'94 KACC (1994. 10. 17 ~ 20)

Comparison of the Traditional and the Neural Networks Approaches

Kil To Chong *
Yeungnam University
Kyungsan, Korea

Alexander G. Parlos
Texas A&M University
College Station, TX

Abstract:
In this paper the comparison between the neural networks and traditional approaches as system identification method are considered. Two model structrues of neuarl networks are the state space model and the input output model neural networks. The traditional methods are the AutoRegressive eXogeneous Input model and the Nonlinear AutoRegressive eXogeneous Input model. The examples considered do not represent any physical system, no a priori knowledge concerning their structure has been used in the identification process. Testing inputs for comparison are the sinusoidal, ramp and the noise ramp.

Key Words:
System Identification, Neural Networks, AutoRegressive eXogeneous Input Model, Nonlinear AutoRegressive eXogeneous Input Model

1. Introduction

Traditional and the neural network approaches to nonlinear system identification will be considered in this paper. In this paper case studies are presented comparing conventional and biologically motivated (Artificial Neural Network based) model structures and parameter estimation algorithms.

Indeed, an attempt has been made to present an objective treatment by using traditional model structures and parameter estimation algorithms which are popular and widely available in commercial software packages, while not considering approaches requiring significant effort to code and considerable problem-specific tuning. Review of the literature reveals the existence and availability of popular traditional model structures for linear system identification[1], whereas algorithms for traditional nonlinear input-output model structures are not easily available [2].

There have been very few reported studies on the use of nonlinear system identification approaches for improving the relative accuracy of traditional linear model structures, hence this has not yet resulted in the wide acceptance of any one particular approach. On the other hand, it is believed that ANN based model structures offer quite a general framework for identifying nonlinear systems with very few tuning parameters [3].

In testing the various approaches in this paper, the focus has been on dynamic systems with structurally unknown nonlinearities. Thus, even though the examples considered do not represent any physical system, no a priori knowledge concerning their structure has been used in the identification process.

2. Traditional Approaches Considered

Three traditional model structures the Auto-Regressive with eXogeneous Input model(ARX) and the Nonlinear ARX(NARX) are considered. These are the model structures chosen for comparison with the biologically inspired model structures presented in the later sections. This, however, must not be considered an exhaustive study. Our aim has been to present a comparison with what appear to be the most prominent traditional model structures, without fine-tuning them for specific applications. We define as traditional model structures parameterization which have not borrowed ideas from developments in the neurobiological disciplines.

In the traditional nonlinear system identification literature, results for two major problem categories have been reported: (1) structure identification of nonlinear dynamic systems, and (2) parameter estimation of an assumed nonlinear structure. Results for structure identification of nonlinear systems are scarce but for the reported simple cases they appear to be encouraging. In a recent survey paper on structure identification of nonlinear systems, encouraging results on the structure detection of systems with linear dynamics and nonlinear output functions have been presented[4]. The major difficulty with the reported approaches has been the large number of possible model structure combinations, and the lack of a systematic procedure to effectively narrow-down the available alternatives. Even though a number of recent results have been reported for the second problem category, the somewhat complex nature of the parameter estimation algorithms for nonlinear model structures have limited their acceptance. There does not yet appear to be a widely accepted nonlinear structure and an associated parameter estimation algorithm as is the case, for example, with the Auto-Regressive Moving Average(ARMA) representation and the linear least squares estimation. Furthermore, the commercially available system identification software packages do not yet include options for nonlinear structures and, as witnessed from this study, any attempt to code such algorithms appears to require significant effort and computational resources. Therefore in this work the algorithms available in the $MATLAB^{TM}$ System Identification Toplbox have been utilized for comparing the traditional linear system identification structures. Comparisons with the traditional nonlinear system identification structures was accomplished via software implementation of the cited algonthms [5].

2.1 Auto-Regressive with eXogenous Input (ARX) Model Structure

One of the simplest input-output model structures selected belongs to the class of black-box models, resulting from the assumption that the function f(·) in equation (1) is a linear combination of past observations, that is an AutoRegressive with eXogeneous input (ARX) model of equation (2).

$$y(k) = f[y(k-1), y(k-2), \cdots, y(k-n_y), u(k-1), u(k-2), \cdots, u(k-n_u)] + e(k). \quad (1)$$

$$y(k) + A_1 y(k-1) + \cdots + A_{ma} y(k-n_a) = B_1 u(k-1) + \cdots + B_{nb} u(k-n_b) + e(k) \quad (2)$$

The input-output delay n_k, present in an ARX structure and determined by trial and error at the model structure selection stage, can be chosen to best fit the data. The ARX model structure is one of the most widely used model structures in the system identification community. Once the parameters n_a, n_b and n_k are chosen, then the coefficient of the ARX model can be determined using, for example, the least squares estimation algorithm. This is accomplished by solving an overdetermined set of linear equations [1]. For simplicity, the ARX model structures used in this study will be denoted by (n_a, n_b, n_k).

