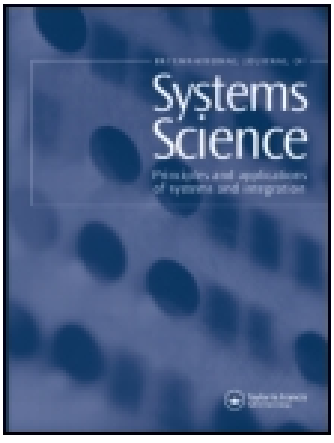

International Journal of Systems Science
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tsys20

Order and delay selection for neural network modelling by identification of linearized models

J. B. Gomm [a] & D. L. Yu [a]
[a] John Moores University
Published online: 26 Nov 2010.

To cite this article: J. B. Gomm & D. L. Yu (2000) Order and delay selection for neural network modelling by identification of linearized models, International Journal of Systems Science, 31:10, 1273-1283, DOI: 10.1080/00207720050165771

To link to this article: http://dx.doi.org/10.1080/00207720050165771

International Journal of Systems Science, 2000, volume 31, number 10, pages 1273–1283

Order and delay selection for neural network modelling by identification of linearized models

J. B. GOMM† and D. L. YU‡

An approach to selecting the order and delay for neural network modelling of nonlinear dynamic systems is proposed by identifying local linear models at points spanning the system operating range. The method is based on relationships between linear and nonlinear models and is developed for three popular nonlinear model structures; nonlinear autoregressive with exogenous inputs (NARX), NARX with a linear noise model and nonlinear autoregressive moving-average with exogenous inputs (NARMAX). Simulation results illustrate the application of the method, and the suitability of the orders and delays selected are demonstrated by nonlinear system identification using radial basis function neural networks. The method is also shown to indicate the suitability of a particular nonlinear model structure for representing a system.

1. Introduction

The suitability of neural networks as general parametric structures for approximating the input–output dynamics of nonlinear systems from measured data has been widely demonstrated. Typical neural networks for nonlinear system modelling are multilayer perceptron (Billings *et al.* 1992, Gomm *et al.* 1996, Song 1998), radial basis function (RBF) (Elanayar and Shin 1994, Yingwei *et al.* 1997, Yu *et al.* 1997) and B-spline (Harris *et al.* 1996, Lightbody *et al.* 1997) networks, mainly because of the proven capability of these networks for approximating general nonlinear functions. A widely used representation for modelling nonlinear systems is to configure a network in a chosen nonlinear aautoregressive moving-average with exogenous inputs (NARMAX) type of model structure, which may or may not include delayed noise terms as inputs (Chen *et al.* 1990, Chen and Billings 1992). In the chosen configuration the order and delay of the model need to be selected, which corresponds to choosing the network input node assignments. Typical selection procedures currently used involve solving a nonlinear search to determine the model order and delay together with the parametrization of the network.

A selection procedure frequently used is to develop several neural network models with different orders and delays and to compare the models by examining their prediction performance (Pottmann and Seborg 1992, Irwin *et al.* 1995, Doherty *et al.* 1997). However, with this method it is often the case that the influence of some experimental factors, such as the number of hidden layer nodes in a network and the value of the width parameter in a Gaussian RBF network, could lead to an inappropriate selection. Correlation tests for assessing the adequacy of neural network models have been developed (Billings *et al.* 1992, Billings and Zhu 1995) and were demonstrated by Billings *et al.* (1992) to provide an indication of the existence of any relevant missing input terms in a neural network model. Another possiblity is to use network pruning techniques (Bhat and McAvoy 1992). Here, a large neural network model is initially developed. Then, using various criteria, the size of the network is gradually reduced, by removing insignificant connections, until further reductions significantly affect the network performance. The procedure can lead to smaller networks with lower model order and delay but suffers from the need to train a large network at the start. Retraining of the net-

Revised 13 September 1999. Accepted 23 September 1999.

†Control Systems Research Group, School of Engineering, Liverpool John Moores University, Byrom Street, Liverpool, L3 3AF, UK. e-mail: j.b.gomm@livjm.ac.uk.

‡Control Systems Research Group, School of Engineering, Liverpool John Moores University, Byrom Street, Liverpool L3 3AF, UK. e-mail: d.yu@livjm.ac.uk.

work during the pruning process is also required because of the nonlinearity of the problem.

These selection methods are time consuming and have a high computational load because of the complexities of estimating and comparing neural network models with different orders and delays. Alternative procedures that have been recently proposed for selecting network inputs include the use of two-stage fuzzy curves and surfaces (Lin *et al.* 1998) and an information theoretic approach (Sridhar *et al.* 1998). The availability of such methods for determining the model order and delay without network training, enable efforts to be concentrated on developing the neural network model with the chosen model inputs. The approach investigated in this paper fits into this category of selection techniques.

Linear system identification theory is well established and understood, and the design of a linear identification experiment follows well-defined procedures. In contrast the identification of nonlinear models is, at present, a relatively immature area compared with the linear case. Many real systems can be approximated within a limited operating range by a linear model. Identified linear models, or data suitable for linear model identification, may be available for a given system as an outcome of a prior system analysis for controller design or other investigation. Linear models at different system operating points can also be identified on line using recursive algorithms and may be available from existing self-tuning controllers. Moreover, with the simpler robust estimation algorithms and the software packages available for linear system identification, it is considerably less time consuming to identify and validate linear models than their nonlinear counterparts at this time. Hence, any inferences from linear system identification techniques that can be applied to the nonlinear domain would be both logical and beneficial.

Identification and integration of linear models within a network, similar to an RBF network, have been recently studied for approximating nonlinear systems (Johansen *et al.* 1998). In this paper, linear model identification is investigated for application to the choice of model order and delay for a particular model structure, which is subsequently identified using a neural network. The paper is organized as follows. In § 2 the three different nonlinear model structures considered and the problem of identifying these models with neural networks are described. The relationships between linear and nonlinear models are developed in § 3, which lead to the proposed approach for model order and delay selection. Application of the method is illustrated in § 4 for modelling three nonlinear systems and the suitability of the selected model orders and delays is demonstrated using RBF networks to identify one of the systems. Extension of the method to the modelling of multiple-input multiple-output systems is examined in § 5, followed by a discussion and conclusions in § 6.

2. Nonlinear models and neural network approximations

Three general discrete-time model structures of nonlinear dynamic systems are considered in this paper. These are as follows.

Model 1. The nonlinear autoregressive with exogenous inputs (NARX) model is given by

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

Model 2. The NARX model augmented by a linear noise model is given by

$$y(t) = f\,[y(t-1),\ldots,y(t-n_y),u(t-k),\\ u(t-k-1),\ldots,u(t-k-n_u+1)]\\ e(t) + p_1 e(t-1) + \cdots + p_{n_e} e(t-n_e). \quad (2)$$

Model 3. The NARMAX model is given by

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

where $u(t) \in U \subset \Re$ and $y(t) \in Y \subset \Re$ are the sampled system input and output respectively, $e(t) \in E \subset \Re$ is random system noise, n_u, n_y and n_e are the number of delayed system input, output and noise terms respectively in the models, k is the system input–output delay, $f(*)$ is a continuous nonlinear function, and $p_1,\ldots,p_{n_e}$ are the real parameters in the linear noise model of model 2.

