Comparison of Traditional and Neural Network Approaches to Stochastic Nonlinear System Identification

Kil To Chong* and Alexander G. Parlos**

(Received July 8, 1996)

A comparison between neural network and traditional approaches to nonlinear system identification is investigated with respect to aspects of model performance. Two neural network models, the state space and input-output model structures, are considered. A global recurrent RMLP and a teacher forcing RMLP are categorized as the state space models, and a global feedback FMLP and a teacher forcing FMLP are considered as the input-output models. In the traditional methods an AutoRegressive eXogeneous (ARX) Input model and a Nonlinear AutoRegressive eXogeneous (NARX) Input model are considered. Basic algorithms of models are described, and simulation results are also presented through the system output response. Performance of models is compared based on the Mean-Square-Errors (MSE). Noise-added sinusoidal, pulse and step signals are chosen as the test inputs for the validation of the obtained models. Two different noise levels are augmented to the chosen input signals.

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

1. Introduction

Traditional and neural network approaches to nonlinear system identification methods will be considered in this paper. Case studies are also presented for 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 comparison by using traditional model structures and parameter estimation algorithms that 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 (Ljung, 1987), whereas algorithms for traditional nonlinear input-output model structures are not easily available (Chen and Billings, 1989a).

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, as of yet there is no widespread acceptance of any one particular approach. On the other hand, it is believed that Artificial Neural Network (ANNs) based model structures offer quite a general framework for identifying nonlinear systems with very few tuning parameters (Naredra and Parthasarathy, 1991).

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 apriori knowledge concerning their structure has been used in the identi-

* Chon Buk National University, School of Electrical Engineering
** Texas A&M University, Department of Nuclear Engineering

fication process.

2. Traditional Approaches

Two traditional model structures of the Auto-Regressive with eXogeneous Input model (ARX) and the Nonlinear ARX (NARX) are considered. These are the model structures chosen for comparison with 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 parameterizations that 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 (Harber and Unbehauen, 1990). 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 has 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 linear least squares estimation. Comparisons with traditional nonlinear system identification structures was accomplished via software implementation of the cited algorithms (Chen and Billings, 1989b).

2.1 Auto-regressive with eXogeneous input 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 $f(\cdot)$ in Eq. (1) is a linear combination of past observations, that is an AutoRegressive with eXogeneous input model of Eq. (2):

$$\begin{aligned} y(k) = f[&y(k-1),\ y(k-2),\cdots,y(k-n_y),\\ &u(k-1),\ u(k-2),\cdots,u(k-n_u)]\\ &+e(k) \end{aligned} \tag{1}$$

$$\begin{aligned} &y(k)+A_1y(k-1)+\cdots+A_{na}y(k-n_a)\\ &=B_1u(k-1)+\cdots+B_{nb}u(k-n_b)\\ &\quad+e(k) \end{aligned} \tag{2}$$

where $y(k)$, $u(k)$, $e(k)$ are the output observations, input term, and a white noise term, respectively. n_y is the output time delay and n_u is the input time delay of the system. 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_y, n_u 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 (Ljung, 1987).

2.2 Polynomial nonlinear auto-regressive with eXogeneous input model structure

Each output component of the Multi-Input Multi-Output (MIMO) NARX model structure depicted by Eq. (3) can also be represented by Eq. (4):

$$\begin{aligned} y(k) = f[&y(k-1),\cdots,y(k-n_y),\\ &u(k-1),\cdots,u(k-n_u)]\\ &+e(k) \end{aligned} \tag{3}$$

$$\begin{aligned} y_i(k) = \theta_o^{(i)} &+ \sum_{j_1=1}^{n}\sum_{j_2=j_1}^{n}\theta_{j_1j_2}^{(i)}x_{j_1}(k)\,x_{j_2}(k)+\cdots\\ &+\sum_{j_1=1}^{n}\cdots\sum_{j_l=j_{l-1}}^{n}\theta_{j_1\cdots j_l}^{(i)}x_{j_1}(k)\cdots x_{j_l}(k)\\ &+e_i(k),\quad i=1,\cdots,m \end{aligned} \tag{4}$$

where l is the degree of the polynomial expansion and where $n=m\times n_y+r\times n_u$. It is suggested that an equal number of past input and past output vector components are used in the expansion of the above model. Furthermore, the following expansion terms are defined:

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

To complete a NARX model, the parameters, θ_{ij}, multiplying the monomials in the expansion (4) 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.(Chen and Billings, 1989b), has been used to identify models with NARX structure. As indicated by Billings et al. (Billings, Chen and Korean berg, 1989), the same algorithm with some modifications can be used to identify polynomial Nonlinear Auto-Regressive Moving Average with eXogeneous Input (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.(Chen and Billings, 1989b; Billings, Chen and Korenberg, 1989).

