Use of Neural Networks to Forecast Time Series: River Flow Modeling

RICHARD CHIBANGA[1], JEAN BERLAMONT[2] AND JOOS VANDEWALLE[3]
[1]Ph.D. student in the Hydraulics Laboratory, Civil Eng. Dept.
[2]Prof. and Head of Hydraulics Laboratory, Civil Eng. Dept.,
[3] Prof. and Head of Electrical Engineering Dept.
Catholic University of Leuven,
Kasteelpark Arenberg 40, 3001 Heverlee (Leuven),
BELGIUM

Abstract: - This paper presents an alternative approach to time series forecasting, through use of artificial neural networks (ANNs), a relatively new concept in hydrological research. Box and Jenkins ARMAX (autoregressive moving average with exogenous inputs) models have been widely used in modeling various time series with satisfactory results. This study shows that ANNs can produce comparable, to ARMAX, and in some cases even, better forecasting results, especially for long-term prediction. By learning, through training, the underlying mapping of the time series, an ANN provides robust forecasting. The results obtained using real-life data from a catchment in Zambia suggest ANNs could be used as an efficient and effective models in forecasting hydrological variables such as river discharge, river stage, and runoff.

Key-Words: - Artificial Neural Networks, feedforward, ARMAX, alternative approach, training, mapping, hydrologic(al), forecasting.

1 Introduction

Man is in constant pursuit of taming or at least being in “control” of naturally occurring phenomena so as to harness them to his advantage. This control is often realized when he is able to predict events with reasonable accuracy, thereby being able to plan in advance what course of action to take. In hydrology forecasting of one process output or the other such as runoff, river discharge, and river stage is fairly commonplace. In cases where the interest is in the understanding of the underlying hydrological processes conceptual models are the best, however, there are many practical situations such as streamflow forecasting where the interest is in making accurate predictions at specific watershed locations.

In the latter situation, which is the subject of this paper, a hydrologist may prefer not to expend the time and effort required to develop and implement a conceptual model and instead use a simpler system theoretic model. In the system theoretic approach, difference or differential equation models are used to identify a direct mapping between the inputs and outputs without detailed consideration of the internal structure of the physical processes [1, 2]. In this paper the system theoretic models used are the artificial neural networks, ANNs. ANN modeling is a technique that compiles and refines the main advantages of the different conventional approaches, generally intended for systems behavior forecasting. In view of Kolmogorov’s theorem [3] and Funahashi’s work [4] it is now universally held that a three layered ANN using sigmoid transfer functions can serve as any continuous function approximator for as long as a sufficient number of neurons are used. River discharge time series as measured at the Kafue Hook bridge (KHB) in the Kafue sub-catchment in Zambia was presented to several three-layer feedforward – backpropagation (FF-BP) ANNs for training. The KHB gauging station captures the main inflow contribution of the Kafue River to a series of three reservoirs; the Itezhi-tezhi, (Itezhi, man-made), the Kafue Flats (NR, natural) and the Kafue Gorge (KafG, man-made) in that order, see Figure 1.

Hydroelectric power generated at KafG is the major reason for the existence two man-made reservoirs; at least 60% of Zambia’s electric energy is produced here. But the storage in these reservoirs has to account for irrigation and water supply needs thereby rendering them multipurpose in their operation. As such it is imperative to devise effective and efficient reservoir management and operation policies to meet the competing needs. To do this it is of utmost importance that one is able to predict, with reasonable certainty and accuracy, the expected inflows over a certain time horizon with minimum consideration of the details of the interacting processes. Traditionally, the linear time series models such as ARMAX have been commonly used in such situations because they are relatively easy to develop and implement, and they have been found to give satisfactory predictions in many applications [5, 6]. But in recent years ANNs have been gaining some ground in being used in the water resources studies [7, 1, 8]. To assess and evaluate the performance of the ANNs that are selected they are compared to best performing ARMAX models on the basis of appropriate performance criteria.

