$\frac{\delta L}{\delta W} = 0 \rightarrow W = \sum_{t=1}^{N} \alpha_t \phi(x_t)$$

$$\frac{\delta L}{\delta k} = 0 \rightarrow \sum_{t=1}^{N} \alpha_t = 0$$

$$\frac{\delta L}{\delta e_t} = 0 \rightarrow \alpha_t = C e_t, \quad t = 1, \dots, N$$

$$\frac{\delta L}{\delta \alpha_t} = W^T \phi(x_t) + k + e_t - y_t\ ,\ \ t = 1, \dots, N$$

The above conditions can be expressed in matrix form as:

$$\begin{bmatrix} 0 & \vec{1}^T \\ \vec{1} & \Omega + C^{-1} I \end{bmatrix} = \begin{bmatrix} K \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ Y \end{bmatrix}$$

where

$\vec{1}$ is a vector of ones and

Ω is used to represent the kernel function (K) satisfying Mercer's theorem (Mercer 1909; Okkan and Serbes 2012):

$$\Omega_{uv} = \phi(x_u)^T \phi(x_v) = K(x_u,\ x_v),\ \ u, v = 1, \dots\ ,\ N,$$

where the kernel function [$K(x_u,\ x_v) \epsilon \mathbb{R}$] may be represented by the commonly used RBF:

$$K(x_t,\ x) = e^{\left(-\frac{\|x-x^2\|}{2\sigma^2}\right)}$$

σ represent the RBF kernel width.

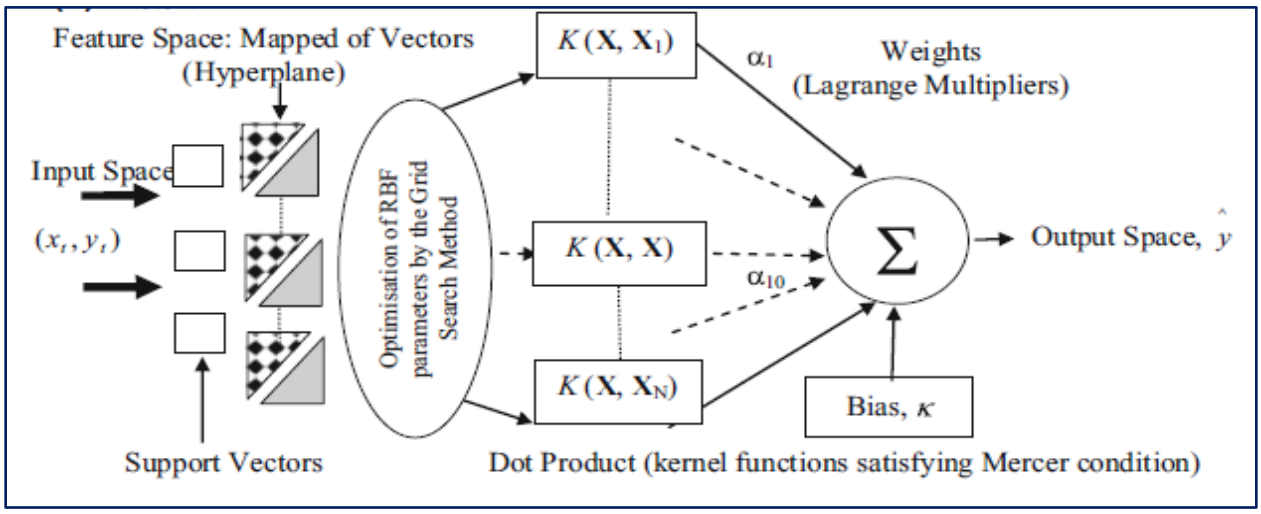

Figure: Schematic representation of LSSVR

2.3 Results and Discussion

2.3.1 Study area and climate data

This study utilised monthly precipitation data of Sagar and Chattarpur district of Bundelkhand region (Fig. 2.1) for a 52-year period (1951–2002) acquired from the India Water Portal website. The choice of test stations was ideal for testing the skill of the drought model. In terms of its climatic conditions, the Sagar district had average annual precipitation 1060 mm while the Chattarpur district had average precipitation 1032 mm with coefficient of variation 0.211 and 0.257 respectively. Hence, the EDI models developed in this study can be useful tools for drought-risk studies in these sites. Figure 2.2 and 2.3 shows plot of precipitation of Sagar and Chattarpur district respectively.

