INNSPUB

Journal of Biodiversity and Environmental Sciences (JBES)
ISSN: 2220-6663 (Print) 2222-3045 (Online)
Vol. 9, No. 1, p. 231-244, 2016
http://www.innspub.net

RESEARCH PAPER OPEN ACCESS

An Efficient Approach of Focused Time Delay Neural Network in Drought Forecasting in Central Iran

Abbasali Vali*, Fatemeh Roustaei

Department of Combating Desertification, University of Kashan, Iran

Article published on July 25, 2016

Key words: Drought forecasting, Precipitation Index, Central Iran.

Abstract

An exact prediction and modeling of drought is essential for watersheds management. The main contribution of this research is in the design, performance and comparison of drought forecasting models using Focused Time Delay Neural Networks (FTDNN). The network was trained to perform one-step-ahead predictions. Standardized Precipitation Index(SPI) were applied in various time scales including 3, 6, 9, 12, 18, 24 and 48 monthly time series in 14 synoptic stations in Central Iran during 1965–2014. Five categories of back-propagation training algorithms namely resilient back propagation (RP), batch gradient descent (GD and GDX), Quasi-Newton (BFGS), conjugate gradient (CGF, CGP, and CGB) and Levenberg-Marquardt (LM) were used. Then, according to the best algorithm, the number of neurons in the hidden layer was optimized and the best performance was identified. The number of epochs, high Correlation Coefficient (R^2), least Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) were considered to evaluate the performance of the FTDNN model at each step. The result showed that the Levenberg-Marquardt (LM) was the best algorithm and node 31 was the most efficient for drought prediction. Finally, the designed network was applied on all of the SPIs time series to determine the best in prediction according to statistical parameters. It was found that better results can be achieved by increasing the duration of the time series. According to the results obtained, FTDNN trained by LM is an efficient tool to model and predict drought events especially in long term time series.

***Corresponding Author:** Abbasali Vali ✉ vali@kashanu.ac.ir

Introduction

Drought monitoring and development of an early cautionary system can lead to effective forecast of future drought events (Belayneh *et al.*, 2014). Several forecasting techniques have been proposed to assess the probable evolution of drought related hydro meteorological factors or drought indices based on either physical/conceptual or data driven models.

Although physical/conceptual models can prepare a vision into catchment processes, they are problematic due to their being laborious in the implementation of forecasting applications, requiring various types of data and resulting in models that are excessively complex (Beven, 2006). Data driven models need minimum information requirements and less development times, and have been found to be perfect in several hydrological predicting applications (Adamowski, 2008).

Among data driven models, artificial Neural Networks (ANN), which are derived from the human nervous system, have been recognized as a very strong tool in dealing with complex problems such as pattern recognition, clustering and function approximation. ANN with adequate complexity can estimate any function to a greater accuracy (Hornik *et al.*, 1989).

Moreover, ANN is considerably able to learn and provide precise output even if the input contains error (Crone, 2004; Şenkal *et al.*, 2012).In several studies, it has been used as a drought forecasting tool (Belayneh and Adamowski, 2012; Belayneh *et al.*, 2014; Dastorani and Afkhami, 2011; Jamshidi *et al.*, 2011; Keskin *et al.*, 2011; Kim and Valdés, 2003; Sepulcre-Canto *et al.*, 2012).

Time delay neural network model is an efficient model for response prediction based on past input. Waibel (1989) is the first to introduce the Time-Delay neural network (TDNN) and adopted time delays in connection with feed forward networks, it has been successfully applied in several studies (Afkhami *et al.*, 2010; Charaniya and Dudul, 2013; Dastorani and Afkhami, 2011; N. A.Charaniya, 2013; Xie *et al.*, 2006; Yazdani *et al.*, 2009).

In their study, Luk *et al.* (2000) used ANNs to forecast the spatial distribution of precipitation for an urban catchment.

They implemented three types of ANNs including multi-layer feed-forward neural networks (MLFN), partial recurrent neural networks and time-delay neural networks (TDNN) to perform one-step-ahead forecasts. Their result demonstrated that MLFN and TDNN could capture the dynamic structure of the precipitation procedure.

There are numerous algorithms for training a multilayer feed-forward artificial neural network. Scientists in every field have surveyed the efficiency of each category of algorithms on the training of neural networks. For instance, Mokbnache and Boubakeur (2002) compared the performance of three algorithms: Levenberg-Marquardt, Back Propagation (BP) with momentum, and BP with momentum and adaptive learning rate to classify/rank the trans-former oil dielectric and cooling state.

They recognized that the BP with momentum and adaptive learning rate improve the accuracy of the BP with momentum and also gives a fast convergence to the network. In order to predict the stream-flow and to discriminate the lateral stress in soils lacking cohesion, Kişi and Uncuoğlu (2005) compared the conjugate gradient, Levenberg-Marquardt and resilient algorithms.

They found that the Levenberg- Marquardt algorithm was quicker and achieved better performance than the other algorithms in training.

Adeoti and Osanaiye (2013) compared the efficiency of various algorithms on the performance of ANN for pattern recognition of bivariate process and found that the Levenberg-Marquardt is the best algorithm process in terms of recognition accuracy and the resilient back propagation is best in terms of speed and mean square error performance.

This study is focused on assessment of time delay neural network on long-term drought forecasting of well-known meteorological drought indices means Standardized Precipitation Index (SPI) in Central Iran. To achieve the goal, a variety of TDNN were used, which consists of a feed forward network with a tapped delay line at the input layer, called the focused time-delay neural network (FTDNN).

In the first step, by considering the same conditions and using the 12-month SPI obtained from the Mehrabad station in Tehran, the best algorithm was obtained. Then according to the most efficient algorithm, hidden layer neurons were optimized, and finally, the designed network was conducted on the SPIs of 14 stations located in Central Iran, to determine the best models for performing one-step-ahead predictions.