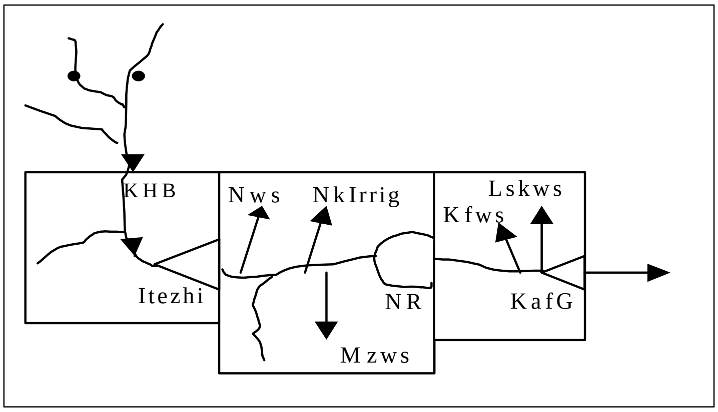

Fig. 1: Kafue sub-catchment study area showing the main features

2 Data series partitioning for ANN training

The measured discharge time series was almost 20 years long of daily averages. The inputs and output series were partitioned in three pattern sets for training, validation and testing in such a way that some earlier, middle and recent data points were included in each set. This should remove or rather to minimize any bias that may be inherent in the series due to long-term weather changes. For instance, it may be that the earlier years are generally wetter than the recent years or the other way round. By sampling several disjointed segments of data from the original series we present the ANN with many possible varied patterns during training. By partitioning the data series this way some impartiality in evaluating the prediction performance is built-in, in that data from "all" periods (earlier, middle and later sections of the series) are used for training (modeling), testing (prediction), and validation. All these sets are of the form $\{x_1(t), x_2(t), \ldots, x_p(t), d(t)\}$, in which $x_i(t)$, are the inputs while $d(t)$ is the desired output -present *Inflow(t)*. Both the inputs and output series were then normalized (to range from 0 to 1) by dividing each variable by the maximum value hitherto obtained. From the correlation matrix of possible exogenous inputs, rainfall and evaporation measured from nearby meteorological stations it was clear that their influence was negligible, but there were very strong correlation amongst the flow at time t and the recent past flows at t-1, t-2 and t-3. The output, flow at t, $Y(t)$ was assumed to be related to the past flows $Y(t-i)$ and since it is also known that ANNs are generally nonlinear we can write the general nonlinear model structure as

$$Y(t) = f_{non}(Y(t-1), \cdot\ \cdot\ \cdot\ , Y(t-n_a)) + e(t) \qquad (1)$$

where $f_{non}(\ )$ is the unknown nonlinear mapping function, $e(t)$ is the unknown mapping error (to be minimized), and n_a is the (unknown) number of past outputs contributing to the present output. It was decided to work with three layer (Input layer:1 Hidden layer: Output layer) FF-BP ANNs, owing to the well known universal approximation property in neural network research that it is capable of mapping and generalizing any continuous mathematical function. Figure 2 shows the general configuration of a three layer feed forward ANN. Note that there is no interconnection between neurons within the same layer, rather neurons in one layer are allowed to have interconnections with those in previous or forward layers. Such an ANN model structure is represented by the notation ANN[n_a, n_h, n_o], where n_a is the number of nodes in the input layer, and is the same n_a as in equation (1), n_h is the number of nodes in the hidden, and n_o is the number of nodes in the output layer (n_o = 1 in our case). Since n_o = 1 is and fixed the notation is further abbreviated to ANN[n_c, n_h]. Input training sets consisting of {*Y(t-1)*}, {*Y(t-1), Y(t-2)*}, and {*Y(t-1), Y(t-2), Y(t-3)*} combinations were presented

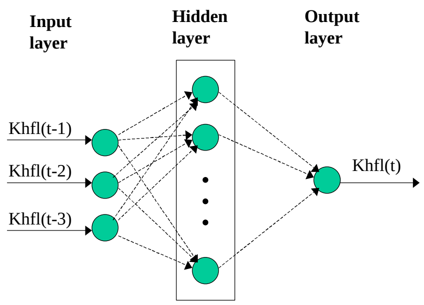

Fig. 2: Three-layer feed forward network

to a host of three layer FF-BP ANNs for training, using MATLAB routines. For each combination of input patterns the number of neurons in the hidden layer was varied from 2 to 10 while the pertinent global 'goodness-of-fit' statistics for each satisfactorily trained network were noted. The training method used is the Levenberg-Marquardt (L-M) in which an early stopping criteria is incorporated to safeguard against overfitting. In each step in the training phase, the network is required to predict the next value in the time sequence. The error between the value predicted (by the network) and the value actually observed is calculated and propagated backwards along the feedforward connections. During back-propagation the network weights are modified by minimizing the error between a target and computed outputs. The objective of weight modification is to find a set of weights (weight matrix) that enables the trained ANN to approximate the target output as closely as is desired. This mode of training is what is known as supervised or teacher-forcing.

