Iranian Journal of Science & Technology, Transaction B, Vol. 28, No. B1
Printed in Islamic Republic of Iran, 2004

"Research Note"

INTELLIGENT FORECASTING OF RAINFALL AND TEMPERATURE OF SHIRAZ CITY USING NEURAL NETWORKS*

P. SETOODEH A. A. SAFAVI[1]** AND M. J. NAZEMOSADAT[2]

[1]Dept. of Electrical Engineering, School of Engineering, Shiraz University, I. R. of Iran
Email: safavi@shirazu.ac.ir
[2]Dept. of Irrigation, School of Agriculture, Shiraz University, I. R. of Iran

Abstract – Most parts of southern Iran have frequently experienced drought and flooding events. The occurrence of these natural disasters is common in Fars province which supplies about 25% of the national wheat product. Shiraz (capital city of the province) is studied as a representative of the province. Estimating rainfall and temperature can help in agricultural water management, protection from water shortages, and flood damage, thus having significant economic impacts. The prediction is, however, a complicated procedure and conventional mathematical methods are not able to easily capture such a relationship. To overcome the problem, neural network-based models were used for forecasting temperature and rainfall in Shiraz (Iran). Various simulation results based on the real data are presented. The results suggest that the applied methodology is suitable and more practical than the previous approaches for the prediction of rainfall and temperature. The developed model is able to predict rainfall and temperature one season ahead with reasonable error.

Keywords –ARX, NNARX, time series prediction, ENSO, north atlantic oscillation, Persian Gulf SST

1. INTRODUCTION

Forecasting is a highly complex application, because we are typically dealing with systems that receive thousands of inputs, which interact in a complex nonlinear fashion. Usually with a very small subset of inputs or measurements available from the system, its behavior is to be estimated and extrapolated to the future. What gives some hope solving some of the difficult forecasting problems is that typically successive events or inputs that affect the time series are serially correlated, and that causes a time series pattern that can give some hint of the future. Time series can exhibit some past random-like cyclical behavior, and can therefore give a hint about the probable position in the cycle at a certain instant in the future [1].

Neural networks (NN's) are powerful computational tools for learning complex mapping. They may be easily developed and solve complicated problems. They only need an historical pattern of the events and a judicious selection of NN parameters and structure. There is substantial literature available on various NN's and their applications [2-4]. Most neural network approaches to the problem of forecasting use a multilayer network trained using the backpropagation algorithm. Consider a time series $x(1),...,x(t-1)$, where it is required to forecast the value of $x(t)$. The inputs to the multilayer network are typically chosen as the previous k values $x(t-k),...,x(t-1)$ and the output will be the forecast. The network is trained and tested on sufficiently large training and testing sets that are extracted from the historical time series. In addition to previous time series values, one can utilize, as inputs, the values or forecasts of other time series (or external variables) that have a correlated or causal relationship with the series to be forecasted [1].

*Received by the editors December 8, 2001 and in final revised form April 27, 2003
**Corresponding author

Although the influence of teleconnection patterns on Shiraz rainfall have already been investigated, a model has not yet been developed to seek the impact of combined indices on the rainfall. Moreover, every member of the hierarchy of classical multivariate methods such as multiple regression, principal component analysis (PCA), and canonical correlation analysis (CCA) are inherently linear, but natural phenomena such as climatic variables are nonlinear with complex interactions. The significantly advanced nonlinear empirical modeling ability of neural networks can be utilized to overcome these two shortcomings.

In this paper, the problem of forecasting the rainfall of Shiraz is investigated using NN approaches. Besides some preprocessing analysis, order identification and model reduction methods are also invoked in order to improve the results. The structure of this paper is as follows: First, problem formulation and order identification are presented. A linear model is then tried in order to investigate the complications. A neural network based model is further developed and optimized and a conclusion will be given in the last section.

2. PROBLEM FORMULATION

Understanding the predictability of the atmosphere at a seasonal-to inerannual time-scale has improved considerably during the past decade. It has been shown that the variability of rainfall and temperature in Iran has been accentuated by the occurrence of El Nino-Southern Oscillation (ENSO), North Atlantic Oscillation (NAO) and the Persian Gulf sea surface temperatures (PGSSTs) [5-8]. Compared to other parts of Iran, the impact of the PGSSTs on surface climate is more meaningful for the southwestern portion of the country [8]. These studies have shown that the variations in PGSSTs account for about 35% of the total variance of winter rainfall in Shiraz. The influence of ENSO and NAO phenomena on the rainfall was, however found to be less than the corresponding values of PGSSTs.

The aim is to forecast temperature and rainfall in Shiraz based on the related and historical data available. The inputs that we used are combinations of the following:

1) The seasonal rainfall at the previous time periods;
2) The seasonal average of the temperature at the previous time periods;
3) The seasonal average of the Persian Gulf Sea Surface Temperature (PGSST);
4) The seasonal average of the Southern Oscillation Index (SOI);
5) The seasonal average of the North Atlantic Oscillation (NAO).

