Material Science Research India Vol. **6**(1), 11-18 (2009)

Inputs selection for artificial neural networks for multivariate time series

Z.H. ASHOUR* and H.A. FAYED

Department of Engineering Mathematics and Physics, Cairo University, Cairo (Egypt).

(Received: September 20, 2008; Accepted: November 14, 2008)

ABSTRACT

A new method in selecting the inputs of the neural networks for multivariate time series is proposed. The input and output time series are analyzed and suitable mathematical models are built in the input-output model parametric representation. The inputs to the best input-output models are chosen as the inputs to the neural network model. The estimates from different models are then combined to obtain an unbiased estimate of the output series. The resulting combined neural network model is used in forecasting the bench marking gas furnace multivariate time series. The new developed procedure outperformed previously used forecasting techniques.

Key words: Artificial neural networks, inputs Selection, multivariate time series.

INTRODUCTION

The key to good forecasts is the building of a suitable model. During the past forty years several techniques have been used in time series modeling[1-3]. Most of the suggested mathematical models were linear models[4]. In general the observable time series of real life physical processes are nonlinear. The comparison between mathematical linear and nonlinear models led researchers to favor the use of nonlinear models to represent time series data[5].

The use of neural network models represented a good alternative to existing Box-Jenkins ARIMA (autoregressive integrated moving average) linear models[6-7], but the neural network structure remained a debatable issue. The choice of the inputs, the structure and the number of layers were left open to the researcher. Ashour[8] proposed the proper selection of the inputs to the neural network by performing statistical analysis of the data .The procedure involved building linear Box-Jenkins ARIMA models for the series under investigation then the inputs to the ARIMA model are used as inputs to the neural network .The proper selection algorithm made use of the statistical properties of the time series hence only the inputs that really affect the output are chosen[9-10]. The proper choice of the inputs to the neural network results in a less complicated network architecture and fewer weights to calculate[11].

In the case of multivariate time series, this approach did not build a good, suitable model for the time series, since not only the current time series output depends upon the past observations, but also it depends on other variables[2].

In this paper a new technique in selecting the inputs of the neural networks for multivariate time series is proposed so that only inputs affecting the output series are selected. The new developed procedure was tested on the famous gas furnace data set and the resulting neural network model outperformed Box-Jenkins transfer function model for the same data.

Combining the output estimates of several neural networks models[12-13] and combining the estimates of linear models with neural networks models estimates[14] are known to improve the efficiency of the model representing the time series under consideration.

In our experimental results we also combined the estimates of the neural network models with the estimates of the autoregressive AR model built for the output series, considering that the input series is not observable, to obtain a better estimate of the output series.

In this paper the formulation of the problem is outlined. The input selection technique for the univariate time series case is presented. The proposed inputs selection approach for multivariate time series is described. Experimental results on the bench marking gas furnace data are detailed. Finally conclusions are presented.

Problem formulation

Consider the nonlinear autoregressive model NAR expressed by:

$$y_t = h(y_{t-1}, y_{t-2}, \ldots, y_{t-p}) + e_t \quad \ldots(1)$$

The residuals e_t are assumed to be a white noise sequence with zero mean and variance σ^2. Equation (1) defines a very large and general class of models. The function *h* is implemented by a multilayer perceptron with a number of hidden units and the well-known sigmoidal activation function $f(x)$. So the output of the neural network can be formulated as:

$$\hat{y}_t = \sum_{i=1}^{I} W_i f\left(\sum_{j=1}^{p} w_{ij} y_{t-j} + \theta_i\right) \quad \ldots(2)$$

The parameters W_i are the hidden neurons output weights and w_{ij} are the hidden neurons inputs weights .The θ_i's are the bias for the neurons, and I is the number of hidden neurons. The neural network structure as presented by equation (2) is a feedforward neural network. Another neural network structure is the recurrent neural network. In this type of networks the output is fed back to the network making it suitable to represent NARMA (non linear ARMA(p,q)) time series 11., 15.. The estimate of the output y is given by:

$$\hat{y}_t = \sum_{i=1}^{I} W_i f\left(\sum_{j=1}^{p} w_{ij} y_{t-j} + \sum_{j=1}^{q} w'_{ij}(y_{t-j} - \hat{y}_{t-j}) + \theta_i\right) \ldots(3)$$

In order to find the best neural network to fit the time series we have many choices to make. We have to choose its interconnection structure whether it is a feedforward or recurrent neural network, the number of layers, the number of hidden units (neurons) in every layer .The number of hidden neurons affects the number of weights to be calculated.