In the absence of one definitive evaluation test, a multi-criteria assessment was used to select a best-fit network for further tests: 1) The ratio of the standard error of estimate (S_e) to the standard deviation (S) of the observed signal, flow, (S_e/S).

$$S_e = \left[\frac{1}{n} \sum_{i=1}^{N} (y_{oi} - y_{pi})^2 \right]^{1/2} \qquad (2)$$

S_e is the unexplained variance and is the standard error of estimate [9]; ν = degrees of freedom, and is the number of observations in the training set minus the number of network weights; y_{oi} and y_{pi} are observed and predicted values of output, respectively. S_e/S, also called the noise-to-signal ratio indicates the degree to which noise hides the information [10]. The smaller the ratio the better the model can provide accurate predictions of the modeled signal. 2) The the percent volume error (%VE), the smaller this is the better is the approximation of observed and predicted volumes over the period in consideration. 3) Then the linear regression correlation between the observed and simulated flows (CORR), the higher the correlation certainly the better the general approximation. Where no optimal architecture emerged, the A information criterion (AIC) and/or its variant the B information criterion (BIC) were used to discriminate.

$$\text{AIC} = m*\ln(\text{RMSE}) + 2npar \qquad (3)$$

$$\text{BIC} = m*\ln(\text{RMSE}) + npar\ \ln(m) \qquad (4)$$

Where m is the number of input-output patterns, and *npar* is the number of parameters to be identified. Since the AIC and BIC statistics penalize the model for having more parameters they therefore tend to select more parsimonious models [1].

3 Results and discussion

From the 21 initial ANN models, with varying inputs, that were trained, a preliminary 9 were short-listed based mainly on S_e/S-ratio and %VE values that were obtained for Testing and Validation. Figure 3 shows the comparison of S_e/S-ratios (bars) along with %VE (lines) for some of the trained models. The input set for each column of three models is also included in the top row(s) of the Figure: from which, using the same criteria as above, all the models but the first three were selected, that is six of them.

These models were first tested for simulating the data series segments that were not used for training; they all performed well wit respect to the mean squared error, MSE and the coefficient of regression, R which were comparable. It is envisaged that the successful model(s) is (are) to be used for forecasting the Kafue Hook bridge inflows to be used in determining reservoir operation policies. To do this, all the 6 models were tested for their prediction capabilities at 1day, 10 days, 50 days and 100 days-ahead forecasting horizons, evaluating them using MSE and R. In multi-step prediction, we append the previously predicted values to our input series and use these 'new' values to predict future values. The best three models were retained as candidates that could be used for actual prediction, namely M3-[3,5], M2-[2,2] and M2-[2,8]; they are listed as 1, 2, and 3 in Table1.

Figures 4 show the 1 d- and 100 d-ahead comparisons of forecasts with observed inflow together with their respective regression correlation plots. By visual inspection of Figures 4 it can be seen that the forecasts closely match the observed data points despite the slight departure in the initial stages of the 100 d-ahead plot. It can be concluded that the selected ANN(s), which may not necessarily be of optimal architecture, seem to have captured the general underlying relationship between the current inflow and the recent past flows.

4 Comparison with ARMAX models

ANNs are a relatively new modeling concept in general and even newer in hydrological modeling. It is therefore appropriate to compare the performance of ANNs to "traditionally" used approaches: [11] talking of ANNs note, "*These new technologies, however, require evaluation against conventional models and statistical tools, in order to determine their relative performance…*" To evaluate the selected ANNs further for their forecasting capabilities they are compared with ARMAX models.