Materials and methods

Study area

Central Iran is bordered by the Alborz Mountains in the north, the Zagros Mountains in the south and west, and the scattered mountains of Khorasan in the east. The average elevation of this area is 1,300m, but it drops to 700 m inside Dasht-eKavir, and further to 300 m in the central part of Lut.

The Alborz and Zagros mountains block the Mediterranean rain-bearing wind from entering into Central Iran, thus lowering the average annual rainfall within this area to less than 300 mm (Ghorbani, 2013).

The climate in the study area is dominantly arid and semiarid based on the De Martonne index (De Marto (de Martonne, 1926; Nne, 1926). The mean annual temperature is variable from 15 to about 30°C within the study area, also the minimum and maximum daily temperatures are–18 and 51°C, respectively(Naderi and Raeisi, 2015).

The monthly rainfall data, measured in millimeter (*mm)*, was obtained from Meteorological Stations in Central Iran with 50 year records (1965-2014). The geographical location and characteristics of stations are given in Fig.1 and Table 1, respectively.

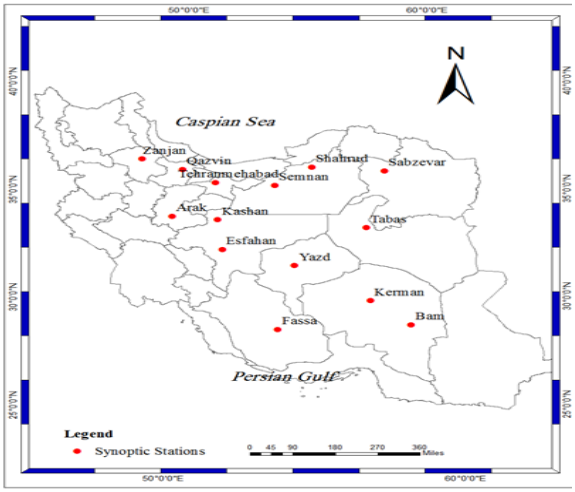

Fig. 1. The spatial distribution of 14 synoptic stations in Central Iran.

Table 1. Synoptic stations characteristics in Central Iran.

Station name	X coordinate	Y coordinate	Elevation (m)
Arak	49.77	34.10	708
Bam	58.35	29.10	66.9
Esfahan	51.67	32.62	1550.4
Fasa	53.68	28.97	1288.3
Ghazvin.	50.05	36.25	279
Kashan	51.45	33.98	982.3
Kerman	56.97	30.25	1753
Sabzevar	57.72	36.20	977.6
Semnan	53.55	35.58	130.8
Shahrud	54.95	36.42	345.3
Tabas	56.92	33.60	711
Tehran	51.32	35.68	190.8
Yazd	54.28	31.90	1237.2
Zanjan	48.48	36.68	663

Standardized precipitation index

The Standardized Precipitation Index is a meteorological drought index (SPI) that was introduced by McKee *et al.* (1993) (McKee *et al.*, 1993). The index is based only on precipitation data hence its evaluation is relatively easy, its value is not affected by the regional geography and finally it is possible to describe drought on multiple time scales. These properties distinguish the SPI from the other indices (Belayneh and Adamowski, 2012).

The SPI is calculated by fitting a probability density function to the frequency distribution of precipitation summed over the time scale of interest. This is performed independently for each month (or any other time-based of the raw precipitation time-series) and for each location in space (Moradi Dashtpagerdi *et al.*, 2014).

Gamma distribution is the most common way to perform the aim (Edossa *et al.*, 2010; Hayes *et al.*, 1999). The gamma distribution is defined by its probability density function as:

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

Where α> 0 and β>0 are the shape factor and the scale factor, respectively, and x>0 is the amount of precipitation. Γ (α) is the gamma function which is defined as:

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

Fitting the distribution to the data needs that and β be estimated. There are different methods for the estimating of these two parameters. For instance Edwards (1997) proposed a method using the calculation of Thom (1958) for maximum likelihood as follows:

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

$$\beta = \frac{\bar{x}}{\alpha} \quad (4)$$

Where for n observations

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

Respect to resulting parameters the cumulative probability of an observed precipitation event for the given month or any other time scale can be obtained:

$$G(x) = \int_0^x g(x)dx = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} = \int_0^x x^{(\alpha-1)} e^{\left(-\frac{x}{\beta}\right)} \quad (6)$$

Can be replaced by t to incomplete gamma function

$$G(x) = \frac{1}{\Gamma(\alpha)} \int_0^x t^{\alpha-1} e^{-t} dt \quad (7)$$

Since Gamma function is not defined for x = 0, and there may be no precipitation, the cumulative probability becomes

$$H(x) = q + (1 - q)G(x) \quad (8)$$

Where q is the probability of zero precipitation. The cumulative probability, H(x) is then altered into a normal standardized distribution with null average and unit variance from which is the value of SPI. However, is neither practical nor numerically simple to use if there are several grid points of numerous stations on which to compute the SPI index. Hereon, an alternative method is described in Edwards (1997) using the technique of approximate conversion developed in Abramowitz and Stegun (1965) that converts the cumulative probability into a standard variable *Z*. The SPI index is then defined as

$$Z = SPI = \begin{cases} -\left(t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t_1^2 + d_3 t^3}\right) & for\ 0 < H(x) \leq 0.5 \\ +\left(t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t_1^2 + d_3 t^3}\right) & for\ 0.5 < H(x) < 1 \end{cases} \quad (9)$$

Where

$$t = \begin{cases} \sqrt{ln\left[\frac{1}{(H(x))^2}\right]} & for\ 0 < H(x) \leq 0.5 \\ \sqrt{ln\left[\frac{1}{(1 - H(x))^2}\right]} & for\ 0.5 < H(x) < 1 \end{cases} \quad (10)$$

where x is the cumulative probability of recorded precipitation, and $c0$, $c1$, $c2$, $d0$, $d1$, $d2$ are constants with the following values: c_0=2.515517, c_1=0.802853, c_2=0.010328, d_1=1. 432788, d_2=0.189269, d_3=0.00-1308 (Mishra and Desai, 2005).