Measurements of the above variables were available from 1951 to 1993. The first two data sets were derived from the yearly weather-books published by the Iranian Meteorological Organization (IMO) for the period 1951-1993. The PGSST data were extracted from the comprehensive oceanic and atmospheric data set (COADS) for the period 1951-1993. The Troup′s (1965) Southern Oscillation Index (SOI) data, which have been used as the ENSO indicator, were supplied by the Australian Bureau of Meteorology. The NAO data were extracted from the web site of the national oceanic and atmospheric administration (NOAA). Time series were created from these data, and then the models were developed based on the data from 1951-1990. Models are tested based on the data from 1991-1993.

As is the case with many neural network applications, preprocessing the inputs may improve the results significantly. Input and output preprocessing means extracting features from the inputs and transforming the target outputs in a way that makes it easier for the network to extract useful information from the inputs and associate it with the required outputs. Preprocessing is considered an art and there is no set of rules in choosing it [1].

The cross-correlation function is a measure for the statistical dependence of two different processes, *x(t) and y(t).* The amount of correlation between *x* and *y* can be expressed by a *coefficient of correlation.* The correlation coefficient is the correlation function normalized so that the autocorrelation will be 1 for lag 0 [9]. Normalized correlation coefficients between Shiraz rainfall and temperature and other variables up to a maximum lag of 25 were calculated. As shown in Fig.1, Shiraz rainfall and temperature are more correlated with PGSST than with NAO or SOI. This supports the idea that rainfall and temperature in Shiraz are more influenced by the surface climate of the Persian Gulf than the other previously studied teleconnections. It is obvious from Fig.1 that temperature is more correlated with PGSST than rainfall. Also, normalized correlation coefficient curves related to temperature have a more regular shape than those related to rainfall. So it is not surprising that temperature can be more accurately forcasted than rainfall.

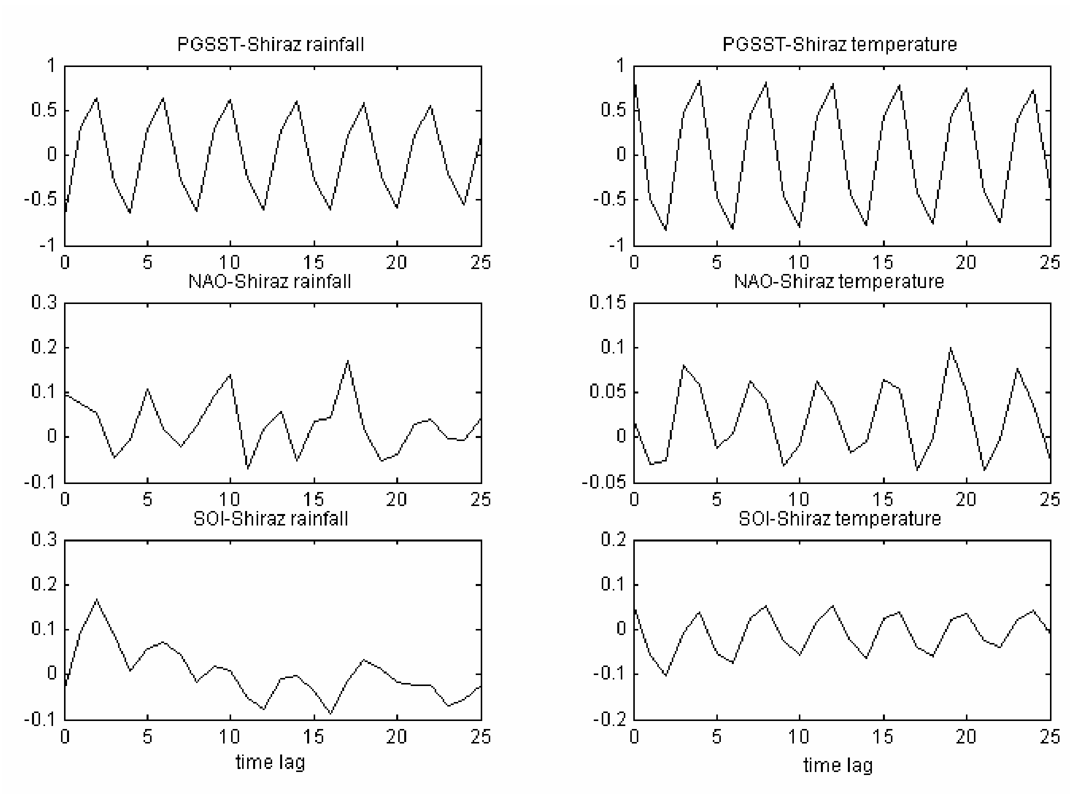


Fig. 1. Normalized correlation coefficients between Shiraz rainfall and temperature and other teleconnections

3. ORDER IDENTIFICATION

In a dynamic system the current outputs depend not only on the current inputs, but also on the inputs and outputs at the previous time periods. As the first step of modeling, a reasonable estimate of the order of the system under consideration should be obtained. In other words, we should reasonably estimate how many previous inputs and outputs have had considerable effects on current outputs. The function *lipschit* of MATLAB *Neural Network Based System Identification* (*NNSYSID*) *Toolbox* can be employed to do this [10].