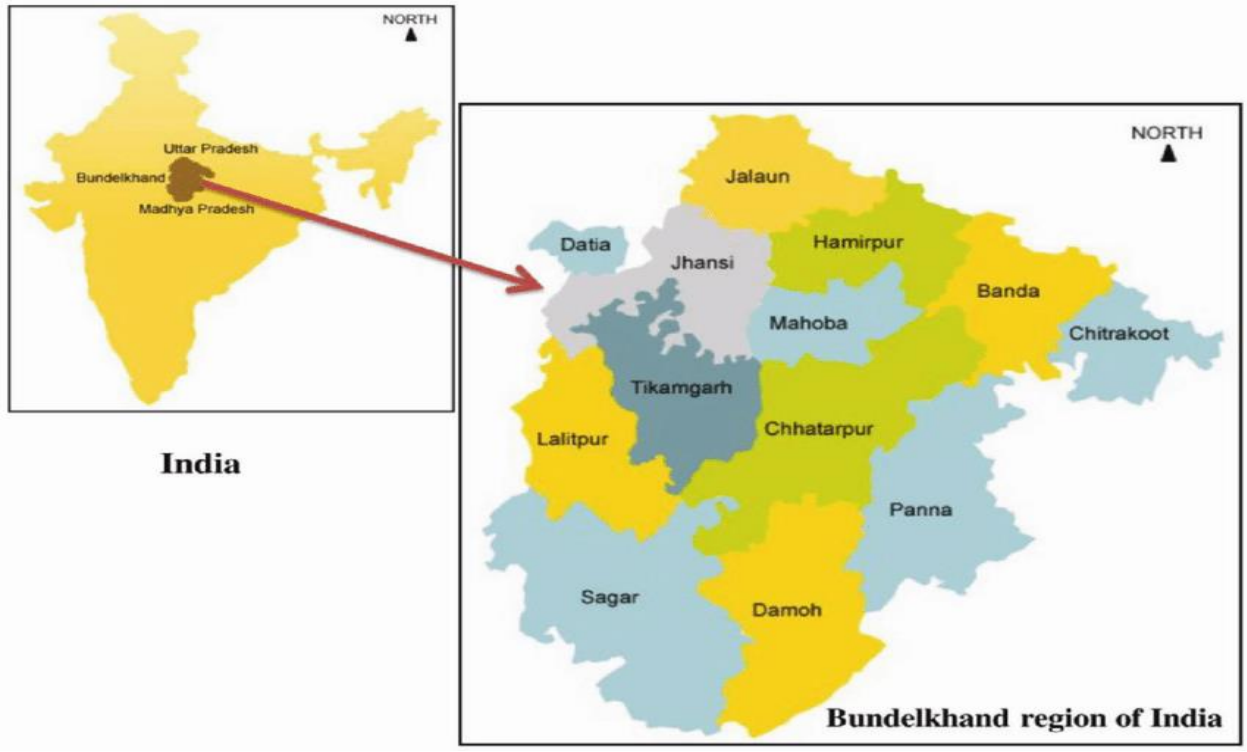

Figure 2.1 Geographical location of Bundelkhand region

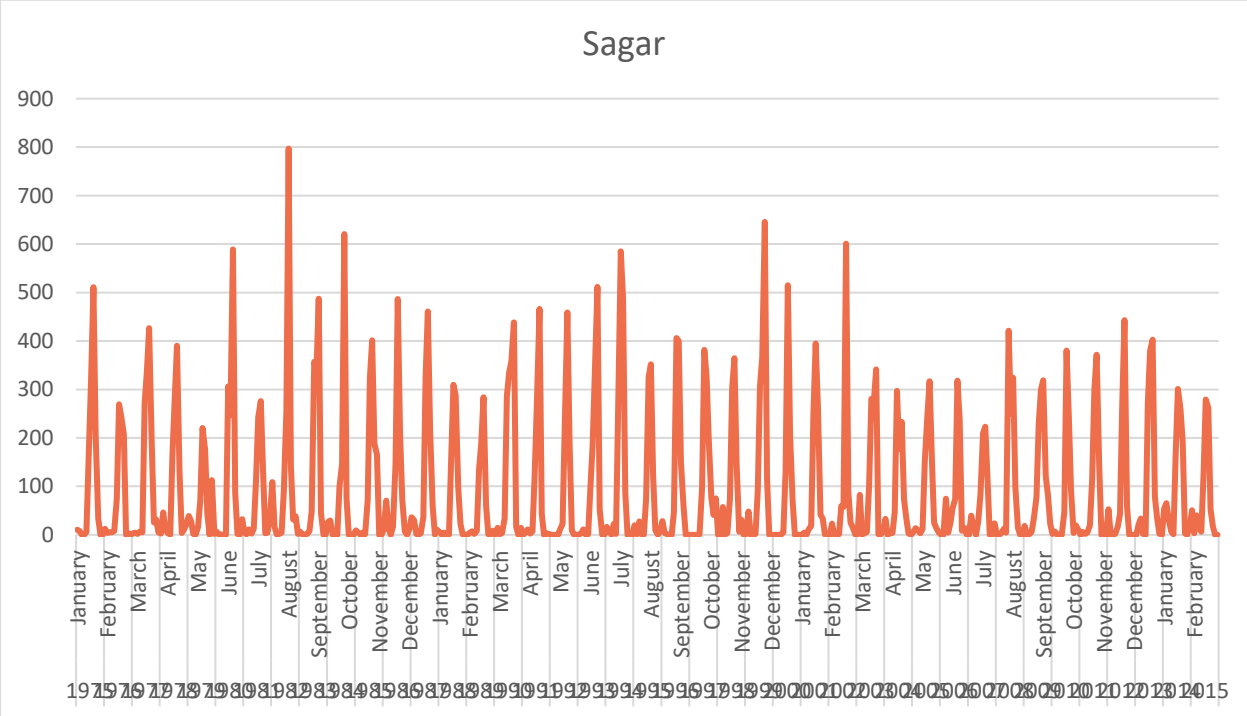

Figure 2.2 Plot of precipitation of Sagar district

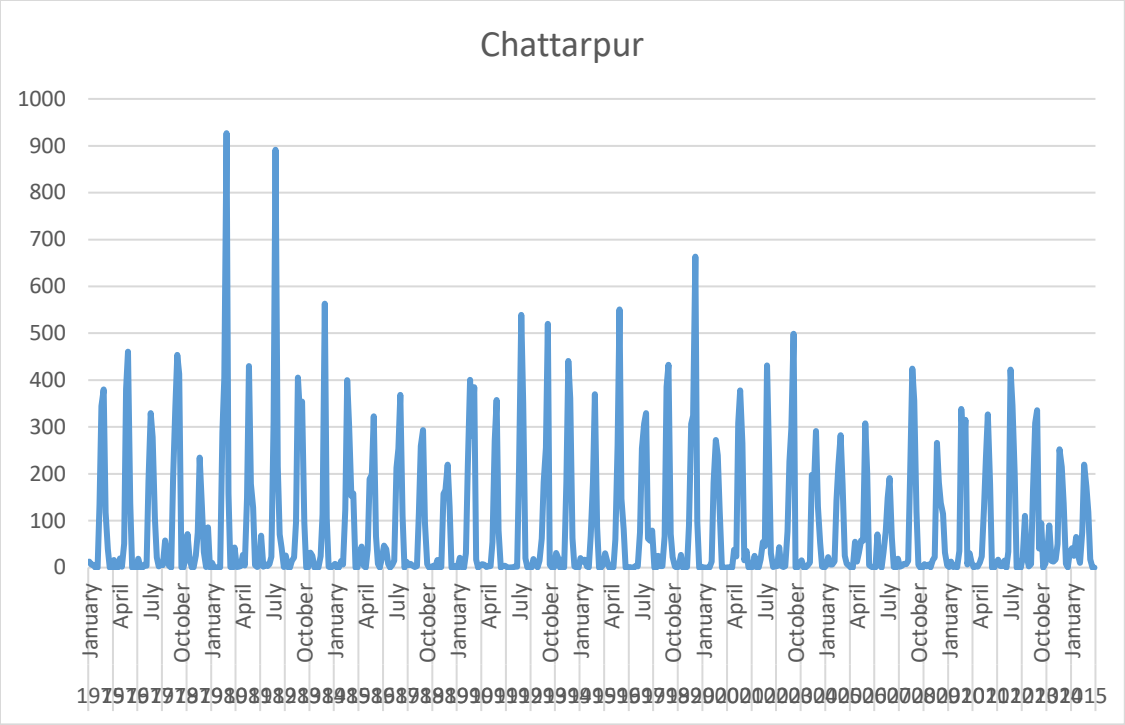

Figure 2.3 Plot of precipitation of Chattatpur district

2.3.2 Calculation of effective drought index (EDI)

In order to develop drought forecasting models, the precipitation data were used to calculate the effective drought index (EDI) (Byun and Wilhite 1999). In this study, the EDI was determined from daily effective precipitation (P_E), with summed precipitation both for the current and the antecedent day determined by a time-dependent reduction function.

Suppose that P_m is the rainfall recorded on any day, m $(1 \leq m \leq 365)$ and N* is the duration of summation of the preceding period, then P_E for that (current i^{th}) day is:

$$P_{E_i} = \sum_{N^*=1}^{D} \left[\frac{\sum_{m=1}^{N^*} P_m}{N^*} \right]$$

Above equation defines the degree to which Pm is converted into P_m for the i^{th} day but importantly, this model considers antecedent precipitation with reduced weighting. That is, the P_E will accumulate 100 % of precipitation received a day before, and 85 % of that received 2 days before, and 77 % of that received 3 days before, and so on, to 0.0423 % of precipitation 365 days before

the ith day (Kim et al. 2009). The model puts the highest weight on present rainfall, whereas the previous days' contributions decrease gradually up to the annual cycle (N = 365 days, ignoring leap year for simplicity). In order to formulate the EDI, two simple equations were utilized:

$$DP_{E_i} = P_{E_i} - MP_E$$

$$EDI_i = \frac{DP_{E_i}}{SP_E}$$

where, MPE is the mean PE, SPE is the standard deviation derived from the hydrological period (1951–2002), and *i* is the particular day. The "drought range" of Effective Drought Index (EDI) are:

Extremely dry conditions at Less than -2

Severe drought at -1.5 to -1.99

Moderate drought at -1 to -1.49.

Near normal conditions -0.99 to 0.99

In this study, for converting precipitation data in to Effective Drought Index (EDI), R code has been developed. By using developed code all precipitation data were converted in to EDI.

2.3.3 Input selection and drought model development

When designing a robust drought model, it is important to consider pertinent factors such as the proper selection of predictor variables or their combinations, training of predictor variables, ensuring the minimum level of over-fitting, selecting the best training algorithm and activation functions, ensuring the smallest generalization error and using the best performance assessment metrics. As there is no 'rule-of-thumb' to determine the significant inputs (Tiwari and Adamowski 2013), patterns engrossed in the EDI signal were recognized by correlation statistics i.e. autocorrelation function (ACF) and partial ACFs (PACFs; Tiwari and Adamowski 2013). The statistical method attempted to extract lagged information from the signal to analyse time periods between the current index value, and the index at a point in the past domain. Subsequently, optimum inputs for each time lag were identified by statistically analysing the lagged combinations and correlation coefficients (r). While the ACF detected seasonality in inputs, the PACF removed

any dependence on intermediate elements (those within lags), interpreted as a regression of the time-series against its past lagged value. Figures 2.1 to 2.4 shows correlograms of ACF and PACF for both district considered with a 95 % confidence interval. As a norm, scaling of predictor variables was applied prior to the simulations in order to avoid the patterns or attributes with larger numeric ranges dominating those of the smaller numeric range.