One of the main problems left for research is the proper selection of the inputs to the neural network[8-11,16]. Weight elimination techniques start by a comfortable overparametrized neural network structure then they eliminate insignificant weights[5], but comfortable is neither a precise number nor a convincing criterion. A large number of inputs increases the number of weights to be calculated and a small number of inputs misses the information needed to estimate the desired output. A network with fewer inputs has fewer adaptive parameters to calculate and will give a neural network with better generalization properties and a network with fewer weights is faster to train[11].

A technique to facilitate the selection of the inputs to the neural networks was proposed in 8. and is presented in the next section. The method relies on the correlation analysis of the data. This approach in selecting the inputs and the configuration of the network led to a better neural network model for real life time series which consequently gave better forecasts[8-11,13], required less computation time and fewer weights to evaluate then starting the procedure with a comfortable length of past data as inputs.

Input selection algorithm for univariate time series

Time series are spaced observations or readings taken with time of many real physical processes[17]. The need is great to fit a suitable model for these series in order to control the process or to forcast future values in the objective of planning future decisions.

We may only have measurements of the output time series in the case of unknown input or when we have difficulty measuring it. Ashour[8] used the statistical analysis of the data series y(k) as the criteria for proper input selection. The series is studied for possible trends and periodicities. The mean μ, the variance σ^2 and the autocorrelation function r_k at different lags are calculated.

$$\rho_k = \frac{E(y_k-\mu)(y_{k+1}-\mu)}{\sigma^2} \quad \text{...(4)}$$

A suitable Box–Jenkins ARIMA*(p,d,q)* model is then built to fit the time series data, it is given by the following equation:

$$\Phi(B)\nabla^d y(k) = \theta(B)a(k) \quad \text{...(5)}$$

where *a(k)* is a zero mean white noise sequence, *B* is the backward shift operator, *Φ(B)* and *θ(B)* are polynomials in *B* containing the autoregressive and moving average parameters respectively

$$\nabla^d = (1-B)^d \quad \text{...(6)}$$

where *d* is the degree of differencing of the series to overcome non stationarity problems.

The inputs to the ARIMA *(p,d,q)* model in equation (5) are then selected as the inputs to the neural network model[8]. The neural network configuration (feedforward or recurrent neural network) is chosen based on the ARIMA model fitting the time series[11].

This discrete dynamic system according to Box-Jenkins[3] is often parsimoniously represented by:

where the noise is modeled as an ARIMA stochastic model, *θ(B), Φ(B)* are the moving average and autoregressive operators respectively for n_t.

Proposed approach: Inputs selection algorithm for multivariate time series

In the case of multivariate time series the output time series depends on a measurable input time series. The ARIMA*(p,d,q)* model only describes the relation between the current output and previous observations taken at different lags. This model is not adequate by itself to model and forecast the output time series well since other measured potential inputs are available.

The transfer function model links the outputs of a given process to its inputs when the relation between the inputs and outputs is linear. The transfer function model itself is a nonparametric representation of the process and requires the identification of a noise model to accommodate for any discrepancies in the model.

Consider the output y_t of the linear filter:

$$y_t = V(B)x_{t-b} + n_t \quad \text{...(7)}$$

where $V(B)=V_0+V_1B+V_2B+\ldots$ is the transfer function and the weights $V_0, V_1,\ldots$ are called the impulse response function of the system.

This discrete dynamic system according to Box-Jenkins 3. is often parsimoniously represented by:

$$y_t = \frac{\omega(B)}{\delta(B)}x_{t-b} + \frac{\theta(B)}{\phi(B)}a_t \quad \text{...(8)}$$

where the noise is modeled as an ARIMA stochastic model, *q(B)*, *f(B)* are the moving average and autoregressive operators respectively for n_t. The Box-Jenkins transfer function model requires a considerable amount of computation to identify the model, estimate the parameters diagnostic checking of the residuals. It does also require expertise to decide for the initial values of all the parameters.