To develop a neural network representation of one of the above model structures, the network inputs $x(t)$ consist of the delayed input, output and noise terms in the nonlinear part $f(*)$ of the model, and the network output provides a prediction of the system output $y(t)$. This is illustrated for each model in figure 1. Using input–output data collected from the system under a suitable excitation and sample time, the adjustable network parameters are determined to minimize a loss function, which is typically the mean of the squared one-step-ahead prediction errors over the identification data set:

$$V = \frac{1}{N}\sum_{t=1}^{N} \varepsilon(t)^2 = \frac{1}{N}\sum_{t=1}^{N} [\,y(t) - \hat{y}(t)]^2, \quad (4)$$

where $\hat{y}(t) = \hat{f}\,(x(t))$ is the one-step-ahead prediction of the system output at time t and N is the length of the identification data record. For models 2 and 3, delayed noise terms, which are used as inputs to the model, can

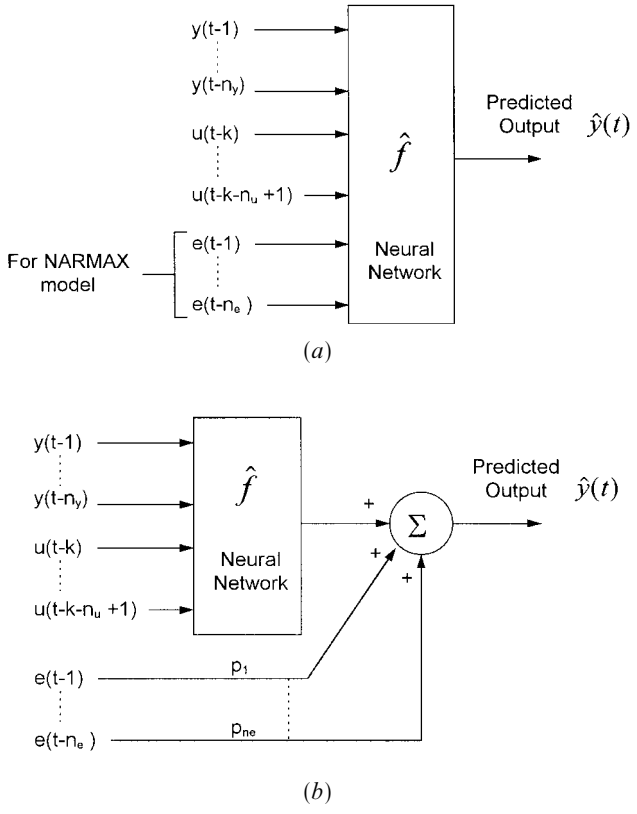

Figure 1. Nonlinear model structures studied: (*a*) NARX and NARMAX models; (*b*) NARX with a linear noise model.

be substituted by delayed prediction errors $\varepsilon(t)$ (Chen *et al.* 1990, Sorensen 1996).

For fixed model order n_u, n_y and n_e and delay k, the difficulty of neural network training for each model increases from model 1 to model 3. Models 1 and 2 are special cases of the more general NARMAX model and are considered because of their ease of development compared with model 3. Of the three models, the first NARX model is the simplest to develop owing to the absence of delayed noise terms in the inputs of the model structure. It has been widely used for neural network modelling (see for example Pottmann and Seborg (1992), Irwin *et al.* (1995) and Gomm *et al.* (1996)). In model 2 the linear noise model can be appended to the network output and the parameters, $p_1, \ldots, p_{n_e}$, can be estimated with the other network parameters (figure 1 (*b*)). An example is the multilayer perceptron neural network where estimation of the noise parameters in model 2 was achieved by modification of the training algorithm (Billings *et al.* 1992). Alternatively, as discussed later, the noise parameters in model 2 could be fixed on the basis of some previous estimates, simplifying the network training procedure. Development of a neural network NARMAX model is the most difficult owing to the delayed noise terms entering the model nonlinearly and has been discussed by Chen *et al.* (1990) and Sorensen (1996).

Given a system that can be represented by one of the models in (1)–(3), it is evident that the model order and delay required for adequate approximation are functions of the system and not of the neural network model. Hence, as investigated in this paper, development of the nonlinear models 1–3 is separated into two parts: firstly model order and delay selection (n_u, n_y, n_e, and k), which corresponds to determining the network inputs $x(t)$ and secondly neural network development to approximate the nonlinear function $f(*)$.

3. Model order and delay selection methodology

The selection procedure for the order and delay of the nonlinear models 1–3 investigated in this paper is based on identifying linearized models of a nonlinear system valid locally for small deviations around different system operating points. The rationale for this is derived from examining the effects of linearizing models 1–3, as discussed in the following. Taking the NARMAX model as the general case for the three models, for a particular operating point i,

$$\begin{aligned} y(t)|_i = f[\, & y(t-1)|_i, \ldots, y(t-n_y)|_i, \quad u(t-k)|_i, \\ & u(t-k-1)|_i, \ldots, u(t-k-n_u+1)|_i, \\ & e(t-1), \ldots, e(t-n_e)\,] + e(t). \end{aligned} \tag{5}$$

Approximation of $f(*)$ about the operating point i using a first-order Taylor series expansion yields

$$\hat{y}(t) = y(t)|_i + \nabla f\,|_i (\psi(t) - \psi|_i), \tag{6}$$

where $\nabla f\,|_i$ is the Jacobian of f with respect to its variables and is a row vector here. Writing $\nabla f\,|_i$ in full gives

$$\begin{aligned} \nabla f\,|_i &= \theta_i^{\mathrm{T}} \\ &= \left[\left.\frac{\partial f}{\partial y(t-1)}\right|_i, \ldots, \left.\frac{\partial f}{\partial y(t-n_y)}\right|_i, \left.\frac{\partial f}{\partial u(t-k)}\right|_i, \ldots, \right. \\ &\quad \left.\frac{\partial f}{\partial u(t-k-n_u+1)}\right|_i, \\ &\quad \left. \left.\frac{\partial f}{\partial e(t-1)}\right|_i, \ldots, \left.\frac{\partial f}{\partial e(t-n_e)}\right|_i \right], \end{aligned} \tag{7}$$

and it is equated to a parameter vector with unknown but definite components θ_i for notational convenience. The data vector is

$$\begin{aligned} \psi(t) = [\, & y(t-1), \ldots, y(t-n_y), \quad u(t-k), \ldots, \\ & u(t-k-n_u+1), \quad e(t-1), \ldots, e(t-n_e)]^{\mathrm{T}}, \end{aligned}$$

and the operating point vector is

$$\psi|_i = [\,y(t-1)|_i, \ldots, y(t-n_y)|_i, u(t-k)|_i, \ldots, u(t-k-n_u+1)|_i, E[e(t-1)], \ldots, E[e(t-n_e)]]^{\mathrm{T}}.$$