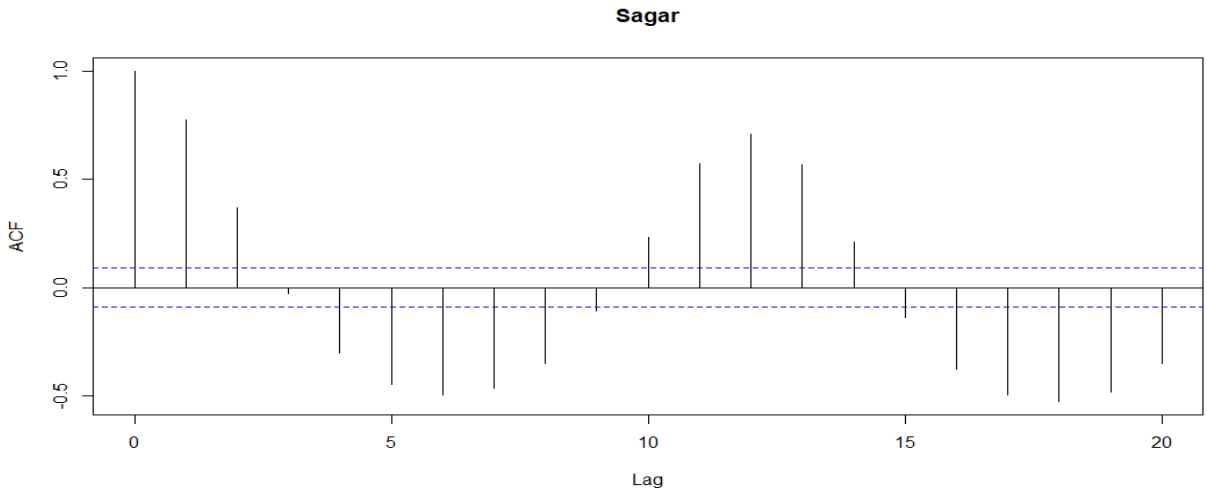

Figure 2.4 ACF of Sagar district

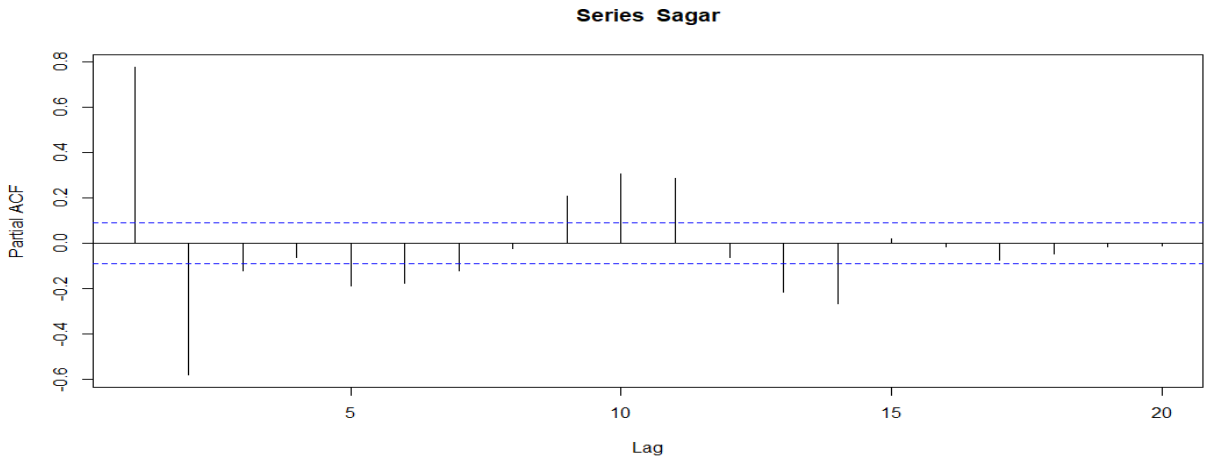

Figure 2.5 PACF of Sagar district

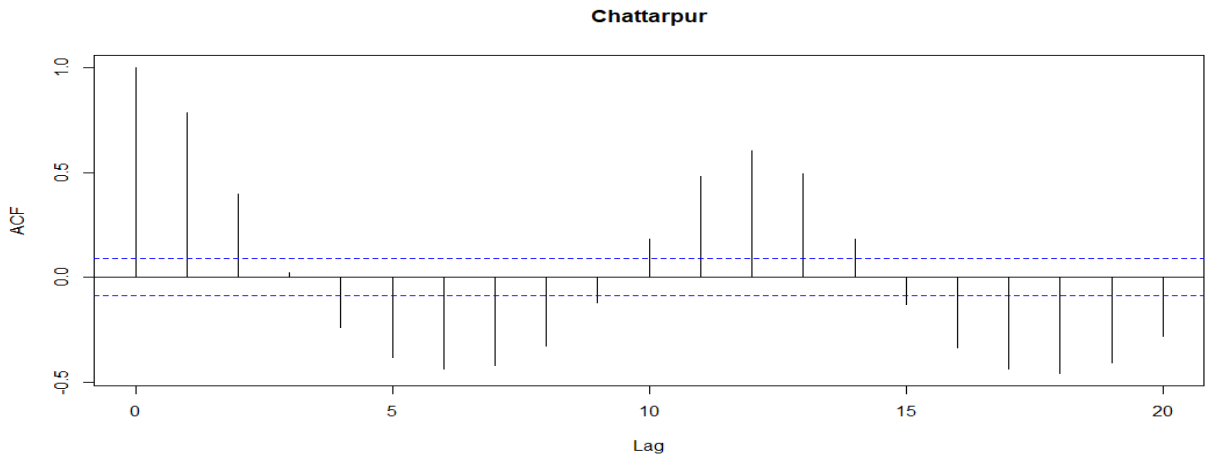

Figure 2.6 ACF of Chattarpur district

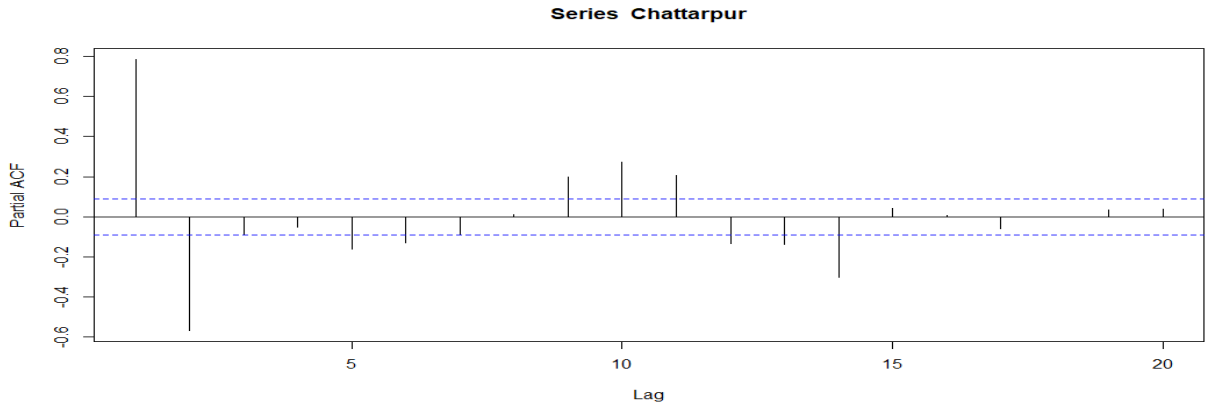

Figure 2.7 PACF of Chattarpur district

A three layer network that contained the input, feature optimisation and the output space was employed with the predictor dataset from training and the testing sets used for developing the present drought models. ELM model in this study was developed using the logarithmic sigmoid activation function. To identify the best ELM network architecture (i.e., number of hidden neurons), the number of hidden neurons was decided upon a priori. The architecture that performed best on a particular partition was then used to justify a particular architecture as optimal. Initially, the ELM model was randomly executed 50–1000 times to explore the effect of the variation of the randomized hidden layer weights and biases on the network's output. The objective was to obtain the smallest mean square error (MSE) for the weights and optimal nodes in the hidden layer.

Finally, this resulted in 100 randomizations that were appropriate for a stable solution of the forecasted monthly EDI. For each station, the time to run drought models was also recorded.

As a benchmark, the ANN model was also developed as this model has been shown to be an effective tool for drought forecasting. The ANN models were developed by considering different architectures with iteratively varying hidden neurons (1–35). During this period, the MSE value was monitored at the iterations of the training phase. The training was stopped when the MSE was at its minimum in comparison to earlier iterations or when the maximum number of epochs was reached (Bishop 1995). As with previous studies (Tiwari and Adamowski 2013; Deo and S¸ ahin 2015a), the secondorder Levenberg–Marquardt training algorithm, which is a fast and efficient algorithm, was employed for the ANN model.

The LSSVR model developed in this study was based on the radial basis kernel function (RBF). The choice of the RBF was appropriate, as it has a good ability to map nonlinear input samples to high dimensional space, it is able to discover non-linear relationships between target data and input data attributes and it has superior performance compared to the linear kernel. Also, the linear kernel is a special case of the RBF and the sigmoid kernel behaves like the RBF for certain parameters. The setting of C and r is important for a model's forecasting accuracy (Hoang et al. 2014). A range of kernel widths, $r = 2^{[15,13,11,9,7,5,3,1,-1,-3,-5]}$ and regularisation constants, $C = 2^{[-15,-13,-11,-9,-7,-5,-3,-1,1,3]}$ were trialled to create the best LSSVR structure. The optimised parameters were chosen based on the smallest MSE. Initially, a grid-search was used to identify plausible parameters that were then fine-tuned using the Nelder–Mead simplex algorithm.

2.3.4 Model Evaluation

All the model simulations using the ELM, SVR and the ANN algorithms were conducted in the R software. Table 2.1 shows the parameters of the ELM, SVR and the ANN model. The 52-years of available data (1951–2002) were portioned into two parts, viz the training (1951–1999) and the testing (2000–2002) phases. The training dataset was used for designing both network models. After training the proposed network, a weight matrix was obtained and applied to the independent inputs in the ''test'' set. Then the final outcomes were compared with the observed (actual) values of the Effective Drought Index.

Table 2.1 The neuronal arrangements for the ELM and the ANN models

Location	Model	Training algorithm or Activation function	Neuron structure
Sagar	ANN	Levenberg-Marquardt	10-26-1
	ELM	Logarithmic sigmoid	11-50-1
Chattarpur	ANN	Levenberg-Marquardt	10-32-1
	ELM	Logarithmic sigmoid	12-45-1

For designing the ELM model three layers were used to build the architecture (see Table 2.1) for predicting monthly EDI trained with data from 1951 to 1999, and tested over 2000 to 2002 with observed Effective Drought Index (EDIo). The ELM output layer had one neuron representing the predicted monthly Effective Drought Index (EDIp) but in hidden layers a maximum of 75 neurons are tested. A taxonomy of activation functions were tried one by one, which included sigmoid, log-sigmoid, hyperbolic-tangent sigmoid, radial bias, triangular bias, hyperbolic-tangent sigmoid and hard-limit. In each trial the numbers of nodes in hidden layer were increased gradually by an interval of five. Then, the nearly optimal node for ELM was selected as 50 with the hard-limit activation function and 11-50-1 neurons in the architecture of the ELM model for Sagar and 12-45-1 architecture for Chattarpur district.

In addition, to show the potential of the proposed ELM model for predicting the Effective Drought Index a performance comparison in terms of the estimation capability was made between the ELM and the conventional feedforward ANN model run with the BP algorithm. Because it is a well-known universal estimator, the ANN model can rather be considered as standard benchmark. In accordance with Maier and Dandy (2000), all data prior to its inclusion into the ANN model was scaled appropriately. In the present investigation, the input neurons were scaled in the range of [-1, 1] and a transfer function was implemented to explain the nonlinear relationship between input and output neurons. For determining the optimum ANN model to be used in this work, the set of five backpropagation training algorithms used were as follows: scaled conjugate gradient, one-step secant, BFGS quasi-Newton, Bayesian regulation and Levenberg-Marquardt. Additionally, the set of two commonly used family of hidden transfer functions (hyperbolic-tangent sigmoid and log-sigmoid) and the three output functions (linear, hyperbolic-tangent sigmoid & log-sigmoid) were all tried on the testing datasets one at a time in order to seek the optimum model for the final

experiments. Like in the case of the ELM model, the number of neurons in the hidden layer was varied gradually but this time over the range 4 to 50 neurons. Consequently, the optimum training algorithm was the Levenberg–Marquardt, the hidden transfer function was the hyperbolic-tangent sigmoid and the output transfer function was linear with an ANN architecture of 10-26-1 for Sagar and 10-32-1 for Chattarpur district.