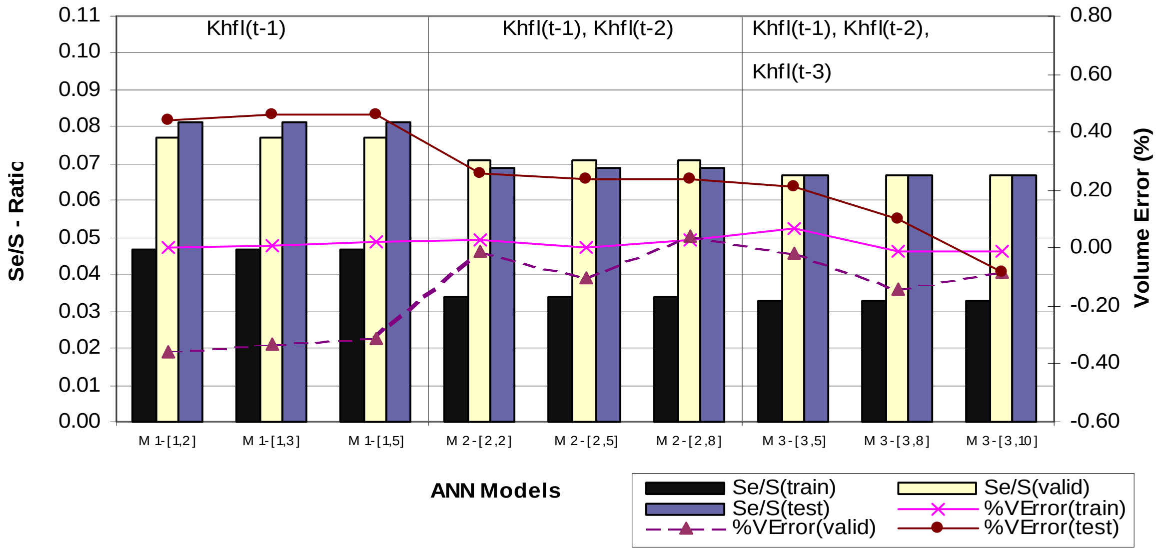

Fig. 3 Comparison of S_e/S-ratios for Training, Validation and Testing along with %VE values

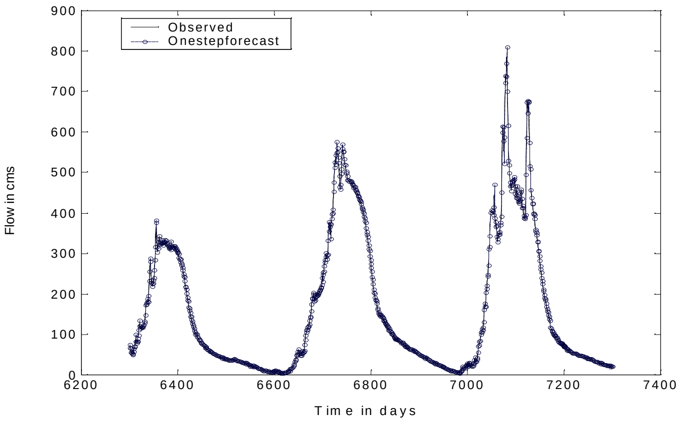

Figure 4a: 1 d-ahead inflow obtained by M3-[3,5]

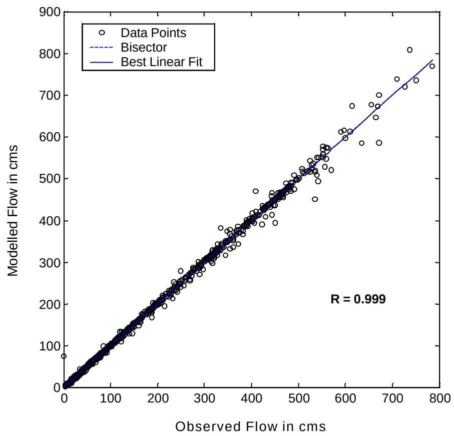

Figure 4b: Regression of the results of Fig. 4a

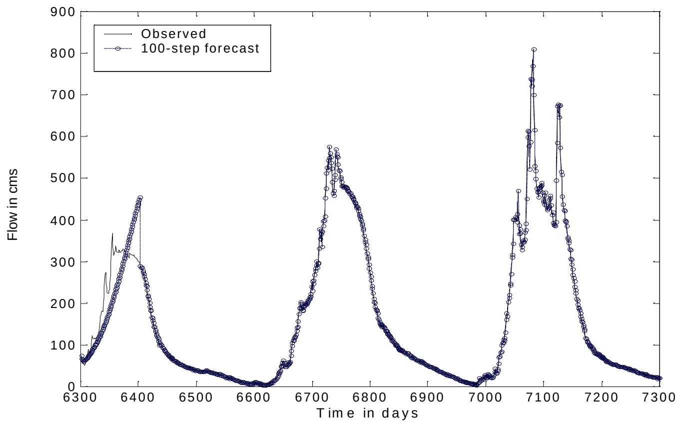

Figure 4c: 100 d-ahead inflow obtained by M3-[3,5]