The proposed approach

A parametric model representation of multi-input single-output systems is chosen to model the time series. It has the advantage of linking the output time series to the inputs in a direct way as:

$$y_t + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \ldots + \alpha_r y_{t-r} = \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \ldots + \beta_m x_{t-m} \quad \text{...(9)}$$

Several techniques can be used to decide for the order of the model, i.e. the structural parameters r and m, like the range error test[18], the residual error technique[4], etc. But this will require extra calculations and the examination of a very large number of possible combinations. The authors propose to study the autocorrelations of the output and the cross correlations between the input and the output to decide for the preliminary inputs that affect the output series. This approach will narrow the modeling procedure to only few models to choose from.

The input-output representation can be translated to the following difference equation:

$$y_t = \sum_{j=0}^{m} \beta_j x_{t-j} - \sum_{i=1}^{r} \alpha_i y_{t-i} + e_t \quad ...(10)$$

where e_t should be a zero mean white noise sequence in case of a good fitting model. So if we have N observations and H is the information matrix containing the samples of the inputs and output data and Φ is the parameters vector, then the vector of all outputs can be expressed as:

$$\underline{Y} = H\,\underline{\Phi} + \underline{E} \quad ...(11)$$

The proposed procedure was tested on a bench marking time series, the gas furnace data set of Box-Jenkins [3]. Then, parameter estimation can be performed using the least squares optimization technique. It can be implemented in one shot requiring matrix inversion, i.e. batch calculations. The other alternative is to use the recursive pseudo inverse algorithm.

The best model is the one which minimizes the sum of squares of the residual errors, *SSE*.

$$SSE = \underline{E}^T\,\underline{E} = (\underline{Y} - H\,\underline{\Phi})^T(\underline{Y} - H\,\underline{\Phi}) \quad ...(12)$$

The inputs to the best input-output model are chosen as inputs to the neural network model. Those inputs are the real inputs that affect the output based on the statistical analysis of the input and output time series. This selection is likely to result in an adequate neural network model for the time series under investigation with fewer inputs than previously anticipated.

RESULTS

The proposed procedure was tested on a bench marking time series, the gas furnace data set of Box-Jenkins 3.. The data was recorded from a combustion process of a Methane-Air mixture. It consists of 296 pairs of input-output readings. The input x_t is the gas flow into the furnace and the output y_t is the CO_2 concentration in outlet gas. The sampling interval is 9 seconds. The data set was split into two subsets: the first 200 samples were used for modeling and the following 96 samples were used for testing.

Best inputs to the neural net using the proposed approach

The autocorrelation and partial autocorrelation functions for the output time series, given in figures (1-2) show the dependence of the current output y_t on the past four output measurements. An autoregressive model AR was built to fit the output series as if the input time series was not observable:

$$(1 - 2.192\,B + 1.401B^2 + 0.06847B^3 - 0.2775B^4)y_t = a_t \quad ...(13)$$

The cross correlation function reveals a strong interdependence between the current output y_t, the current input x_t and a long string of previous inputs, see figure 3. Since every observation is a function of the past observations up to the fourth lag and that every observation incorporates the effect of all past inputs, we considered to only include x_t and x_{t-1} in our modeling procedure in the linear input-output model.

Two input–output models M1, M2 were built to fit 200 samples of the gas furnace time series based on the correlations analysis results. The inputs to M1 are $y_{t-1}, y_{t-2}, y_{t-3}, y_{t-4}$ and x_{t-1} and the inputs to M2 are $y_{t-1}, y_{t-2}, y_{t-3}, y_{t-4}, x_t$ and x_{t-1}. In the two models, the residual errors passed the whiteness test, but the statistical diagnostic tests indicated the inflation of the p-value for some parameters and a large standard error (S.E.). The tests revealed that y_{t-3} should be dropped as an input and that x_t and x_{t-1}

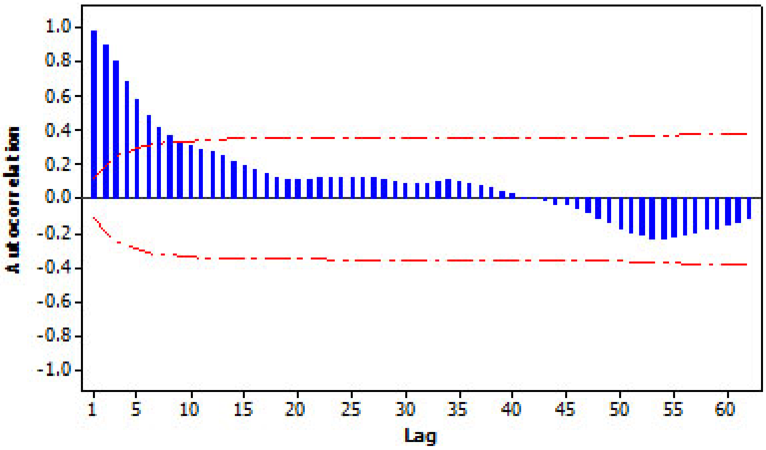

Fig. 1: Autocorrelation function of the output series y_t with 5% significance limits