The estimation capability of the monthly EDI from the all machine learning algorithms were statistically evaluated using the following score metrics or prediction error indicators: Root-Mean Square Error (RMSE) and Mean Absolute Error (MAE). Table 2.2 shows the performance capability of the ELM, SVR and the ANN models used in this study. The results confirms that ELM model perform better compare to the ANN and SVR model in all the evaluation criteria.

Table 2.2 performance capability of the ELM, SVR and the ANN models

Location	**Predictive Model**	**RMSE**	**MAE**	**Correlation coefficient**
Sagar	ELM	0.679	0.387	0.835
	ANN	0.690	0.390	0.818
	SVR	0.699	0.435	0.801
Chattarpur	ELM	0.725	0.418	0.794
	ANN	0.743	0.429	0.761
	SVR	0.778	0.457	0.750

Chapter-3

Multiple kernel extreme learning machine model for drought index

3.1 Introduction

Accurate forecasting of drought index is an appealing yet difficult activity in the modern world. Many factors influence the behavior of the drought index, therefore, drought forecasting is regarded as one of the most challenging topics for researchers. In the past, methods based on statistics were proposed for tackling this problem, such as the autoregressive (AR) model, the autoregressive moving average (ARMA) model and the autoregressive integrated moving average (ARIMA) model. These are linear models which are, more than often, inadequate for drought index forecasting. Recently, nonlinear approaches have been proposed, such as artificial neural network, fuzzy neural networks (FNN), support vector regression (SVR), Extreme learning machine etc.

ANN has been widely used machine learning techniques due to its universal approximation property. Previous researchers indicated that ANN, which implements the empirical risk minimization principle, outperforms traditional statistical models. However, ANN suffers from local minimum traps and difficulty in determining the hidden layer size and learning rate. On the contrary, SVR, proposed by Vapnik and his co-workers, has a global optimum and exhibits better prediction accuracy due to its implementation of the structural risk minimization principle which considers both the training error and the capacity of the regression model. However, the practitioner has to determine in advance the type of kernel function and the associated kernel hyperparameters for SVR. Unsuitably chosen kernel functions or hyperparameter settings may lead to significantly poor performance. Most researchers use trial-and-error to choose proper values for the hyperparameters, which obviously takes a lot of efforts. In addition, using a single kernel may not be sufficient to solve a complex problem satisfactorily.

Extreme learning machine (ELM) was first designed for single hidden layer feed forward neural network and then extended to generalized single hidden layer feed forward networks (SLFN) which did not necessarily resemble neurons. Different from traditional neural SLFN learning algorithms ELM aims to minimize both training error and the norm of output weights. Due to its (1) high efficiency, (2) easy-implementation, (3) unification of classification and regression and (4) unification of binary and multi-class classification, ELM has been an active research topic over

the past a few years. In addition the ELM has also been successfully applied to many applications such as imbalance learning, missing data learning and activity recognition etc.

Although researchers have made great progress from both a theoretical and a practical point of view, ELM has still not well considered the following two issues. The first one is how to choose an optimal kernel for a specific application when the kernel trick is applied to ELM such as in previous work. The other one is how to handle information fusion in ELM when multiple heterogenous data sources are available. Liu et al. 2015 proposed a general framework by borrowing the idea of multiple kernel learning (MKL) to handle the above two issues. They called their framework a multiple kernel extreme learning machine (MK-ELM). In the MK-ELM, the optimal kernel is assumed to be a linear combination of a group of base kernels, and the base kernel combination weights and structural parameters of ELM are jointly optimized in the learning process. Though sharing the same assumption that the optimal kernel is a linear combination of base kernels, the proposed MK-ELM and the widely studied SVM based MKL algorithms have important differences. (1) In MK-ELM, the binary and multi-class classification problems are unified in to one common formula. In contrast, the one- against-one (OAO) and one-against-all (OAA) strategies are usually adopted in SVM based MKL algorithms to handle the multi-class classification problems. (2) The optimization problem for MK-ELM is much simpler than the one used in SVM based MKL algorithms. The structural parameter of MK-ELM can be analytically obtained by a matrix inverse operation, while a constrained quadratic programming (QP) solver is required to solve the optimization problems of SVM based MKL algorithms.

The development of better forecasting models for forecasting of drought is an appealing problem as drought events in many part of India are known to significantly impact sustainable economic growth, infrastructure, daily living, agricultural industry and natural ecosystems. However, in predicting future drought the application of the Effective Drought Index (EDI) for drought assessment has been considered superior to the many other indices used in literature. In particular, the merits of the EDI were emphasized in a study by Pandey et al., (2008) that investigated drought in Orissa (India) showed the greater capability of the EDI compared to the other drought indices (e.g., Standardized Precipitation Index, SPI and the Rainfall Decile Drought Index, RDDI) in quantifying water resources in relation to drought. Another study by Morid et al. (2006) in Tehran (Iran) who showed the significantly better performance of EDI, and its better response in detecting

the start of drought when compared to other approaches (e.g., percent of normal, SPI, China-Z index, Z-Score).

The purpose of this investigation to develop multiple kernel Extreme Learning Machine (MK-ELM) model for the forecasting of effective drought Index and compared the output with ANN and ELM models. The prediction capabilities of the proposed model were assessed by performance metrics like the Mean Absolute Error, Root-Mean Square Error.

3.2 Materials and Methods

ELM was proposed for ''generalized'' single-hidden layer feedforward networks where the hidden layer need not be neuron alike (Huang, et al., 2012). The output function of ELM for generalized SLFNs is

$$f_L(\mathbf{X}) = \sum_{i=1}^{L} \boldsymbol{\beta}_i h_i(\mathbf{X}) = \boldsymbol{h}(\mathbf{X})\boldsymbol{\beta}$$

where $\boldsymbol{\beta} = [\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_L]^T$ is the output weight vector between the hidden layer of L nodes to the m≥ 1 output nodes, and $\boldsymbol{h}(\boldsymbol{X}) = [h_1(\mathbf{X}), \dots, h_L(\mathbf{X})]$ is ELM nonlinear feature mapping. The output functions of hidden nodes may not be unique. Different output functions may be used in different hidden neurons. In particular, in real applications $h_i(\mathbf{X})$ can be

$$h_i(\mathbf{X}) = G(\mathbf{a}_i, b_i, \mathbf{X}), \qquad \mathbf{a}_i \in \mathbf{R}^d, b_i \in \mathrm{R}$$

where G(**a**, b, **x**) (with hidden node parameters (a, b)) is a nonlinear piecewise continuous function satisfying ELM universal approximation capability theorems (Huang, et al., 2006).

ELM trains an SLFN in two main stages: (1) random feature mapping and (2) linear parameters solving. In the first stage, ELM randomly initializes the hidden layer to map the input data into a feature space (called ELM feature space1) by some nonlinear mapping functions. The random feature mapping stage differs ELM from many existing learning algorithms such as SVM, which uses kernel functions for feature mapping, or deep neural networks, which use Restricted Boltzmann machines (RBM) or Auto-Encoders/Auto-Decoders for feature learning. The nonlinear mapping functions in ELM can be any nonlinear piecewise continuous functions.

3.2.1 Basic ELM

In the second stage of ELM learning, the weights connecting the hidden layer and the output layer, denoted by β, are solved by minimizing the approximation error in the squared error sense:

$$\underset{\beta \in R^{L\times m}}{min} \ \|\mathbf{H}\beta - \mathbf{T}\|^2$$

where **H** is the hidden layer output matrix (randomized matrix):

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}(\mathbf{X}_1) \\ \vdots \\ \boldsymbol{h}(\boldsymbol{X}_N) \end{bmatrix} = \begin{bmatrix} h_1(\mathbf{X}_1) & \dots & h_L(\boldsymbol{X}_1) \\ \vdots & \vdots & \vdots \\ h_1(\boldsymbol{X}_N) & \cdots & h_L(\boldsymbol{X}_N) \end{bmatrix}$$

and **T** is the training data target matrix:

$$T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix} = \begin{bmatrix} t_{11} & \dots & t_{1m} \\ \vdots & \vdots & \vdots \\ t_{N1} & \cdots & t_{Nm} \end{bmatrix}$$

Where $\|\cdot\|$ denotes the Frobenius norm.

The optimal solution to (3) is given by

$$\beta^* = \mathbf{H}^{\tau}\mathbf{T}$$

where $\mathbf{H}^{\tau}$ denotes the Moore–Penrose generalized inverse of matrix H. There are many efficient ethod, iterative method, and single value decomposition (SVD).