As shown in Table 2, a drought event occurs at the time when the value of the SPI is constantly negative; the event finishes when the SPI becomes positive. Since SPI is normalized, it can be used to track climate changes in any region, regardless of its average rainfall. This gives the advantage of comparing climate changes in various places and time scales (Illeperuma and Sonnadara, 2009).

Table 2. Classification scale of SPI (U.S. Drought Monitor Classification Scheme).

SPI values	Ranges
-0.5 to -0/7	Abnormally Dry
-0.8 to -1.2	Moderate Drought
-1.3 to -1.5	Severe Drought
-1.6 to -1.9	Extreme Drought
-2.00 and less	Exceptional Drought

As presented earlier, the SPI values in various time scales (3, 6, 9, 12, 18, 24 and 48 months) were calculated for the period of 1965 to 2014, using monthly precipitation data from 14 synoptic stations.

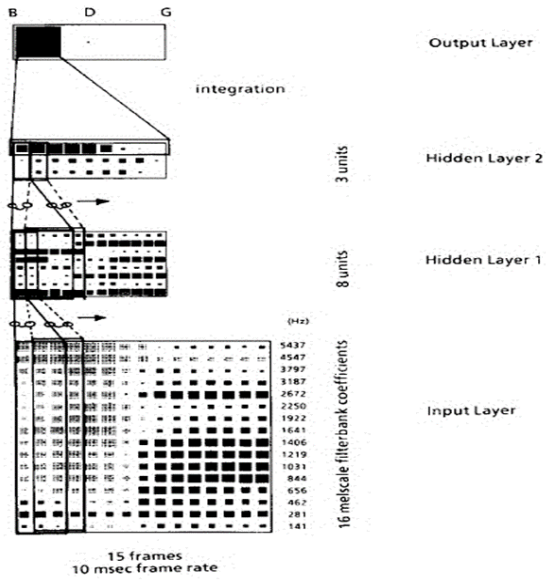

Fig. 2. Time Delay Neural Network as proposed by Alexander Waibel *et al.* (1989).

Focused time delay neural network

Focused time-delay neural network (FTDNN) is the most straight-forward dynamic network. This is part of an overall class of dynamic networks, named focused networks, in which the dynamics appear just at the input layer of a static multilayer feed forward network. This network is appropriate to time-series prediction. The following Fig. 2. demonstrates a two-layer FTDNN.

Training algorithms

Back propagation is used as a common method of training an Artificial Neural Network to perform a determined task. The aim of the training is to discover the set of weights among neurons that specify the global minimum of error function. The back propagation algorithm is applied in layered feed forward ANNs. Each layer receives its inputs from the previous layer and forwards its outputs to the next and errors are propagated backwards. In this learning process, the network which uses the algorithm, computes the outputs according to the examples of the input and thereafter, the error is calculated. The principal objective of the back propagation algorithm is to minimize this error, until the ANN learns the training data (Ike and Adoghe, 2013).

All training algorithms apply the gradient of the performance function to define in what way the weight can be adjusted to minimize performance.

The gradient is determined by a technique called back propagation, which involves performing calculations backwards through the network. The basic back propagation training algorithm, in which the weights are moved in the direction of the negative gradient is called Gradient Descent (GD) (Wilson and Martinez, 2003).

This algorithm is too slow for practical problems. There are several high performance algorithms that analyze ten to one hundred times faster than GD. They are divided into two main fold; the first fold uses heuristic techniques developed from analyzing the performance of the standard steepest descent algorithm,

which involves the gradient descent with adaptive learning rate, gradient descent with momentum, gradient descent with momentum and adaptive learning rate, as well as the resilient algorithm.

The second category applies standard numerical optimization techniques which includes the quasi-Newton, conjugate gradient, and Levenberg-Marquardt (LM) algorithm (Noori *et al.*, 2010). In the standard steepest descent, the learning rate is constant during the training process.

The performance of the algorithm is very sensitive to the proper setting of the learning rate. The performance of the algorithm can be improved if the learning rate is allowed to change throughout the training process. In the conjugate gradient algorithms, a search is performed along conjugate directions; therefore the convergence is faster than the steepest descent directions.

The Quasi-Newton method often converges faster than the conjugate gradient methods, since it does not require the calculation of second derivatives (Lahmiri, 2011). For instance, it updates an approximate Hessian matrix at each iteration.

Finally, the Levenberg-Marquardt algorithm merges the steepest descent method and the Gauss-Newton algorithm, so that it takes the rapidity of the Gauss-Newton and steadiness of the steepest descent method. For example, it uses the Jacobian which requires less calculation in comparison to the Hessian matrix.

Designing the network

In this study, 15 month time delay in the input data and 20 epochs selected by trial and error, were used for the training phase. Using cross validation technique to partition the data sets, the best proportion of data was determined as 70% for training while the remaining 30% of the data with equal proportion was selected for testing and validation of models. In order to calculate the error gradient and update the network weights and biases, the training set was applied. The error from the validation set was used to monitor the training process. However, when the network began to overfit the data, the error in the validation set increased. As the validation for a specified number of repetitions increased, the training was stopped, and the weights and biases at the minimum validation error were returned (Belayneh and Adamowski, 2012). The testing data set as an independent data set was used to confirm the performance of the model.

Before applying the ANN algorithm, the data was normalized to [0-1], using the following transformation function(Dastorani and Afkhami, 2011):

$$SPI_n = \frac{SPI_O - SPI_{Min}}{SPI_{Max} - SPI_{\min}} \quad (11)$$

Where: SPI_n is the normalized value, SPI_O is the measured value, SPI_{Min} and SPI_{Max} are the minimum and maximum values of SPI_O, respectively. If SPI_{Min} is equal to SPI_{Max} then normalized SPI_n is set as 0.5.

Using the 12-month SPI obtained from the Mehrabad station in Tehran, the network architectures, having various algorithms and a single neuron in the hidden layer, were evaluated and compared by means of statistical indices and the number of epochs.