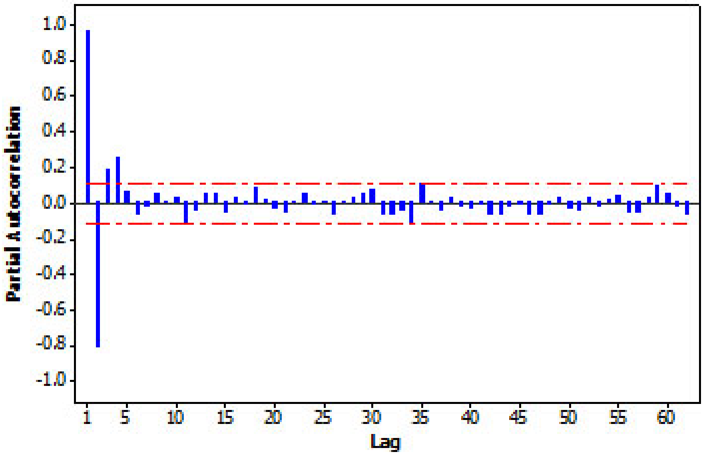

Fig. 2: Partial autocorrelation function of the output series y_t with 5% significance limits

contain redundant information and only one of the two should be selected as input. Based on this conclusion model M3 was built with inputs x_{t-1}, y_{t-1}, y_{t-2} and y_{t-4} (see Table 1).

According to the proposed approach the inputs to models M3 are the suitable inputs to the neural network model to represent the gas furnace time series.

The model M3 is given by the following difference equation:

$$y_t = -0.009\, x_{t-1} + 2.2070 y_{t-1} - 1.406 y_{t-2} + 0.2590 y_{t-4} \quad ...(14)$$

Best neural network model using the old technique

The best neural network (NN) model is obtained as follows

The training samples were split equally into 2 subsets, the first one is used for training the neural network while the second is used to validate the neural network to assure its generalization capabilities. The best architecture found (with respect to the validation set) is the recurrent neural network with one hidden layer consisted of 2 nodes with sigmoid transfer function and an output layer with linear transfer function. Several NN's were trained with different inputs, selected as $y_{t-1},\ldots,y_{t-r}$ and $x_t,\ldots x_{t-m}$ where r is varied from 1 to 4 and m is varied from 0 to 3 generating 16 models to train. For every model 50 NN's with different initial weights

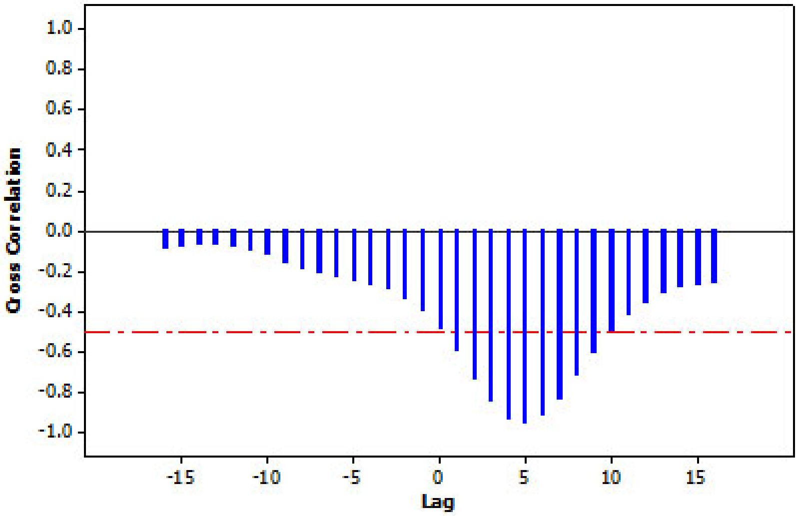

Fig. 3: Cross correlation function of the input and output series

Table 1: Diagnostic test for M3

Predictor	Coefficient	S.E.	p-value
x_{t-1}	-0.0090	0.0045	0.0485
y_{t-1}	2.2070	0.0440	≈ 0
y_{t-2}	-1.4606	0.0626	≈ 0
y_{t-4}	0.2590	0.0235	≈ 0

