Drought forecasting using feed-forward recursive neural network

Article *in* Ecological Modelling · September 2006

DOI: 10.1016/j.ecolmodel.2006.04.017

2 authors:

Ashok Mishra
Clemson University

V. R. Desai
Indian Institute of Technology Kharagpur

ECOLOGICAL MODELLING 198 (2006) 127–138

available at www.sciencedirect.com

ScienceDirect

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

Drought forecasting using feed-forward recursive neural network

A.K. Mishra, V.R. Desai*

Department of Civil Engineering, Indian Institute of Technology, Kharagpur 721302, India

ARTICLE INFO

Article history:
Received 5 January 2005
Received in revised form 25 April 2006
Accepted 27 April 2006
Published on line 14 June 2006

Keywords:
Kansabati catchment
Neural networks
ARIMA
SPI
Forecasting

ABSTRACT

Drought affects natural environment of an area when it persists for a longer period. So, drought forecasting plays an important role in the planning and management of natural resources and water resource systems of a river basin. During last decade neural networks have shown great ability in modeling and forecasting nonlinear and non-stationary time series. This paper compares linear stochastic models (ARIMA/SARIMA), recursive multi-step neural network (RMSNN) and direct multi-step neural network (DMSNN) for drought forecasting. The models were applied to forecast droughts using standardized precipitation index (SPI) series as drought index in the Kansabati River Basin, which lies in the Purulia district of West Bengal, India. The results obtained from three models and their potential to forecast drought over different lead times are presented in this paper.

1. Introduction

Drought is a normal feature of climate and its recurrence is inevitable. However, there remains much confusion within the scientific and policy-making community about its characteristics. Research has shown that the lack of a precise and objective definition in specific situations has been an obstacle to understanding drought which has led to indecision and inaction on the part of managers, policy-makers, and others (Wilhite and Glantz, 1985; Wilhite et al., 1986). The global climate change in recent years is likely to enhance the number of incidences of droughts. While much of the weather that we experience is brief and short-lived, drought is a more gradual phenomenon, slowly taking hold of an area and tightening its grip with time. In severe cases, drought can last for many years, and can have devastating effects on agriculture and water supplies. It may be difficult to determine when a drought begins or ends. A drought can be short, lasting just a few months, or persist for years before climatic conditions return to normal. Because the impacts of a drought accumulate slowly at first, a drought may not even be recognized until it has become well established.

Between 1967 and 1992, droughts have affected 50% of the 2.8 billion people who suffered from all natural disasters. Because of direct and indirect impacts of droughts, 1.3 million human lives were lost, out of a total number of 3.5 million people killed by disasters (Obasi, 1994). Nearly 50% of the world's most populated areas are highly vulnerable to drought. More importantly, almost all of the major agricultural lands are located there (USDA, 1994). Drought produces a complex web of impacts that spans many sectors of the economy and reaches well beyond the area experiencing physical drought. In India drought is common and these drought areas are mainly confined to the Peninsular and Western parts of the country and there are few pockets in other parts of India. Out of 795 million ha of geographical area in India about 260

* *Corresponding author*. Tel.: +91 9434368750.
E-mail addresses: akmishra@civil.iitkgp.ernet.in (A.K. Mishra), venkapd@civil.iitkgp.ernet.in (V.R. Desai).
0304-3800/$ – see front matter
doi:10.1016/j.ecolmodel.2006.04.017

million ha of land are subjected to different degrees of water stress and drought conditions.

Drought forecasting plays an important role in the mitigation of impacts of drought on water resources systems. Traditionally, statistical models have been used for hydrologic drought forecasting based on time series methods. Regression models and autoregressive moving average (ARMA) models are typical models for statistical time series methods for forecasting. However they are basically linear models assuming that data are stationary, and have a limited ability to capture non-stationarities and nonlinearities in the hydrologic data. Univariate Box–Jenkins ARIMA analysis (Box et al., 1994) has been extensively used for forecasting hydrologic variables of interest such as, annual and monthly stream flows, precipitation, etc., which have been generally accepted by practitioners during the past several decades. However, it is necessary for hydrologists to consider alternative models when nonlinearity and non-stationarity play a significant role in the forecasting. In recent decades, artificial neural networks (ANNs) have shown great ability in modeling and forecasting nonlinear and non-stationary time series in hydrology and water resource engineering due to their innate nonlinear property and flexibility for modeling. Some of the advantages of ANNs are (ASCE, 2000a).

(1) They are able to recognize the relation between the input and output variables without explicit physical considerations. (2) They work well even when the training sets contain noise and measurement errors. (3) They are able to adapt to solutions over time to compensate for changing circumstances. (4) They possess other inherent information-processing characteristics and once trained are easy to use. An application of ANN to solve civil engineering problems began in the late 1980s (Flood and Kartam, 1994a,b). Preliminary concepts of artificial neural networks (ANNs) and their adaptability to hydrology are well explained in ASCE (2000a) and Govindaraju and Rao (2000). An exhaustive list of references on ANN applications in hydrology can be referred to ASCE (2000b). Its application to simulation and forecasting problems in water resources has shown great ability and some of the applications are mention here. Karunanithi et al. (1994), Hsu et al. (1995), attempted to predict flow at the current catchment outlet with inputs such as rainfall, upstream flow, and/or temperature only. Nagesh Kumar et al. (2004) used recurrent neural networks for river flow forecasting. Some of the applications of ANN used in ecological modeling are: Acharya (2006) used a three layer feed-forward neural network model consisting of an input layer, one hidden layer and an output layer to predict the extent of sulphur removal from three types of coal using native cultures of Acidithiobacillus ferrooxidans. Pasini et al. (2006) used a feed-forward neural network, trained by means of a backpropagation strategy using a generalised Widrow–Hoff rule for updating the connection weights for the analysis of forcings/temperatures relationships at different scales in the climate system. Melesse and Hanley (2005) used a feed-forward neural network using a back propagation algorithm to three different ecosystems (forest, grassland and cropland) using partitioned energy fluxes, air and soil temperature as input variables to predict carbon flux is presented. Sahoo et al. (2005) used ANN for prediction of pesticide occurrence in rural domestic wells from the available limited information. Among the three ANN models (a feed-forward back propagation (BP), a radial basis function (RBF) and an adaptive neural network-based fuzzy inference system (ANFIS)) employed for this investigation, the BP neural network was found to be superior to RBF and ANFIS type networks for the detection of pesticide occurrences in wells. Gevrey et al. (2003) used multi-layer feed-forward network using an error backpropagation training algorithm studied the explanatory capacity of the variables to identify environmental factors affecting trout abundance and how these factors contribute to trout abundance. Antonic et al. (2001) used feed-forward ANN with multilayer perceptions (MLP) for empirical model development using seven climatic variables (monthly mean air temperature, monthly mean daily minimum and maximum air temperature, monthly mean relative humidity, monthly precipitation, monthly mean global solar irradiation and monthly potential evapotranspiration). Recknagel (1997) used neural network shell explorer, which is a feed-forward network with backpropagation algorithm for modeling of algal blooms in four different freshwater systems. Lae et al. (1999) used multilayer feed-forward neural network approach for modelling and prediction of fish yield with relation to the environmental characteristics developed from the combination of six variables: catchment area over maximum area, fishing effort, conductivity, depth, altitude and latitude for 59 lakes in Africa. Walter (2001) used two type of model: (i) SALMO which is driven by process-based differential equations model, (ii) ANNA is designed as recurrent feed-forward neural network trained by time series data in prediction of phytoplankton abundance in the Burrinjuck Reservoir.