First the training set is scaled to zero mean and variance one, and then the test set is scaled with the same constants. Given corresponding input and output sequences, the function *lipschit* calculates a matrix of indices that can be helpful for determining a proper lag space structure before identifying a model of a dynamic system. An insufficient lag space structure leads to a large index. While increasing the lag space the index will decrease until a sufficiently large lag space structure is reached. Increasing the lag space further will not change the index significantly. In other words, we are looking for the knee-point of the plot. The indices for system orders from 1 to 40 were investigated. Results for rainfall are shown in Fig.2. It is not unreasonable to use a seventh order model for rainfall prediction, because the slope of the curve decreases for model orders equal or greater than 7. Also, a fourth order model may be reasonable too.

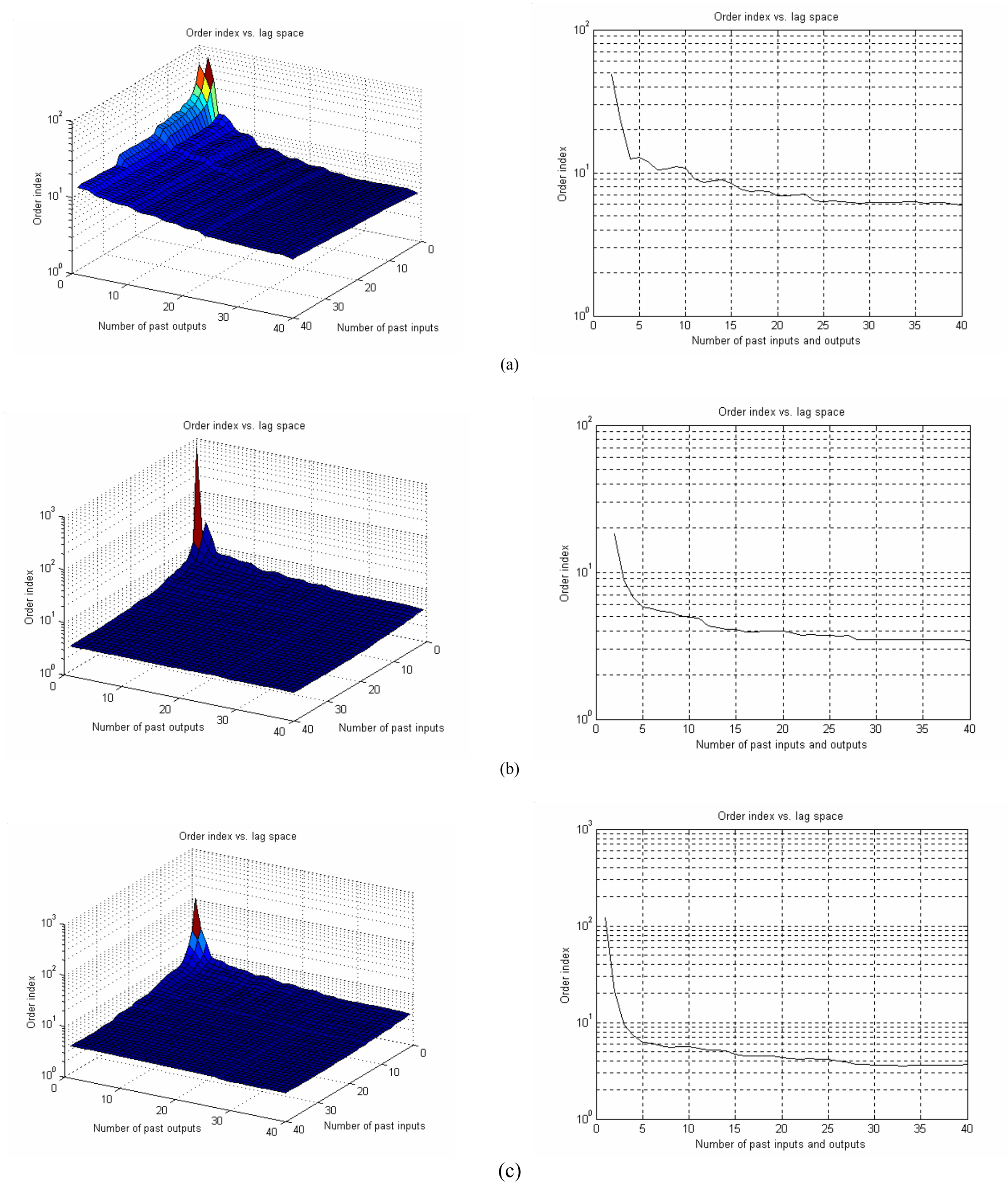


Fig. 2. Order identification for Shiraz rainfall; (a) input: PGSST, output: Shiraz rainfall
(b) input: NAO, output: Shiraz rainfall (c) input: SOI, output: Shiraz rainfall