If $\psi|_i$ is chosen to be a system equilibrium point, where $u(t-k)|_i = \cdots = u(t-k-n_u+1)|_i = u|_i$ and $y(t-1)|_i = \cdots = y(t-n_y)|_i = y|_i$ which is usual for the identification of linear models, then the linearized model at the equilibrium point can be written as

$$\hat{y}(t) - y|_i = \theta_i^{\mathrm{T}}(\psi(t) - \psi|_i) + e(t). \tag{8}$$

Equation (8) is a linear autoregressive moving-average with exogenous inputs (ARMAX) model, which is valid locally for small deviations around the operating point $(u|_i,\ y|_i)$ and can be easily identified using well-known algorithms. One of the important issues is that all the terms in the data vector for this linearized ARMAX model are present in the original NARMAX model. It is also recognized that terms in the data vector of the NARMAX model may not appear in an identified ARMAX model if the corresponding elements in $\nabla f\,|_i$ are zero or close to zero.

Based on the above observations the following approach is proposed to determine the order and delay of a NARMAX model.

Step 1. ARMAX models are identified about several steady-state operating points $[u|_i,\ y|_i]$, $i = 1, \ldots, N_s$, spanning the system operating range. Identification at more than one operating point reduces the possibility of missing relevant terms in the NARMAX model which cause zero elements in the Jacobian at a particular operating point.

Step 2. The ARMAX model representation at each operating point, n_{ui}, n_{yi}, n_{ei} and k_i, is selected from a set of candidate models by applying the principle of model parsimony to choose the model with the fewest parameters which adequately describes the system. This leads to an overall parsimonious model order and delay for a NARMAX model valid over the whole system operating range.

Step 3. The order and delay for a NARMAX representation of the system is determined by combining the terms in the data vectors of each ARMAX model, $n_u = \max(k_i + n_{ui} - 1)$, $n_y = \max(n_{yi})$, $n_e = \max(n_{ei})$ and $k = \min(k_i)$, $i = 1, \ldots, N_s$.

If the partial differentiation of the true nonlinear system function $f(*)$ with respect to one or more of the input variables is zero, or approximately zero, over the operating range considered, then the function does not significantly change along the direction of these inputs. It follows, therefore, that these inputs may not be identified in linearized models because of the occurrewnce of zeros in the Jacobian. In this case the order and delay selected for the nonlinear model structure, using the proposed method, can be less than those of the true system. However, these inputs contribute constant terms in the system nonlinearity. The effect of the missing constant terms can be realized by a constant bias in the neural network output. Thus, the selected order and delay should still be sufficient for approximating the dynamics of the system over the operating range considered.

The above approach can also be applied to selecting the order and delay of system models 1 and 2 where linearization yields the following remarks. An ARMAX representation also results for model 2 with the difference that the parameters $p_1, \ldots, p_{n_e}$ in the linear noise model also appear as the corresponding parameters in the ARMAX models at all operating points. The development of a neural network configured as model 2 could then be simplified by fixing the noise model parameters to estimates obtained from the identified linear models. Linearization of model 1 results in an autoregressive with exogenous inputs (ARX) model where the data vector does not contain delayed noise terms. Thus, the order and delay of a NARX model can be chosen using a similar procedure to that described above, but estimating ARX models at each operating point and omitting the selection of n_e in step 3.

In addition to selecting a suitable nonlinear model order and delay, identifying linearized models can also provide an indication as to which nonlinear model structure (1)–(3) is most appropriate for a system. Consistent estimates of $p_1, \ldots, p_{n_e}$ in identified ARMAX models at different operating points indicate the suitability of a linear (model 2) rather than a nonlinear (model 3) noise model to represent the system. A comparison of the adequacy of ARX and ARMAX models for the same operating points provides an indication as to the suitability of a NARX model (1) for representing the nonlinear system.

The selection of a linear model order and delay, from a set of estimated candidate models, for each operating point and the adequacy of a particular linearized model structure (ARX or ARMAX) can be guided by various well-developed tests in linear system identification theory (Ljung 1987). In the experiments described in this paper the parameters of ARX and ARMAX models were estimated using the well-known least-squares and maximum-likelihood algorithms respectively. Model order and delay selection was performed using Akaike's final prediction error (FPE) criterion. The adequacy of selected models was evaluated by an

autocorrelation test on the model prediction errors. The FPE is a metric for selecting a parsimonious model and is a weighted form of the loss function V in (4), which penalizes reductions in V for increasing numbers of model parameters. The FPE for an identified model is given by

$$\mathrm{FPE} = \frac{V(1+\alpha n_{\mathrm{p}}/N)}{1-\alpha n_{\mathrm{p}}/N}, \tag{9}$$

where N is the number of data points used to identify a model, $n_{\mathrm{p}}(=n_{ui}+n_{yi}+n_{ei})$ is the number of model parameters and α is a weighting term. The original FPE used $\alpha = 1$ (Akaike 1969) however, it has been reported that this metric can result in the choice of overparameterized models (Leontaritis and Billings 1987) which was also noted in this research. The improved metric with $\alpha = 2$ suggested by Leontaritis and Billings (1987) was therefore used in the investigations described in this paper.

4. Simulation examples

The proposed methodology is illustrated in applications to model order and delay selection for identifying three simulated nonlinear dynamic systems S1, S2 and S3, conforming to the three model structures considered equations (1)–(3) respectively. These systems were chosen for their highly nonlinear steady-state and dynamic properties. Figure 2 illustrates the significant nonlinearity of these systems and also that the systems are nonmonotonic over the operating range considered.

System S1 is given by

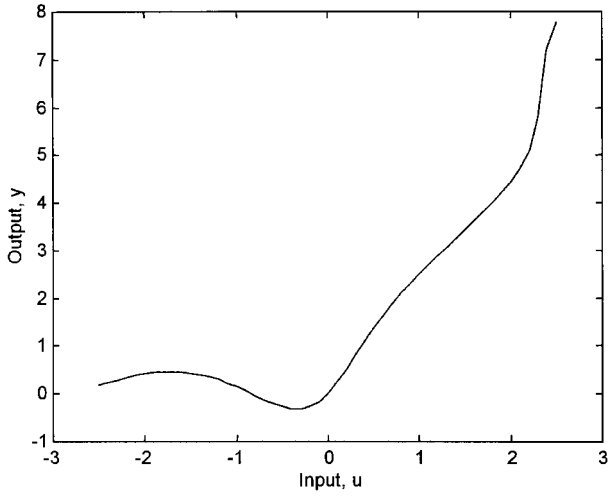

Figure 2. Steady states of systems S1–S3 over the considered operating range.

$$\begin{aligned} y(t) = {} & 0.4\sin[y(t-1)] + 0.3y(t-2) + 0.6u(t-1) \\ & + 0.1u(t-2)y(t-3) \\ & + 1.8\tanh[0.7u(t-3)]^2 + e(t). \end{aligned} \tag{10}$$

System S2 is given by (10) except $e(t)$ is replaced by $e(t) - 1.2e(t-1) + 0.35e(t-2)$.