Five groups of ANN training algorithms namely resilient back propagation (RP), batch gradient descent (GD and GDX), Quasi-Newton (BFGS), conjugate gradient (CGF,CGP,CGB) and Levenberg-Marquardt (LM) were used (Table 3).

According to the best algorithm, the number of neurons in the hidden layer was optimized. This was with a view of finding the best results with the number of epochs, high Correlation Coefficient (R^2), least Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE):

$$R^2 = (SPI_{Pi} - SPI_{Oi})_t^2 \quad (12)$$

Table. 3. List of algorithms applied in FTDNN (Lahmiri, 2011).

Algorithm	Adaptation	Description
Gradient descent (standard)(GD)	$\Delta w_k = -\alpha_k . g_k$	The weights and biases value are updated in the leading of the negative gradient of the performance function
Gradient Descent with Momentum and Adaptive Learning Rate Back propagation (GDX)	$\Delta w_k = p.\Delta w_{k-1} + \alpha . p$	Weight and bias deal updates pursuant to gradient descent momentum and an adaptive learning rate.
Resilient Back propagation(RP)	$\Delta w_k = -\text{sign}\left[\frac{\Delta E_k}{\Delta w_k}\right].\Delta$	Only the sign of the derivative can determine the direction of the weight update multiplied by the step size
Conjugate Gradient Back propagation with Fletcher-Reeves Update (CGF)	$p_0 = -g_0$ $\Delta w_k = -\alpha_k . p_k$ $p_k = -g_k + \beta_k p_{k-1}$ $\beta_k = \frac{g'_k g_k}{g'_{k-1} g_{k-1}}$	Iteration onset via probing in the steepest descent direction. A search line method1 is applied to discover the optimal current search direction Next (update) search direction is calculated as it is conjugate to prior search directions.
Conjugate Gradient Back propagation with Polak-Ribiére Update (CGP)	$p_0 = -g_0$ $\Delta w_k = \alpha_k . p_k$ $p_k = -g_k + \beta_k p_{k-1}$ $\beta_k = \frac{\Delta g'_{k-1} g_k}{g'_{k-1} g_{k-1}}$	Update is performed through calculating the product of the earlier change in the gradient with the current gradient divided by the square of the previous gradient.
Conjugate Gradient Back propagation with Powell-Beale Restarts (CGB)	$\left\|g'_{k-1} g_k\right\| \geq 0.2 \|g_k\|^2$	Update of search direction is reset to the negative of the gradient just when just when there is this suitable condition.
Quasi-Newton Back propagation (BFGS)	$\Delta w_k = -H'_k . g_k$	The update is calculated as a function of the gradient. H is the Hessian (second derivatives) matrix.
Levenberg-Marquardt (LM)	$\Delta w_k = -H'_k . g_k$ $H' = J'J$ $g = J'_e$	J is the Jacobian matrix (first derivatives) and e is a vector of network errors. Update is performed like Quasi-Newton

Where SPI_{Pi} and SPI_{Oi} are the predicted and observed SPI in period t, respectively.

$$MAE = \frac{1}{N}\sum_{i=1}^{n}\left|(SPI_{Pi} - SPI_{Oi})_t\right| \quad (13)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{n}(SPI_{Pi} - SPI_{Oi})_t^2} \quad (14)$$

Results and discussion

Fig. 3. shows the changes in different time series (3, 6, 9, 12, 18, 24 and 48 monthly SPI and precipitation) from 1965 to 2014 in one of the selected stations (Tehran Mehrabad synoptic station). As shown in the Fig., the generally linear that fitted the curve was also indicated for each time series. According to the Fig., by increasing the time series, there was a slight increase in the steepness of the fitted curves and a downward trend was seen at all-time series. In other words, as the SPI time series increases, the negative steepness of the fitted curves increased.

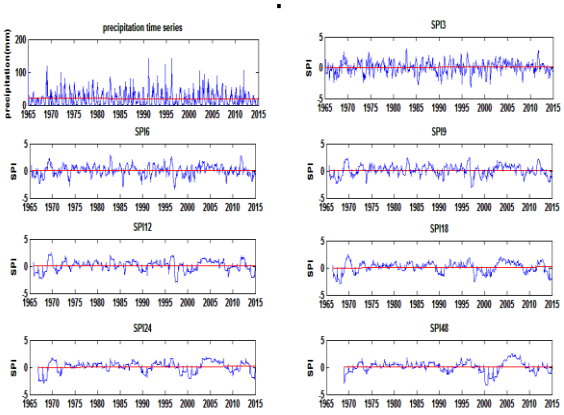

Fig. 3. The change in the variation of monthly mean of the precipitation and SPI (3, 6, 9, 12, 18, 24 and 48 months) during 1965 to 2014 in Tehran Mehrabad synoptic station.

The trained neural network was used to test the performance of the network on the test dataset. The testing results of the networks, based on a delay time of 15 months in the input layer and 25 neurons in the

hidden layer and 12-month SPI for the Mehrabad station, are shown in Table 4. The number of iterations (epoch), Root Mean of Squared errors (RMSE), Mean Absolute Error (MAE), Correlation Coefficient(R^2) were considered to determine the effect of training algorithms on the performance of the FTDNN model.

As shown in Table 4 and Fig. 4. in terms of statistical parameters, Levenberg-Marquart with the highest R^2 and Lowest MAE and RMSE performed the best.

The Conjugate Gradient Back propagation algorithms, include the CGF, CBP and CGB, as well as Quasti-Newton with the same RMSE, R and MAE had acceptable performance after the L-M. Whereas gradient descent and gradient descent with momentum and adaptive Learning with high MAE, RMSE and low correlation coefficient had the worst performance. Also, the performance of Resilient back propagation in drought forecasting, owing to low R, is unacceptable.

Table. 4. Performance of training algorithms on FTDNN model.