4. FITTING A LINEAR MODEL

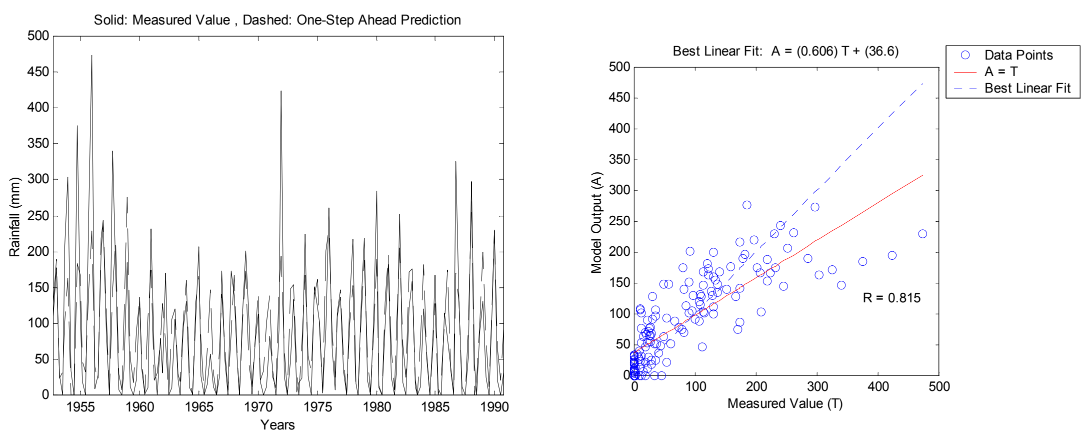
The golden rule in identification (and in most other matters) is to try simple things first. If a linear model does a job properly, one should not bother wasting time on neural network based models. The *ARX* (auto regression with eXtra inputs, called eXogeneous variables in econometrics) model was used to identify a seventh order linear model with three inputs (PGSST, NAO and SOI) and one output (rainfall) [11-12]. Results are shown in Fig.3.

The normalized root mean square error (NRMSE) is used for model performance investigation

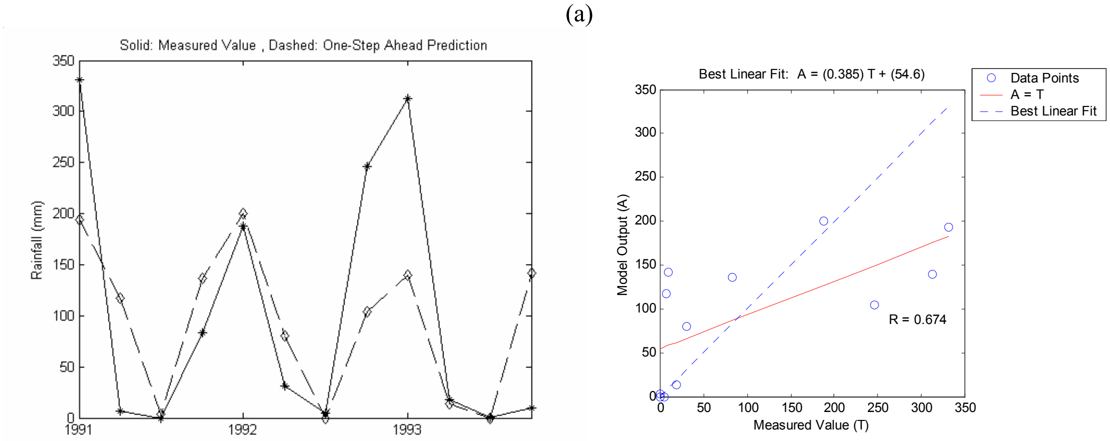
$$NRMSE = \sqrt{\frac{\sum [x(t) - \hat{x}(t)]^2}{\sum x^2(t)}}$$

where $\hat{x}$ is the model output, x is the target and time increment, t, is season. NRMSE is 0.4432 for training set and 0.5774 for test set.

It is often useful to perform a regression analysis between the model response and the corresponding targets. Here we have three parameters. The first two correspond to the slope and the y-intercept of the best linear regression relating targets (T) to model outputs (A). If we had a perfect fit (outputs exactly equal to targets), the slope would be 1, and the y-intercept would be 0. The third variable is the correlation coefficient (R) between the outputs and targets. It is a measure of how well the variation in the targets is explained by the outputs. If this number is equal to 1, then there is perfect correlation between targets and outputs [13]. R is 0.815 for training set and 0.674 for test set. The best previous prediction of Shiraz rainfall has resulted in a correlation coefficient of R=0.425, which was obtained by Nazemosadat for winter rainfall. So, the ARX model, because of its ability to consider the impact of combined indices on rainfall has significantly improved predictions. Although the ARX model, because of its ability to investigate the impact of combined indices on the rainfall, has significantly improved predictions, it is apparent that the linear model has severe problems. It is thus concluded that the underlying system's behavior may be described more accurately by a nonlinear model.