3. Neural Network Approaches Considered

3.1 Feedforward multilayer perceptron model structure

One of the model structures that has been motivated by the resurgence of ANN is that of an Feedforward Multilayer Perceptron(FMLP) reported by Narendra et. al.(1991) 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 Eq. (6):

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

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 Eq. (8), the input -output equations for a single hidden layer network can be expressed by Eqs. (9) and (10). Equations (9) and (10) can be combined in the compact form (11), which is in the form of a NARX model structure:

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

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

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

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

$$y_i(k)=f_i(y(k-1),\cdots,y(k-n_y),\ u(k),\ u(k-1),\cdots,u(k-n_u))+e_i(k) \tag{11}$$

where, $x_{[l]}(k)$ is the input to the neural network, and $x_{[l,i]}(k)$ is the output of the j[th] node at the l[th] layer. The $w_{[l,i][k,j]}$ is the weight connection between i[th] node of the l[th] layer to the j[th] node of the k[th] layer., $b_{[l,i]}$ is bias the j[th] node of the l[th] layer, $F_{[l]}$ is the activation function of the l[th] layer, and $y_i(k)$ is the output of the neural network.

The same argument can be extended to a network with multiple hidden layers. Therefore a TFFMLP and an RFMLP network can be considered as a NARX model structure of the form depicted by Eq. (11). This ANN is used as a

nonlinear, input-output black-box model structure.

3.2 Recurrent multilayer perceptron model structure

This model is based on the RMLP network reported by Parlos et. al.(1994). 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(Chong, 1993). 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 Eqs. (12) and (13):

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

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

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:

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

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

respectively. $\hat{y}_k(k)$ is the estimate of the neural network and $y_k(k)$ is the observation.

For a single hidden layer GRRMLP and TFRMLP, the input-output equations can be expressed as follows:

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

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

for $j=1,\ \cdots,\ N(2)$, and $i=1,\ \cdots,\ N(3)$, where $N(l)$ is the number of nodes in the lth layer. Now defining state vector as:

$$x(k)=[x_{[2,1]}(k),\cdots,x_{[2,N(2)]}(k)]^T \tag{18}$$

and

$$x_1(k)=[x_{[1,1]}(k),\cdots,x_{[1,N(1)]}(k)]^T \tag{19}$$

Equations (16) and (17) again can be rewritten compactly as Eqs. (20) and (21):

$$\begin{aligned} x(k)&=g(x(k-1),\ x_1(k))\\ &=g(x(k-1),\ u(k-1)) \end{aligned} \tag{20}$$

$$\begin{aligned} y_i(k)&=x_{[3,i]}(k)+e_i(k)\\ &=W_{2,3i}x(k)+e_i(k)\\ &=h_i(x(k),\ e_i(k)) \end{aligned} \tag{21}$$

for $i=1,\ \cdots,\ m$, where $g(\cdot)$ and $f(\cdot)$ are general functions and $W^i_{2,3}$ is the i^{th} row vector of the matrix $W_{2,3}$ containing all of the feedforward weights connecting the hidden layer to the output layer. Equations (20) and (21) however, are in the state-space form of Eqs. (22) and (23), though the state vector x(k) defined in Eq. (18) consists of artificial states, characterizing this empirical state-space model structure:

$$x(k+1)=f(k,x(k),u(k),w(k),\theta) \tag{22}$$

$$y(k)=h(k,x(k),u(k),v(k),\theta) \tag{23}$$

The same arguments can be extended to a network with multiple hidden layers. Therefore, a GRRMLP and a TFRMLP network can be considered as an empirical state-space model structure of the form depicted by Eqs. (22) and (23).

4. Computer Simulation

The example presented in this paper is for demonstrating the system identification capabilities of several conventional and neural network 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 Mean Square Error(MSE) have been calculated using the following equation:

$$MSE(\varrho_i) \equiv \frac{\textit{Mean-Squared Error}}{\textit{Target Mean-Squared Deviation}} = \frac{\sum_{k=1}^{NP} (x_{[L,i]}(k) - y_i(k))^2}{\sum_{k=1}^{NP} (y_i(k) - \bar{y}_i)^2} \tag{24}$$

When selecting the data set, it is important to consider the relative mix of the steady state of various transient responses. The analytical nonlinear system considered is stochastic multi -input multi-output system.

A stochastic MIMO nonlinear system is expressed by the following states and outputs:

$$x_1(k) = 0.5(x_1^2(k-1))^{\frac{1}{3}} + 0.3x_2(k-1)x_3(k-1) + 0.2u_1(k-1) + n_{x1}^{pr}(k) \tag{25}$$

$$x_2(k) = 0.5(x_2^2(k-1))^{\frac{1}{3}} + 0.3x_3(k-1)x_1(k-1) + 0.2u_1(k-1) + n_{x2}^{pr}(k) \tag{26}$$

$$x_3(k) = 0.5(x_3^2(k-1))^{\frac{1}{3}} + 0.3x_1(k-1)x_2(k-1) + 0.2u_2(k-1) + n_{x3}^{pr}(k) \tag{27}$$

$$y_1(k) = 0.5(x_1(k) + x_2(k) + x_3(k)) \tag{28}$$

$$y_2(k) = 2(x_1(k))^2 \tag{29}$$

where $n_{x_i}^{pr}(k)$ for $i=1, 2, 3$ is the process noise of the system.