2.2 Polynomial Nonlinear Auto-Regressive with eXogenous Input (NARX) Model Structure

Each output component of the MIMO NARX model structure depicted by equation (3) can also be represented by equation (4).

$$y(k) = f[y(k-1), \cdot, y(k-n_y), u(k-1), \cdot, u(k-n_u)] + e(k). \quad (3)$$

$$y_i(k) = \theta_o^{(i)} + \sum 1^{n}_{j1=1} \sum_{j2=j1}^{n} \theta_{j1j2}^{(i)} x_{j1}(k) x_{j2}(k) + \cdots + \sum_{j1=1}^{n} \cdots \sum_{jl=jl-1}^{n} \theta_{j1 \cdots jl}^{(i)} x_{j1}(k) \cdots x_{jl}(k) + e_i(k), \quad i=1, \cdots, m \quad (4)$$

$$x_1(k) = y_1(k-1), x_2 = y_1(k-2), \cdots, x_{m \times n_y}(k) = y_m(k-n_y), \quad (5)$$
$$x_{m \times n_y + 1}(k) = u_1(k-1), \cdots, x_n(K) = u_r(k-n_u).$$

To complete a NARX model, the parameters, θ_{ij}, multiplying the monomials in the expansion (5) must be estimated. It should be noted that even though the utilized model structure is nonlinear the parameterization is linear in the parameters. The forward-regression orthogonal parameter estimation algorithm, a least-squares estimator with a model structure selection criterion, reported by Billings et al[6]. has been used to identify models with NARX structure. As indicated by Billings *et al.* the same algorithm with some modifications can be used to identify NARMAX structures. However, for consistency with the biologically inspired model structures only NARX structures have been considered in this paper. Further details on the reasons for implementing a structure selection criterion in addition to parameter estimation can be found in a number of papers by Billings et al. [5], [6].

3 Neural Networks Approaches Considered

3.1 Feedforward Multilayer Perceptron Model Structure

One of the model structures that has been motivated by the resurgence of ANNs is that of an Feedforward Multilayer Perceptron(FMLP) with or without teacher forcing. In this model structure, past observations are used for the teacher forcing FMLP (TFFMLP), and past estimates are used for the recurrent FMLP (RFMLP), in the approximation of function $f_i(\cdot)$ in equation (6).

$$y_i(k) = f_i(y(k-1), \cdots, y(k-n_y), u(k), u(k-1), \cdots, u(k-n_u)) + e_i(k), \quad (6)$$

It is assumed that the input and the output layers have linear discriminatory functions and no biases. The inputs to the first layer, i.e. the inputs to the network, can be defined by the vector (7). Considering the special structure of the input and output layers, and in view of equation (8), the input-output equations for a single hidden layer network can be expressed by equations (9) and (10). equation (9) and (10) can be combined in the compact form (11), which is in the form of a NARX model structure.

$$x_{[1]}(k) = [y(k-1), \cdot \cdot, y(k-n_y), u(k), u(k-1), \cdot \cdot, u(k-n_u)]^T. \quad (7)$$

$$x_{[l,i]}(k_l) = F_{[l]}\left(\sum_{j=1}^{N(l-1)} w_{[l-1,j][l,i]} x_{[l-1,j]}(k) - b_{[l,i]} \right). \quad (8)$$

$$x_{[2,j]}(k) = F_{[2]}\left(\sum_{n=1}^{N(1)} w_{[1,n][2,j]} x_{[1,n]}(k) + b_{[2,j]} \right) \quad (9)$$

$$f_i(\cdot) = x_{[3,i]}(k) = \sum_{j=1}^{N(2)} w_{[2,j][3,i]} x_{[2,j]}(k) \quad (10)$$

$$y_i(k) = f_i(y(k-1), \cdot \cdot, y(k-n_y), u(k), u(k-1), \cdot \cdot, u(k-n_u)) + e_i(k), \quad (11)$$

The same argument can be extended to a network with multiple hidden layers, and therefore a TFFMLP and an RFMLP network can be considered as a NARX model structure of the form depicted by equation (11). This ANN is used as a nonlinear, in put-output black-box model structure.