System S3 is given by (10) except $0.4\sin[y(t-1)]$ is replaced by $0.4\sin[y(t-1)][1-e(t-1)]^2$. For all these systems, $e(t) \in N(0, 0.01)$. The signal-to-noise ratio ($\mathrm{SNR} = \sigma_y^2/\sigma_e^2$) during linear identification of the systems varied at different operating points because of the system nonlinearities. The SNR for the three systems was in the range 8–50, with a typical value of 15, which was considered a realistic amount of noise for linear identification that may be encountered in practice. The input operating range for the systems was chosen as $u \in [-2.5, 2.5]$, to cover a significant portion of the system nonlinearity. Comparing the system equations with the respective model structures (1)–(3), the true system orders and delays are determined to be as follows: for S1, $n_y = 3$, $n_u = 3$ and $k = 1$; for S2, $n_y = 3$, $n_u = 3$, $n_e = 2$ and $k = 1$; for S3, $n_y = 3$, $n_u = 3$, $n_e = 1$ and $k = 1$.

For all three systems, five steady-state operating points spanning the system operating ranges, defined by the inputs $u|_{i=1,\ldots,5} = [-2, -1, 0, 1, 2]$, were chosen to demonstrate the procedure fully. Linear models were identified at each operating point using 100 input–output data generated by a small-amplitude random binary sequence input excitation signal $u(t) = u|_i \pm 0.05$. This caused output variations from steady state of approximately ± 0.1 at all operating points. Estimated steady-state input and output levels were subtracted from the data to form the estimation data set at each operating point. Corresponding to the system structures, ARX models were estimated for system S1 and ARMAX models for systems S2 and S3. Estimation of all candidate models for an operating point took at most a few minutes on a standard personal computer. The FPEs for estimated models were normalized at each operating point to assist comparison. There are several ways of choosing the order and delay for linear models (Ljung, 1987). In the approach used here, the search for delay is not included in the search for order to reduce the number of models identified.

4.1. *Selection of model delay*

The procedure used to determine the nonlinear model delay k was the same for all three systems. Fixed high-order linear models with $n_{ui} = n_{yi} = n_{ei} = 3$ were estimated with various delays $k_i = 1, 2, 3, 4$ and 5 at all operating points. Similar results were obtained for all systems but are only illustrated in figure 3 for system

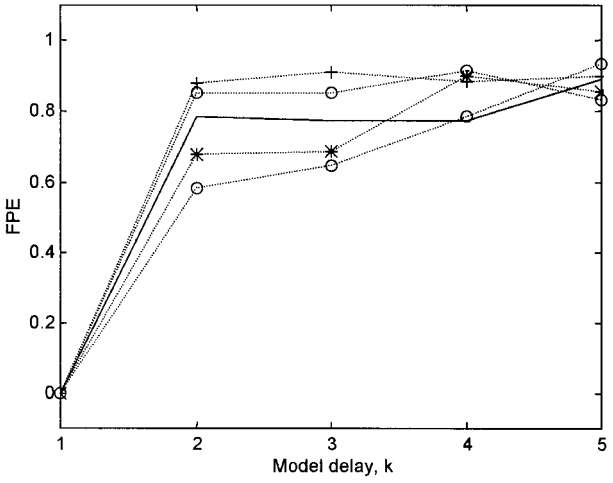

Figure 3. Selection of model delay for system S3 showing normalized FPEs of linear ARMAX models estimated at different operating points: (+), $u|_1$; (*), $u|_2$; (○), $u|_3$; (×), $u|_4$; (—), $u|_5$.

S3. The delay for nonlinear modelling was chosen as $k = 1$ as this clearly shows the lowest FPE for all operating points. This choice corresponds to the true delay present in each of the three systems.

To demonstrate further the effectiveness of this procedure for selecting a nonlinear model delay, the approach was applied to the simulation of system S1 with the system delay increased to three samples. The results are presented in figure 4, which shows that the FPEs for all operating points again converge to their lowest values at the true system delay $k = 3$.

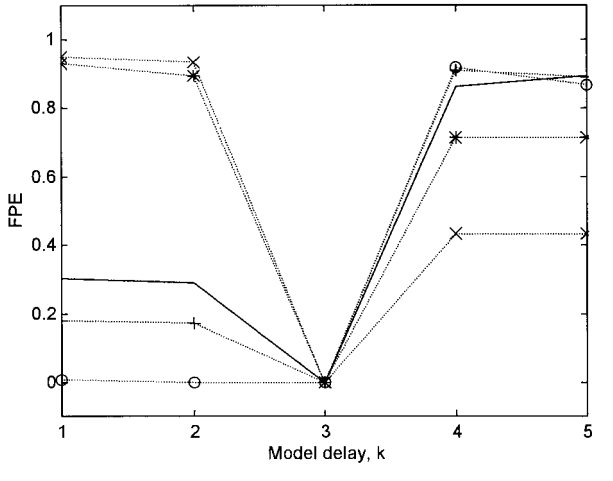

Figure 4. Selection of model delay for system S1 with an increased delay of three samples showing normalized FPEs of linear ARX models estimated at different operating points: (+), $u|_1$; (*), $u|_2$; (○), $u|_3$; (×), $u|_4$; (—), $u|_5$.

4.2. *Model order selection*

To select model orders for the three systems, the delay k_i in the estimated linear models was fixed at unity, as selected previously in §4.1, and linear models with different orders were identified at all system operating points. For system S1, the estimated ARX model orders contained the order combinations $[n_{yi} = 1, 2, 3, 4\ (n_{ui} = 1, 2, 3, 4)]$, giving a candidate set of 16 models. For systems S2 and S3, ARMAX models were estimated from a candidate model set of 48 models comprising the order combinations $\{n_{yi} = 1,2,3,4\ [n_{ui} = 1,2,3,4\ (n_{ei} = 1,2,3)]\}$.

Figure 5 shows the lowest FPEs for ARX models with the same number of parameters identified from system S1 at the five operating points. It is clear that for four of the operating points $(u|_i = -2, -1, 1, 2)$ a suitable model order is one containing six parameters, which achieves the minimum FPE in these cases. This model for all four operating points has the order $n_{yi} = 3$ and $n_{ui} = 3$. For the operating point $u|_3 = 0$, a model with three parameters is indicated and corresponds to a model order $n_{y3} = 2$ and $n_{u3} = 1$. Following step 3 of the procedure, the order of a NARX model for approximating the system over the whole operating range is chosen as $n_y = 3$ and $n_u = 3$. For a smaller operating range around $u|_3 = 0$, a NARX model order of $n_y = 2$ and $n_u = 1$ may be adequate.