Algorithms	Epoch	MAE	RMSE	R^2
Gradient descent (standard)(GD)	20.00	0.50	0.62	-0.20
Gradient Descent with Momentum and Adaptive Learning Rate Back propagation (GDX)	20.00	0.26	0.34	0.15
Resilient Back propagation(RP)	19.00	0.09	0.11	0.77
Conjugate Gradient Back propagation with Fletcher-Reeves Update (CGF)	20.00	0.06	0.09	0.85
Conjugate Gradient Back propagation with Polak-Ribiére Update (CGP)	16.00	0.06	0.09	0.85
Conjugate Gradient Back propagation with Powell-Beale Restarts (CGB)	20.00	0.06	0.09	0.86
Quasi-Newton Back propagation (BFGS)	19.00	0.06	0.08	0.86
Levenberg-Marquardt (LM)	10.00	0.05	0.07	0.91

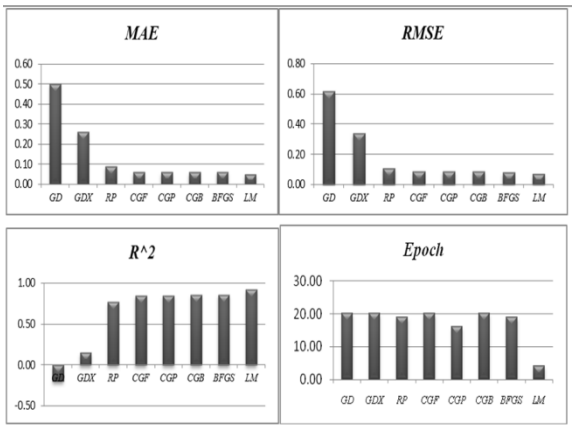

Fig. 4. Comparison of statistical parameters (MAE, RMSE and R) and number or iteration (Epoch) for the effect of training algorithms.

To compare algorithms in terms of the number of epochs, the Levenberg-Marquardt algorithm was found to be better than others. Subsequently, most speed of convergence is related to the Polak-Ribiére algorithm. The Gradient descent, gradient descent with momentum and adaptive Learning, Fletcher-Reeves and Powell-Beale with the same number of epochs were found to be worst.

The findings of this study clearly show that the Levenberg-Marquardt (LM) provides the best accuracy not only for the time required for training. This technique is the lowest in terms of epoch numbers, but also had low RMSE, MAE and highest R^2 statistics.

The prediction performance of the FTDNN models using LM training algorithm on the 600 independent test, training, validation and all datasets are shown in Fig. 5. As shown in this Fig., R^2between the predicted and measured values for FTDNN trained with L-M was found to be 93.42% (training) and 85.09% (testing), respectively. Therefore, the LM training algorithm could be used reliably with FTDNN for drought forecasting.

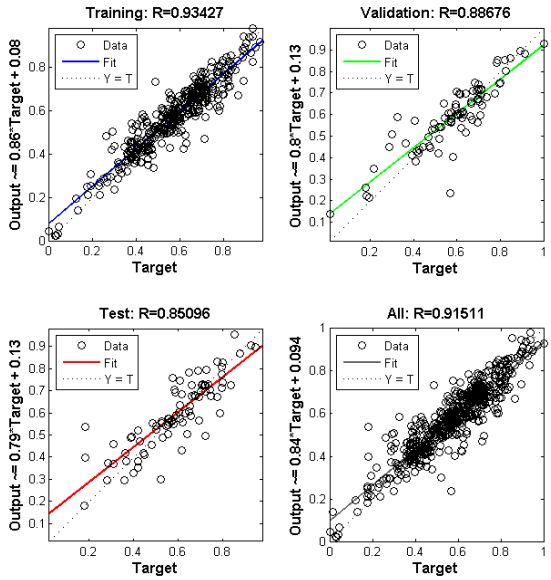

Fig. 5. Focused Time Delay Neural Networks prediction accuracies using Lavenberg-Marquardt (LM) training algorithm.

This result is also in agreement with the results of some previous studies that assessed the role of learning algorithms in ANN models for predicting. For instance,Benmahdjoub *et al.* (2013) in their study on the prediction of precipitation using Time Delay Neural Network in Tizi-Ouzou (Algeria), compared two learning algorithms including recursive gradient with constraint of shared weights and algorithm of Lavenberg-Marquardt.

After evaluating the correlation between the inputs and TDNN outputs, they found that the Lavenberg-Marquardt yields better description and forecast of rainfall. In a study by Lahmiri (2011), the accuracy of back propagation neural networks trained using various heuristic and numerical algorithms was measured for assessment purpose in financial prediction. Consequently, BFGS and Levenberg-Marquardt were determined as the best in terms of accuracy. Generally, Levenberg–Marquardt algorithm is regarded as one of the most efficient training algorithm sowing to its ability to resolve problems existing in both the gradient descent method and the Quasi–Newton method for neural-networks training, by the combination of these two algorithms (Hagan and Menhaj, 1994).

In the gradient descent method, the sum of the squared errors is reduced by updating the parameters in the direction of the greatest reduction of the least squares objective. In the Quasi -Newton method, the sum of the squared errors is reduced by assuming that the least squares function is locally quadratic, and finding the minimum of the quadratic. When the parameters are away from their optimum, the performance of the Levenberg-Marquardt method is similar to the gradient-descent method, and its action is similar to the Quasti-Newton method when the parameters are close to their optimal value. Indeed, the LM converges faster than other methods since the Hessian matrix is not computed, but only approximated and the Jacobian needs less computation than the Hessian matrix.

Then the number of nodes in the hidden layer was optimized for the FTDNN model. Hence, the number of iteration and statistical parameters for the Levenberg-Marquart algorithm, as the most efficient training algorithm, were identified by varying the number of nodes in the hidden layer chosen empirically between 25 to 35, since there is no reliable method for systematically determining them (Adeoti and Osanaiye, 2013; Aparisi *et al.*, 2006; Demuth *et al.*, 2008). The results are shown in Table 5. The number of nodes in the hidden layer that gave the lowest Epoch, RMSE, MAE and highest R was determined by focused time delay neural network.