3.2 Recurrent Multilayer Perceptron Model Structure

This model is based on the Recurrent Multilayer Perceptron(RMLP) which has been reported by Chong[7]. The RMLP model structure allows for feedforward links among the nodes of neighboring layers, and recurrent and cross-talk links within the hidden layers which carry time delayed signals. If additionally the observations are provided to the input layer, the model structure becomes a teacher forcing RMLP (TFRMLP) model, otherwise if the estimates are fed back then it is a globally recurrent RMLP (GRRMLP) model structure. The nodes of the GRRMLP and TFRMLP network are both governed by the equations (12) and (13).

$$z_{[l,i]}(k) = \sum_{j=1}^{N(l)} w_{[l,j][l,i]} x_{[l,j]}(k-1) + \sum_{j=1}^{N(l-1)} w_{[l-1,j][l,i]} x_{[l-1,j]}(k) + b_{[l,i]} \quad (12)$$

$$x_{[l,i]}(k) = F_{[l]}(z_{[l,i]}(k)), \quad (13)$$

The input and output layers have linear discriminatory functions and no biases, no recurrency, and no cross-talk. The inputs to the first layer of the GRRMLP and TFRMLP are defined by the following vectors:

$$x_{[1]}(k) = [u_1(k-1), u_2(k-1), \cdots, u_{N(1)}(k-1), \widehat{y}_1(k-1), \cdots, \widehat{y}_{1N(L)}(k-1)]^T, \quad (14)$$

$$x_{[1]}(k) = [u_1(k-1), u_2(k-1), \cdots, u_{N(1)}(k-1), y_1(k-1), \cdots, y_{N(L)}(k-1)]^T, \quad (15)$$

respectively.

For a single hidden layer GRRMLP and TFRMLP, the input-output equations can be expressed as equations (16) and (17). Equations (16) and (17) again can be rewritten compactly as equations (18) and (19). Equations (18) and (19) however, are in the state-space form of equations (20) and (21), though the state vector x(k) defined in equation (22) consists of artificial states, characterizing this empirical state-space model structure.

$$x_{[2,j]}(k) = F_{[2]}\left(\sum_{n=1}^{N(2)} w_{[2,n][2,j]} x_{[2,n]}(k-1) + \sum_{p=1}^{N(1)} w_{[1,p][2,j]} x_{[1,p]}(k) + b_{[2,j]} \right), \quad (16)$$

$$x_{[3,i]}(k) = \sum_{j=1}^{N(2)} w_{[2,j][3,i]} x_{[2,j]}(k) \quad (17)$$

$$x(k) = g(x(k-1), x_1(k)) = g(x(k-1), u(k-1)) \quad (18)$$

$$y_i(k) = x_{[3,i]}(k) + e_i(k) = W^i_{2,3} x(k) + e_i(k) = h_i(x(k), e_i(k)), \quad (19)$$

$$x(k+1) = f(k, x(k), u(k), w(k); \theta), \quad (20)$$

$$y(k) = h(k, x(k), u(k), v(k); \theta), \quad (21)$$

$$x(k) = [x_{[2,1]}(k), \cdots, x_{[2,N(2)]}(k)]^T \quad (22)$$

The same arguments can be extended to a network with multiple hidden layers, and therefore, a GRRMLP and a TFRMLP network can be considered as an empirical state-space model structure of the form depicted by equations (20) and (21).

4 Case Studies

The example presented in this paper is for demonstrating the system identification capabilities of several conventional and neural networks model structures. In identifying models for these systems, however, an attempt has been made to use only information that would be available when investigating a complex nonlinear system. Therefore, in this study no information about the system order and the nature or severity of the nonlinearities being identified has been explicitly used in choosing the structure and size of the neural network or of the conventional model structure.

There are some additional general comments which can be offered, applicable to all of the examples presented in this study. The relative MSE have been calculated using the following equation:

$$MSE(e_i) \equiv \frac{Mean\text{-}Squared\ Error}{Target\ Mean\text{-}Squared\ Deviation} = \frac{\sum_{k=1}^{NP}(x_{[L,i]}(k) - y_i(k))^2}{\sum_{k=1}^{NP}(y_i(k) - \bar{y}_i)^2} \quad (23)$$

When selecting the data set, it is important to consider the relative mix of steady state versus transient response.