(a)



(b)

Fig. 3. ARX model results for rainfall forecasting in Shiraz: (a) train set (b) test set

5. SELECTING A NONLINEAR MODEL STRUCTURE

A neural network ARX (NNARX) model structure with the same parameters is attempted. *NNARX* determines a nonlinear ARX model of a dynamic system by training a hidden layer neural network with the

Levenberg-Marquardt algorithm [10]. Initially a fully connected network architecture with 15 hyperbolic tangent hidden nodes, one linear output node and 496 weights was selected. A schematic diagram of such a network is shown in Fig.4. We trained the network for 300 iterations with a small weight decay of 0.001. NRMSE is 3.0097e-004 for the training set and 0.5147 for the test set. R is 1 for the training set and 0.789 for the test set. Results are shown in Fig. 5. Performance of this nonlinear model is compared to that of the linear model in Table 1, which shows noticeable improvement. The NNARX model has increased correlation (R) and decreased NRMSE with respect to the linear ARX model. Also, results of a fourth order NNARX model are summarized in Table 1. It can be concluded that the number of data in the train and test sets is not sufficient.

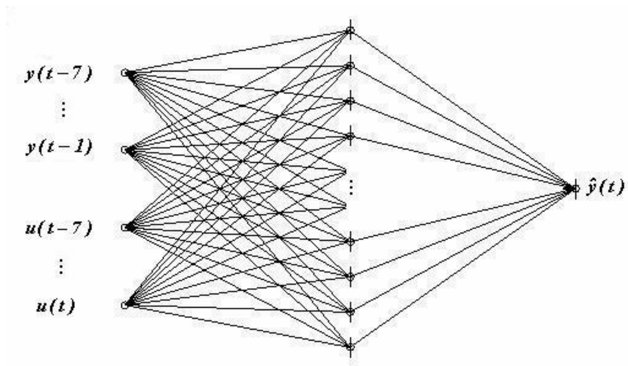
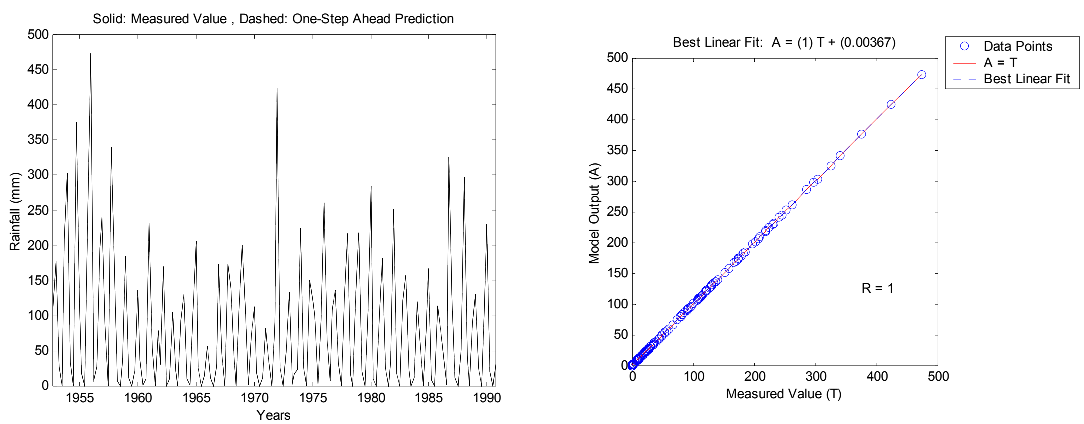
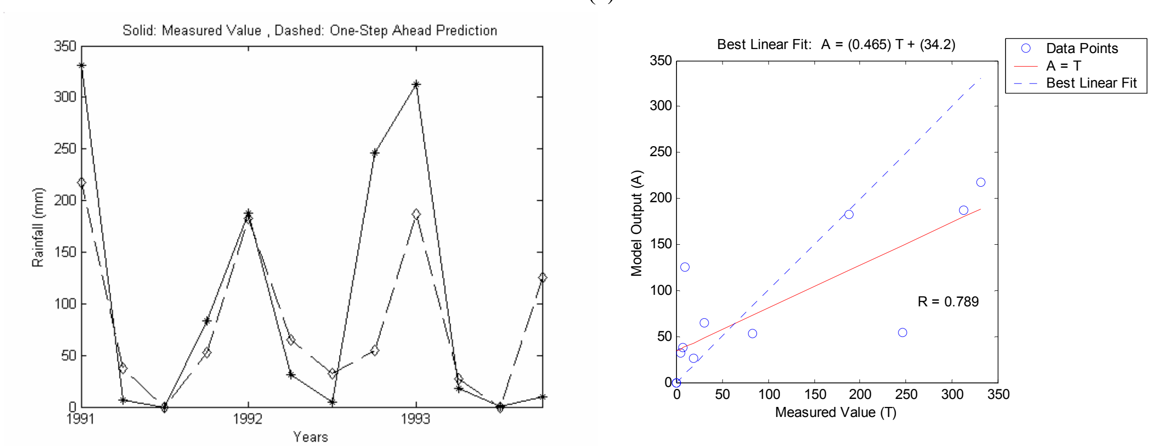


Fig. 4. Neural network schematic diagram



(a)