Scardi (2001) used the error backpropagation (EBP) algorithm in all neural networks for phytoplankton primary production modeling. Karul et al. (2000) used a three-layer feed-forward neural network using a tangent-sigmoid transfer function between the input layer and the hidden layer, and a linear transfer function was selected between the hidden layer and the output layer to model the eutrophication process in three water bodies of Turkey. Jeong (2001) used a recurrent artificial neural network for time series modelling of phytoplankton dynamics in the hypertrophic Nakdong River system considering meteorological, hydrological and limnological parameters as input variables and chl. a concentration as output variable.

It is observed that most of the neural networks are based on the multilayer feed-forward neural network using back propagation algorithm. Most of the papers, where the neural network are used for prediction of events over different lead time are based on direct approach. In the present paper a different approach is used which is known as recursive multistep approach to forecast drought events over different lead time. The differences between two approaches are discussed in the present paper.

Accurate drought forecasting would enable optimal operation of irrigation systems. The ARMA models, pattern recognition techniques, physically based models using Palmer drought severity index (PDSI), standardized precipitation index (SPI), a moisture adequacy index involving Markov chains, or the notion of conditional probability, seems to offer a potential to develop reliable and robust forecasts towards this goal (Panu and Sharma, 2002). Rao and Padmanabhan

(1984) investigated the stochastic nature of yearly and monthly Palmer's drought index (PDI) and to characterize them using valid stochastic models to forecast and to simulate PDI series. Sen (1990) predicted the possible critical drought durations that may result from any hydrologic phenomenon during any future period using second order Markov chain. Kim and Valdes (2003) used PDSI as drought parameter to forecast drought in the Conchos River Basin in Mexico.

The Neural network models presented in this paper are based on SPI as drought index. The SPI is used in this study for the following advantages, which are discussed by Hayes et al. (1999). The primary reason is that SPI is based on rainfall alone, so that drought assessment is possible even if other hydrometeorological measurements are not available. The SPI is also not adversely affected by topography. Another advantage of SPI is its variable timescale, which allows it to describe drought conditions important for a range of meteorological, hydrological and agricultural applications. The third advantage of SPI comes from its standardization, which ensures that the frequencies of extreme events at any location and on any time scale are consistent. SPI can also detect moisture deficit more rapidly than PDSI, which has a response time scale of approximately 8–12 months (Hayes et al., 1999).

The main objective of present study is to calculate time series of standardized precipitation index (SPI) for multiple time scales and to compare neural networks model with linear stochastic models to forecast drought-using SPI as drought index. The potential of models to forecast drought over different lead times are discussed here.

2. Database

The physical area considered in this study is the portion of Kansabati River Basin upstream of Kangsabati Dam, in the extreme western part of West Bengal state in eastern India. The region has an area of 4265 km^2. The major crops grown in the catchment are paddy, maize, pulses and vegetables. It is considered a drought prone area with irregular rainfall and the soil is mostly laterite having low water holding capacity. About 50–60% of the study area is upland, which is managed by the poor farmers. Lands are mostly mono-cropped having limited surface irrigation facilities. Irrigated crops are not widespread because water is not enough for that purpose always. For this study, five raingauge stations were considered as shown in Table 1. Monthly rainfall data was procured for the period from 1965 to 2001 for these stations. The basin was affected by severe droughts in the years 1965–1967 and around 1980s, which was for a longer duration. The severity of drought in 1990s was less though frequency of short term drought oscillations was more (Mishra and Desai, 2005a). This indicates more frequent droughts in the basin. SPI time series for multiple time scales were derived for the average rainfall over the basin and these SPI values were used as drought index for forecasting the drought.

Table 2 – Drought classification based on SPI

SPI values	Class
>2	Extremely wet
1.5–1.99	Very wet
1.0–1.49	Moderately wet
−0.99 to 0.99	Near normal
−1 to −1.49	Moderately dry
−1.5 to −1.99	Severely dry
<−2	Extremely dry

2.1. *Development of SPI series in Kansabati catchment*

A deficit of precipitation impacts on soil moisture, stream flow, reservoir storage, and ground water level, etc., on different time scales. McKee et al. (1993) developed the SPI to quantify precipitation deficits on multiple scales. The nature of the SPI allows an analyst to determine the rarity of a drought or an anomalously wet event at a particular time scale for any location in the world that has a precipitation record. A drought event occurs at the time when the value of SPI is continuously negative. The event ends when the SPI becomes positive. Table 2 provides a drought classification based on SPI.

Bussay et al. (1999) and Szalai and Szinell (2000) assessed the utility of SPI for describing drought in Hungary. They concluded that SPI was suitable for quantifying most types of drought event. Stream flow was best described by SPIs with time scale of 2–6 months. Strong relationships to ground water level were found at time scales of 5–24 months. Agricultural drought (i.e., deficit of soil moisture content) was replicated by the SPI on a scale of 2–3 months. Lana et al. (2001) used the SPI to investigate patterns of rainfall over Catalonia, Spain. Hughes and Saunders (2002) studied drought climatology for Europe based on monthly SPIs at time scales of 3, 6, 9, 12, 18, and 24 months for the period 1901–1999.

Calculation: The SPI is computed by fitting a probability density function to the frequency distribution of precipitation summed over the time scale of interest. This is performed separately for each month (or whatever the temporal basis is of the raw precipitation time series) and for each location in

Table 1 – Raingauge stations in the Kansabati River Basin

Station no.	Raingauge station	Area (km^2)	Elevation (m) (a.m.s.l)	Geographic coordinates	
				Latitude	Longitude
1	Simulia	1279.5	220.97	23°10′	86°22′
2	Rangagora	1151.55	222.92	23°4′	86°24′
3	Tusuma	554.45	158.6	23°08′	86°43′
4	Kharidwar	682.4	135.96	23°00′	86°38′
5	Phulberia	597.1	144.32	22°55′	86°37′

space. Each probability density function is then transformed in to the standardized normal distribution.

The gamma distribution is defined by its frequency or probability density function is defined as

$$g(x) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}, \quad \text{for } x > 0 \tag{1}$$

where $\alpha > 0$ is a shape factor, $\beta > 0$ is a scale factor, and $x > 0$ is the amount of precipitation. $\Gamma(\alpha)$ is the gamma function which is defined as

$$\Gamma(\alpha) = \int_0^{\infty} y^{\alpha-1} e^{-y}\, dy \tag{2}$$

Fitting the distribution to the data requires α and β to be estimated. Edwards and McKee (1997) suggest estimating these parameters using the approximation of Thom (1958) for maximum likelihood as follows:

$$\hat{\alpha} = \frac{1}{4A}\left(1 + \sqrt{1 + \frac{4A}{3}}\right) \tag{3}$$

$$\hat{\beta} = \frac{\bar{x}}{\hat{\alpha}} \tag{4}$$

where for n observations

$$A = \ln(\bar{x}) - \frac{\sum \ln(x)}{n} \tag{5}$$

The resulting parameters are then used to find the cumulative probability of an observed precipitation event for the given month and time scale:

$$G(x) = \int_0^x g(x)\, dx = \frac{1}{\hat{\beta}^{\hat{\alpha}}\Gamma(\hat{\alpha})} = \int_0^x x^{\hat{\alpha}-1} e^{-x/\hat{\beta}}\, dx \tag{6}$$

Substituting t for $x/\hat{\beta}$ reduces equation to incomplete gamma function. McKee et al. (1993) use an analytic method along with suggested software code from Press et al. (1986). Since the gamma function is undefined for $x = 0$ and a precipitation distribution may contains zeros, the cumulative probability becomes

$$H(x) = q + (1 - q)G(x) \tag{7}$$

where q is the probability of zero precipitation.