The data set consisted of all possible 25 combinations of steps with magnitudes 0.125, 0.25, 0.375, 0.5, as well as the zero step input for both input channels(See Fig. 1 for the input signals). Each signal consists of 15 samples, so that the total number of the training dataset becomes 375 samples. For simulations tried with this data, however, model validation shows poor performance of the obtained model. The reason for this result is insufficient data/set capacity to represent the system. So five different pulses of suitable

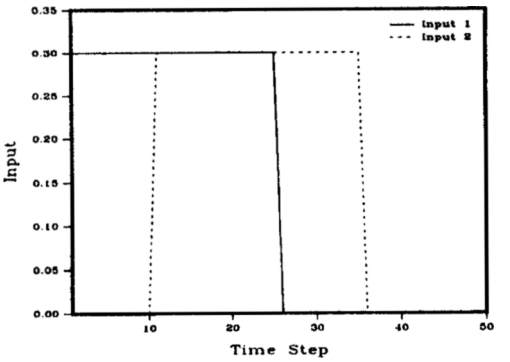

Fig. 1 Pulse inputs

shapes were added to the above data/set, each containing 40 samples. As a result the total number of samples becomes 575 in the training data/set. The process noise is zero mean, white Gaussian noise with 0.02 standard deviation.

Three tests are performed with unknown signals during identification for investigating the models' predictive performance. The first test signal is $u_1(k) = 0.3 + 0.2\sin\left(\frac{\pi k}{8}\right)$, $u_2(k) = 0.2$, where the step input in the second channel is delayed by 5 time steps with respect to the input in the first channel. The pulse inputs used to test the identified models are shown in Fig. 1. The final test set consists of a step augmented with zero mean, white Gaussian noise of 0.1 standard deviation. The magnitudes for the steps are 0.3 and 0.2 for the two input channels, respectively. Each test signal is augmented with zero mean, white Gaussian noise with 0.02 standard deviation for simulating a low noise environment, and 0.08 standard deviation for simulating a high noise environment. For testing the performance of the identified models, the system given by Eqs. (25) through (29) is simulated with zero mean, white Gaussian noise with 0.01 standard deviation (low noise). Performance of each model is measured through the MSEs which are shown in Table 1.

4.1 ARX Model

The ARX model structure used in this example has 406 parameters involved at the initial iteration. The final structure is as follows:

Table 1 Relative mean-squared-errors for models

Model Structure	Sinusoidal Input	Pulse Input	Noise Step Input
ARX(σ=0.02);y_1	3.58E−2	1.53E−2	5.35E−2
ARX(σ=0.02);y_2	0.20E+0	0.13E+0	0.50E+0
ARX(σ=0.08);y_1	0.15E+0	0.14E+0	0.46E+0
ARX(σ=0.08);y_2	0.88E+0	1.04E+0	4.45E+0
NARX(σ=0.02);y_1	3.11E−2	1.55E−2	4.81E−2
NARX(σ=0.02);y_2	9.74E−2	5.62E−2	0.30E+0
NARX(σ=0.08);y_1	0.16E+0	0.16E+0	0.47E+0
NARX(σ=0.08);y_2	0.58E+0	0.63E+0	3.31E+0
GRRMLP(σ=0.02);y_1	7.67E−2	5.66E−2	0.16E+0
GRRMLP(σ=0.02);y_2	6.74E−2	3.42E−2	0.22E+0
GRRMLP(σ=0.08);y_1	7.67E−2	5.66E−2	0.16E+0
GRRMLP(σ=0.08);y_2	6.74E−2	3.42E−2	0.22E+0
TFRMLP(σ=0.02);y_1	4.04E−2	1.80E−2	8.80E−2
TFRMLP(σ=0.02);y_2	8.41E−2	4.46E−2	0.24E+0
TFRMLP(σ=0.08);y_1	0.32E+0	0.19E+0	1.25E+0
TFRMLP(σ=0.08);y_2	0.49E+0	0.34E+0	3.38E+0
RFMLP3(σ=0.02);y_1	4.05E−2	0.36E+0	0.22E+0
RFMLP3(σ=0.02);y_2	6.38E−2	0.19E+0	0.16E+0
RFMLP3(σ=0.08);y_1	4.05E−2	0.36E+0	0.22E+0
RFMLP3(σ=0.08);y_2	6.38E−2	0.19E+0	0.16E+0
TFFMLP3(σ=0.02);y_1	8.50E−2	0.13E+0	0.16E+0
TFFMLP3(σ=0.02);y_2	0.15E+0	9.93E−2	0.45E+0
TFFMLP3(σ=0.08);y_1	0.60E+0	0.63E+0	0.88E+0
TFFMLP3(σ=0.08);y_2	0.97E+0	0.94E+0	2.47E+0
RFMLP5(σ=0.02);y_1	0.22E+0	0.26E+0	0.27E+0
RFMLP5(σ=0.02);y_2	0.15E+0	0.15E+0	0.26E+0
RFMLP5(σ=0.08);y_1	0.22E+0	0.26E+0	0.27E+0
RFMLP5(σ=0.08);y_2	0.15E+0	0.15E+0	0.26E+0
TFFMLP5(σ=0.02);y_1	0.13E+0	0.13E+0	0.21E+0
TFFMLP5(σ=0.02);y_2	0.12E+0	6.14E−2	0.28E+0
TFFMLP5(σ=0.08);y_1	0.45E+0	0.35E+0	0.97E+0
TFFMLP5(σ=0.08);y_2	0.64E+0	0.37E+0	2.05E+0