As shown in Table 5, for node 26, the RMSE and MAE were at minimum, the highest R was seen, but the number of epochs was at maximum; and conversely, less speed, more MAE and RMSE and less R were found for node 27. Also, for nodes 32 and 33, similar conditions can be seen (less epoch and R), other nodes were shown to be of either acceptable statistical parameters or appropriate epochs, but the best node in the hidden layer was found to be 31 based on the fact that it is more appropriate in both statistical parameters and number of epochs, and theR^2between the predicted and measured values was found to be 92.96% for training data and 87.75% for testing data.

Table. 5. Efficiency of the nodes number in hidden layer on FTDNN model trained with Levenberg-Marquart.

NN Configuration	Epoch	MAE	RMSE	R^2
1-25-1	4	0.05	0.07	0.92
1-26-1	9	0.04	0.06	0.93
1-27-1	2	0.06	0.08	0.87
1-28-1	4	0.05	0.07	0.91
1-29-1	4	0.05	0.07	0.91
1-30-1	5	0.04	0.06	0.92
1-31-1	4	0.04	0.06	0.92
1-32-1	3	0.06	0.08	0.89
1-33-1	3	0.05	0.07	0.90
1-34-1	4	0.05	0.07	0.92
1-35-1	4	0.05	0.07	0.91

Finally, the designed Focused Time Delay Neural Network was applied for all SPIs in all stations to be determined, the performance of time series was more accurate in prediction, therefore, only the statistical parameters were compared and the results are shown in Tables 6 and 7. According to the results, SPI 48 followed by SPI 24, were recognized as the best SPI time series for forecasting in all stations and SPI 3 was observed as the worst. In other words, the prediction performed better in long term time series.

The results in Table 7 showed that R^2between the predicted and measured values for training and testing data was improved by increasing the time series. Fig. 5 shows a comparison between the response of the network for SPI 3 and SPI48. A1 and B1 show the difference between output and target; also, A2 and B2 represent the error of response of FTDNN for SPIs time series. Based on the aforementioned, the difference between the output and target and consequently the error of prediction for SPI 48 was smaller than SPI 3.

Table. 6. Statistical parameters of FTDNN model for 3, 6 and 9 monthly SPI time series in Central Iran.

Station name	3 monthly SPI			6 monthly SPI			9 monthly SPI		
	MAE	RMSE	R^2	MAE	RMSE	R2	MAE	RMSE	R2
Arak	0.11	0.15	0.64	0.86	0.11	0.81	0.0.7	0.09	0.87
Bam	0.07	0.09	0.68	0.07	0.09	0.75	0.6	0.08	0.86
Esfahan	0.14	0.18	0.44	0.11	0.15	0.71	0.09	0.11	0/83
Fasa	0.09	0.12	0.60	0.07	0.09	0.78	0.05	0.07	0.92
Kashan	0.11	0.14	0.65	0.08	0.11	0.79	0.06	0.08	0.89
Kerman	0.11	0.14	0.62	0.09	0.11	0.74	0.06	0.09	0.86
Qazvin	0.08	0.11	0.71	0.07	0.09	0.84	0.05	0.07	0.90
Sabzevar	0.09	0.11	0.64	0.08	0.10	0.81	0.06	0.09	0.89
Semnan	0.09	0.12	0.65	0.11	0.13	0.73	0.08	0.10	0.85
Shahrud	0.08	0.11	0.67	0.08	0.10	0.80	0.06	0.08	0.89
Tabas	0.82	0.11	0.67	0.07	0.1	0.78	0.07	0.09	0.87
Tehranmerhabad	0.09	0.12	0.68	0.07	0.10	0.76	0.06	0.07	0.89
Yazd	0.10	0.14	0.62	0.09	0.12	0.79	0.08	0.10	0.88
Zanjan	0.09	0.12	0.68	0.06	0.08	0.85	0.06	0.08	0.90

In a study carried out byRezaeian-Zadeh and Tabari (2012) on MLP-based drought forecasting using SPI (3,6,9,12 and 24) in different climatic regions, it was concluded that MLPs can forecast SPI24 and SPI12 more accurately than the other SPIs. Also, Belayneh *et al.* (2014) forecasted the long-term SPI (12 and 24 month) drought in the Awash River Basin in Ethiopia using wavelet neural network and wavelet support vector regression models.

The results obtained from all the data driven models generally showed that SPI 24 forecasts are more accurate than SPI 12 forecasts.

Moreira *et al.* (2012) and Vicente-Serrano (2006), in their respective studies, emphasized that large time scales of SPI recognizes anomalous dry periods of relatively long duration and relates well with the conditions assumed in the adopted definition.

This can result in the limitation of SPI analysis in defining drought in low precipitation areas at shorter time scales, where zero or small amount of droughts occurred for the whole time period using the 3-month and sporadically 6-months SPI test. This was due to the median precipitation being zero in these regions, for the periods on record. By increasing the time scales and considering additional months, this problem can be solved (Bari Abarghouei *et al.*, 2011; Kangas and Brown, 2007).

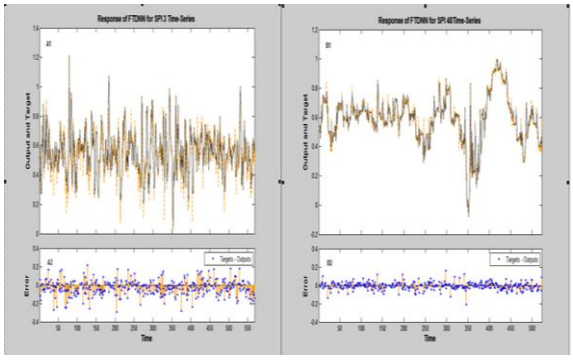

Fig. 6. The response of FTDNN for Tehran Mehrabad SPIs series: In A1 and B1 the dashed line shows the error and the solid line showed the response. In B1 and B2 the error of response has been represented.

Table. 7. Statistical parameters of FTDNN model for 12, 18 , 24 and 48 monthly SPI time series in Central Iran.