3.2.2 Regularized ELM and ELM kernels

Huang, Zhou, et al. (2012) especially studied the stability and generalization performance of ELM with σ1 = σ2 = p = q = 2:

$$\underset{\beta \in R^{L\times m}}{min} \ \frac{1}{2}\|\boldsymbol{\beta}\|^2 + \frac{C}{2}\sum_{i=1}^{N}\|e_i\|^2 \qquad s.t.\ \mathbf{h}(\mathbf{X}_i)\beta = \mathbf{t}_i^T - \mathbf{e}_i^T,\ \ i = 1, \dots, N.$$

$$\underset{\beta \in R^{L\times m}}{min} \ L_{ELM} = \frac{1}{2}\|\boldsymbol{\beta}\|^2 + \frac{C}{2}\|\mathbf{T} - \mathbf{H}\boldsymbol{\beta}\|^2$$

The above problem is widely known as the ridge regression or regularized least squares. By setting the gradient of LELM with respect to β to zero, we have

$$L_{ELM} = \beta^* - \mathrm{C}\mathbf{H}^T(\mathbf{T} - \mathbf{H}\beta^*) = 0$$

If H has more rows than columns (N > L), which is usually the case where the number of training patterns is larger than the number of the hidden neurons, we have the following closed form solution for β:

$$\beta^* = \left(\boldsymbol{H}^T\boldsymbol{H} + \frac{\boldsymbol{I}}{C}\right)^{-1}\mathbf{H}^T\mathbf{T}$$

where **I** is an identity matrix of dimension L.

Note that in practice, rather than explicitly inverting the L × L matrix in the above expression, we can instead solve a set of linear equations in a more efficient and numerically stable manner. If the number of training patterns is less than the number of hidden neurons (N < L), then **H** will have more columns than rows, which usually gives an under-determined least squares problem. Moreover, it is less efficient to invert a L×L matrix in this case. To handle this problem, we restrict β to be a linear combination of the rows in **H**: $\boldsymbol{\beta} = \boldsymbol{H^T\alpha}(\boldsymbol{\alpha} \in \mathbf{R}^{N\times m})$. Notice that when N < L and **H** is of full row rank, then $\mathbf{HH}^T$is invertible. Substituting $\boldsymbol{\beta} = \mathbf{H}^T\boldsymbol{\alpha}$into (10), and multiplying both sides by $(\mathbf{HH}^T)^{-1}\mathbf{H}$, we get

$$\boldsymbol{\alpha}^* - C(\mathbf{T} - \mathbf{HH}^T\boldsymbol{\alpha}^*) = 0$$

This yield

$$\boldsymbol{\beta}^* = \boldsymbol{H}^T\boldsymbol{\alpha}^* = \mathbf{H}^T\left(\mathbf{HH}^T + \frac{I}{C}\right)^{-1}\mathbf{T}$$

where **I** is an identity matrix of dimension N.

3.2.3 Multiple Kernel Learning

It is well known that the choice of kernels is crucial for kernel-based algorithms Cortes and Vapnik (1995). Much effort has been devoted to tuning an optimal kernel for a specific application. MKL

provides an elegant way to handle such an issue by optimizing a data-dependent kernel. In MKL, the optimal kernel is assumed to be a linear combination of a group of base kernels, and the optimal combination coefficients and the structural parameters of classifiers are jointly learned by maximizing the margin. Specifically, MKL takes the form of

$$k(\cdot,\cdot;\gamma) = \sum_{p=1}^{m} \gamma_p k_p(\cdot,\cdot)$$

where $\{k_p(\cdot,\cdot)\}_{p=1}^{m}$ are *m* pre-defined base kernels, and $\{\gamma_p\}_{p=1}^{m}$ are the base kernel combination coefficients.

3.3 Results and Discussion

Extreme learning machine (ELM) has been an important research topic over the last decade due to its high efficiency, easy-implementation, unification of classification and regression, and unification of binary and multi-class learning tasks. Though integrating these advantages, existing ELM algorithms pay little attention to optimizing the choice of kernels, which is indeed crucial to the performance of ELM in applications. More importantly, there is the lack of a general framework for ELM to integrate multiple heterogeneous data sources for the prediction. In this study, we propose a multiple kernel extreme learning machines (MK-ELM), to improve the forecasting accuracy of drought index of the extreme learning machine. In the proposed MK-ELM, the optimal kernel combination weights and the structural parameters of ELM are jointly optimized.

In this study, we assume that the optimal kernel can be expressed as a linear combination of base kernels, and jointly learn the structural parameters of ELM and the optimal kernel combination coefficients. Several kernel combination has been empirically specified and optimized and found that Gaussian kernel and a polynomial kernel are the best kernel combination.

Details about the data and input selection and drought model development process has been discussed in the section 2.3.1 and 2.3.3 respectively in the previous chapter. To evaluate the performance of the model, we used RMSE and MAE. The RMSE and MAE provide different types of information about the forecasting capabilities of the model in the overall test set. Basically, RMSE measures the goodness-of-fit relevant to high flow values whereas MAE is not weighted towards high(er) magnitude or low(er) magnitude events, but instead evaluates all deviations from

the observed values, in both an equal manner and regardless of sign. Table 3.1 summarizes the performance metrics of the MK-ELM model in terms of its mean statistical metrics. Also, the comparative statistics of the ANN and ELM models used as benchmarks for the MK-ELM are shown. It is found that the statistical performance metrics of the MK-ELM model were better in magnitude to those of the other models.

Table 3.1 Performance metrics of the MK-ELM model compared with the ANN and the ELM models

Location	**Predictive Model**	**RMSE**	**MAE**	**Correlation coefficient**
Sagar	MK-ELM	0.639	0.352	0.856
	ELM	0.679	0.387	0.835
	ANN	0.690	0.390	0.818
Chattarpur	MK-ELM	0.683	0.392	0.831
	ELM	0.725	0.418	0.794
	ANN	0.743	0.429	0.761

A comparison of the performance based on statistical analysis of errors of the predicted output with the observed values for each model for the two district are shown in Table 3.1. The MK-ELM outperformed the ELM and ANN model by all means of performance indicators between the predicted and observed Effective Drought Index. The RMSE and MAE was lower for the MK-ELM compared to the ELM and ANN models. If the comparison of the correlation coefficient is made, then MK-ELM model exhibited the best prediction performance with 0.856 and 0.831 for Sagar and Chattarpur district respectively. The proposed MK-ELM model outperforms to the ELM and ANN model in all the evaluation criteria.

Chapter-4

Wavelet based multiple kernel extreme learning machine model for forecasting drought index

4.1 Introduction

The popularity of discrete wavelet transformation (DWT) in water resources research has grown in recent years (Adamowski and Karapataki 2010; Tiwari and Adamowski 2013, Partal et al. 2015, etc) as wavelet-conjunction forecasting models have been found to be more accurate than classical (non-wavelet) ML counterparts. In the previous chapter, drought forecasting in Sagar and Chattarpur district, the classical MK-ELM based model was developed for the forecasting of the EDI. However, in that study the entire signal was utilized instead of its wavelet decomposed equivalents. Hence in this study, we have used DWT whereby wavelet sub-series with various frequency components of the raw signal were used to develop the W-MK-ELM model. Like other signals of a hydro-meteorological origin, a set of input time-series that represent drought evolution in the temporal domain is likely to exhibit localized high and low frequency components with dynamical features of nonlinearity and non-stationarities. Due to the localization properties in time and scale, wavelet transformation of such signals can allow a ML model to better analyze the time evolution of hydrological processes at different scales. Hence, wavelet decomposition of inputs is considered a multi-resolution tool for pre-processing non-stationary signals; for instance, prior to developing ML models (Adamowski et al. 2012) and analysing trends or periodicities. This technique is similar to short time Fourier transformation as a windowing tool where inputs are decomposed into shifted and scaled versions according to a mother wavelet to extract frequency information.

4.2 Material and Methods

Discrete wavelet transformation

Mathematically, coefficients of wavelet transformation $W_f(r,s)$ of a continuous signal *f(t)* are defined by a linear integral operator (Tiwari and Adamowski 2013):

$$W_f(r,s) = |r|^{\frac{1}{2}} \int_{-\infty}^{+\infty} f(t)\Psi * \left(\frac{t-s}{r}\right) dt$$

where

* is a complex conjugate, the mother wavelet function $\Psi(t)$ is real or complex,

r is the scale/frequency factor controlling the dilation $(r > 1)$ and contraction $(r < 1)$ and

s is the time factor affecting the temporal translation of that function.

It is assumed that the mother wavelet has finite energy (i.e., a condition of admissibility) and therefore, is of the form $\int_{-\infty}^{+\infty} \Psi(t)dt = 0$ with successive wavelet function $\Psi_{r,s}(t)$ written as:

$$\Psi_{r,s}(t) = |r|^{-\frac{1}{2}}\Psi\left(\frac{t-s}{r}\right) \quad r \in R,\ r \neq 0, \quad s \in R$$

Note that when the input signal has infinite energy it will be impossible to cover its frequency spectrum and its time duration with wavelets. However, generally most real-life signals (including those in hydrology) are constrained with finite energy, so the assumption of admissibility can be considered as reasonable. By way of its action, seeks to identify the level of similarity between input time series and mother wavelets at different scales and translations. This generates a set of wavelet coefficient contour maps (or scalogram). In order to transform input data by wavelet analysis, the transformation will search for correlations between the signal and wavelet function at different scales of r and locally around the time of s to produce a scalogram. As it is not feasible to compute wavelet coefficients at every resolution level of r and s, a dyadic grid arrangement can be made where coefficients with values corresponding to powers of two are chosen. This simple, yet efficient method for practical purposes is defined as (Mallat 1998):

$$\Psi_{m,n}\left(\frac{t-s}{r}\right) = r_0^{\frac{-m}{2}}\Psi * \left(\frac{t - ns_0r_0^m}{r_0^m}\right),$$

where m and n are the integers that denote the magnitude of wavelet dilation and translation, respectively ($r_0 > 1$) is a specified dilation step and ($s_0 > 0$) is the location parameter. Typically, it is reasonable to choose ($r_0 = 2$) and ($s_0 = 1$) (Tiwari and Adamowski 2013).