$$\begin{aligned} y_1(k) = & 0.0801 + 0.6049 y_1(k-1) \\ & + 0.1237 y_2(k-1) \\ & + 0.3448 u_1(k-1) \\ & + 0.2676 u_2(k-1) \end{aligned} \tag{30}$$

$$\begin{aligned} y_2(k) = & -0.0365 y_1(k-38) \\ & - 0.1489 u_1(k-5) \\ & + 0.9409 y_2(k-1) \\ & + 0.3336 u_1(k-1) \end{aligned} \tag{31}$$

The first output response of the identified ARX model and of the reference model for low noise environment test signals are shown in Fig. 2, while the second output responses are shown in Fig. 3. For the high noise environment test signals, the first output responses of the identified ARX model and of the reference model are shown in Fig. 4, and the second output responses are shown

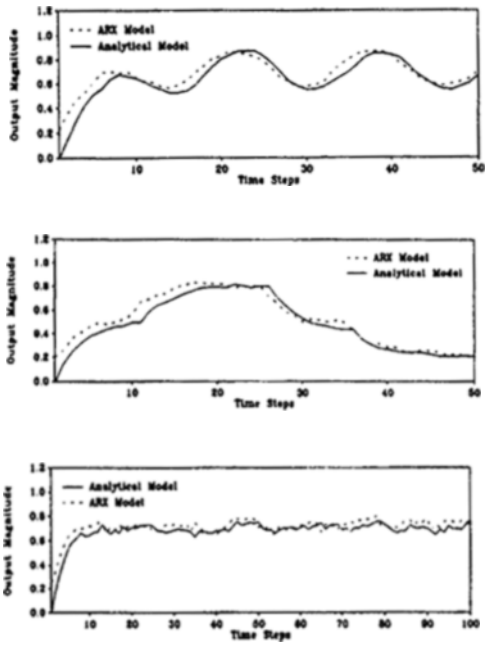

Fig. 2 Response for low noise environment using ARX model first output; top: sinusoidal input; middle: ramp input; bottom: step input.

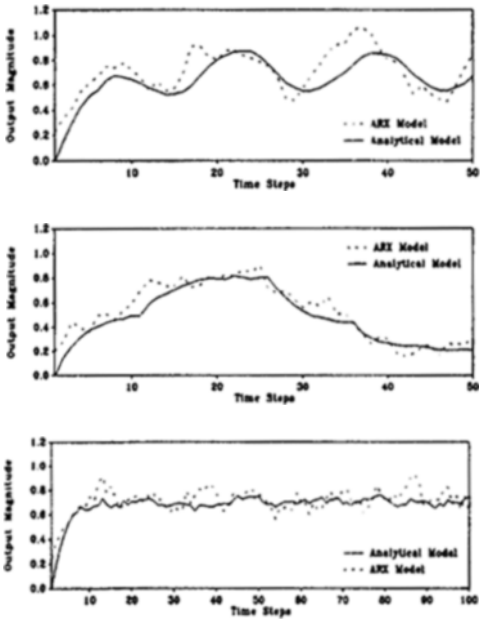

Fig. 4 Response for high noise environment using ARX model first output; top: sinusoidal input; middle: ramp input; bottom: step input.