The cumulative probability, $H(x)$, is then transformed to the standard normal random variable Z with mean zero and variance one, which is the value of SPI. Following Edwards and McKee (1997), Hughes and Saunders (2002), we employ the approximate conversion provided by Abramowitz and Stegun (1965) as an alternative:

$$Z = \text{SPI} = -\left(t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3}\right), \quad \text{for } 0 < H(x) \le 0.5 \tag{8}$$

$$Z = \text{SPI} = +\left(t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3}\right), \quad \text{for } 0.5 < H(x) < 1 \tag{9}$$

Table 3 – Correlation matrix SPI vs. hydrological variables

SPI series	Monthly discharge in river	Monthly reservoir storage
SPI 1	0.718	0.3166
SPI 3	0.555	0.661
SPI 6	0.359	0.589
SPI 9	0.236	0.224
SPI 12	0.22	0.188
SPI 24	0.099	0.0844

where

$$t = \sqrt{\ln\left[\frac{1}{(H(x))^2}\right]}, \quad \text{for } 0 < H(x) \le 0.5 \tag{10}$$

$$t = \sqrt{\ln\left[\frac{1}{(1 - H(x))^2}\right]}, \quad \text{for } 0.5 < H(x) < 1 \tag{11}$$

and

$$c_0 = 2.515517,\; c_1 = 0.802853,\; c_2 = 0.010308,\; d_1 = 1.432788,$$

$$d_2 = 0.189269,\; d_3 = 0.001308$$

The SPI series for different timescales are shown in Fig. 1. The correlation coefficient between average discharge in river and reservoir storage over different months with SPI calculated over multiple time scales is shown in Table 3. It is observed that SPI 1 and SPI 3 have significant correlations with Kansabati river flow discharge and SPI 3, SPI 6 with the storage in the reservoir which lies in the downstream of Kansabati River.

3. Methodology

3.1. *ARIMA models*

The stochastic models which are often known as time series models (ARIMA) have been used in scientific, economic and engineering applications for the analysis of time series. Time series modeling techniques have been shown to provide a systematic empirical method for simulating and forecasting the behavior of uncertain hydrologic systems and for quantifying the expected accuracy of the forecasts which can be found in lot of research literatures. There are two classes of stochastic models, which are described below.

3.1.1. *Non-seasonal models*

Autoregressive (AR) models can be effectively coupled with moving average (MA) models to form a general and useful class of time series models called autoregressive moving average (ARMA) models. In an ARMA model the current value of the time series is expressed as a linear aggregate of p previous values and a weighted sum of q previous deviations (original value minus fitted value of previous data) plus a random parameter. However they can be used when the data

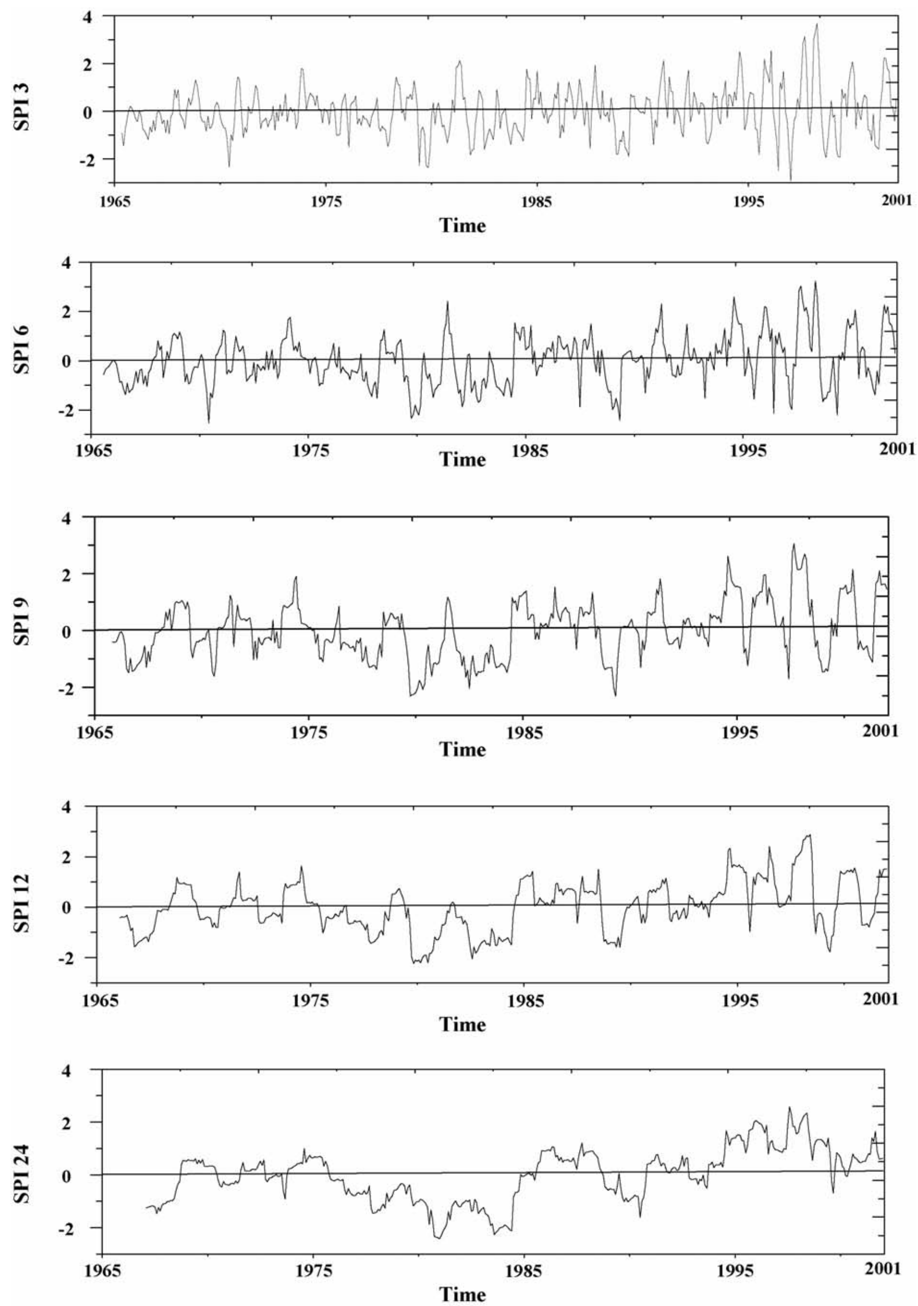

Fig. 1 – SPIs series over different time scale based on average rainfall over Kansabati basin.

are stationary. This class of models can be extended to non-stationary series by allowing differencing of data series. These are called autoregressive integrated moving average (ARIMA) models. Box and Jenkins (1976) popularized ARIMA models. The general non-seasonal ARIMA model is autoregressive to order p and moving average to order q and operates on dth difference of the time series z_t; thus a model of the ARIMA family is classified by three parameters (p, d, q) that can have zero or positive integral values.

The general non-seasonal ARIMA model may be written as

$$\phi(B)\nabla^d z_t = \theta(B)a_t \tag{12}$$

where $\phi(B)$ and $\theta(B)$ are polynomials of order p and q, respectively:

$$\phi(B) = (1 - \phi_1 B - \phi_2 B^2 - \ldots \phi_p B^p) \tag{13}$$

and

$$\theta(B) = (1 - \theta_1 B - \theta_2 B^2 - \cdots \theta_q B^q) \tag{14}$$

3.1.2. *Seasonal models*

Box et al. (1994) have generalized the ARIMA model to deal with seasonality, and define a general multiplicative seasonal ARIMA model, which are commonly known as SARIMA models. In short, the SARIMA model can be described as ARIMA $(p, d, q)(P, D, Q)_s$, where (p, d, q) is the non-seasonal part of the model and $(P, D, Q)_s$ is the seasonal part of the model, which is mentioned below:

$$\phi_p(B)\Phi_P(B^s)\nabla^d\nabla_s^D z_t = \theta_q(B)\Theta_Q(B^s)a_t \qquad (15)$$

p is the order of non-seasonal autoregression, d the number of regular differencing, q the order of non-seasonal moving average, P the order of seasonal autoregression, D the number of seasonal differencing, Q the order of seasonal moving average and s is the length of the season.