The FPEs for the ARMAX models estimated at different operating points for systems S2 and S3 are shown in figures 6 and 7 respectively. Following a similar line of enquiry as described above for system S1, the orders chosen for the linear models at each operating point and for NARMAX modelling of the systems are summarized in table 1. At the operating point $u|_3 = 0$ for system

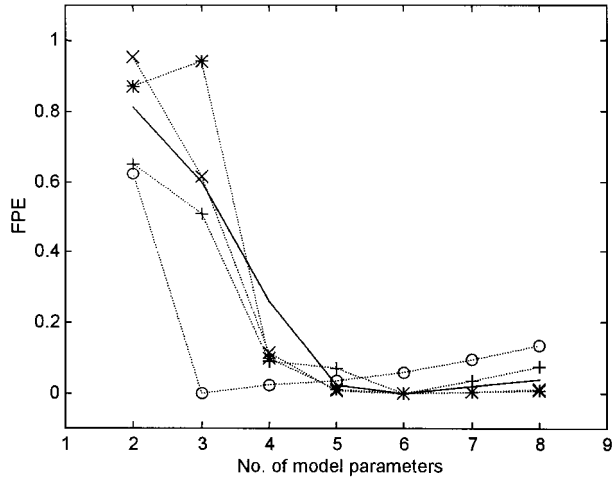

Figure 5. Selection of model order for system S1, showing normalized FPEs of linear ARX models estimated at different operating points: (+), $u|_1$; (*), $u|_2$; (○), $u|_3$; (×), $u|_4$; (—), $u|_5$.

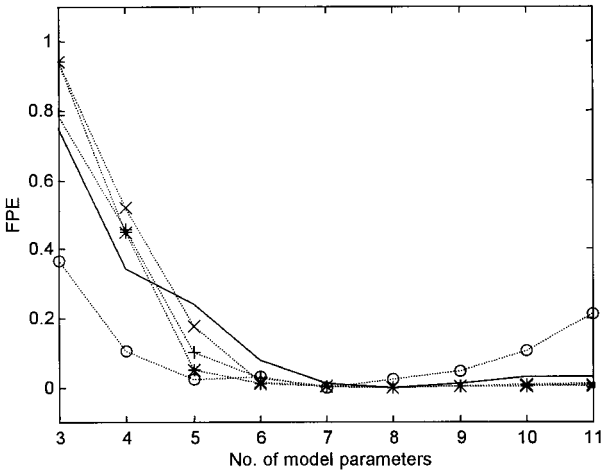

Figure 6. Selection of model order for system S2, showing normalized FPEs of linear ARMAX models estimated at different operating points: (+), $u|_1$; (*), $u|_2$; (○), $u|_3$; (×), $u|_4$; (—), $u|_5$.

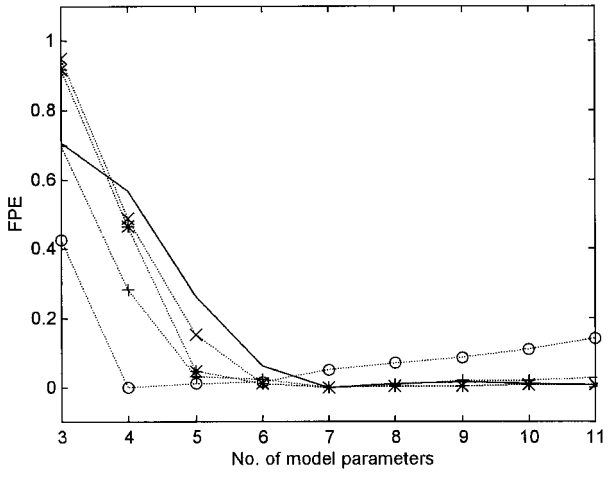

Figure 7. Selection of model order for system S3, showing normalized FPEs of linear ARMAX models estimated at different operating points: (+), $u|_1$; (*), $u|_2$; (○), $u|_3$; (×), $u|_4$; (—), $u|_5$.

S2 (figure 6) the lowest FPE occurred for a model with seven parameters and the order $n_{y3} = 3$, $n_{u3} = 2$ and $n_{e3} = 2$. Further examination of this model showed that a pole and a zero cancelled in the model, which resulted in a selected model with five parameters and the order in table 1.

The results in table 1 show that the proposed approach to nonlinear model order and delay selection results in a choice corresponding to the true system order and delay for modelling all three simulated systems over the whole operating range considered. Furthermore, the true system orders and delays are selected with results from any two of the five considered operating points. Hence, only two or three operating points may suffice in practice. The selection of reduced orders, for both linear and nonlinear models, for approximating the systems around the operating point $u|_3 = 0$ is due to terms in the Jacobian, corresponding to those terms in the system equations containing $y(t-3)$, $u(t-2)$ and $u(t-3)$, tending to zero at this operating point.

4.3. *Model structure selection*

This section describes how the proposed method can indicate a suitable model structure (1)–(3) to represent a nonlinear system. Figures 8 and 9 show the variation in the estimates of the noise parameters in the chosen ARMAX models for systems S2 and S3 at the different operating points. The standard deviations of the estimates, obtained from the estimation algorithm, are also included. The estimates for system S2 (figure 8) are fairly consistent for the whole operating range, correctly indicating that a linear noise model may be appropriate for approximating this system. The estimates are also close to the true values in system S2 of −1.2 and 0.35. To simplify the neural network training in this case, the parameters in the linear noise model could be fixed to values representative of those estimated in these linear ARMAX models. In contrast with this, figure 9

Table 1. Selected linear and nonlinear model orders and delays for systems S1, S2 and S3.

System	Operating point $u\|_i$	Selected linear model					Selected nonlinear model			
		n_p	n_{yi}	n_{ui}	n_{ei}	k_i	n_y	n_u	n_e	k
S1	−2, −1, 1, 2	6	3	3		1	3	3		1
	0	3	2	1		1	2	1		1
S2	−2, −1, 1, 2	8	3	3	2	1	3	3	2	1
	0	5	2	1	2	1	2	1	2	1
S3	−2, −1, 1, 2	7	3	3	1	1	3	3	1	1
	0	4	2	1	1	1	2	1	1	1

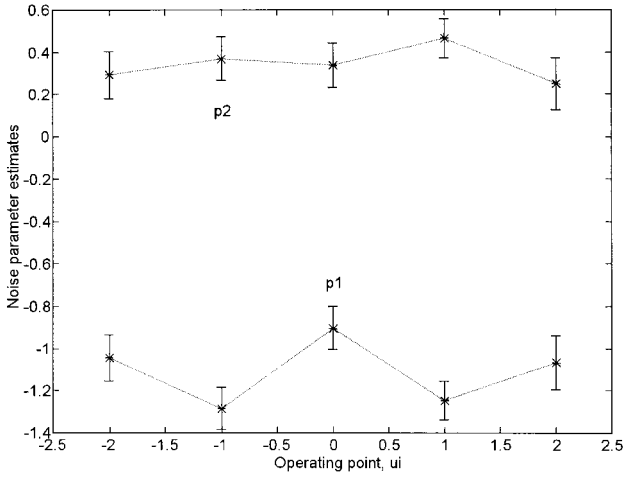

Figure 8. Noise parameters in selected linear ARMAX models of system S2 estimated at different operating points $u|_i$.