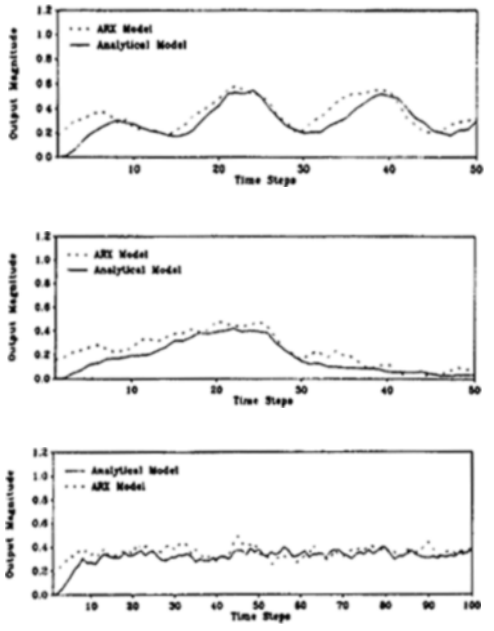

Fig. 3 Response for low noise environment using ARX model second output; top: sinusoidal input; middle: ramp input; bottom: step input.

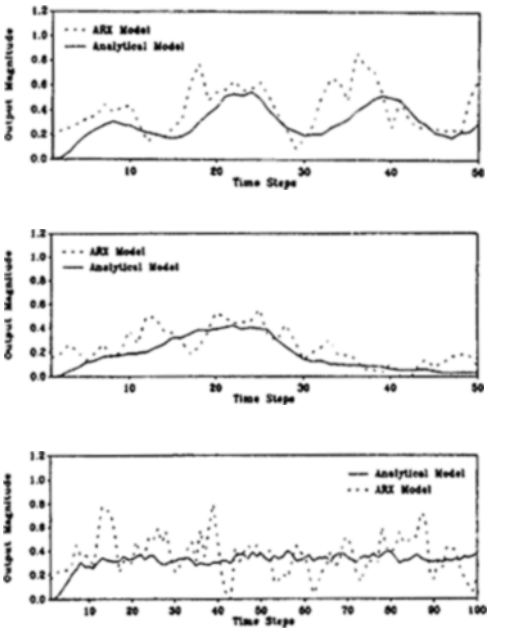

Fig. 5 Response for high noise environment using ARX model second output; top: sinusoidal input; middle: ramp Input; bottom: step input.

in Fig. 5. The relative MSEs for the first output of the ARX model for low noise environment test signals, shown in Table 1, are 3.58×10^{-2}, 1.53×10^{-2}, and 5.35×10^{-2} for the sinusoid, pulse and noisy step inputs, respectively. The relative MSEs for the second output are 0.20, 0.13, and 0.50 for the sinusoid, pulse, and noisy step inputs, respectively. Also, the relative MSEs for the second output of the ARX model for the high noise environment test signal are shown in Table 1. The ARX model gives the highest MSE among the proposed models, which indicates the lowest performance.

4.2 NARX model

The NARX model structure used in this model identification was obtained with the NARX model using the forward-regression orthogonal method. The number of parameters involved at the initial iteration was 695. The model obtained is as follows:

$$y_1(k)=0.0388+0.7505y_1(k-1)+0.33359u_1(k-1)$$

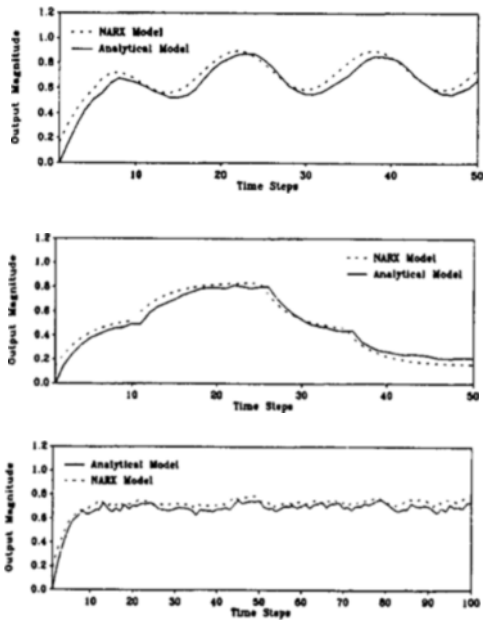

Fig. 6 Response for high noise environment using NARX model first output; top: sinusoidal input; middle: ramp input; bottom: step input.

$$+0.2349u_2(k-1) \quad (32)$$

$$y_2(k)=0.7575y_2(k-1)+0.4536y_1(k-1)u_1(k-1) \quad (33)$$

Every model used in this paper has four different output responses; the first and second output for both the low noise and the high noise case (as in the ARX model). Due to page limit constraints, only the second output of the high noise environment is presented for each model. The first output response of the identified NARX model and of the reference model using the high noise environment test signals are shown in Fig. 6. The relative MSEs of the identified NARX of the low noise environment for the first output are 3.11×10^{-2}, 1.55×10^{-2}, 4.81×10^{-2}, while for the second outputs they are 9.74×10^{-2}, 5.62×10^{-2}, and 0.30 for the test signals. The performance of the NARX model is as good as the RMLP model, and is superior to the other models.

4.3 RMLP models

Because this is a stochastic identification problem, a predictor form of the RMLP must be used. This is an RMLP network with the latest output in its input layer. The sensed or predicted latest output could be used, resulting in two different networks.