As an example the following SISO system is identified:

$$y(k) = 0.8y(k-1) + u^3(k) \quad (24)$$

The data set consists of the following signals: a number of steady-state samples with the input and output at 0, 10 steps starting from 0, with magnitudes 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, respectively, 10 ramps of slopes 0.01, 0.02, 0.06, 0.125, 0.175, 0.2, 0.225, 0.25, 0.3, 0.45, respectively, starting from 0 and ending at 1, and 10 ramps with slopes -0.01, -0.02, -0.06, -0.125, -0.t75, -0.2, -0.225, -0.3, -0.375, -0.5, respectively, starting from 1 and ending at 0. Each of the step and ramp signals in the training set contains 20 samples, for a training set total of 620 samples.

Three tests are performed with signals unknown during identification, for investigating the predictive performance of the model. The first test signal was of the following sinusoidal form: $0.6 + 0.3\sin(\frac{\pi t}{40})$.

The second test signal consisted of a sequence of ramps., as shown in Figure (6). Additionally, white Gaussian noise with zero mean and 0.1 standard deviation was superimposed on the same sequence of ramps and used as a test signal.

Since the analytical model is a NARX model structure, the parameter estimation algorithm for ARX is used to obtain a model with the System Identification Toolbox available on the commercial package $MATLAB^{TM}$. The first step towards modeling the system is obtaining a specific model structure, i.e. the number of past inputs and outputs to be used. The subroutines called ARXSTRUC and SELSTRUC are available in the System Identification Toolbox for this purpose. So the selected model structures contained 5 past outputs, 6 past inputs including a feedforward term i.e. n_a=5, n_b=6, n_k= 0, and 3 past outputs, 2 past inputs with no feedforward term, i.e. n_a=3, n_b=2, n_k=1. The first model structure is denoted ARX(5,6,0) and the second model structure is denoted ARX(3,2,1) in this study. Parameters are estimated iteratively using subroutine ARX in the System Identification Toolbox, for both ARX(5,6,0) ARX(3,2,1). The obtained ARX(3,2,1) model is as follows:

$$\begin{aligned} y(k) = {} & 1.831y(k-1) - 1.1706y(k-2) + 0.312y(k-3) \\ & - 0.2077u(k-1) + 0.3022u(k-2). \end{aligned} \quad (25)$$

With the identified models, three tests are performed for investigating the ARX predictive performance. The responses of the ARX(3,2,1) and the analytical model for all the three test signals are shown in Figure 1. The relative MSEs for ARX(3,2,1) model are 0.499, 0.434, 0.449 for sine, ramp and noise ramp inputs, respectively. Also parameters are estimated iteratively using subroutine the ARX, for ARX(5,6,0), until the parameters converged and the obtained model is as follows:

$$\begin{aligned} y(k) = {} & 1.9106y(k-1) - 1.1820y(k-2) + 0.2424y(k-3) \\ & + 0.0745y(k-4) - 0.0590y(k-5) + 0.9908u(k) \\ & - 1.1552u(k-1) + 0.3757u(k-2) \\ & - 0.0433u(k-3) - 0.1015u(k-4) \\ & - 0.0098u(k-5) \end{aligned} \quad (26)$$

The responses of the ARX(5,6,0) and the analytical model for all the three signals are also shown in Figure 2. The relative MSEs computed for the ARX(5,6,0) model are 0.818, 0.435, 0.451 for sinusoid, ramp and noisy ramp inputs, respectively.

The third order NARX model with two past inputs and outputs has 56 terms involved at the beginning of the iterations. However, because this example is also a third order NARX, the exact model can be identified. Hence repeating relative MSEs is not be very meaningful in this case. So a second order NARX structure with four delays is assumed, and the following model is identified:

$$y(k) = 0.80218737y(k-1) + 0.93263932u^2(k) \quad (27)$$

The responses of the second order NARX and the analytical model to all the test signals are shown in Figure 3. The relative MSEs for the NARX are also tabulated in Table 1.

The RMLP used in this example consists of an input layer with 1 node, 2 hidden layers with 5 and 3 nodes, respectively, and an output layer with 1 node. The 1-5-3-1 RMLP network which has 66 connection links was trained for 3800 cycles, where one cycle (iteration) consists of one presentation of the whole data set, using 0.005 learning rate for the weights and 0.001 for the biases. The responses of the RMLP network and of the analytical model are shown in Figure 4. The relative MSEs for the RMLP, also shown in Table 1.