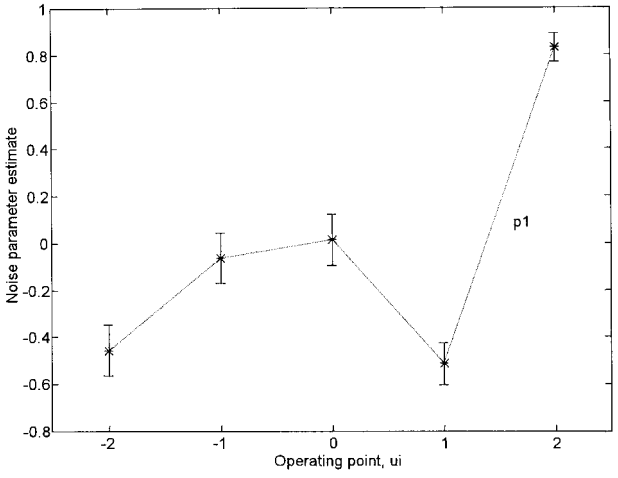

Figure 9. Noise parameters in selected linear ARMAX models of system S3 estimated at different operating points $u|_i$.

shows that the parameter in the linear noise models for system S3 varies considerably over the operating range, correctly indicating that a NARMAX model would be more appropriate for approximating this system.

To illustrate the choice between the NARX model 1 and models 2 and 3, the results of autocorrelation tests on the prediction errors of ARX and ARMAX models identified at one operating point of system S3 are shown in figure 10. The ARMAX model had the order given in table 1 and the ARX model order was $n_{y5} = n_{u5} = 3$. The model prediction errors can be considered as statistically uncorrelated if the test lies within the confidence limits for lags greater than zero. The ARX model test exceeds the confidence limits at lags 1 and 3, thus failing the test. This result and the successful test by the ARMAX model correctly indicate that model 2 or 3 is a better choice for modelling this system. ARX models

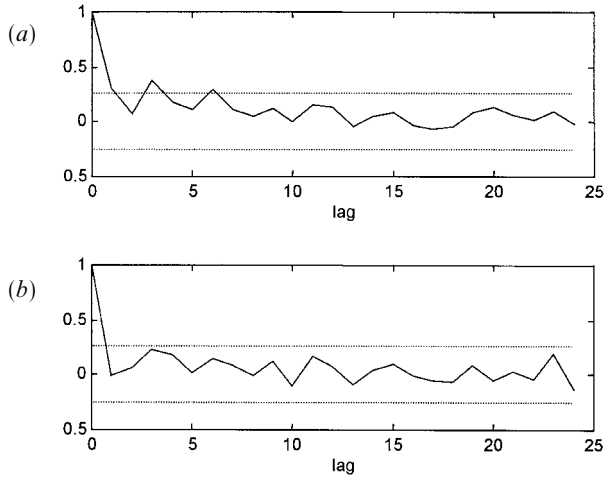

Figure 10. Autocorrelation test of prediction errors for linear models estimated at operating point $u|_5$ of system S3. (*a*) ARX model; (*b*) ARMAX model. The dotted lines show the 95% confidence limits.

identified for system S1 satisfied the autocorrelation test, thus confirming the choice of a NARX model structure for this system.

4.4. *Identification using radial basis function networks*

The suitability of the model orders and delays selected in §§ 4.1 and 4.2 is demonstrated in this section with the application of RBF networks to approximating system S1. The networks were of the form

$$\hat{y}(t) = \sum_{i=1}^{n_h} w_i \phi_i(\|x(t) - c_i\|), \qquad (11)$$

where $\hat{y}(t)$ is the network prediction of the system output at time t, n_h is the number of hidden layer nodes, w_i is the weight connecting the ith hidden node output to the network output, ϕ_i and c_i denote the output and centre vector respectively for the ith hidden node and $\|*\|$ denotes the Euclidean distance. The nonlinear basis function was chosen to be the thin-plate spline (Chen and Billings 1992)

$$\phi_i[\|x(t) - c_i\|] = [\|x(t) - c_i\|]^2 \log_e [\|x(t) - c_i\|], \quad (12)$$

and the network input vector was set to configure the networks as NARX models (figure 1 (*a*)) $x(t) = [y(t-1), \ldots, y(t-n_y), u(t-k), u(t-k-1), \ldots, u(t-k-n_u+1)]^T$. An initial set of 60 network centres were determined using batch k-means clustering applied to the input training data (Moody and Darken 1989). The orthogonal least-squares (OLS) algorithm (Chen *et al.* 1990, Chen and Billings 1992) was then used to select sequentially the most significant centres from this initial set and to determine the output layer weights. The FPE criterion was used to obtain networks with a compro-

Table 2. Performance of RBF networks for approximating system S1.

Input range	Network	n_u, n_y, k	n_h	Training data		Test data	
				$V \times 10^{-2}$	FPE $\times 10^{-2}$	$V \times 10^{-2}$	FPE $\times 10^{-2}$
[−2.5, 2.5]	A	3, 3, 1	26	1.29	2.78	1.15	5.1
	B	1, 2, 1	21	7.08	9.96	8.58	12.07
[−0.25, 0.25]	A	3, 3, 1	20	0.035	0.062	0.04	0.072
	B	1, 2, 1	31	0.03	0.05	0.035	0.059

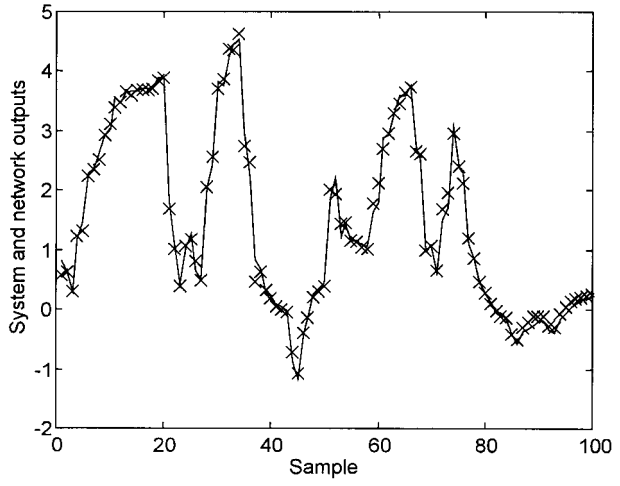

Figure 11. System 1 output (—) and predictions (×) of network A on test data over full operating range.