(randomly selected) have been trained and the best performer over the validation set is selected as the best NN. All NN's were trained using conjugate gradient algorithm ("traincgb" in Matlab) for 1000 epochs. The weights for each NN are stored at each epoch during the training process. The weights for each NN are selected as the ones corresponding to the minimum validation error along the 1000 epochs. The best NN obtained has inputs x_t, x_{t-1}, y_{t-1}… y_{t-4}.

DISCUSSION

The proposed approach for the selection of inputs to the neural network for multivariate time series suggested fewer inputs than those found in by the old technique as the inputs to the best neural network model. Unlike the previous custom of fitting a large number of inputs to the neural network or trying different inputs combinations, the proposed approach based on the correlations analysis reduces the choice of the inputs to a limited number of candidates. This approach indicates the proper inputs that really affect the output series, consequently fewer weights are calculated and computer time and effort are minimized. Most important of all we obtain a better mathematical model to fit the time series and hence better forecasts.

The performance of the neural network models built for the gas furnace data is tested in forecasting the last 96 output readings. A comparison is performed involving the models built by the different techniques and the ARMA model obtained based on Box-Jenkins transfer function approach[3] in forecasting the same output readings in Table 2.

The comparison measure is the mean squared error of the one-step ahead prediction:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(y_1^i - y_2^i\right)^2 \quad ...(15)$$

where N is the number of samples , y_1^i and y_2^i are the actual and model outputs of the i-th sample.

The new developed procedure outperformed previously used forecasting techniques. In a way to benefit from the above two

Table 2: One-step ahead prediction error for test set. NN1: Best NN with inputs x_t, x_{t-1} and y_{t-1},…, y_{t-4}. NN2: Best NN with inputs x_{t-1}, y_{t-1}, y_{t-2}, y_{t-4} (Proposed Approach)

Model	*MSE*
Box-Jenkins Transfer Function model	0.2020
AR model of the output series	0.1763
NN1	0.1512
Linear Regression ofNN1 and AR (NN1-AR)	0.1458
NN2	0.1448
Linear Regression ofNN2 and AR (NN2-AR)	0.1426

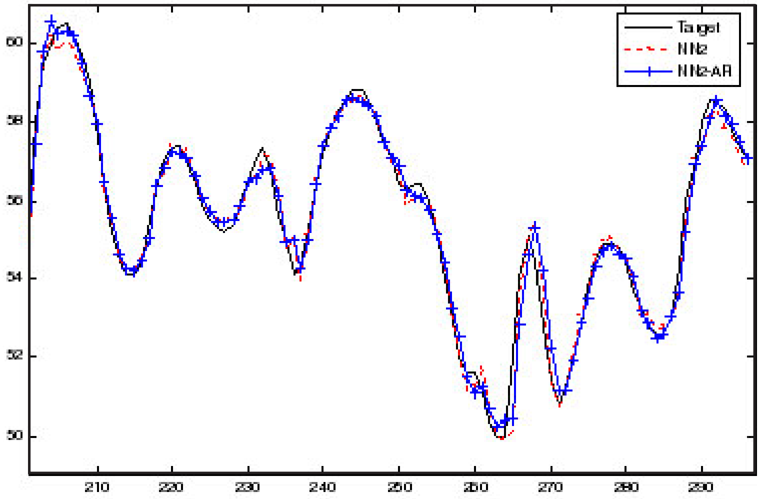

Fig. 4: One-step ahead prediction of NN2 and the combined model of NN2 with AR for the gas furnace time series

methods to provide a better estimate of the output series, a combined model is built by the linear combination of the outputs of the neural network and the autoregressive model using linear regression (LR). Fig. 4 shows the performance of the neural network model with the selected inputs and the LR combined model in forecasting the last 96 samples of the gas furnace data.