The TFFMLP which has three past outputs and two past inputs, with no feedforward term, and it is denoted by TFFMLP(3,2,1). The TFFMLP(3,2,1) consists of an input layer with 6 nodes, 2 hidden layers with 6 and 4 nodes, respectively, and an output layer with 1 node. The 6-6-4-1 TFFMLP(3,2,1) network has 69 connection links. The architecture was chosen based on other model parameters, such as the order, the delays and the number of the unknown parameters of the ARX(5,6,0), ARX(3,2,1) and RMLP. The TFFMLP(3,2,1) was trained for 45000 cycles using varied learning rates varying from 0.01 to 0.00125, for the weights and the biases. The responses of the TFFMLP(3,2,1) network and of the analytical model for the test signals are shown in Figure 5. The relative MSEs for the TFFMLP(3,2,1) also shown in Table 1. A TFFMLP(5,6,0) network with 5 past past outputs, 6 past inputs, and a feed-forward term is also used. The TFFMLP(5,6,0) consists of an input layer with 11 nodes, 2 hidden layers with 3 and 7 nodes, respectively, and an output layer with 1 node. The 11-3-7-1 TFFMLP(5,6,0) network has 75 connection links, and it was trained for 45000 cycles using varied learning rates varying from 0.01 to 0.00125 for the weights and for the biases, which is same as in the case of the TFFMLP(3,2,1) network. The responses of the TFFMLP(5,6,0) network and the analytical model for three test signals are shown in Figure 6. The relative MSEs for the TFFMLP(5,6,0) are also shown in Table 1.

5. Discussion

Traditional and biologically inspired model structures are compared for their effectiveness to identify nonlinear systems. The ARX, the NARX are the conventional model structures used in the comparisons. The FMLP and the RMLP with and without teacher forcing are the biologically motivated nonlinear model structures. For the identification of an ARX model structure the System Identification Toolbox has been used which is available in the commercial software package $MATLAB^{TM}$. However, an algorithm for estimating the parameters of the NARX model structure has been programmed and used for this comparison. Comparisons of the chosen model structures is accomplished through a number of examples. The responses of the identified models are obtained for three different test signals unknown during the identification process. Relative mean squared errors are calculated for the numerical comparison. For a sake of a fare comparison, the parameters such as the orders, the number of delays, the number of connection links, and the number of iterations are carefully chosen for all model structures.

From the deterministic numerical simulations, it is possible to postulate that the NARX, the RMLP and the FMLP models are good candidate structures for nonlinear system identification. However, the FMLP model structures is not as effective as the RMLP and the NARX models. Further works will be done for more complicate systems like MIMO and even for stochastic environments. Thus we can further compare the NARX and the RMLP structures in a better way.

Reference

[1] L.Ljung, *System Identification Theory for the Users*, Prentice Hall, 1987.

[2] S.Chen and S.S.Billings, "Modeling and Analysis of Nonlinear Time Series," *Int. J. Control*, Vol. 50, No. 6, pp. 2151-2171, 1989.

[3] K.S.Narendra and K.Parthasarathy, "Neural Networks in Control Systems," Workshop on Neural Networks in Control Systems, *American Control Conference*, Workshop Manual, 1991.

[4] R.Harber, and H.Unbehauen, "Structure Identification of Nonlinear Dynamic Systems-A Survey on Input/Output Approaches," *Automatica*, Vol.26, No 4-A, pp.651-677, 1990.

[5] S.Chen, and S.S.Billings, "Orthogonal Least Square Methods and their Application to Nonlinear System Identification, " *Int. J. Control*, Vol. 50, No.5, pp.1873-1896, 1989.

[6] S.S.Billings, S.Chen and Korenberg, Identification of MINO Nonlinear Systems Using a Forward-regression Orthogonal Estimator," *Int.J. Control*, Vol. 49, pp. 2157-2189, 1989.

[7] Kil To Chong, ''Nonlinear System Identification Using Recurrent Neural Networks'', PhD Dissertation, Texas A&M University, College Station, Tx, May 1993.

Table 1. Relative Mean-Squared-Errors.