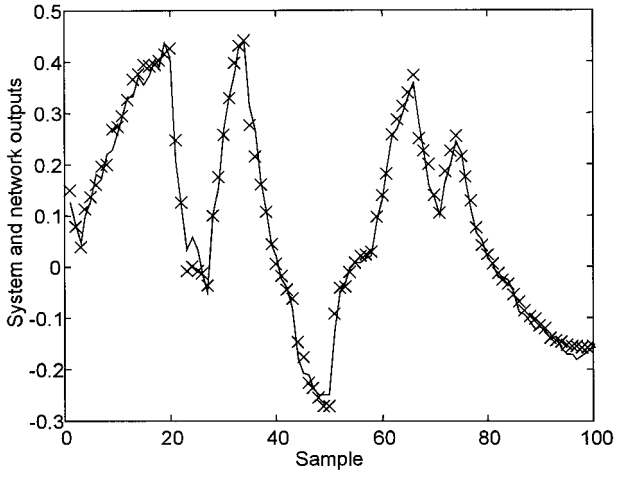

Figure 12. System 1 output (—) and predictions (×) of network B on test data over a small range around operating point $u|_3 = 0$.

mise between accuracy and complexity. Training of the networks with the OLS algorithm was stopped when a minimum FPE for a network was achieved.

Two experiments were performed to compare the performance of two RBF networks, first for approximating the system S1 over the whole nonlinear operating space and secondly for a smaller operating range about the point $u|_3 = 0$. The model orders and delays for network A were $n_u = 3$, $n_y = 3$ and $k = 1$ and for network B were $n_u = 1$, $n_y = 2$ and $k = 1$, as previously selected (§4.2). Input–output data for network training and testing in the two experiments were collected by applying system input signals with random amplitudes in the ranges $[-2.5, 2.5]$ and $[-0.25, 0.25]$ respectively and with random transition periods in the range $[1, 10]$ samples. All networks were trained and tested on 500 and 100 data points respectively. The results are summarized in table 2.

The results support the orders and delays selected by the proposed method. Table 2 shows that to approximate the system over the whole nonlinear operating space the larger model order (network A) is appropriate with much lower loss functions and FPEs on both the training and the testing data than the network with the reduced model order (network B). However, to approximate the system in a smaller operating space about $u|_3 = 0$, then the reduced model order has a marginally better performance than the higher-order network based on the loss functions and has lower FPEs. Thus, on the parsimony principle, the network with the reduced model order should be chosen as a suitable fit for this limited operating range. The satisfactory predictions of both networks in these two situations on the test data are illustrated in figures 11 and 12.

5. Extension to multiple-input multiple-output systems

Multivariable forms of the three model structures considered in this paper are possible (Chen and Billings 1992). Extension of the proposed order and time delay selection method to multiple-input multiple-output systems is examined here by considering the modelling of a system with the most widely used of the three models, namely a multivariable NARX model. A similar approach is applicable to multivariable forms of models 2 and 3, but with the additional requirement of estimating the noise parameters in the corresponding line-

arized multivariable models. A multi-variable NARX model is described by

$$\boldsymbol{Y}(t) = f\,[\,\boldsymbol{Y}(t-1),\ldots,\boldsymbol{Y}(t-n_y),\boldsymbol{U}(t-k),\ldots,\boldsymbol{U}(t-k-n_u+1)] + \boldsymbol{E}(t), \qquad (13)$$

with $\boldsymbol{Y}(t) = [y_1(t),\ldots,y_m(t)]^{\mathrm{T}}$, $\boldsymbol{U}(t) = [u_1(t),\ldots,u_r(t)]^{\mathrm{T}}$, $\boldsymbol{E}(t) = [e_1(t),\ldots,e_m(t)]^{\mathrm{T}}$ and where m and r are the number of system outputs and inputs respectively. Without loss of generality, the model (13) can be decomposed into a set of subsystem models, one for each system output y_j, as follows

$$\begin{aligned} y_j(t) = f_j[\,& y_1(t-1),\ldots,y_1(t-n^j_{y1}),\ldots,y_m(t-1),\ldots,\\ & y_m(t-n^j_{ym}),u_1(t-k^j_1),\ldots,\\ & u_1(t-k^j_1-n^j_{u1}+1),\ldots,u_r(t-k^j_r),\ldots,\\ & u_r(t-k^j_r-n^j_{ur}+1)] + e_j(t). \end{aligned} \qquad (14)$$

Linearization of (14) at a system operating point yields a linear multiple-input, single-output ARX model with the same orders and delays, $n^j_{y1},\ldots,n^j_{ym}$, $n^j_{u1},\ldots,n^j_{ur},k^j_1,\ldots,k^j_r$, as the nonlinear subsystem in (14). Thus, suitable orders and delays for each nonlinear subsystem model in (14) can be determined by estimating corresponding linearized models at different system operating points, in a similar way to that described in § 3 for the single-input single-output case. The applicability of the proposed method for multivariable systems is further discussed in the next section.

6. Discussion and conclusions

The use of identifying local linear models for selecting the order and delay for neural network modelling of nonlinear systems is described. Simulation results illustrate the applicability of this method for determining a suitable model order and delay for approximating a nonlinear system over a considered operating range. In all investigated cases, the chosen order and delay matched those of the true system. The method was also demonstrated to provide an indication of a suitable model structure for a neural network: NARX, NARX with a linear noise or NARMAX. An important point is that no assumption of a particular modelling technique has been made in the development of the method. Therefore, the method could also be applicable to other nonlinear modelling techniques, such as fuzzy and polynomial models. The possibility of extending the method to the modelling of multiple-input multiple-output systems was examined. The analysis suggests the method can be extended quite straightforwardly to the multivariable case. Further work is needed to investigate this in more detail.

Linear models contain significantly fewer parameters than nonlinear neural network models do and the identification algorithms for linear models are less complex than the training algorithms for neural networks. Consequently, the proposed selection method is less computationally intensive than methods based on training neural networks, such as the trial and error method often used for training and comparing several neural networks with different orders and delays. The identified linear models can provide further valuable insight to the system and the identification experiment. For example, comparing the gains and time constants of the linear models at different operating points can provide information on the system nonlinearity. Also, some indication on the suitability of the sampling time used can be obtained by examining the performance and poles of the linear models (Doherty *et al.* 1994). This information is not readily available from other proposed model input selection techniques (see for example Lin *et al.* (1998) and Sridhar *et al.* (1998)). It is acknowledged, however, that the method is likely only to be successful for those systems that can be adequately approximated in a limited operating region by a linear model.

