International Journal of Control
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713393989

Practical identification of NARMAX models using radial basis functions

S. Chen [a]; S. A. Billings [b]; C. F. N. Cowan [a]; P. M. Grant [a]
[a] Department of Electrical Engineering, University of Edinburgh, Edinburgh, Scotland, U.K
[b] Department of Control Engineering, University of Sheffield, Sheffield, England, U.K

Online Publication Date: 01 December 1990
To cite this Article: 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, 52:6, 1327 - 1350
To link to this article: DOI: 10.1080/00207179008953599
URL: http://dx.doi.org/10.1080/00207179008953599

INT. J. CONTROL, 1990, VOL. 52, NO. 6, 1327–1350

Practical identification of NARMAX models using radial basis functions

S. CHEN†, S. A. BILLINGS‡, C. F. N. COWAN† and P. M. GRANT†

A wide class of discrete-time non-linear systems can be represented by the non-linear autoregressive moving average (NARMAX) model with exogenous inputs. This paper develops a practical algorithm for identifying NARMAX models based on radial basis functions from noise-corrupted data. The algorithm consists of an iterative orthogonal-forward-regression routine coupled with model validity tests. The orthogonal-forward-regression routine selects parsimonious radial-basis-function models, while the model validity tests measure the quality of fit. The modelling of a liquid level system and an automotive diesel engine are included to demonstrate the effectiveness of the identification procedure.

1. Introduction

Most systems encountered in the real world are non-linear to some extent and in many practical applications non-linear models may be required to achieve an acceptable prediction accuracy. The NARMAX model (Leontaritis and Billings 1985, Chen and Billings 1989 a) provides a unified representation for a wide class of discrete-time non-linear systems. In a NARMAX description the system is modelled in terms of a non-linear functional expansion of lagged inputs, outputs and prediction errors. Two considerations are of practical importance for the application of the NARMAX approach. The functional describing a real-world system can be very complex and the explicit form of this functional is usually unknown, so that any practical modelling of a real-world process must be based upon a chosen model set of known functions. Obviously this model set should be capable of approximating the underlying process within an acceptable accuracy. Secondly, an efficient identification procedure must be developed for the selection of a parsimonious model structure, because the dimension of a non-linear model can easily become extremely large. Without efficient subset selection, the identification would entail excessive computation and the resulting model would have little practical value.

Previous research (Billings *et al.* 1989 b, Chen *et al.* 1989) has investigated the polynomial NARMAX model. The set of polynomial functions is known to be dense in the space of continuous functions. This means that a system can be approximated within an arbitrary accuracy by a polynomial model. By augmenting and modifying well-known orthogonal least-squares methods, an iterative orthogonal-forward-regression (OFR) routine (Chen *et al.* 1989) has been derived to select parsimonious polynomial models. This identification algorithm, when coupled with some model validity tests (Billings and Voon 1986, Bohlin 1978, Leontaritis and Billings 1987),

Received 18 April 1990.

† Department of Electrical Engineering, University of Edinburgh, Mayfield Road, Edinburgh EH9 3JL, Scotland, U.K.

‡ Department of Control Engineering, University of Sheffield, Mappin Street, Sheffield S1 3JD, England, U.K.

0020-7179/90 $3.00

provides a practical means for fitting parsimonious polynomial models to structure-unknown real systems.

The present study considers an alternative approach for fitting NARMAX models based on the radial basis function (RBF). The RBF method has traditionally been used for strict interpolation in multidimensional space and has been shown to be superior to many existing methods (Powell 1985, Micchelli 1986). The generalization due to Broomhead and Lowe (1988) provides a more suitable basis for system modelling. The major problem remains of how to select an appropriate set of RBF centres. In order to utilize the advantages of the linear-in-the-parameters formulation, the centres are often chosen to be a subset of the data. Although researchers are well aware that the fixed centres should suitably sample the input domain, most published results simply assume that some mechanism exists to select centres from data points and do not offer any real means for choosing them. However, by interpreting the RBF as an extended model set (Billings and Chen 1989 b), Chen *et al.* (1991) have shown that a suitable set of centres can be readily identified in a relatively straightforward manner by employing the OFR algorithm. Because these results were only valid for the unrealistic case of additive white noise, the present study extends the previous results by the authors, and develops an identification procedure for fitting parsimonious NARMAX models using the RBF expansion in the presence of correlated noise. The techniques employed in the current study are very similar to those used for the polynomial NARMAX models using the RBF expansion in the presence of correlated noise. The techniques employed in the current study are very similar to those used for the polynomial NARMAX model. The OFR routine is utilized for the selection of subset RBF models and the adequacy of the fitted model is determined using the model validity tests developed previously for other non-linear models. A brief comparison with the polynomial approach is also given. Finally the identification results for two real processes are presented to illustrate the approach of RBF modelling.

The single-input single-output case is studied throughout for notational simplicity. However the NARMAX representation and the RBF model are both valid for multi-input multi-output systems (Leontaritis and Billings 1985, Broomhead and Lowe 1988).