The time series model development consists of three stages, i.e., identification, estimation and diagnostic check (Box et al., 1994) which are available in literatures and time series books.

3.2. *Artificial neural networks*

Neural networks are a class of flexible nonlinear models that can discover patterns adaptively from the data. Theoretically, it has been shown that given an appropriate number of nonlinear processing units, neural networks can learn from experience and estimate any complex functional relationship with high accuracy. Although many types of neural network models have been proposed, the most popular one for time series forecasting is the feed-forward model. Fig. 2 shows a typical three-layer feed-forward model used for forecasting purposes. The input nodes are the previous lagged observations while the output provides the forecast for the future value. Hidden nodes with appropriate nonlinear transfer functions are used to process the information received by the input nodes. In the present paper two different approaches of neural networks for forecasting several time steps ahead are discussed below.

(a) *Recursive multi-step neural network approach* (RMSNN): This forecasting technique is similar to ARIMA models in forecasting approach which has single output node. A recursive multi-step approach based on one output node, forecasting a single step ahead, and the network is applied recursively, using the previous predictions as inputs for the subsequent forecasts (Fig. 2). In this way it was carried out recursively for six steps to find drought events over 6-month lead time.

(b) *Direct multi-step neural network approach* (DMSNN): This approach is based on the advantages of neural networks over ARIMA models. It is based on the multiple outputs, when several nodes are included in the output layer, and each output node represents one time step to be forecasted, which is shown in Fig. 3. In the present study there are six out put nodes, indicating 1–6-month lead time.

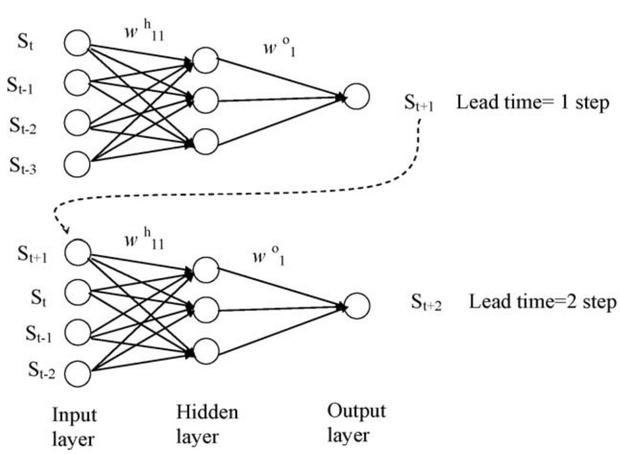

Fig. 2 – Feed-forward recursive multi-step Neural Network approach.

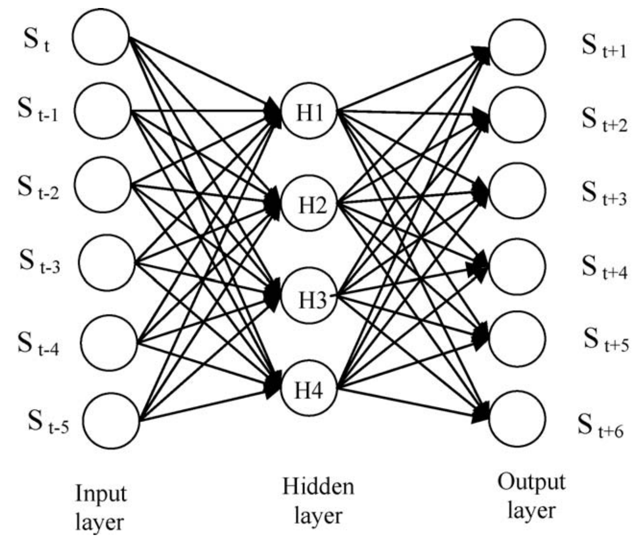

Fig. 3 – Direct multi-step Neural Network approach.

To build a model for forecasting, the network is processed through three stages: (1) The training stage where the network is trained to predict future data, based on past and present data. (2) The testing stage where the network is tested to stop training or to keep in training. (3) The evaluation stage where the network ceases training and is used to forecast future data and to calculate different measures of error. Back propagation algorithm, which is essentially a steepest gradient descent method is used in the present study.

3.2.1. *Back propagation training algorithm for three layered neural networks*

Back propagation network (BPN), developed by Rumelhart et al. (1986) is the most prevalent of the supervised learning models of ANN. BPN uses the steepest gradient descent method to correct the weight of the interconnectivity neuron. BPN easily solves the interaction of the processing of processing elements by adding hidden layers. In the learning process of BPN, the interconnection weights are adjusted using error convergence technique to obtain a desired output for a given input. In general, the error at the output layer in the BPN model propagates backward to the input layer through the hidden layer in the network to obtain the final desired output. The gradient descent method is utilized to calculate the weight of the network and adjusts the weight of interconnections to minimize the output error. The error function at the output neuron is defined as

$$E = \frac{1}{2}\sum_{k}(T_k - A_k)^2 \quad (16)$$

In which T_k and A_k represent the actual and predicted values of output neuron, k.

The gradient descent algorithm adapts the weights according to the gradient error, which is given by

$$\nabla W_{ij} = -\eta \times \frac{\partial E}{\partial W_{ij}} \quad (17)$$

where η is the learning rate and the general form of the $\frac{\partial E}{\partial W_{ij}}$ is expressed by the following form (Rumelhart et al., 1986):

$$\frac{\partial E}{\partial W_{ij}} = -\delta_j^n A_i^{n-1} \quad (18)$$

Substituting (14) into (13), we have the gradient error as

$$\nabla W_{ij} = \eta \delta_j^n A_i^{n-1} \quad (19)$$

in which A_i^{n-1} is the output value of sub-layer related to the connective weight (W_{ij}).

δ_j^n is the error signal, which is computed based on whether or not the neuron j is in the output layer. If neuron j is one of the output neurons, then

$$\delta_j = (T_j - Y_j)Y_j(1 - Y_j) \quad (20)$$

If neuron j is a neuron of the hidden layer

$$\delta_j = \left[\sum_{j}\delta_j (W_{hy})_{hj}]H_h(1 - H_h)\right] \quad (21)$$

where H_h is the value of hidden layer.

Finally, the value of weight of the inter-connective neuron can be expressed as follows:

$$W_{ij}^m = W_{ij}^{m-1} + \nabla W_{ij}^m = W_{ij}^{m-1} + \eta\delta_j^n A_i^{n-1} \quad (22)$$

To accelerate the convergence of the error in learning procedure, Jacobs (1988) proposed the momentum term with momentum gain, α, in Eq. (18):

$$W_{ij}^m = W_{ij}^{m-1} + \eta\delta_j^n A_i^{n-1} + \alpha\nabla W_{ij}^{m-1} \quad (23)$$

in which, value for α is between 0 and 1.

3.2.2. *Design of network*

The use of an ANN for forecasting time series implies that the input nodes reconnected to a number of past-observed values to identify the processes at future time steps. The activation function determines the relationship between input and outputs of a node and a network. In the present work sigmoid function $\left(\frac{1}{1+e^{-x}}\right)$ is used, which is the most popular choice. Data sets are normalized before the training begins using the following equation:

$$X_n = \frac{X_0 - X_{min}}{X_{max} - X_{min}} \quad (24)$$

where X_n and X_0 represent the normalized and original data. X_{min} and X_{max} represent the minimum and maximum value among original data.