In this study, DWT was applied to select translation and location parameters of discrete input signals. Subsequently, discrete wavelet coefficients (DWCs) were acquired to represent the minimum number of components needed to reflect the time-series according to the mother wavelet. Several wavelet families that have proven useful for various applications are described in Mallat (1998). In practice, hydrologists need to analyse a discrete signal (e.g., monthly, seasonal or annual) rather than a continuous signal as signals in this field of study are mostly discrete in the time domain. For a discrete time-series, $W_f(t)$, that could also represent the input variable of a drought model, assuming ($r_0 = 2$) and ($s_0 = 1$), the DWT function simplifies as:

$$W_f(m,n) = 2^{-m/2} \sum_{t=0}^{N-1} f(t)\Psi * (2^{-m}t - n),$$

where $W_f(m,n)$, is the wavelet coefficient of the DWT process, f(t) is a finite time series (t = 0, 1, 2,…,N – 1), n is the time translation parameter between$0 < n < 2^{M-m} - 1$ and m is the magnitude dilation parameter (1<). In this way, a DWT process is able to perform a multilevel resolution decomposition of time-series by choosing a discrete scale (integer) for m and n to develop a set of wavelet coefficients. In this study, the DWCs were used to create time-series (input) variables in order to forecast the EDI.

Considering that in wavelet analysis, a time-series of length N can be decomposed into M components with zero redundancy, the inverse discrete transform in terms of a signal smoothed component ($\bar{W}$) to denote the entire signal mean is described as:

$$f(t) = \bar{W} + \sum_{m=1}^{M} \sum_{n=0}^{2^{M-m}-1} W_f(m,n) 2^{\frac{-m}{2}} \Psi * (2^{-m}t - n)$$

In simplified form, above Eq. can be written as:

$$f(t) = \bar{W}(t) + \sum_{m=1}^{M} W_m(t),$$

where $\bar{W}(t)$ is the approximation sub-signal at level M, and $W_m(t)$ is the detailed sub-signal at each level $m = 1, 2,...,M$. As the mother wavelet translates across the input signal during the wavelet transformation process, it generates wavelet coefficients that represent similarity between the signal and mother wavelet (at specific scales). The wavelet coefficients, $W_i(t)$ ($i = 1, 2...,M$) show details of the input signal that capture small features of interpretational values in data (generally, the fast changing nature of time-series). The residual (or approximation) term shows the background information of data (long-term or low frequency information). In this study, the DWT technique was adopted as a robust tool without the need for erroneous assumptions or parametric procedures for analysis of the input signal. Because it permits the analysis of more detailed information [W1(t), W2(t),…,Wl(t)] with low frequencies as well as approximated signals with high frequency, the relevant characteristics in a given hydrologic dataset (e.g., periods, hidden period, dependence and jumps) can be easily diagnosed by closely observing the DWCs. Consequently, the accuracy of the ELM model was expected to be higher by pre-processing of the inputs.

3. Results and Discussion

As further improvement to the MK-ELM model, the DWT was applied on the predictor signals to achieve a time-scale representation of the localized and transient phenomenon at different scales in the data series (Tiwari and Adamowski 2013). Figure 4.1 shows a schematic view of the stages in wavelet based model development. Given that the EDI signal was derived from precipitation data which had distinct seasonality and specific frequency components, the DWT process aimed to achieve a time-scale realisation of both the localized and transient phenomena at various frequencies.

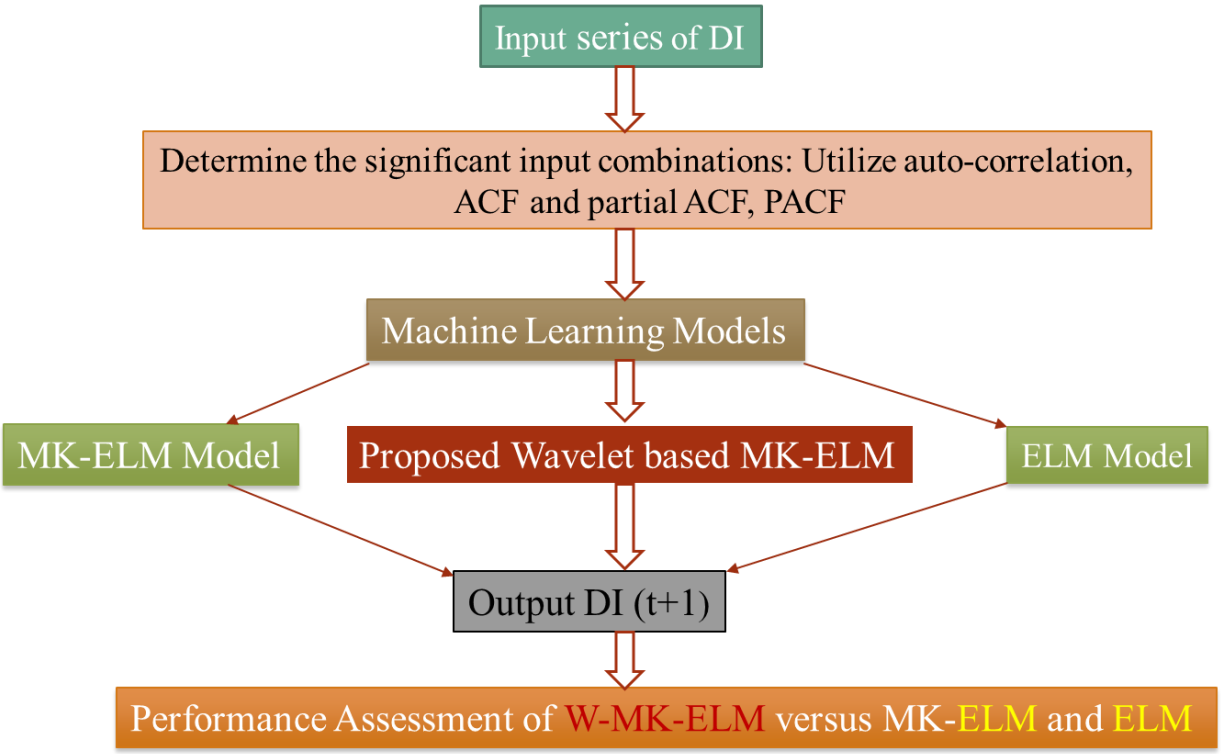

Figure 4.1 schematic view of the stages in wavelet based model development

The wavelet function utilised in this study was adopted from the family of Daubechies mother wavelets, where the DWT process operated as two sets of functions with a high-pass and a low-pass filter. The predictor variables were passed through the high- and low-pass filters to acquire detail (D1, D2, D3) in terms of high frequency components and the approximation coefficients (A3) in terms of low frequency components of the signal. When developing wavelet-based forecast models, it is an important requirement during the decomposition stage to avoid incorporating information from future data (that is to be used in the testing phase) in the calibration set. This is achieved in this study by performing wavelet decomposition on each partition (training, and testing) independently. If this requirement is not followed and the complete dataset (training and testing) is decomposed together (as is sometimes done in other studies in the literature), then future data (that is not truly available to the forecaster at a particular time step) would be used in the calculation of the wavelet and scaling coefficients for a particular time step, unintentionally introducing bias into the forecast. As the performance of the Daubechies (db5) wavelet with three levels of decomposition was the best, for illustration purposes, only three levels of decomposition (D1, D2, D3) and one approximation (A3) for the EDI data over the tested period. The low-

frequency components reflected by A3 showed the broad-scale patterns in the predictor dataset including its periodicity and trends, and were closely in-phase with the predictor signal, whereas the high-frequency components (D1, D2, D3) appeared to replicate greater details of the subtle but significant patterns in the input time-series.

It is noteworthy that although earlier studies demonstrated the better performance of wavelet-based models, the way in which wavelet sub time-series are included in model development can vary. Some studies have used all of their wavelet sub-series, whereas others have removed the db1 sub-series and added the remaining series, considering the former series as noise due to low correlation with their original data. In other studies, new wavelet time-series were developed by adding up the effective DWCs based on regression correlation.

Table 4.1 lists correlation coefficients (r) of wavelet-decomposed signals with the original signal. Interestingly, the approximation component A3 of the predictor time-series appeared to constitute the primary wavelet series for monthly EDI-forecasting. However, considering that the correlation with the original dataset was relatively large in its magnitude (0.313–0.836) compared with previous studies (Tiwari and Adamowski 2013), we used all wavelet sub-series for drought model development in the present study.

Table 4.1 Correlation coefficients (r) of wavelet-decomposed signals with the original signal

Discrete wavelet component	Sagar	Chattarpur
A3	0.836	0.786
D1	0.313	0.318
D2	0.347	0.351
D3	0.375	0.419

In this study, a primary objective was to test the usefulness of the wavelet based MK-ELM model relative to the other ML models considered for monthly forecasting of the EDI. There was a harmonic variation in the evolution of the forecasted and observed value of the DI. The scatter plots revealed significant differences in the performance of the W-MK-ELM model for all stations considered for EDI forecasting. For all three stations, the W-MK-ELM was superior in its performance for EDI forecasting relative to the ELM and the MK-ELM models.