Station name	12 monthly SPI			18 monthly SPI			24 monthly SPI			48 monthly SPI		
	MAE	RMSE	R2	MAE	RMSE	R2	MAE	RMSE	R2	MAE	RMSE	R2
Arak	0.05	0.07	0.91	0.05	0.07	0.94	0.04	0.06	0.96	0.28	0.04	0.98
Bam	0.05	0.08	0.91	0.05	0.07	0.92	0.04	0.06	0.95	0.03	0.03	0.98
Esfahan	0.06	0.08	0.92	0.04	0.06	0.94	0.04	0.05	0.96	0.03	0.05	0.97
Fasa	0.05	0.07	0.92	0.04	0.06	0.95	0.03	0.05	0.97	0.03	0.04	0.99
Kashan	0.06	0.08	0.92	0.05	0.07	0.94	0.05	0.07	0.93	0.04	0.06	0.94
Kerman	0.05	0.07	0.92	0.05	0.06	0.92	0.03	0.05	0.96	0.03	0.05	0.98
Qazvin	0.07	0.09	0.89	0.05	0.07	0.94	0.04	0.06	0.95	0.04	0.06	0.96
Sabzevar	0.07	0.09	0.92	0.05	0.07	0.95	0.05	0.06	0.95	0.04	0.05	0.96
Semnan	0.05	0.09	0.90	0.06	0.08	0.91	0.04	0.06	0.94	0.04	0.05	0.96
Shahrud	0.05	0.07	0.93	0.04	0.06	0.95	0.04	0.05	0.96	0.03	0.04	0.98
Tabas	0.05	0.07	0.92	0.05	0.07	0.90	0.04	0.06	0.95	0.03	0.05	0.97
Tehranmerhabad	0.05	0.07	0.90	0.06	0.07	0.91	0.04	0.06	0.95	0.03	0.06	0.97
Yazd	0.06	0.08	0.91	0.05	0.08	0.93	0.04	0.06	0.95	0.03	0.05	0.97
Zanjan	0.05	0.07	0.90	0.04	0.06	0.94	0.05	0.06	0.95	0.04	0.05	0.97

Table. 8. R^2 between the predicted and measured values for training, testing and all data by FTDNN for Tehranmehrabad SPIs series.

SPI time series	Testing data	Training data	all
3 monthly SPI	0.50	0.74	0.68
6 monthly SPI	0.59	0.52	0.76
9 monthly SPI	0.83	0.91	0.88
12 monthly SPI	0.84	0.92	0.90
18 monthly SPI	0.92	0.91	0.91
24 monthly SPI	0.95	0.94	0.95
48 monthly SPI	0.95	0.98	0.97

Conclusion

Focused time-delay neural networks were designed in order to predict various time series of SPI in central Iran. As such, the performance of the algorithms was assessed on optimized network with delay time of 15 months in the input layer and 25 neurons in the hidden layer. It was found that the LM algorithm provide the greatest performance in terms of speed of convergence and statistical parameters for the prediction of SPIs.

The designed networks were applied for SPIs time series in 14 stations in study area. The accuracy of the predictions was higher for the long term time series (R>0.90) compared to the short term time series (R<0.90). Concerning the statistical parameters, by increasing the time scale, the accuracy of forecasting increased.

The developed neural network model can be a very helpful tool for water resource planners to take necessary actions timely when there is water scarcity which may ultimately develop into drought conditions.

Future work in different climatic stations with SPI and other drought indices is suggested to analyze the efficiency of FTDNN in the forecasting of drought in various time scales.

Also, a survey, based on the simulation and prediction of result using other artificial neural networks and back propagation algorithms for SPI forecasting and comparing the forecasted result accurately with current forecasting methods, can further improve the efficiency of the Focused Time Delay method.

Acknowledgement
This research is part of a PhD thesis that was supported by the University of Kashan. The authors are grateful to the University for this kind gesture.

References

Abramowitz M, Stegun IA. 1965. Handbook of mathematical functions. Dover New York.

Adamowski JF. 2008. Development of a short-term river flood forecasting method for snowmelt driven floods based on wavelet and cross-wavelet analysis. Journal of Hydrology **3533,** 247-266.

Adeoti OA, Osanaiye PA. 2013. Effect of Training Algorithms on the Performance of ANN for Pattern Recognition of Bivariate Process. International Journal of Computer Applications **6920,** 8-12.

Afkhami H, Dasturani M, Malekinejad H, Mobin M. 2010. Effect of climatic factors on accuracy of ann-based drought prediction in yazd area. Water and Soil Science(in Persian) **1451,** 157-170.

Aparisi F, Avendaño G, Sanz J. 2006. Techniques to interpret T 2 control chart signals. IIE Transactions **388,** 647-657.

Bari Abarghouei H, Zarch M AA, Dastorani MT, Kousari MR, Zarch MS. 2011. The survey of climatic drought trend in Iran. Stochastic Environmental Research and Risk Assessment **256,** 851-863.

Belayneh A, Adamowski J. 2012. Standard Precipitation Index Drought Forecasting Using Neural Networks, Wavelet Neural Networks, and Support Vector Regression. Applied Computational Intelligence and Soft Computing **2012,** 1-13.

Belayneh A, Adamowski J, Khalil B, Ozga-Zielinski B. 2014. Long-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet neural network and wavelet support vector regression models. Journal of Hydrology **508,** 418-429.

Benmahdjoub K, Ameur Z, Boulifa M. 2013. Forecasting of Rainfall using Time Delay Neural Network in Tizi-Ouzou (Algeria). Energy Procedia **36,** 1138-1146.

Beven K. 2006. A manifesto for the equifinality thesis. Journal of Hydrology **3201,** 18-36.

Charaniya N, Dudul S. 2013. Time Lag recurrent Neural Network model for Rainfall prediction using El Niño indices. International Journal of Scientific and Research Publications 367.

Crone SF. 2004. A business forecasting competition approach to modeling artificial neural networks for time series prediction pp. 207-213.