If the predicted output is used, then the GRRMLP is obtained. In this example the GRRMLP consists of an input layer with 4 nodes of two inputs and the two outputs, 2 hidden layers with 12 and 10 nodes, respectively, and an output layer with 2 nodes (4-12-10-2). The 4-12-10-2 GRRMLP network was trained for 3600 cycles, using 0.01 learning rate for both the weights and the biases. Training was continued for 1500 cycles using a learning rate of 0.001. The second output responses of the identified GRRMLP model and of the reference model using the high noise environment test signals are shown in Fig. 7. The relative MSEs of the identified GRRMLP of the low noise environment for both the first output are 7.67×10^{-2}, 5.66×10^{-2}, 0.16, and for the second output are 6.74×10^{-2}, 3.42×10^{-2}, 0.22 for the sinusoid, pulse and noisy step inputs, respectively. The relative MSEs for the first output of

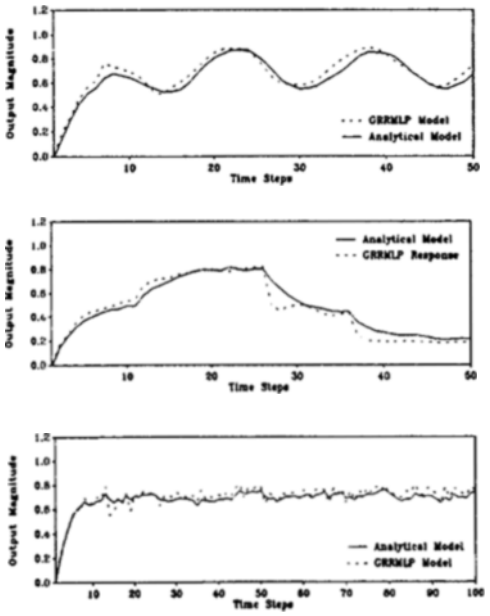

Fig. 7 Response for high noise environment using GRRMLP model first output; top: sinusoidal input; middle: ramp input; bottom: step input.

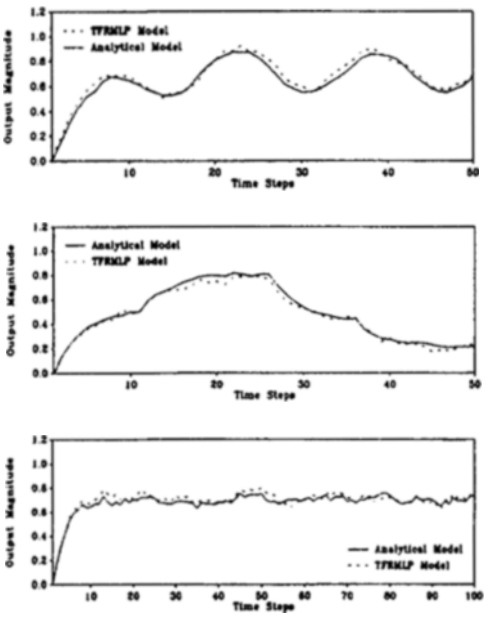

Fig. 8 Response for high noise environment using TFRMLP model first output; top: sinusoidal input; middle: ramp input; bottom: step input.

the identified GRRMLP model for the high noise environment test signals are the same as the ones reported above, because the GRRMLP is not using the sensed but only the predicted outputs.

As previously mentioned, an RMLP with teacher forcing has exactly the same model structure as the GRRMLP, except that instead of the predicted output, the latest sensed output is used in the network input layer. This is the TFRMLP network. The TFRMLP used in this example has the same architecture as the GRRMLP discussed (4-12-10-2). The 4-12-10-2 TFRMLP network was trained for 500 cycles, using 0.01 learning rate for both the weights and the biases, and it was further trained for another 100 cycles with 0.01 learning rate. The second output responses of the identified TFRMLP and of the reference model for high noise environment test signals are shown in Fig. 8. The relative MSEs of the identified TFRMLP of the low noise environment for both the first output are 4.04×10^{-2}, 1.80×10^{-2}, 8.80×10^{-2} and the second outputs are 8.41×10^{-2}, 3.46×10^{-2}, 0.24 for the sinusoid, pulse and noisy step inputs, respectively. Also, the relative MSEs of the identified TFRMLP of the high noise environment for both the first output are 0.43, 0.19, 1.25 and for the second output are 0.72, 0.34, 3.38 for the sinusoid, pulse and noisy step inputs, respectively.