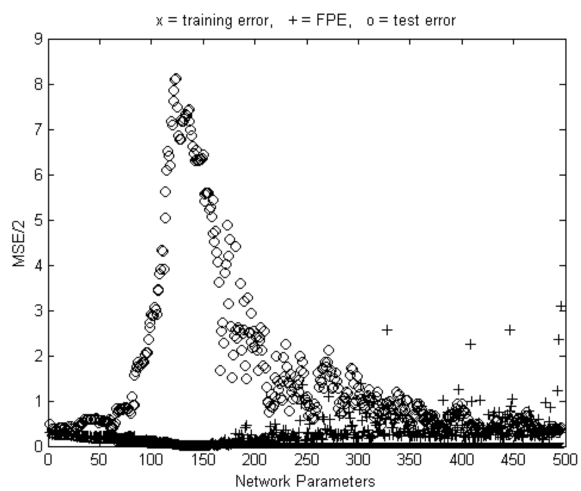
(b)

Fig. 5. NNARX model results for rainfall forecasting in Shiraz: (a) train set (b) test set

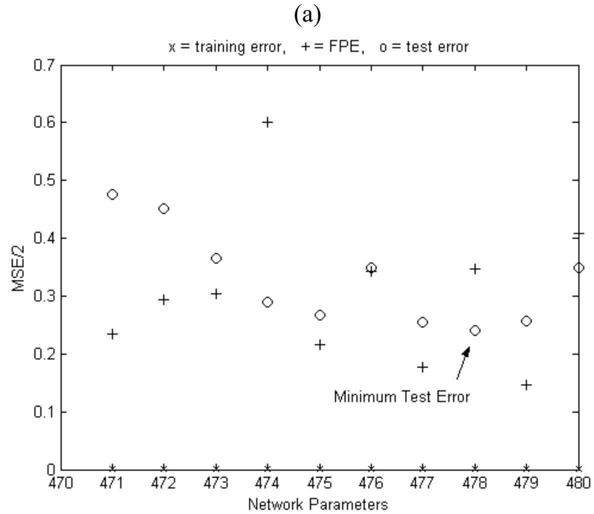
Table 1. Rainfall forecasting

MODEL STRUCTURE	Model Order	Train		Test	
		NRMSE	R	NRMSE	R
ARX	7	0.4432	0.815	0.5774	0.674
NNARX	4	3.0249e-006	1	0.4249	0.87
	7	3.0097e-004	1	0.5147	0.789
Pruned.NNARX	7	1.9945e-005	1	0.3939	0.907

Though better performance of the seventh order NNARX model with respect to the ARX model was obvious, it showed overfitting of data to some extent. Therefore, to remove the superfluous weights from the network, the function *NNPRUNE of NNSYSID Toolbox* was used. This function determines the optimal network architecture by pruning the network according to the optimal brain surgeon strategy [10]. The network was retrained with a maximum of 50 iterations after each weight elimination (i.e. eliminate one weight, retrain, eliminate one weight, retrain,...). This is the slowest but safest strategy. It is reasonable to select the network with the smallest test error as the final one. *NNPRUNE* produces the plot shown in Fig. 6, which displays training error, test error, and final prediction error estimate of generalization (FPE) of each of the intermediate networks. The plot reveals that the minimum of the test error occurs when there are only 478 weights left in the network. Regularization is helpful when pruning neural networks.



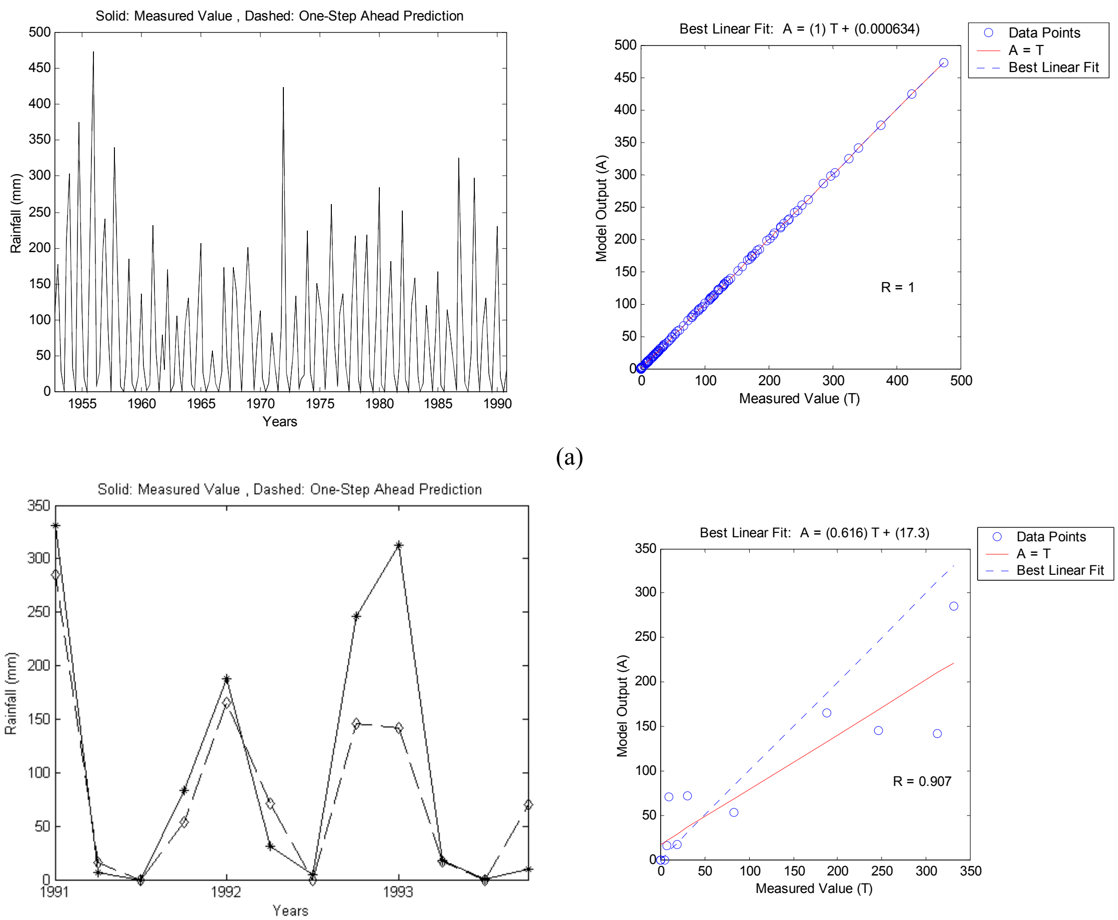
(a)