The practical applicability of the method is supported by previous work where linear models have been partially investigated but a formal procedure for their use has not been developed. Doherty *et al.* (1994) identified linear models around several operating points of a chemical pH process, which is well known to exhibit significant nonlinearity. The most appropriate order and delay for the linear models were the same at all operating points. This order and delay were also found to be the most suitable for a neural network model of the process when various networks with different model structures were trained and compared. Independently, Irwin *et al.* (1995) identified linear models of a turbogenerator for comparison with neural network models. It was reported that the input structure of the linear models was also the most suitable for neural network modelling. Suitable orders and delays for neural network modelling of a multivariable chemical process, involving temperature, pH and dissolved oxygen, have also been found by identifying linear models at a few operating points (Yu *et al.* 1997). All these studies involved the use of NARX model structures. This paper provides an explanation as to why these observations were made and a formal order and delay selection procedure based on linear model identification is proposed. The paper also extends the application to model structures containing delayed noise terms, such as the NARMAX model.

The practical studies mentioned above suggest that the orders and delays for modelling many real nonlinear systems do not vary at different operating points,

although the parameter values in identified linear models do vary. Thus, the choice of the operating points used in the selection method may not be too critical for many systems. It is possible, however, that the choice of operating points for some systems could significantly affect the suitability of the orders and delays chosen. A more systematic approach to choosing the operating points may be desirable, such as the techniques recently developed for operating regime decomposition when building local linear model networks (Johansen *et al.* 1998).

A disadvantage of the method is the need to provide system excitation and to collect data from different system operating points for linear modelling. To remove this as an extra experimental stage, the excitation signal for nonlinear modelling could be modified to include some periods of small amplitude excitation, which excite the system at the required operating points. Thus, the data for both order and delay selection using the proposed method and neural network development are collected at the same time. Excitation signals similar to that just described have previously been suggested as more realistic for use in practice than pure random amplitude signals are (Pottmann and Seborg 1992, Irwin *et al.* 1995).

Acknowledgements

The authors gratefully acknowledge support of this research by the UK Engineering and Physical Sciences Research Council (grant GR/K 35815).

References

AKAIKE, H., 1969, Fitting autoregressive models for prediction. *Annals of the Institute of Statistical Mathematics,* **21**, 425–439.

BHAT, N. V., and MCAVOY, T. J., 1992, Determining model structure for neural models by network stripping. *Computers and Chemical Engineering,* **16**, 271–281.

BILLINGS, S. A., JAMALUDDIN, H. B., and CHEN, S., 1992, Properties of neural networks with applications to modelling non-linear dynamical systems. *International Journal of Control,* **55**, 193–224.

BILLINGS, S. A., and ZHU, Q. M., 1995, Model validity tests for multivariable nonlinear models including neural networks. *International Journal of Control,* **62**, 749–766.

CHEN, S., BILLINGS, S. A., COWAN, C. F. N., and GRANT. P. M., 1990, Practical identification of Narmax models using radial basis functions. *International Journal of Control,* **52**, 1327–1350.

CHEN, S., and BILLINGS, S. A., 1992, Neural networks for nonlinear dynamic system modelling and control. *International Journal of Control,* **56**, 319–346.

DOHERTY, S. K., GOMM, J. B., WILLIAMS, D., and EARDLEY, D. C., 1994, Design issues in applying neural networks to model highly nonlinear processes. *Proceedings of the IEE International Conference on Control,* Vol. 2, Coventry, UK, 1994, pp. 1478–1483.

DOHERTY, S. K., GOMM, J. B., and WILLIAMS, D., 1997, Experiment design considerations for nonlinear system identification using neural networks. *Computers and Chemical Engineering,* **21**, 327–346.

ELANAYAR V. T., S., and SHIN, Y. C., 1994, Radial basis function neural network for approximation and estimation of nonlinear stochastic dynamic systems. *IEEE Transactions on Neural Networks,* **5**, 594–603.

GOMM, J. B., WILLIAMS, D., EVANS, J. T, DOHERTY, S. K., and LISBOA, P. J. G., 1996, Enhancing the nonlinear modelling capabilities of MLP neural networks using spread encoding. *Fuzzy Sets and Systems,* **79**, 113–126.

HARRIS, C. J., BROWN, M., BOSSLEY, K. M., MILLS, D. J., and FENG, M., 1996, Advances in neurofuzzy algorithms for real-time modelling and control. *Engineering Applications of Artificial Intelligence,* **9**, 1–16.

IRWIN, G. W., BROWN, M., HOGG, B., and SWIDENBANK, E., 1995, Neural network modelling of a 200 MW boiler system. *Proceedings of the Institution of Electrical Engineers,* Pt D, **142**, 529–536.

JOHANSEN, T. A., HUNT, K. J., and FRITZ, H., 1998, A software environment for gain scheduled controller design. *IEEE Control Systems,* **18**, 48–60.

LEONTARITIS, I. J., and BILLINGS, S. A., 1987, Model selection and validation methods for nonlinear systems. *International Journal of Control,* **45**, 311–341.

LIGHTBODY, G., O'REILLY, P., IRWIN, G. W., KELLY, K., and MCCORMICK, J., 1997, Neural modelling of chemical plant using MLP and *B*-spline networks. *Control Engineering Practice,* **5**, 1501–1515.

LIN, Y. H., CUNNINGHAM, G. A., COGGLESHALL, S. V., and JONES, R. D., 1998, Nonlinear system input structure identification: two stage fuzzy curves and surfaces. *IEEE Transactions on Systems, Man and Cybernetics* A, **28**, 678–684.

LJUNG, L., 1987, System Identification, Theory for the User (Englewood Cliffs, New Jersey: Prentice-Hall).

MOODY, J., and DARKEN, C. J., 1989, Fast learning in networks of locally-tuned processing units. *Neural Computation,* **1**, 281–294.

POTTMANN, M., and SEBORG, D. E., 1992, Identification of nonlinear processes using reciprocal multiquadric functions. *Journal of Process Control,* **2**, 189–203.

SONG, Q., 1998, Robust training algorithm of multilayered neural networks for identification of nonlinear dynamic systems. *Proceedings of the Institution of Electrical Engineers,* Pt D, **145**, 41–46.

SORENSEN, O., 1996, Nonlinear pole-placement control with a neural network. *European Journal of Control,* **2**, 36–43.

SRIDHAR, D. V., BARTLETT, E. B., and SEAGROVE, R. C., 1998, Information theoretic subset selection for neural network models of chemical processes. *Computers and Chemical Engineering,* **22**, 613–626.

YINGWEI, L., SUNDARARAJAN, N., and SARATCHANDRAN, P., 1997, Identification of time-varying nonlinear systems using minimal radial basis function neural networks. *Proceedings of the Institution of Electrical Engineers,* Pt D, **144**, 202–209.

YU, D. L., GOMM, J. B., and WILLIAMS, D., 1997, RBF network modelling of a multivariable chemical process with ROLS weight updating. *Preprints of the 11th IFAC Symposium on System Identification,* Vol. 2, Fukuoka, Japan, 1997, pp. 565–570.