Dastorani M, Afkhami H. 2011. Application of artificial neural networks on drought prediction in Yazd (Central Iran). Desert **161,** 39-48.

De Martonne E. 1926. L'indice d'aridité. Bulletin de l'Association de géographes français **39,** 3-5.

Demuth H, Beale M, Hagan M. 2008. Neural network toolbox™ 6. User's guide.

Edossa DC, Babel M S, Gupta AD. 2010. Drought analysis in the Awash river basin, Ethiopia. Water resources management **247,** 1441-1460.

Edwards DC. 1997. Characteristics of 20th century drought in the United States at multiple time scales, DTIC Document.

Ghorbani M. 2013. The economic geology of Iran: mineral deposits and natural resources. Springer Science & Business Media.

Hagan MT, Menhaj MB. 1994. Training feedforward networks with the Marquardt algorithm. Neural Networks, IEEE Transactions on **56,** 989-993.

Hayes MJ, Svoboda MD, Wilhite DA, Vanyarkho OV. 1999. Monitoring the 1996 drought using the standardized precipitation index. Bulletin of the American Meteorological Society **803,** 429-438.

Hornik K, Stinchcombe M, White H. 1989. Multilayer feedforward networks are universal approximators. Neural networks **25,** 359-366.

Ike DU, Adoghe A. 2013. Back-Propagation Artificial Neural Network Techniques for~ Optical Character Recognition-A Survey. International Journal of Computers and Distributed Systems **3II,** 1-6.

Illeperuma G, Sonnadara U. 2009. Forecasting Droughts using Artificial Neural Networks. Promoting Knowledge Transfer to Strengthen Disaster Risk Reduction & Climate Change Adaptation 100.

Jamshidi H, Arian A, Rezaeian-Zadeh M. 2011. Drought forecasting by Multilayer Perceptron network in Different climatological regions pp. 15-23.

Kangas RS, Brown TJ. 2007. Characteristics of US drought and pluvials from a high - resolution spatial dataset. International Journal of Climatology **2710,** 1303-1325.

Keskin ME, Terzi O, Taylan ED, Küçükyaman D. 2011. Meteorological drought analysis using artificial neural networks. Scientific Research and Essays **621,** 4469-4477.

Kim TW, Valdés JB. 2003. Nonlinear model for drought forecasting based on a conjunction of wavelet transforms and neural networks. Journal of Hydrologic Engineering **86,** 319-328.

Kişi Ö, Uncuoğlu E. 2005. Comparison of three back-propagation training algorithms for two case studies. Indian journal of engineering & materials sciences **125,** 434-442.

Lahmiri S. 2011. On Simulation Performance of Feedforward and NARX Networks Under Different Numerical Training Algorithms.

Luk K, Ball J, Sharma A. 2000. A study of optimal model lag and spatial inputs to artificial neural network for rainfall forecasting. Journal of Hydrology **2271,** 56-65.

McKee TB, Doesken NJ, Kleist J. 1993. The relationship of drought frequency and duration to time scales pp. 179-183, American Meteorological Society Boston MA USA.

Mishra A, Desai V. 2005. Drought forecasting using stochastic models. Stochastic Environmental Research and Risk Assessment **195,** 326-339.

Mokhnache L, Boubakeur A. 2002. Comparison of different back-propagation algorithms used in the diagnosis of transformer oil pp. 244-247.

Moradi Dashtpagerdi M, Kousari MR, Vagharfard H, Ghonchepour D, Hosseini M E, Ahani H. 2014. An investigation of drought magnitude trend during 1975–2005 in arid and semi-arid regions of Iran. Environmental Earth Sciences **733,** 1231-1244.

Moreira E, Mexia J, Pereira L. 2012. Are drought occurrence and severity aggravating? A study on SPI drought class transitions using log-linear models and ANOVA-like inference.

N.A.Charaniya SVD. 2013. Time Lag recurrent Neural Network model for Rainfall prediction using El Niño indices. International Journal of Scientific and Research Publications 31, 5.

Naderi M, Raeisi E. 2015. Climate change in a region with altitude differences and with precipitation from various sources, South-Central Iran. Theoretical and Applied Climatology **1243,** 529-540.

Noori R, Khakpour A, Omidvar B, Farokhnia A. 2010. Comparison of ANN and principal component analysis-multivariate linear regression models for predicting the river flow based on developed discrepancy ratio statistic. Expert Systems with Applications **378,** 5856-5862.

Rezaeian-Zadeh M, Tabari H. 2012. MLP-based drought forecasting in different climatic regions. Theoretical and Applied Climatology **1093-4**, 407-414.

Şenkal O, Yıldız BY, Şahin M, Pestemalcı V. 2012. Precipitable water modelling using artificial neural network in Cukurova region. Environmental monitoring and assessment **1841,** 141-147.

Sepulcre-Canto G, Horion S, Singleton A, Carrao H, Vogt J. 2012. Development of a Combined Drought Indicator to detect agricultural drought in Europe. Natural Hazards and Earth System Science **1211,** 3519-3531.

Thom HC. 1958. A note on the gamma distribution. Monthly Weather Review **864,** 117-122.

Vicente-Serrano SM. 2006. Differences in spatial patterns of drought on different time scales: an analysis of the Iberian Peninsula. Water resources management **201,** 37-60.

Waibel A. 1989. Modular construction of time-delay neural networks for speech recognition. Neural computation **11,** 39-46.

Wilson DR, Martinez TR. 2003. The general inefficiency of batch training for gradient descent learning. Neural networks **1610,** 1429-1451.

Xie JX, Cheng CT, Chau KW, Pei YZ. 2006. A hybrid adaptive time-delay neural network model for multi-step-ahead prediction of sunspot activity. International Journal of Environment and Pollution **283-4,** 364-381.

Yazdani M, Saghafian B, Mahdian M, Soltani S. 2009. Monthly runoff estimation using artificial neural networks. Journal of Agricultural Science and Technology **11,** 335-362.