CONCLUSIONS

A new approach for the proper inputs selection of the neural network model for multivariate time series is proposed. The approach is based on studying the correlation functions of the output and its dependence on the inputs to select the actual inputs that affect the output. The new procedure narrows the selection of inputs from a large number of past input and output observations to just few candidates. The limited number of models suggested by the approach is tested to insure their adequacy, and then the inputs of the best model are elected for use as inputs to the neural network model for the multivariate time series. Unlike the previous custom of fitting a large number of inputs to the neural network, the proposed approach based on the correlations analysis reduces the choice of the inputs to a limited number of candidates. This approach indicates the proper inputs that really affect the output series, consequently fewer weights are calculated and computer time and effort are minimized. Most important of all we obtain a better mathematical model to fit the time series and hence better forecasts.

REFERENCES

1. Young, P., “Some Observations on Instrumental Variable Methods of Time Series Analysis”, *International Journal of Control,* **23**(5): 593-612 (1976).
2. Montgomery, D. C., L. A. Johnsons, J. S. Gardiner, Forecasting and Time Series Analysis, McGraw-Hill (1990).
3. Box, G. E. P., and G. M. Jenkins, Time Series Analysis- Forecasting and Control, Holden Day (1976).
4. El-Sherif, H., and N. K. Sinha, "Identification and Modeling for Linear Multivariable Discrete-Time Systems: A Survey", J*ournal of Cybernetics*, **9**: 43-71 (1979).
5. Weigend, A. S. and N. A. Gershenfeld, Time Series Prediction: Forecasting the Future and Understanding the Past, Addison-Wesley, Reading (1994).
6. Haykin, S., "Neural Networks: A Comprehensive Foundation", Maxwell MacMillan Canada (1994).
7. Suykens, J. A. K. and J. Vandewalle, Non Linear Modeling, Advanced Black-Box Techniques, Kluwer Academic Publishers, Boston (1998).
8. Ashour, Z. H., "Inputs selection of Artificial Neural Networks using time series Analysis", Proceedings of the fourth IEEE International: Conference on *Electronics, Circuits, and Systems ICECS,* **1**: 19-22 (1997).
9. Ashour, Z. H., "Artificial Neural Network Models for Forecasting Ozone Data", The thirty annual conference ISSR, Cairo University, **30**(3): 83-96, Cairo, Egypt (1995).
10. Hashem, S., Z. H. Ashour, E. F. Agawad, , and A. A. Hakeem, "A Novel Approach for Training Neural Networks for Long Term Prediction", Proc. of the IEEE International Joint Conference on Neural Networks (IJCNN99), 1594-1599, IEEE Press (1999).
11. Ashour, Z. H., S. Hashem and A. Abdel Hakeem Said, "Modeling and Forecasting Ozone Data using Recurrent Neural Networks", The 3rd ICEMP, 290-293, Faculty of Engineering, Cairo University, Dec (1997).
12. Cooper, L., "Hybrid Neural Network Architectures: Equilibrium Systems that Pay Attention", In R.J. Mammone & Y.Y. Zeevi (Eds), Neural Networks: Theory and Applications. New York: Academic Press, pp. 81-96 (1991).
13. Ashour, Z. H., S. R. Hashem and H. A. Fayed, "A New Approach for Combining Neural Networks During Training for Time Series Modeling", *International Review of Electrical Engineering IREE*, **2**(5): 745-750 (2007).
14. Ashour, Z. H., S. Hashem and A. S. A. Mohamed, "Short-term Electric Load Forecasting using Mixtures of Neural Networks and ARIMA Models", Proceedings of LESCOPE-00, Large Engineering Systems Conference On Power Engineering, Halifax, Nova Scotia, Canada, **23-26**: 198-202 (2000).
15. Conner, J. T., R. D. Martin, and L. E. Atlas, "Recurrent Neural Networks and Robust Time Series Prediction". IEEE Transactions on Neural Networks, **5**(2): 240-253 (1994).
16. Dwinell, W, "Modeling Methodologys: Model Input Selection", PCAI, **12**(1): 23-26 (1998).
17. El-Fandy, M. G., Z. H. Ashour and S. M. M. Tail, "Time Series Models Adoptable for Forecasting Nile Floods and Ethiopian Rainfalls", *Bulletin of the American Meteorological Society,* **75**(1): 1-12 (1994).
18. Guidorzi, R. P., M. P. Losito and T. Muratori, "The Range Error in the Structural Identification of Linear Multivariable Systems", *IEEE Transaction on Automatic Control,* AC-27 No **5**: 1044-1054 (1982).