(b)

Fig. 6. Network pruning result: (a) training error, test error and final prediction error estimate of generalization (FPE) of each of the intermediate networks, where half of mean square error (MSE/2) was chosen as the performance index. (b) a close-up of (a), which shows the minimum test

However, when the optimal network architecture was found, it had to be retrained without the weight decay [10]. Now NRMSE is 1.9945e-005 for training set and 0.3939 for test set. R is 1 for training set and 0.907 for test set. It is concluded that the performance of the network is improved very much by pruning. Fig.7 shows the optimal network results. Ideally it is expected that the training error increase as the weights are pruned, while test error decrease until a certain minimum, and then start increasing. However, such strange results are not uncommon because results depend heavily on the local minimum from which one starts out [10].



(b)

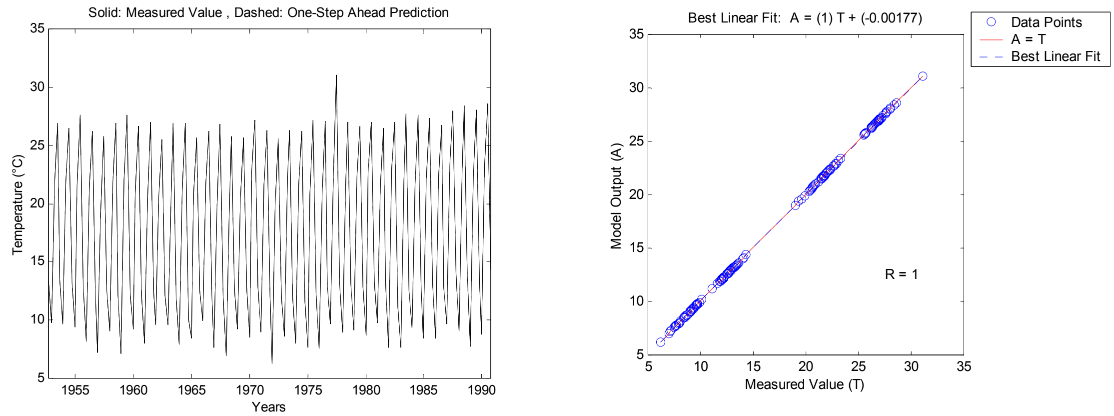
Fig. 7. Optimized network results for rainfall forecasting in Shiraz: (a) train set (b) test set

6. TEMPERATURE FORECASTING

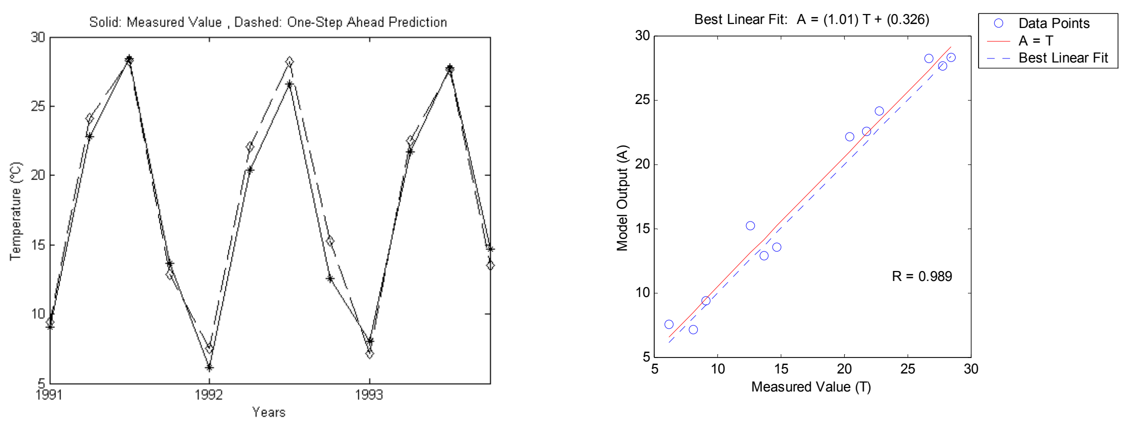
According to our experience and the relatively high correlation between rainfall and temperature an NNARX model structure with the same parameters (except the maximum iteration of 50) was employed for temperature prediction. NRMSE is 1.6300e-004 for training set and 0.0669 for test set. R is 1 for training set and 0.989 for test set, which is very satisfactory. Results are shown in Fig.8 and Tables 1 and 2.