4.4 FMLP models

The RFMLP(3, 2, 1) uses the feedback of the predicted output instead of the sensed output. It consists of an input layer with 10 nodes, 2 hidden layers with 15 and 10 nodes, respectively, and an output layer with 2 nodes. The 10-15-10-2 RFMLP(3, 2, 1) network has 347 connection links, and was trained for 60000 cycles using learning rates varying from 0.01 to 0.0001 for the weights and the biases. The second output responses for high noise environment test signals using the obtained RFMLP(3, 2, 1) model and of the reference model are shown in Fig. 9. The relative MSEs of the identified RFMLP(3, 2, 1) of the low noise environment for both the first output are 0.22, 0.26, 0.27 and for the second output are 0.15, 0.15, 0.26 for the sinusoid, pulse and noisy step inputs, respectively. The relative

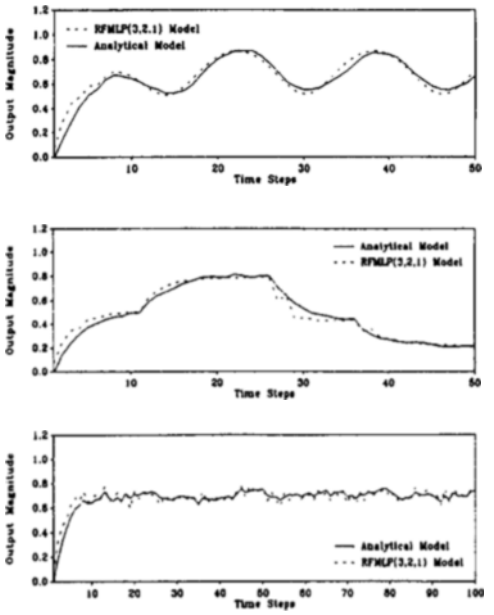

Fig. 9 Response for high noise environment using RFMLP(3, 2, 1) model first output; top: sinusoidal input; middle: ramp input; bottom: step input.

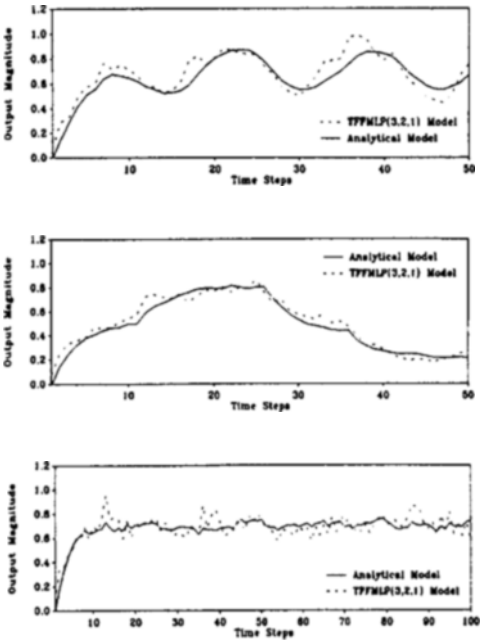

Fig. 10 Response for high noise environment using TFFMLP(3, 2, 1) model first output; top: sinusoidal input; middle: ramp input; bottom: step input.

MSEs for the high noise test signals are the same as the ones in the low noise environment, since RFMLP(3, 2, 1) is not using the sensed but only the predicted outputs.

The TFFMLP, which uses three past inputs with no feedforward term and two past outputs, is denoted TFFMLP(3, 2, 1). The 10-15-10-2 TFFMLP(3, 2, 1) was trained for 45000 cycles using learning rates varying from 0.01 to 0.001 for the weights and for the biases. The second output responses for high noise environment test signals using the identified TFFMLP(3, 2, 1) model and of the reference model are shown in Fig. 10. The relative MSEs of the identified TFFMLP(3, 2, 1) of the low noise environment for both the first output are 0.13, 0.13, 0.21 and for the second output are 0.12, 6.14×10^{-2}, 0.28 for the sinusoid, pulse and noisy step inputs, respectively. The relative MSEs of the high noise environment for both the first output are 0.36, 0.35, 0.98 and for the second output are 0.52, 0.37, 2.05 for the sinusoid, pulse and noisy step inputs, respectively.

The RFMLP(5.6.0) consists of an input layer with 22 nodes, 2 hidden layers with 12 and 10

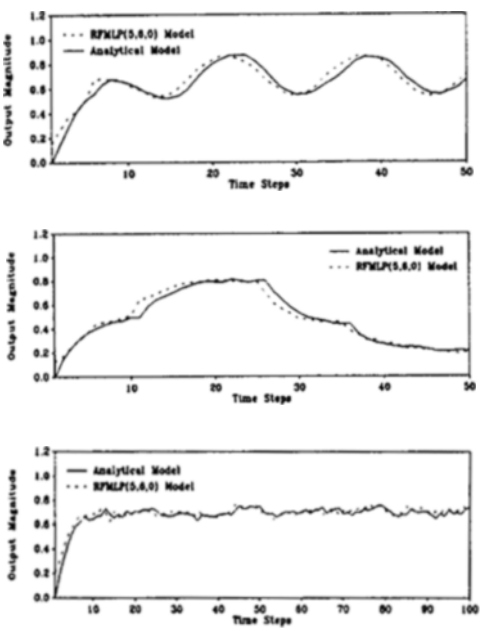

Fig. 11 Response for high noise environment using RFMLP(5, 6, 0) model first output; top: sinusoidal input; middle: ramp input; bottom: step input.