Table 4.2 displays the performance score metrics of the W-MK-ELM, MK-ELM and ELM models. The EDI forecasts generated by the W-MK-ELM were in close agreement with their observed values compared to the simulations by the MK-ELM and ELM models. This was demonstrated by the consistently lower values of the RMSE and MAE for the W-MK-ELM model. When comparing the model performance via the RMSE and MAE, the magnitude of metrics must be as small as possible to reflect small deviations of simulations from observed data. However, the MAE value is less sensitive to extreme values in the simulations than the RMSE value. For the best model the r, which is determined by a scatter plot of observed and forecasted EDI, is expected to be close to unity.

Table 4.2 Performance score metrics of the W-MK-ELM, MK-ELM and ELM models

Location	**Predictive Model**	**RMSE**	**MAE**	**Correlation coefficient**
Sagar	W-MK-ELM	0.621	0.334	0.871
	MK-ELM	0.639	0.352	0.856
	ELM	0.679	0.387	0.835
Chattarpur	W-MK-ELM	0.663	0.375	0.854
	MK-ELM	0.683	0.392	0.831
	ELM	0.725	0.418	0.794

Interestingly, the filtering of the predictor (input) signals by DWT produced a general increase in the magnitude of r and a corresponding reduction in the magnitude of RMSE and MAE for all models. In spite of this improvement, the performance of W-MK-ELM remained superior when compared with MK-ELM and ELM models.

A time-series of the absolute value of prediction errors |PE| generated by the W-MK-ELM, MK-ELM and the ELM models in the test period (Jan 2000–December 2002) is shown in Fig. 4.5. Although mean prediction errors for each month varied significantly, there was a general agreement on overall harmonic variations when compared on a month-by-month basis. Nonetheless, the fluctuations in magnitude of |PE| were generally larger for MK-ELM and ELM models compared to the W-MK-ELM model for all three stations.

A closer examination of the mean forecasted error in different error brackets is exhibited in Table 4.3, where the frequency distribution of the forecast error in various error brackets was analysed using data from all stations for W-MK-ELM compared to the MK-ELM and ELM models.

Table 4.3 The % frequency of the prediction error |PE| in various error brackets for pooled data

Range of prediction error	**W-MK-ELM**	**MK-ELM**	**ELM**
$\mathbf{0 \leq PE < 0.5}$	**93**	**91.2**	**90**
$\mathbf{0.5 \leq PE < 1}$	**7**	**8.0**	**9.3**
$\mathbf{1 \leq PE < 1.5}$	**0**	**0.8**	**0.7**

It is obvious that the frequency of |PE| for the W-MK-ELM model was the highest (≈93 %) for the smallest range of prediction errors (i.e. $0 \leq |PE| \leq 0.5$), whereas that for the MK-ELM model was ≈ 91.2 % and the ELM model was ≈90 %, thus indicating that the wavelet-based MK-ELM model incurred smaller errors than the two counterparts. Notably, in the next upper range of forecasted errors ($0.5 \leq |PE| \leq 1$), approximately 7.0 % of the forecasted errors fell within this range for the W-MK-ELM model compared to 8.0 % for MK-ELM and 9.3 % for ELM models. Likewise, in the next upper band categorized as$(1 \leq |PE| \leq 1.5)$, the W-MK-ELM model did not produce any errors whereas approximately 0.8 % and 0.7% of all forecasted errors were generated by the MK-ELM and ELM models respectively. The wavelet-based MK-ELM model was more accurate than the MK-ELM and ELM models for the forecasting of the EDI.

सार

सूखा-जोखिम प्रबंधन के लिए सूखा पूर्वानुमान मॉडल एक व्यावहारिक उपकरण है। सूखे मॉडल का उपयोग सूखा सूचकांकों (DIs) का अनुमान लगाने के लिए किया जाता है, जो भविष्य के सूखे के प्रभावों की निगरानी और मूल्यांकन करने के लिए इसकी शुरुआत, समाप्ति, और बाद की संपत्तियों जैसे कि गंभीरता, अवधि और चरम तीव्रता से सूखे की मात्रा निर्धारित करते हैं। सूखा सूचकांक सामान्य रूप से वर्षा और / या तापमान, वाष्पीकरण या अन्य औसत दर्जे का चर पर निर्भर करता हैं। सूखा सूचकांकों की गणना के लिए वर्षा के डेटा का व्यापक रूप से उपयोग किया जाता है, क्योंकि लंबे समय तक वर्षा रिकॉर्ड अक्सर उपलब्ध होते हैं। सूखे के आकलन के लिए कई डीआई में, प्रभावी सूखा सूचकांक (ईडीआई) को बेहतर तकनीक माना जाता है। EDI एक गहन सूचकांक है जो सूखे-जोखिम का विश्लेषण करने के लिए समय पर आधारित होने के साथ वर्षा के आंकड़ों के लिए लागू जल संचय पर विचार करता है। इस अध्ययन में प्रभावी सूखा सूचकांक के पूर्वानुमान के लिए मल्टीपल कर्नेल एक्सट्रीम लर्निंग मशीन (MK-ELM) और वेवलेट आधारित MK-ELM एल्गोरिदम प्रस्तावित किए गए हैं। एमके-ईएलएम आधारित तरंगिका के विकास के लिए, इनपुट डेटा को पहले तरंगिका पूर्व प्रसंस्करण के माध्यम से जांचा जाता है। ऑटोकॉरेलेशन फ़ंक्शन (एसीएफ) और आंशिक एसीएफ द्वारा निर्धारित, चालू और पिछले महीनों के लिए अंतरालित ईडीआई संकेतों को 1 महीने के लीड-टाइम ईडीआई पूर्वानुमान के लिए महत्वपूर्ण इनपुट के रूप में उपयोग किया जाता है। सूखे मॉडल विकास के लिए, बुंदेलखंड क्षेत्र के सागर और छतरपुर जिले के डेटा के 52 वर्ष (1951-2002) का उपयोग किया गया है। असतत वेवलेट ट्रांसफ़ॉर्मेशन (डीडब्ल्यूटी) भविष्यवक्ता डेटासेट पर अपने समय-आवृत्ति घटकों में इनपुट विघटित करने के लिए लागू होता है जो आवधिकताओं पर महत्वपूर्ण जानकारी कैप्चर करते हैं। डब्ल्यूडी-एमके-ईएलएम मॉडल के लिए इनपुट के रूप में नई ईडीआई उप-श्रृंखला विकसित करने के लिए डीडब्ल्यूटी उप-श्रृंखला का उपयोग किया गया है। एमके-ईएलएम एवं डब्ल्यू-एमके-ईएलएम की पूर्वानुमान क्षमता को ईएलएम, कृत्रिम तंत्रिका नेटवर्क (एएनएन), LSSVR से तुलना किया गया है। सांख्यिकीय मेट्रिक्स जैसे आर स्कवेर, RMSE एवं MAPE का प्रयोग पूर्वानुमानित और देखे गए ईडीआई के बीच संबंध के लिए किया गया है। परिणाम सूखे मॉडल के उन्नत पूर्वानुमान कौशल को प्रदर्शित करते हैं जो भविष्यवक्ता डेटासेट के वेवलेट प्री-प्रोसेसिंग का उपयोग करते हैं। सांख्यिकीय माप डंडो के आधार पर, डब्ल्यू-एमके-ईएलएम ने पारंपरिक ईएलएम, एलएसएसवीआर, एएनएन मॉडल की तुलना में बेहतर प्रदर्शन किया है। इस अध्यन के परिणाम से पता चलता है कि एमके-

ईएलएम और वेलेट आधारित डब्ल्यू-एमके-ईएलएम मॉडल कि विषेशता तथा सूखे पूर्वानुमान मॉडल के प्रदर्शन को बेहतर बनाने के लिए इनपुट डेटा के तरंग परिवर्तन के लाभों को प्रदर्शित करता हैं।

Abstract

Drought forecasting models are practical tools for drought-risk management. Drought models are used to forecast drought indices (DIs) that quantify drought by its onset, termination, and subsequent properties such as the severity, duration, and peak intensity in order to monitor and evaluate the impacts of future drought. Drought indices are normally continuous functions of precipitation and/or temperature, evapotranspiration or other measurable variable. Precipitation data are widely used to calculate drought indices, because long-term precipitation records are often available. Out of several DIs for drought assessment, Effective Drought Index (EDI) is considered as superior technique. The EDI is an intensive index that considers water accumulation with a weighting function applied to rainfall data with the passage of time in order to analyze the drought-risk. In this study, multiple kernel extreme learning machine (MK-ELM) and wavelet based MK-ELM algorithms are proposed for the forecasting of effective drought index. For the development of wavelet based MK-ELM, the input data are first screened through the wavelet pre-processing. Determined by the autocorrelation function (ACF) and partial ACFs, the lagged EDI signals for the current and past months are used as significant inputs for 1 month lead-time EDI forecasting. For drought model development, 52 years (1951-2002) of data of Sagar and Chattarpur district of Bundelkhand region are used. The discrete wavelet transformation (DWT) is applied to the predictor datasets to decompose inputs into their time–frequency components that capture important information on periodicities. DWT sub-series are used to develop new EDI sub-series as inputs for the W-MK-ELM model. The forecasting capability of MK-ELM and W-MK-ELM is benchmarked with ELM, artificial neural network (ANN), least squares support vector regression (LSSVR) models. Statistical metrics based on agreement between the forecasted and observed

EDI, including the coefficient of determination, root-mean square error and mean absolute error are used to assess the effectiveness of the models. The results demonstrate enhanced forecast skill of the drought models that use wavelet pre-processing of the predictor dataset. Based on statistical measures, W-MK-ELM outperformed traditional ELM, LSSVR, ANN models. The results demonstrate the usefulness of MK-ELM and wavelet based W-MK-ELM over ELM, ANN and LSSVR models and the benefits of wavelet transformation of input data to improve the performance of drought forecasting models.