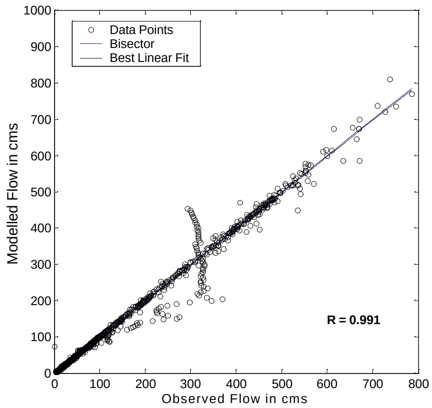

Figure 4d: Regression of the results of (Fig. 4c)

In general, the ARMAX model is expected to simulate very well the behavior of system whose input-output characteristics are approximately linear. ARMAX models have been widely used for watershed modeling because of the ease with which they can be developed [12]; and they have been found to provide satisfactory predictions in many applications [5]. The model structure is represented by the notation ARMAX[n_a, n_b, n_c], where n_a, n_b, and n_c are the (unknown) number (order) of past outputs, inputs, and error terms respectively. The orders (n_a, n_b, and n_c) and the model parameters were estimated using the MATLAB Identification Toolbox.

For consistence it was decided to use the MSE and R as the evaluation criteria. The selection of the best models is based on how these test statistics faired in their prediction mode at the same prediction horizons as used for ANNs, that is at 1 day, 10 days, 50 days and 100 days-ahead forecasting horizons. Here again only the best three models were retained as candidates; these being ARMAX[2, 0, 0], ARMAX[3, 0, 0], and ARMAX[1, 0, 3]. The performance statistics of these models together with those of the ANNs selected earlier are summarized in Table 1. Figures 5 show the comparisons of 1 d-ahead forecasts given by ARMAX[2, 0, 0] with observed inflow. Visual inspection of Figures 5 shows that the forecasts at 1 d-ahead are consistently overestimated and it would not get any better as the forecasting horizons increase. This clearly gives inferior forecasts compared with what was obtained by ANNs. The statistics of Table 1 further show this lackluster performance of ARMAX models in this particular case: the R values show a significant reduction for successive multi-step forecasts and there is even a considerably large increase in the MSE for each higher step of the forecasting horizons considered.

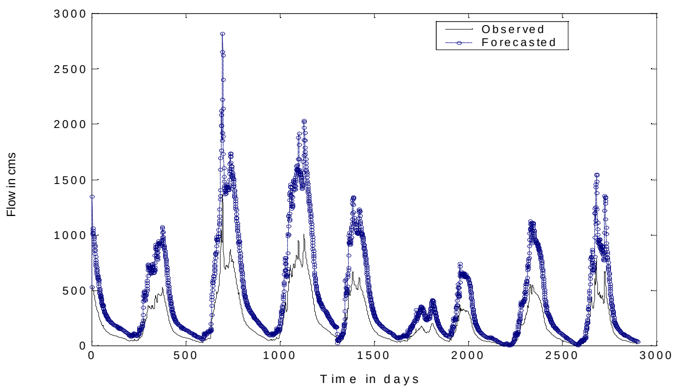

Figure 5a: 1 d-ahead inflows obtained by ARMAX[2,0,0]

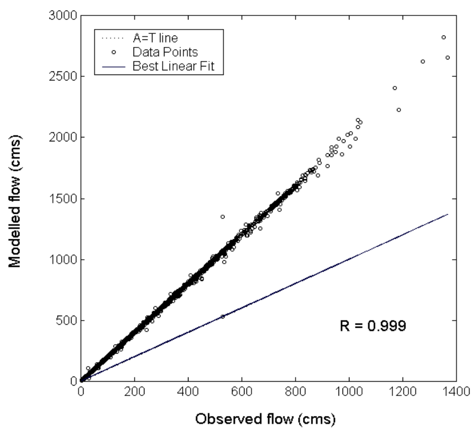

Figure 5b: Regression of the results of (Fig. 5a)