In time series problems the number of input nodes corresponds to the number of lagged observations used to discover the underlying pattern in a time series and to make forecasts for future values. The hidden layer and nodes play very important roles for many successful applications of neural networks. It is the nodes in the hidden layer that allow neural networks to detect the feature, to capture the pattern in the data, and to perform complicated nonlinear mapping between input and output variables. It has been proved that only one layer of hidden units is sufficient for ANNs to approximate any complex nonlinear function with any desired accuracy (Cybenko, 1989; Hornik et al., 1989). The hidden nodes also allow taking into account the presence of non-stationarities in the data, such as trends and seasonal variations (Maier and Dandy, 1996). In the case of the popular one hidden layer networks, several practical guidelines exist. These include using "$2n+1$" (Lippmann, 1987; Hecht-Nielsen, 1990), "$2n$" (Wong, 1991), "n" (Tang and Fishwick, 1993) hidden neurons for better forecasting accuracy, where n is the number of input nodes. Fewer neurons in the hidden layer than in the input layer has worked well in the past (Fletcher and Goss, 1993; Zhang and Dong, 2001). In order to determine the optimal network architecture, the number of neurons in the input and hidden layer were determined by experimentation. Tang and Fishwick (1993), claim that the number of input nodes is simply the number of autoregressive (AR) terms in the Box–Jenkins model for a univariate time series. This is not true because: (1) for moving average (MA) processes, there are no AR terms; (2) Box–Jenkins models are linear models. The number of AR terms only tells the number of linearly correlated lagged observations and it is not appropriate for the nonlinear relationships modeled by neural networks (Zhang et al., 1998). In the present study input neurons (n) ranging from 1 to 20 were tested. For each input layer dimension, the numbers of hidden nodes (h) were progressively increased from 1 to $2n+1$, where n is the corresponding input neurons. The coefficient of correlation for each combination of input and hidden neurons is calculated. The combination having maximum coefficient of correlation and minimum root mean square error is chosen as optimal network. The network is trained for 5000 epochs using back propagation algorithm with learning rate of 0.01 and momentum coefficient 0.9.

The performance of the predictions resulting from the neural network models is evaluated by the following measure for goodness-of-fit:

$$\text{Root mean square error (RMSE)} = \sqrt{\frac{1}{p}\sum_{i=1}^{p}[(X_m)_i - (X_s)_i]^2} \quad (25)$$

Table 4 – Comparison of forecasting measures between observed and predicted data

Model	Forecasting measures	1-month lead time	2-month lead time	3-month lead time	4-month lead time	5-month lead time	6-month lead time
(a) SPI 3 over different lead time							
Stochastic (ARIMA(5,0,2))	R	0.801	0.661	0.4459	0.3234	0.2782	0.2073
	MAE	0.733	0.9874	1.1453	1.2503	1.5636	1.613
	RMSE	0.927	1.2116	1.4301	1.4463	1.4605	1.4784
RMSNN (6-9-1)	R	0.83	0.67	0.462	0.352	0.302	0.232
	MAE	0.6417	0.8965	1.0724	1.18	1.4025	1.5537
	RMSE	0.8382	1.1567	1.383	1.4325	1.4465	1.4691
DMSNN (11-5-6)	R	0.821	0.75	0.689	0.65	0.59	0.54
	MAE	0.653	0.843	0.987	0.998	1.1025	1.1537
	RMSE	0.8478	1.052	1.156	1.1985	1.232	1.283
(b) SPI 6 over different lead time							
Stochastic (ARIMA (1,0,0)(1,1,1)$_6$)	R	0.799	0.6176	0.411	0.352	0.283	0.219
	MAE	0.6595	0.884	1.0022	1.1088	1.1684	1.19
	RMSE	0.9009	1.1335	1.2762	1.3408	1.3839	1.4008
RMSNN (7-4-1)	R	0.8509	0.658	0.483	0.36	0.295	0.254
	MAE	0.5054	0.7806	0.8985	1.008	1.1324	1.1446
	RMSE	0.7139	0.9685	1.0994	1.1468	1.2649	1.3749
DMSNN (16-7-6)	R	0.842	0.78	0.719	0.673	0.621	0.587
	MAE	0.5112	0.7806	0.8985	1.008	1.1324	1.1446
	RMSE	0.718	0.868	0.897	0.957	0.986	1.021
(c) SPI 9 over different lead time							
Stochastic (ARIMA (1, 0, 0) (3, 1, 1)$_9$)	R	0.877	0.7298	0.595	0.5078	0.444	0.388
	MAE	0.4199	0.5854	0.7269	0.8165	0.8755	0.9563
	RMSE	0.5783	0.8297	0.9875	1.0692	1.1111	1.1421
RMSNN (9-3-1)	R	0.905	0.766	0.606	0.52	0.462	0.398
	MAE	0.3759	0.5036	0.5843	0.7106	0.7851	0.8825
	RMSE	0.5262	0.7656	0.9092	0.972	1.08	1.0926
DMSNN (16-8-6)	R	0.898	0.802	0.776	0.683	0.621	0.601
	MAE	0.3871	0.4832	0.5121	0.573	0.632	0.687
	RMSE	0.5291	0.7165	0.821	0.868	0.897	0.931
(d) SPI 12 over different lead time							
Stochastic (ARIMA (1,0,0) (2,1,0)$_{12}$)	R	0.925	0.828	0.73	0.648	0.56	0.476
	MAE	0.2931	0.4546	0.5831	0.6861	0.7928	0.8859
	RMSE	0.4284	0.6423	0.7929	0.8921	0.9817	1.0551
RMSNN (9-3-1)	R	0.93	0.84	0.742	0.68	0.572	0.491
	MAE	0.2731	0.3608	0.4947	0.5372	0.6313	0.7872
	RMSE	0.4101	0.6244	0.7734	0.828	0.9152	1.0075
DMSNN (16-8-6)	R	0.926	0.87	0.81	0.76	0.692	0.623
	MAE	0.294	0.347	0.398	0.51	0.582	0.663
	RMSE	0.419	0.563	0.645	0.711	0.778	0.801
(e) SPI 24 over different lead time							
Stochastic (ARIMA (1,0,0)(0,1,1)$_{24}$)	R	0.9055	0.797	0.714	0.654	0.619	0.588
	MAE	0.236	0.3445	0.4185	0.4698	0.4957	0.5167
	RMSE	0.3445	0.4953	0.577	0.6242	0.6435	0.6578
RMSNN (7-4-1)	R	0.921	0.801	0.721	0.658	0.622	0.591
	MAE	0.2203	0.3172	0.3874	0.4561	0.4782	0.4915
	RMSE	0.3176	0.4645	0.5397	0.6085	0.6264	0.6423
DMSNN (14-5-6)	R	0.916	0.865	0.819	0.786	0.724	0.631
	MAE	0.229	0.281	0.334	0.368	0.391	0.473
	RMSE	0.321	0.423	0.438	0.536	0.568	0.598

R = coefficient of correlation; MAE = mean absolute error; RMSE = root mean square error.

$$\text{Mean absolute error (MAE)} = \frac{1}{p}\sum_{i=1}^{p}\left|(X_m)_i - (X_s)_i\right| \tag{26}$$

where the subscripts m and s represent the observed and simulated SPI values, respectively; p = total number of events considered.