References

Abbot J, Marohasy J (2012) Application of artificial neural networks to rainfall forecasting in Queensland, Australia. Adv Atmos Sci 29:717–730

Abbot J, Marohasy J (2014) Input selection and optimisation for monthly rainfall forecasting in Queensland, Australia, using artificial neural networks. Atmos Res 138:166–178. doi:10.1016/j.atmosres.2013.11.002

Abdullah S S, Malek M, Abdullah NS, Kisi O, Yap KS (2015) Extreme learning machines: a new approach for prediction of reference evapotranspiration. J Hydrol 527:184–195

Acharya N, Shrivastava N, Panigrahi B, Mohanty U (2013) Development of an artificial neural network based multi-model ensemble to estimate the northeast monsoon rainfall over south peninsular India: an application of extreme learning machine. Clim Dyn 43:1303–1310

Adamowski J, Sun K (2010) Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds. J Hydrol 390:85–91

Adamowski J, Fung Chan H, Prasher SO, Ozga-Zielinski B, Sliusarieva A (2012) Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. Water Resour Res 48:W01528. doi:10.1029/2010WR009945

Belayneh A, Adamowski J (2012) Standard precipitation index drought forecasting using neural networks, wavelet neural networks, and support vector regression. Appl Comput Intell Soft Comput 2012:6. doi:10.1155/2012/794061

Belayneh A, Adamowski J, Khalil B, Ozga-Zielinski B (2014) Longterm SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet neural network and wavelet support vector regression models. J Hydrol 508:418–429

Belayneh A, Adamowski J, Khalil B, Quilty J (2016) Coupling machine learning methods with wavelet transforms and the bootstrap and boosting ensemble approaches for drought prediction. Atmos Res 172–173:37–47

Bishop CM (1995) Neural networks for pattern recognition. Oxford University Press, Oxford

Byun H-R, Wilhite DA (1999) Objective quantification of drought severity and duration. J Clim 12:2747–2756

Daubechies I (1990) The wavelet transform, time–frequency localization and signal analysis. IEEE Trans Inf Theory 36:961–1005

Dayal, K., Deo, R.C. and Apan, A. A. (2017) Drought Modelling Based on Artificial Intelligence and Neural Network Algorithms: A Case Study in Queensland, Australia, Climate Change Adaptation in Pacific Countries, 177-198.

Deo RC, S¸ ahin M (2015a) Application of the Artificial Neural Network model for prediction of monthly Standardized Precipitation and Evapotranspiration Index using hydro-meteorological parameters and climate indices in eastern Australia. Atmos Res 161–162:65–81

Deo RC, S¸ ahin M (2015b) Application of the extreme learning machine algorithm for the prediction of monthly Effective Drought Index in eastern Australia. Atmos Res 153:512–525. doi:10.1016/j.atmosres.2013.11.002

Deo, R.C. Tiwari, M.K., Adamowski, J.F. and quity, J.M. (2016) Forecasting effective drought index using a wavelet extreme learning machine (W-ELM) model, Stoch Environ Res Risk Assess, doi:10.1007/s00477-016-1265-z

Deo, R. C., Kisi, O. and Singh, V. P. (2017) Drought forecasting in eastern Australia using multivariate adaptive regression spline, least square support vector machine and M5Tree model, Atmospheric Research, 184, 149-175.

Field CB (2012) Managing the risks of extreme events and disasters to advance climate change adaptation: special report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge

Goyal MK, Bharti B, Quilty J, Adamowski J, Pandey A (2014) Modeling of daily pan evaporation in sub tropical climates using ANN, LS-SVR, Fuzzy Logic, and ANFIS. Expert Syst Appl 41:5267–5276

Hayes MJ, Svoboda MD, Wilhite DA, Vanyarkho OV (1999) Monitoring the 1996 drought using the standardized precipitation index. Bull Am Meteorol Soc 80:429–438

Huang G-B, Zhu Q-Y, Siew C-K (2006) Extreme learning machine: theory and applications. Neurocomputing 70:489–501

Huang GB, Zhou H, Ding X, Zhang R (2012) Extreme learning machine for regression and multiclass classification. IEEE Trans Syst Man Cybern B 42:513–529

Huang G, Huang G-B, Song S, You K (2015) Trends in extreme learning machines: a review. Neural Netw 61:32–48

Karran D, Morin E, Adamowski J (2014) Multi-step streamflow forecasting using data-driven non-linear methods in contrasting climate regimes. J Hydroinform 16(3):671–689

Kim D-W, Byun H-R, Choi K-S (2009) Evaluation, modification, and application of the Effective Drought Index to 200-Year drought climatology of Seoul, Korea. J Hydrol 378:1–12

Kim D-W, Byun H-R, Choi K-S, Oh S-B (2011) A spatiotemporal analysis of historical droughts in Korea. J Appl Meteorol Climatol 50:1895–1912

Koutsoyiannis D (2007) Discussion of ‘‘Generalized regression neural networks for evapotranspiration modelling’’. Hydrol Sci J 52:832–839

Li B, Cheng C (2014) Monthly discharge forecasting using wavelet neural networks with extreme learning machine. Sci China Technol Sci 57:2441–2452

Liu X, a,n, Wang L, Huang G, Zhang J, Yin J, 2015 Multiple kernel extreme learning machine, 149; 253-264.

Lippman, R., 1987. An introduction to computing with neural nets, IEEE ASSP Magazine 4, 4-22.

Maier HR, Dandy GC (2000) Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. Environ Model Softw 15:101–124

Mallat S (1998) A wavelet tour of signal processing. Academic, New York

McCulloch, W.S., Pitts, W., 1943. A logical calculus of the ideas immanent in nervous activity, The bulletin of mathematical biophysics 5, 115-133.

Mekanik F, Imteaz M, Gato-Trinidad S, Elmahdi A (2013) Multiple regression and Artificial Neural Network for long-term rainfall forecasting using large scale climate modes. J Hydrol 503:11–21

Mishra AK, Singh VP (2010) A review of drought concepts. J Hydrol 391:202–216

Mishra AK, Singh VP (2011) Drought modeling—a review. J Hydrol 403:157–175

Montanari A, Rosso R, Taqqu MS (1996) Some long-run properties of rainfall records in Italy. J Geophys Res Atmos (1984–2012) 101:29431–29438

Moustris, K.P., Larissi, I.K., Nastos, P.T., Paliatsos, A.G., 2011. Precipitation forecast using artificial neural networks in specific regions of Greece, Water resources management 25, 1979-1993.

Morid S, Smakhtin V, Moghaddasi M (2006) Comparison of seven meteorological indices for drought monitoring in Iran. Int J Climatol 26:971–985

Morid, S., Smakhtin, V., Bagherzadeh, K., 2007. Drought forecasting using artificial neural networks and time series of drought indices, International Journal of Climatology 27, 2103- 2111.

Nalley D, Adamowski J, Khalil B (2012) Using discrete wavelet transforms to analyze trends in streamflow and precipitation in Quebec and Ontario (1954–2008). J Hydrol 475:204–228

Ortiz-Garcı´a E, Salcedo-Sanz S, Casanova-Mateo C (2014) Accurate precipitation prediction with support vector classifiers: a study including novel predictive variables and observational data. Atmos Res 139:128–136

Pandey R, Dash B, Mishra S, Singh R (2008) Study of indices for drought characterization in KBK districts in Orissa (India). Hydrol Process 22:1895–1907

Partal T, Cigizoglu HK, Kahya E (2015) Daily precipitation predictions using three different wavelet neural network algorithms by meteorological data. Stoch Environ Res Risk Assess 29:1317–1329

Rathinasamy M, Khosa R (2012) Comparative study of different wavelets for hydrologic forecasting. Comput Geosci 46:284–295

Saadat H, Adamowski J, Bonnell R, Sharifi F, Namdar M, Ale-Ebrahim S (2011) Land use and land cover classification over a large area in Iran based on single date analysis of satellite imagery. ISPRS J Photogramm Remote Sens 66:608–619

Sahin M (2012) Modelling of air temperature using remote sensing and artificial neural network in Turkey. Adv Space Res 50:973– 985.

Sahin M, Kaya Y, Uyar M (2013) Comparison of ANN and MLR models for estimating solar radiation in Turkey using NOAA/ AVHRR data. Adv Space Res 51:891–904.

Sahin M, Kaya Y, Uyar M, Yıldırım S (2014) Application of extreme learning machine for estimating solar radiation from satellite data. Int J Energy Res 38:205–212

Suykens JA, De Brabanter J, Lukas L, Vandewalle J (2002) Weighted least squares support vector machines: robustness and sparse approximation. Neurocomputing 48:85–105

Tiwari, M.K., Adamowski, J., 2013. Urban water demand forecasting and uncertainty assessment using ensemble wavelet-bootstrap-neural network models, Water Resources Research 49, 6486-6507.

Vapnik V (1995) The nature of statistical learning theory. Springer, New York