Table 1: Statistics of three best performing (prediction) ANN and ARMAX models

Model	Inputs			MSE & R for different prediction horizons							
	Khfl (t-1)	Khfl (t-2)	Khfl (t-3)	1 d-ahead MSE	R	10 d-ahead MSE	R	50 d-ahead MSE	R	100 d-ahead MSE	R
ANN[3,5,1]	xx	xx	xx	76.22	0.999	76.60	0.999	172.22	0.997	539.03	0.991
ANN[2,2,1]	xx	xx		78.46	0.999	79.22	0.999	370.31	0.994	1954.80	0.967
ANN[2,8,1]	xx	xx		77.29	0.999	78.08	0.999	444.71	0.992	2702.61	0.955
ARMAX[2,0,0]	xx	xx		219.30	0.998	5643.30	0.966	37013	0.802	75306	0.690
ARXMA[3,0,0]	xx	xx	xx	222.57	0.998	5630.20	0.967	37272	0.802	76930	0.692
ARMAX[1,0,3]	xx			231.80	0.998	5724.90	0.967	37555	0.802	74014	0.692

5 Conclusion

In all cases presented in this paper the ANN models generated better global "goodness-of-fit statistics" than ARMAX models as the prediction horizon increased. This is because the latter are sensitive to noise, and since they build their forecasts on previous observations, thus is only good for short term forecasting. An ANN (with a hidden layer) bases its forecasting on the approximated (learnt) underlying mapping. Hence it is more robust and better in the case of long term forecasting [13]. But just as [1] noted also the ANN approach presented here does not provide models that have physically realistic components and parameters... However, the results suggest that the ANN may provide a superior alternative to the ARMAX time series approach for developing input-output simulation and forecasting models. It, therefore, can be seen that ANNs are a viable alternative and/or complementary approach to conventional watershed modeling techniques. Especially in cases where the main interest is in making accurate predictions at specific watershed locations rather than in understanding the hydrologic processes. Notwithstanding the time taken in training an ANN model, the benefit of robust long-term forecasting would make it a favored choice over an ARMAX model, especially in long term planning such as in reservoir operations.

References

[1] Hsu, K., Gupta, V. H., and Sorooshian, S., *Artificial neural network modelling of rainfall-runoff process*, ***Water Resour. Res., 31(10)***, 1995, 2517 – 2530.

[2] Furundzic, D. Application example of neural networks for time sries analysis: Rainfall-runoff modelling, ***Signal Processing, 64***, 1997, 383 – 396.

[3] Kolmogorov, A. N., On the representation of continuous functions of many variables by superposition of continuous function of one variable and addition, ***Dokl. Ak ad. Nauk USSR 114*** 1957, 953 – 956, Russia.

[4] Funahashi, K., On the approximate of continuous mappings by neural networks***, Neural Networks 2***, 1989,183 – 187.

[5] Bras, R., and Rodriguez-Iturbe, I., *Random functions and Hydrology,* Addison_Wesley, Reading, Mass. 1985.

[6] Wood, E. F. (Ed), *Workshop on Real Time Forecasting/Control of Water Resources Systems,* Pergamon, New York. 1980.

[7] Chow, T. W. S., and Leung, C. T. Neural Network based short-term load forecasting using eather compensation, ***IEEE Trans. Power Systems 11(4)***, 1996, 1736 – 1742.

[8] Singh, V. P., Hydrologic Systems: Rainfall-Runoff Modeling, vol., I Prentice-Hall, Englewood Cliffs, N. J., 1988.

[9] Tokar, S. A., and Johnson, P. A., Rainfall-runoff modelling using artificial neural networks, ***J. Hydrol*. Eng., 4(3)**, 1999, 232-239.

[10] Gupta, V. K. and Sorooshian, S., The relationship between data and precision of parameter estimates of hydrologic models, ***J.Hydrol., 81,*** 1985, 57 – 77.

[11] Abraharty, J. R., and See, L., *Comparing neural network and autoregressive moving average techniques for the provision of continuous river flow forecasts in two contrasting catchments,* **Hydrol. Process*.,14***, 2000, 2157-2172.

[12] Young, P., and Wallis, S., *Recursive estimation: A unified approach to the identification, estimation, and forecasting of hydrological systems,* ***Appl. Math. Comp., 17***, 1985, 299-334.

[13] Tang, Z., de Almeida, C., and Fishwick, P. A., *Time series forecasting using Neural Networks vs. Box-Jenkins methodology*, **Simulation, 57(5),** 1991, 303-310.