Table 2. Temperature forecasting

Model	Train		Test	
Structure	NRMSE	R	NRMSE	R
NNARX	1.63300E-004	1	0.0669	0.989



(a)



(b)

Fig. 8. NNARX model results for temperature forecasting in Shiraz: (a) train set (b) test set

7. CONCLUDING REMARKS

Most parts of Iran in general frequently suffer from water shortage, southern parts in particular. For periods when the temperature is above normal, the demand for water increases for both domestic and irrigation purposes. Severe drought is, therefore, expected when a prolonged shortage of rainfall coincides with hot weather and heat waves. The ever-increasing population of Iran causes an increase in the demand for fresh water, and competent management of water resources is needed to mitigate the hazards of drought and flooding. Accurate forecasting of rainfall and temperature a season or even a few seasons ahead, is becoming more and more necessary in the region to accomplish the desired criteria for the management of water resources. Such forecasting is also important for resource management in agriculture, range land, soil conservation, tourism, housing and industry.

This paper utilized the time series as a benchmark to compare several black-box modeling methods such as linear system identification methods and neural networks. Models were applied to forecast rainfall and temperature of the city of Shiraz one season ahead. Simulation results show the capability of neural networks for time series prediction and suggest that the applied methodology is more efficient than the previous approaches. As shown, rainfall was not predicted as accurately as temperature because rainfall depends on numerous factors, of which only three were considered. Prediction will be improved by utilizing other variables that have a causal relationship with rainfall as network inputs. In a future work, some orthogonalization methods to reduce the number of inputs will be studied.

ABBREVIATIONS USED IN THIS PAPER

ARX atoRegression with eXtra inputs (called eXogeneous variables in econometrics)
CCA Canonical correlation analysis
ENSO El Nino-southern oscillation
FPE final prediction error
NAO north atlantic oscillation
NN neural network
NNARX neural network ARX
NRMSE normalized root mean square error
PCA principal component analysis
PGSST persian gulf sea surface temperature
R correlation coefficient
SOI southern oscillation Index

REFERENCES

1. Atiya, A. F., El-Shoura, S. M., Shaheen, S. I. & El-Sherif, M. S. (1999). A Comparison between neural-network forecasting techniques-case study: River flow forecasting. *IEEE Trans. on Neural Networks*, *10*(2), 402-409.
2. Safavi, A. A. (1996). *Wavelet-based neural networks and multiresolution analysis with applications to process systems engineering*. PhD Thesis, Dept. of Chemical Eng., The University of Sydney, Australia.
3. Haykin, S. (1999). *Neural networks-a comprehensive foundation*. 2nd ed, Prentice-Hall.
4. Ding, X., Canu, S. & Denceux, T. (1996). Neural network models for forecasting. *Neural networks and their applications*, In J. G. Taylor, Ed., John Wiley and Sons, 153-167.
5. Nazemosadat, M. J. & Cordery, I. (2000). On the relationship between ENSO and autumn rainfall in Iran. *International Journal of Climatology*, *1*, 47-62.
6. Nazemosadat, M. J. & Cordery, I. (2000). The impact of ENSO on winter rainfall in Iran. *The 26th National and 3rd International Hydrology and Water Resources Symposium*, Inst. Engs. Australia.
7. Nazemosadat, M. J. (1999). The impact of ENSO on autumn rainfall in Iran. *Drought News Network*, *11*, 15-19.
8. Nazemosadat, M. J. (1998). The Persian Gulf sea surface temperature as a drought diagnostic for southern parts of Iran. *Drought News Network*, *10*, 12-14.
9. Siouris, G. M. (1996). *An engineering approach to optimal control and estimation theory*. John Wiley & Sons, INC.
10. Norgaard, M. (2000). *Neural network based system identification toolbox.* Ver. 2, Tech. report 00-E-891, Department of Automation, Technical University of Denmark, (http://www.iau.dtu.dk/nnbook/software.html).
11. Ljung, L. (1999). *System identification-theory for the user*. 2nd ed, Prentice-Hall.
12. Ljung, L. (1997). *System identification toolbox user's guide*. The MathWorks Inc.
13. Demuth, H. & Beale, M. (1998). *Neural network toolbox user's guide*. Ver.3, The MathWorks Inc.