Model Structure	Sinusoidal Input	Ramp Input	Noise Ramp Input
ARX(3,2,1)	0.499E+00	0.434E+00	0.449E+00
ARX(5,6,0)	0.818E+00	0.435E+00	0.451E+00
NARX	1.249E-02	3.308E-03	5.023E-03
RMLP	1.995E-03	3.255E-03	1.160E-02
TFFMLP(3,2,1)	5.116E-03	1.656E-02	0.161E+00
TFFMLP(5,6,0)	6.847E-04	1.938E-03	2.134E-02

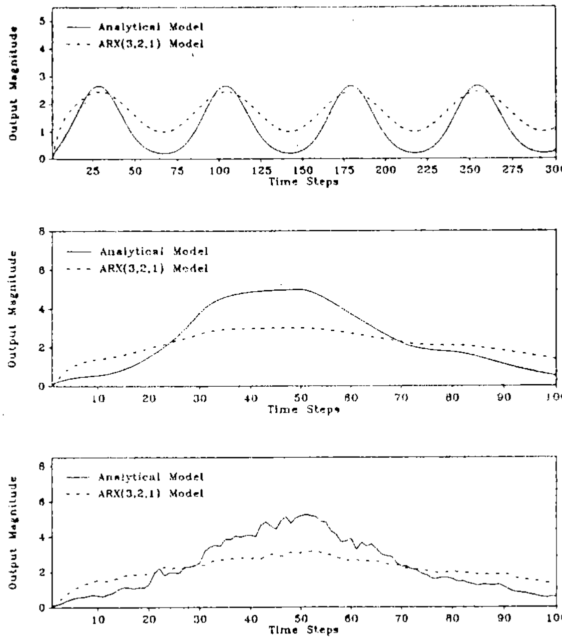

Figure 1. Response for ARX(3,2,1) Model: Top: Sinusoidal Input: Middle: Ramp Input; Botton: Ramp Input Noise Added.

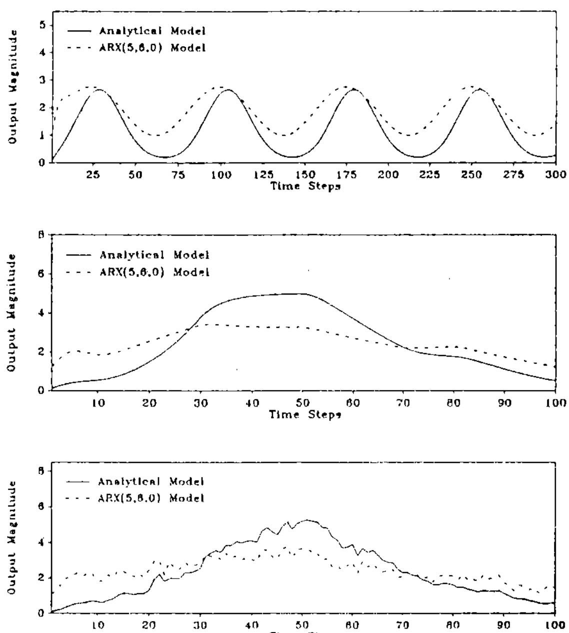

Figure 2. Response for ARX(5,6,0) Model: Top: Sinusoidal Input: Middle: Ramp Input: Botton: Ramp Input Noise Added.

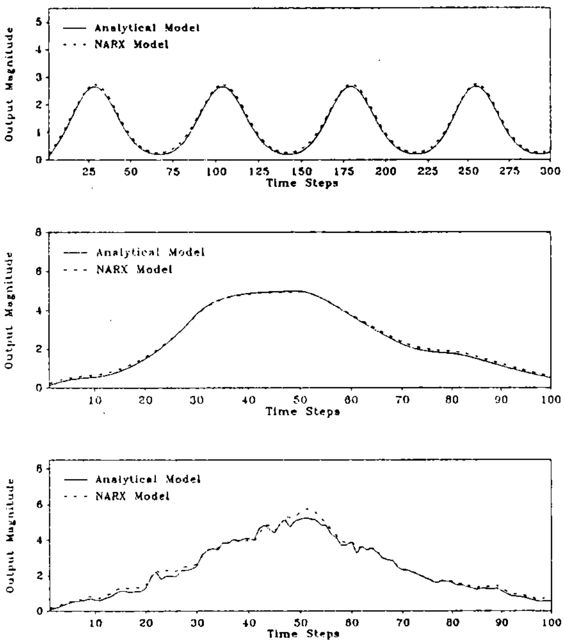

Figure 3. Response for NARX Model: Top: Sinusoidal Input: Middle: Ramp Input: Botton: Ramp Input Noise Added.