4. Results and discussion

In the present paper SPI is used as drought quantification parameter because drought forecasting remains a difficult but vitally remains an important task for water resource managers. Three different types of models are compared in the present paper considering the importance of lead time. The ARIMA/SARIMA models were developed for different SPI series using the correlation methods of Box and Jenkins based on AIC and SBC structure as selection criteria, which can be referred to authors' paper (Mishra and Desai, 2005b). The Neural network models are developed to forecast drought in this study using recursive multi-step approach and direct multi-step approach. The available data are split into two parts, the data set from 1965 to 1994 is used to estimate the model parameters and the data from 1995 to 2001 is used to check

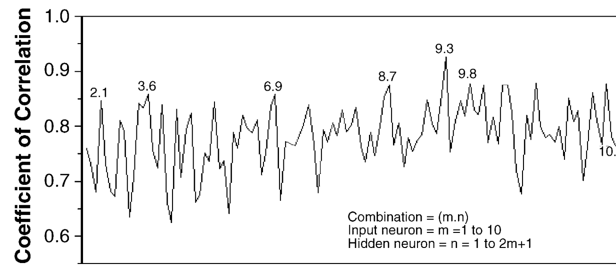

Fig. 4 – Different combination of input and hidden neurons for SPI 12 based on RMSNN approach.

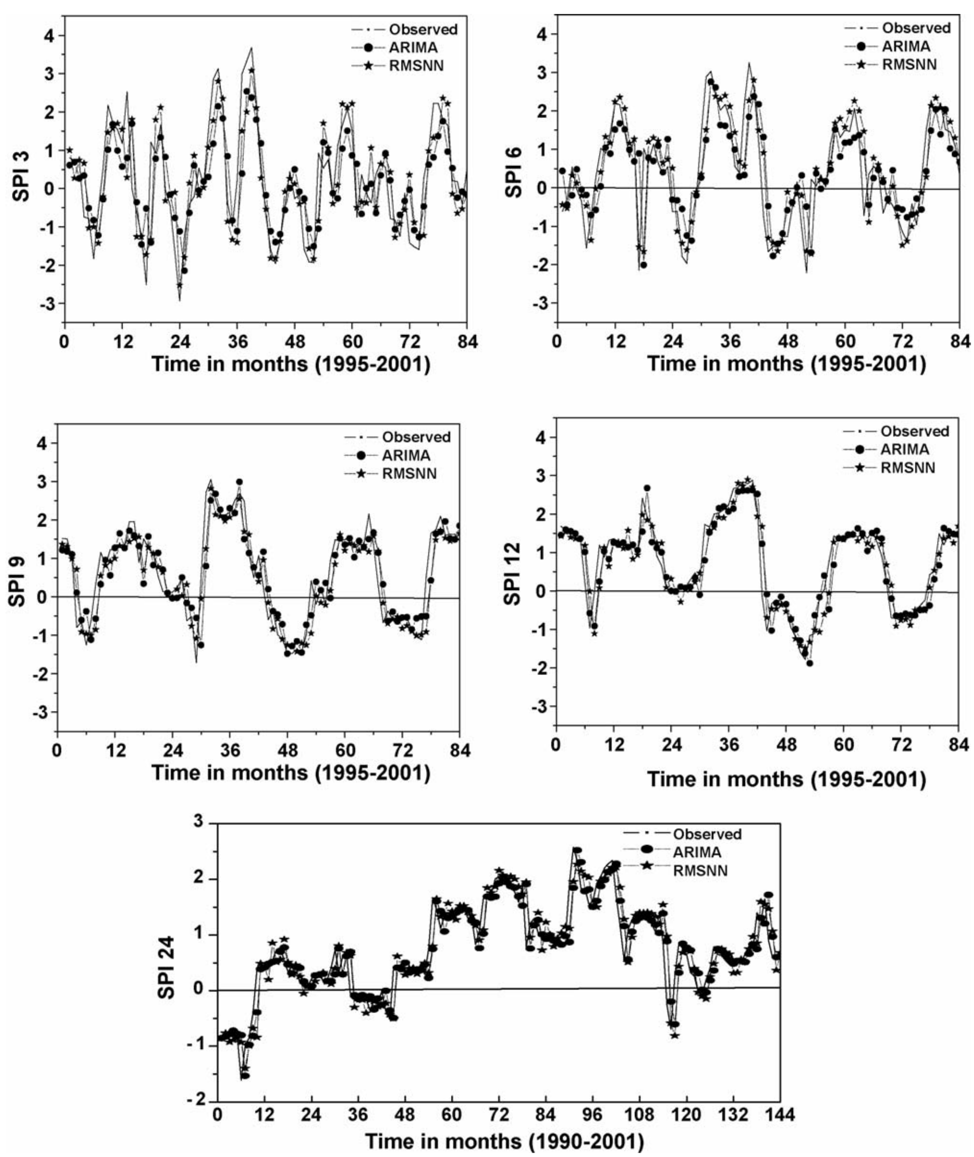

Fig. 5 – A comparison between observed data and predicted data over 1-month lead time using ARIMA and RMSNN for all SPI series.

the forecast accuracy. For SPI 24 the data from 1965 to 1989 is used to estimate model parameters and the data from 1990 to 2001 is used to check the forecast accuracy. The data set is different for SPI 24, so as to include more drought incidences, as drought incidences for this time scale are rare. For model development input data differ for different SPI series, which was found based on experimentation. The coefficient of correlation for each combination taking different number of input and hidden neurons between observed and simulated data is calculated. The input nodes (n) were varied from 1 to 20 and the corresponding hidden nodes varied from 1 to $2n+1$. The combination having maximum coefficient of correlation and minimum root mean square error is chosen as optimal network. Coefficient of correlation with different combination of input neurons and hidden neurons for SPI 12, based on the recursive multi-step neural network approach is shown in Fig. 4. In a similar way the optimal architecture for other SPI series are calculated shown in Table 4. The SPI 3, SPI 6, SPI 9, SPI 12 and SPI 24 were forecasted over different lead times (1, 2, 3, 4, 5 and 6 months) using optimal networks. Finally the results of both type neural network models are compared with ARIMA models developed by Mishra and Desai (2005b). The quantitative evaluation of the different model performance is carried out using correlation coefficient (CC), root mean square error (RMSE), and mean absolute error (MAE) over different lead time for all SPI series shown in Table 4(a)–(e). The numbers of hidden neurons were varied corresponding to each input neuron and it is observed that performance of ANN architecture increases when the number of hidden neuron is approximately half the number of input neurons. It is observed that the number of input neurons increases in direct multi-step approach to forecast over a 6-month lead time in