nodes, respectively, and an output layer with 2 nodes. The 22-12-10-2 RFMLP(5, 6, 0) network has 428 connection links, and it was trained for 75000 cycles using learning rates varying from 0.01 to 0.001 for the weights and for the biases. The second output responses for high noise environment test signals using the identified RFMLP(5, 6, 0) model and of the reference model are shown in Fig. 11. The relative MSEs of the identified RFMLP(5, 6, 0) of the low noise environment for both the first output are 5.22×10^{-2}, 0.40, 0.22 and for the second output are 7.08×10^{-2}, 0.22, 0.13 for the sinusoid, pulse, and noisy step inputs, respectively. The relative MSEs for the first output of the identified RFMLP(5, 6, 0) of high noise environment test signals are same as the ones reported above, because the RFMLP(5, 6, 0) is not using the sensed but only the predicted output.

The TFFMLP(5, 6, 0) consists of an input layer with 22 nodes, 2 hidden layers with 12 and 10 nodes, respectively, and an output layer with 2 nodes (22-12-10-2). The TFFMLP(5, 6, 0) was trained for 60000 cycles using learning rates varying from 0.01 to 0.001 for the weights and for the biases. The second output responses for low noise environment test signals using the identified TFFMLP(5, 6, 0) model and of the reference model are shown in Fig. 12. The relative MSEs of the low noise environment for both the first output are 5.22×10^{-2}, 0.40, 0.22 and for the second output are 7.08×10^{-2}, 0.22, 0.13 for the sinusoid, pulse and noisy step inputs, respectively. Also the relative MSEs of the high noise environment for both the first output are 5.22×10^{-2}, 0.40, 0.22 and for the second output are 7.07×10^{-2}, 0.22, 0.13 for the sinusoid, pulse and noisy step inputs, respectively.

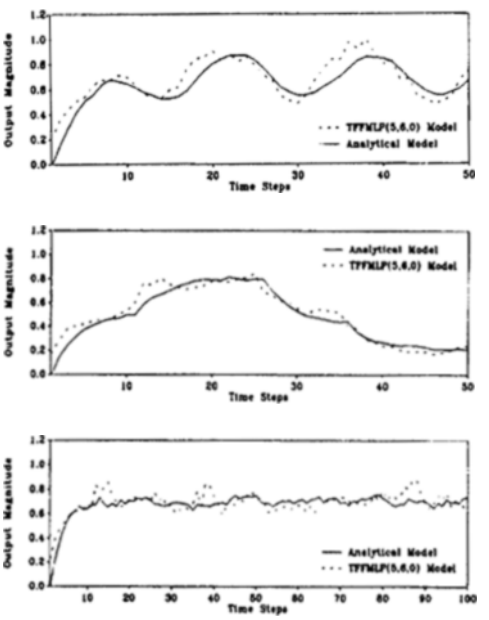

Fig. 12 Response for high noise environment using TFFMLP(5, 6, 0) model first output; top: sinusoidal input; middle: ramp input; bottom: step input.

5. Conclusion

Basic algorithms of the ARX, NARX, FMLP and RMLP type are described with a view toward the nonlinear system identification of input-output and state-space model structures. Traditional and biologically inspired model structures are compared for their effectiveness to identify a complex stochastic multi-input multi-output nonlinear system. The ARX and the NARX are the conventional model structures used in the comparison. The RMLP(FMLP) and the RMLP (FMLP) with and without teacher forcing are used as the biologically motivated nonlinear model structures. For the identification of an ARX model structure and NARX model, algorithms for estimating the parameters have been programmed and used for this comparison instead of using commercially available software. Comparisons of the chosen model is accomplished through a stochastic example. 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 the sake of a fair comparison, the number of connection links, and the number of iterations, are carefully chosen for all model structures.

From the stochastic numerical simulations, it is possible to postulate that the NARX, the FMLP and the RMLP models are good candidate struc-

tures for low noise nonlinear system identification, while the FMLP model structures is not as effective as the RMLP and the NARX models. However, for the high noise environment case the RMLP model is the most effective among the models considered in this paper.

References

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

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

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

Chong, K. T., 1993, "Nonlinear System Identification Using Recurrent Neural Networks," PhD Dissertation, Texas A&M University, College Station, Tx, May.

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

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

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

Parlos A. G., Chong K. T. and Atiya A., 1994, "Application of the Recurrent Multilayer Perceptron in Modeling Complex Process Dynamics," *IEEE Trans. Neural Network*, Vol. 5, No. 2, pp. 255~266, Mar.