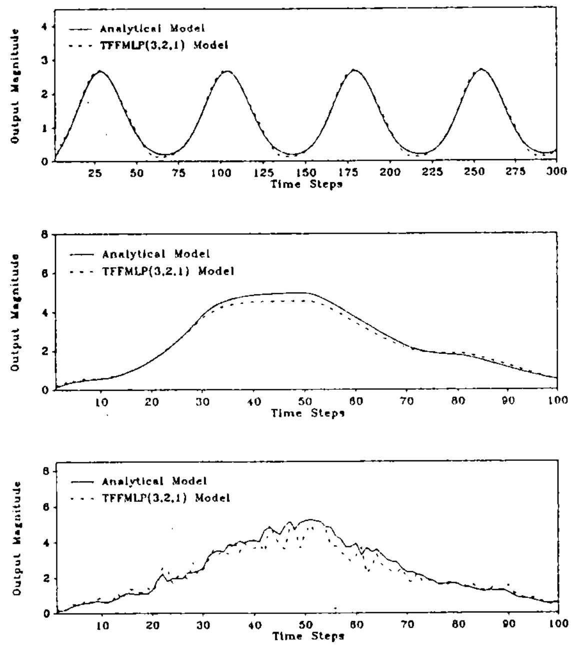

Figure 5. Response for TFFMLP(3,2,1) Model: Top: Sinusoidal Input: Middle: Ramp Input: Botton: Ramp Input Noise Added.

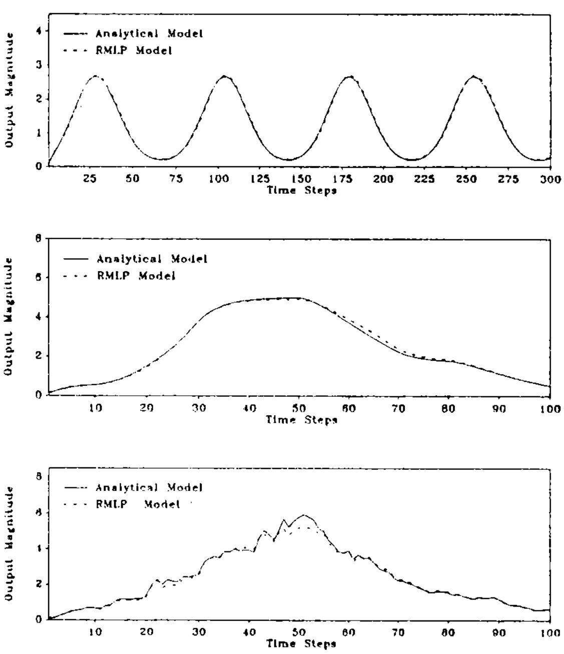

Figure 4. Response for RMLP Model: Top: Sinusoidal Input: Middle: Ramp Input: Botton: Ramp Input Noise Added.

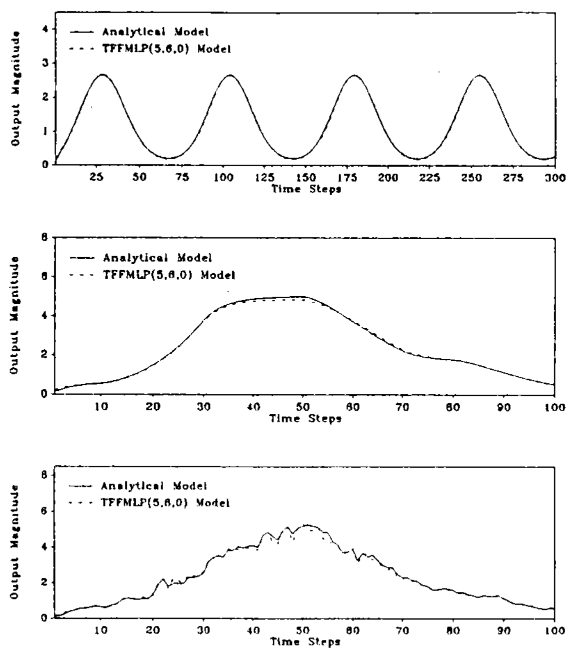

Figure 6. Response for TFFMLP(5,6,0) Model: Top: Sinusoidal Input: Middle: Ramp Input: Botton: Ramp Input Noise Added.