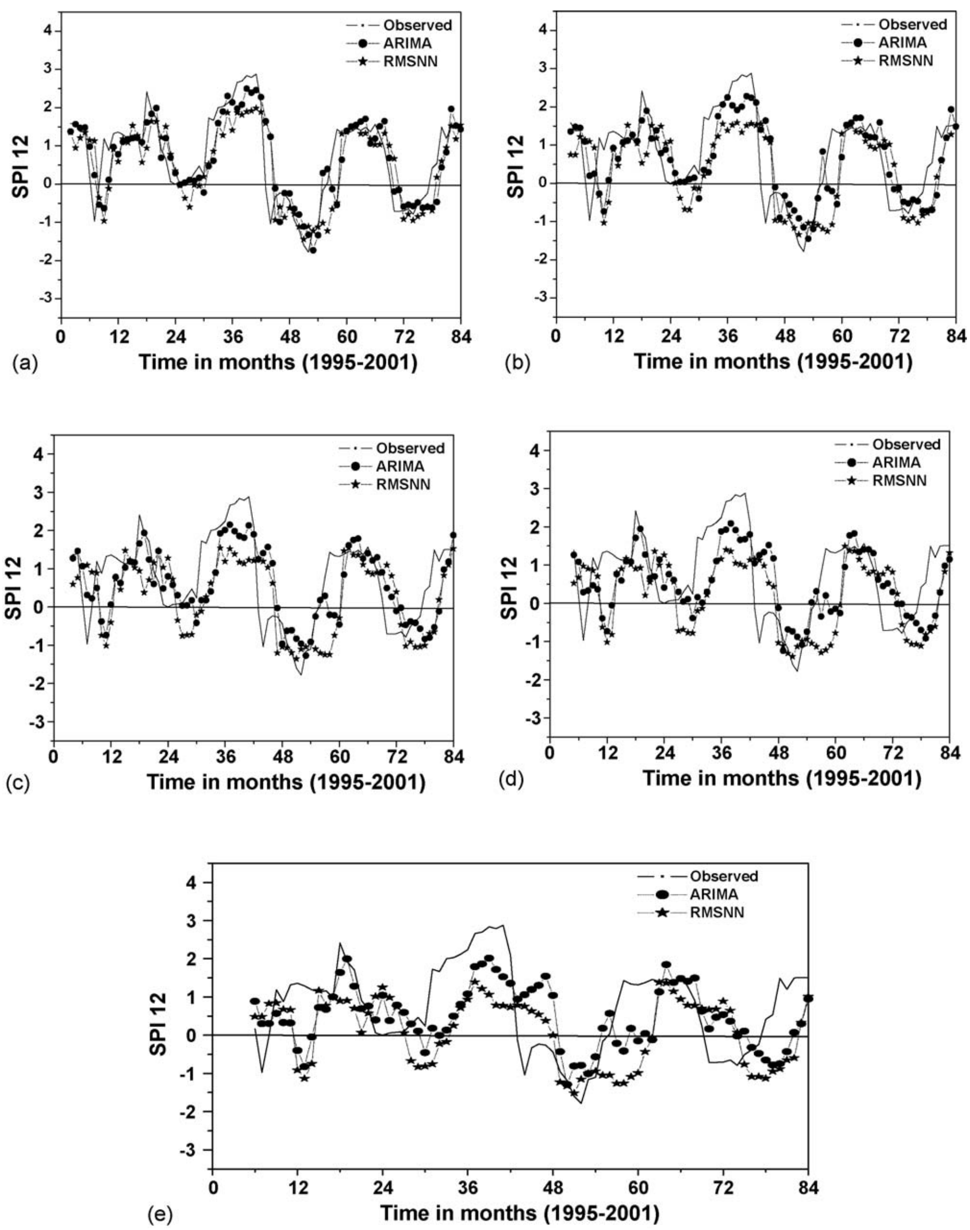

Fig. 6 – A comparison between ARIMA and RMSNN over 2–6-month lead time for SPI 12 series (a) 2-month lead time, (b) 3-month lead time, (c) 4-month lead time, (d) 5-month lead time and (e) 6-month lead time.

Table 5 – Statistics result for 1-month lead time of all SPI series using recursive multi-step neural network (RMSNN) model

SPI series	Mean observed	Mean forecasted	Decision, $\lvert z_{cal} \rvert < 1.96$	Variance observed	Variance forecasted	Decision, $F_{cal} < F_{tab}$
SPI 3	0.2619	0.1932	0.3231	2.06	1.7423	0.8458 < 1.462
SPI 6	0.4498	0.4963	0.2701	1.7446	1.6934	0.9706 < 1.462
SPI 9	0.6279	0.56004	0.713	1.4613	1.2215	0.8359 < 1.462
SPI 12	0.7307	0.6954	0.3209	1.2663	1.2209	0.9611 < 1.462
SPI 24	0.7540	0.7553	0.34	0.6419	0.6731	0.9419 < 1.00

comparison to recursive multi-step approach. The time series of the observed and 1 month ahead simulated values between ARIMA and RMSNN for all SPI series shown in Fig. 5. The time series of the observed and two to 6-month lead time forecasted values between ARIMA and RMSNN for SPI 12 is shown in Fig. 6. It is observed that with longer lead time the forecast accuracy decreases between observed and predicted data. The simulated values up to 1-month lead time were not significantly different from observed values for all three models. The recursive multi-step approach seems to provide very good result up to 1-month lead time in comparison to direct multi-step and ARIMA models. The basic statistical properties are compared between observed and forecasted data for 1-month lead time using recursive multi-step approach, based on *Z*-test for the means and *F*-test for standard deviation shown in Table 5. Since Z_{cal} values found to be less than Z-critical table values (±1.96 for two tailed at a 5% significance level), the data shows that there is no significant difference between the mean values of observed and predicted data. Similarly, the F_{cal} values of standard deviation were smaller than the *F*-critical values at a 5% significance level. Thus, the result shows that predicted data preserves the basic statistical properties of the observed series. When longer lead time of 4 months is considered direct multi-step approach outperforms recursive multi-step and ARIMA models.

5. Conclusions

In this paper the application of an NN has been successfully demonstrated on drought forecasting in Kansabati River Basin, India. The objective of the study was two-fold: first the SPI time series, which is used to quantify drought, are generated over multiple durations in the basin based on the average rainfall. It is observed that SPI 1 and SPI 3 is having good correlations with River flow discharge and SPI 3, SPI 6 with the storage in the reservoir over different months. The second objective was to develop NN models and to compare with ARIMA models for prediction of drought using SPI as drought index. Different ARIMA and NN architecture were applied to forecast SPI series and the best models were found by comparing observed and predicted data. The result of the neural network models are compared with linear stochastic (ARIMA/SARIMA) models. This paper highlights the importance of neural network models for comparison of forecasting results over shorter and longer lead time comparing different forecasting measures. The results obtained from the models show that recursive multi-step approach is best suited for 1 month ahead prediction. When longer lead time of 4 months is considered direct multi-step approach outperforms recursive multi-step and ARIMA models. The performance of the ARIMA models provides good result up to 2-month lead time but inferior in comparison to direct multi-step approach. The performance of ARIMA and recursive multi-step approach decreases over longer lead time because of accumulation of error between observed and predicted values at each time steps. As drought is a creeping phenomenon, so understanding drought is a difficult task. Hence drought forecasting is a real challenge for the researchers, so the present paper will be highly useful for sustainable development of river basins which will prevent environmental degradation of ecosystem. These neural network models can be very useful for local administrations and water resource planners to take precautions considering the severity of drought known in advance.

Acknowledgments

The authors would like to thank two anonymous reviewers for giving valuable suggestions for improving the quality of the paper. The authors would also like to acknowledge the editor for the timely handling the review processes and valuable comments for the improvement of paper.

REFERENCES

Abramowitz, M., Stegun, A. (Eds.), 1965. Handbook of Mathematical Formulas, Graphs, and Mathematical Tables. Dover Publications, Inc., New York, USA.

Acharya, C., et al., 2006. Prediction of sulphur removal with *Acidithiobacillus* sp. using artificial neural networks. Ecol. Model. 190, 223–230.

Antonic, et al., 2001. Spatio-temporal interpolation of climatic variables over large region of complex terrain using neural networks. Ecol. Model. 138, 255–263.

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology, 2000a. Artificial neural networks in hydrology. I. Preliminary concepts. J. Hydrol. Eng. 5 (2), 124–137.

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology, 2000b. Artificial neural networks in hydrology. II. Hydrologic applications. J. Hydrol. Eng. 5 (2), 115–123.

Box, G.E.P., Jenkins, G.M., 1976. Time Series Analysis, Forecasting and Control. Holden-Day, San Francisco.

Box, G.E.P., Jenkins, G.M., Reinsel, G.C., 1994. Time Series Analysis, Forecasting and Control. Prentice Hall, Englewood Cliffs, NJ, USA.

Bussay, A., Szinell, C., Szentimery, T., 1999. Investigation and Measurements of Droughts in Hungary. Hungarian Meteorological Service, Budapest, Hungary.

Cybenko, G., 1989. Approximation by superpositions of a sigmoidal function. Math. Contr. Signals Syst. 2, 303–314.
Edwards, D.C., McKee, T.B., 1997. Characteristics of 20th century drought in the United States at multiple timescales. Colorado State University, Fort Collins. Climatology Report No. 97-2, CO, USA.
Fletcher, D., Goss, E., 1993. Forecasting with neural networks: an application using bankruptcy data. Info. Manage. 24, 159–167.
Flood, I., Kartam, N., 1994a. Neural networks in civil engineering. I. Principles and understanding. J. Comput. Civil Eng. ASCE 8 (2), 131–148.
Flood, I., Kartam, N., 1994b. Neural networks in civil engineering. I. Principles and understanding. J. Comput. Civil Eng. ASCE 8 (2), 149–162.
Gevrey, M., Dimopoulos, I., Lek, S., 2003. Review and comparison of methods to study the contribution of variables in artificial neural network models. Ecol. Model. 160, 249–264.
Govindaraju, R.S., Rao, A.R. (Eds.), 2000. Artificial Neural Networks in Hydrology. Kluwer Academic Publishers, Amsterdam, The Netherlands.
Hayes, M.J., Svoboda, M.D., Wilhite, D.A., Vanyarkho, O.V., 1999. Monitoring the 1996 drought using the standardized precipitation index. Bull. Am. Meterol. Soc. 80, 429–438.
Hecht-Nielsen, R., 1990. Neurocomputing. Addison-Wesley, Menlo Park, CA, USA.
Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. Neural Netw. 2, 359–366.
Hsu, K.L., Gupta, H.V., Sorooshian, S., 1995. Artificial neural network modeling of the rainfall-runpff processes. Water Resour. Res. 31 (10), 2517–2530.
Hughes, B.L., Saunders, M.A., 2002. A drought climatology for Europe. Int. J. Climatol. 22, 1571–1592.
Jacobs, R.A., 1988. Increased rates of convergence through learning rate adaption. Neural Netw., 295–308.
Jeong, K.S., et al., 2001. Prediction and elucidation of phytoplankton dynamics in the Nakdong River (Korea) by means of a recurrent artificial neural network. Ecol. Model. 146, 115–129.
Karul, C., et al., 2000. Case studies on the use of neural networks in eutrophication modeling. Ecol. Model. 134, 145–152.
Karunanithi, N., Grenney, W.J., Whitely, D., Bovee, K., 1994. Neural networks for river flow prediction. J. Comput. Civil Eng. ASCE 8 (2), 201–220.
Kim, T., Valdes, J.B., 2003. Nonlinear model for drought forecasting based on a conjunction of wavelet transforms and neural networks. J. Hydrol. Eng. ASCE 8 (6), 319–328.
Lae, R., Lek, S., Moreau, J., 1999. Predicting fish yield of African lakes using neural networks. Ecol. Model. 120, 325–335.
Lana, X., Serra, C., Burgueno, A., 2001. Patterns of monthly rainfall shortage and excess in terms of the standardized precipitation index. Int. J. Climatol. 21, 1669–1691.
Lippmann, R.P., 1987. An introduction to computing with neural nets. IEEE ASSP Mag., 4–22.
Maier, H.R., Dandy, G.C., 1996. The use of artificial neural networks for the prediction of water quality parameters. Water Resour. Res. 32 (4), 1013–1022.
McKee, T.B., Doesken, N.J., Kliest, J., 1993. The relationship of drought frequency and duration to time scales. In: Proceedings of the Eighth Conference on Applied Climatology, Anaheim, CA, January 17–22. American Meteorological Society, Boston, MA, USA, pp. 179–184.
Melesse, A.M., Hanley, R.S., 2005. Artificial neural network application for multi-ecosystem carbon flux simulation. Ecol. Model. 189, 305–314.
Mishra, A.K., Desai, V.R., 2005a. Spatial and temporal drought analysis in the Kansabati River Basin, India. Int. J. River Basin Manage. IAHR 3 (1), 31–41.
Mishra, A.K., Desai, V.R., 2005b. Drought Forecasting using Stochastic Models Journal of Stochastic Environmental Research and Risk Assessment (Earlier Stochastic Hydrology and Hydraulics), Springer Verlag, Published online: http://dx.doi.org/10.1007/s00477-005-0238-4.
Nagesh Kumar, D., Raju, K.S., Satish, T., 2004. River flow forecasting using recurrent neural networks. Water Resour. Manage. 18, 141–163.
Obasi, G.O.P., 1994. WMO's role in the international decade for natural disaster reduction. Bull. Am. Meteorol. Soc. 75 (1), 655–1661.
Panu, U.S., Sharma, T.C., 2002. Challenges in drought research: some perspectives and future directions. Hydrol. Sci., 47S.
Pasini, A., Lore, M., Ameli, F., 2006. Neural network modelling for the analysis of forcings/temperatures relationships at different scales in the climate system. Ecol. Model. 191, 58–67.
Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T., 1986. Numerical Recipes. Cambridge University Press, Cambridge, UK.
Rao, A.R., Padmanabhan, G., 1984. Analysis and modelling of palmers drought index series. J. Hydrol. 68, 211–229.
Recknagel, F., et al., 1997. Artificial neural network approach for modelling and prediction of algal blooms. Ecol. Model. 96, 11–28.
Rumelhart, D.E., Hilton, G.E., Willams, R.J., 1986. Learning representations by back-propagating errors. Nature 323, 533–536.
Sahoo, G.B., Ray, C., Wade, H.F., 2005. Pesticide prediction in ground water in North Carolina domestic wells using artificial neural networks. Ecol. Model. 183, 29–46.
Scardi, M., 2001. Advances in neural network modeling of phytoplankton primary production. Ecol. Model. 146, 33–45.
Sen, Z., 1990. Critical drought analysis by second order markov chain. J. Hydrol. 120, 183–202.
Szalai, S., Szinell, C., 2000. Comparison of two drought indices for drought monitoring in Hungary a case study. In: Vogt, J.V., Somma, F. (Eds.), Drought and Drought Mitigation in Europe. Kluwer, Dordrecht, pp. 161–166.
Tang, Z., Fishwick, P.A., 1993. Feedforward neural nets as models for time series forecasting. ORSA J. Comput. 5 (4), 374–385.
Thom, H.C.S., 1958. A note on gamma distribution. Monthly Weather Rev. 86, 117–122.
USDA, 1994. Major world crop areas and climatic profiles, World Agricultural Outlook Board, US department of Agriculture, Agricultural Handbook No. 664, pp. 157–170.
Walter, M., et al., 2001. Predicting eutrophication effects in the Burrinjuck Reservoir (Australia) by means of the deterministic model SALMO and the recurrent neural network model ANNA. Ecol. Model. 146, 97–113.
Wilhite, D.A., Glantz, M.H., 1985. Understanding the drought phenomenon: the role of definations. Water Int. 10, 111–120.
Wilhite, D.A., Rosenberg, N.J., Glantz, M.H., 1986. Improving federal response to drought. J. Climate Appl. Meteorol. 25, 332–342.
Wong, F.S., 1991. Time series forecasting using backpropagation neural networks. Neurocomputing 2, 147–159.
Zhang, B.L., Dong, Z.Y., 2001. An adaptive neural-wavelet model for short term load forecasting. Electric Power Syst. Res. 59, 121–129.
Zhang, G., Patuwo, B.E., Hu, M.Y., 1998. Forecasting with artificial neural networks: the state of arts. Int. J. Forecast. 14, 35–61.