2. NARMAX representation

The successful development of identification and many other application procedures for discrete-time linear systems is largely based upon the linear difference equation model

$$y(t) = \sum_{i=1}^{n_y} b_{yi}\, y(t-i) + \sum_{i=1}^{n_u} b_{ui}\, u(t-i) + \sum_{i=1}^{n_e} b_{ei}\, e(t-i) + e(t) \qquad (1)$$

where $y(t)$, $u(t)$ and $e(t)$ are the system output, input and noise, respectively; n_y, n_u and n_e are the maximum lags in the output, input and noise, respectively, and $\{e(t)\}$ is assumed to be a white sequence. The model (1) is commonly known as the ARMAX model.

It is natural to extend the highly successful approach of the difference equation model to the non-linear case. Under some mild assumptions it has been rigorously proved that a wide class of discrete-time non-linear systems can be represented by the following non-linear difference equation model (Leontaritis and Billings 1985):

$$y(t) = f(y(t-1), \ldots, y(t-n_y), u(t-1), \ldots, u(t-n_u), e(t-1), \ldots, e(t-n_e)) + e(t) \qquad (2)$$

where $f(\cdot)$ is some non-linear function. The model (2) is often referred to as the NARMAX model because of its resemblance to (1). Note that even in the case that a noise source enters the system purely as measurement white noise, the system model cannot be simplified as

$$y(t) = f(y(t-1), \ldots, y(t-n_y), u(t-1), \ldots, u(t-n_u)) + e(t) \tag{3}$$

This can be readily shown as follows. Assume that the noise-free underlying system is governed by

$$y_s(t) = f_s(y_s(t-1), \ldots, y_s(t-n_y), u(t-1), \ldots, u(t-n_u)) \tag{4}$$

and the measured system output is a noise corruption of $y_s(t)$

$$y(t) = y_s(t) + e(t) \tag{5}$$

Substituting (5) into (4) yields the model (2), rather than (3).

The input–output relationship (2) is obviously very general. The functional form $f(\cdot)$ for a real-world system can be very complicated and is rarely known *a priori*. A model must therefore be constructed based on some known simpler functions, and some examples are the output-affine model (Sontag 1979, Chen and Billings 1988), the polynomial model (Chen *et al.* 1989) and the rational model (Sontag 1979, Billings and Chen 1989 a). The present study employs an alternative RBF expansion to model the input–output relationship (2).

The parsimonious principle is particularly critical in non-linear systems identification because a non-linear model can easily involve an excessive number of parameters. Without an efficient subset selection, the identification problem is almost certainly ill-conditioned and the parameter estimation will require excessive computation. Furthermore, the resulting model has little practical value owing to its high dimensionality. It is shown below that the linear-in-the-parameter structure of the RBF expansion can be exploited for subset model selection, and the OFR algorithm (Chen *et al.* 1989) can readily be applied to the RBF model.

3. Radial basis functions

The RBF approximation is a traditional technique for interpolating in multi-dimensional space (Powell 1985, Micchelli 1986) and the theoretical properties of the method have been carefully investigated. By relaxing the original restriction on strict interpolating, Broomhead and Lowe (1988) introduced a generalized form of the RBF expansion. For many practical applications, such as image processing, signal processing and control problems, this generalization provides a more convenient basis to work on.

An RBF expansion with n inputs and a scalar output implements a mapping $f_r: \mathbb{R}^n \to \mathbb{R}$ according to

$$f_r(\mathbf{x}) = \lambda_0 + \sum_{i=1}^{n_r} \lambda_i \phi(\|\mathbf{x} - \mathbf{c}_i\|) \tag{6}$$

where $\mathbf{x} \in \mathbb{R}^n$, $\phi(\cdot)$ is a function from $\mathbb{R}^+$ to $\mathbb{R}$, $\|\cdot\|$ denotes the euclidean norm, λ_i, $0 \leqslant i \leqslant n_r$, are the weights or parameters, $\mathbf{c}_i \in \mathbb{R}^n$, $1 \leqslant i \leqslant n_r$, are the RBF centres and n_r is the number of centres. The functional form $\phi(\cdot)$ is assumed to have been given and the centres $\mathbf{c}_i$ are some fixed points in n-dimensional space. The centres $\mathbf{c}_i$ must appropriately sample the input domain, and they are usually chosen either to be a subset of the data or distributed uniformly in the input domain. Typical choices

for $\phi(\cdot)$ are the thin-plate-spline function

$$\phi(v) = v^2 \log(v) \tag{7}$$

the gaussian function

$$\phi(v) = \exp(-v^2/\beta^2) \tag{8}$$

the multiquadric function

$$\phi(v) = (v^2 + \beta^2)^{1/2} \tag{9}$$

and the inverse multiquadric function

$$\phi(v) = \frac{1}{(v^2 + \beta^2)^{1/2}} \tag{10}$$

where β is a real constant. Why such choices of non-linearity form good RBF approximations is shown, for example, by Powell (1987).

3.1. *Connections with neural networks*

Current interest in RBF within the engineering community is largely motivated by the realization that there are strong connections between RBF and neural networks. Indeed eqn. (6) can be implemented in the two-layered network structure depicted in Fig. 1. Given fixed centres, the first layer performs a fixed non-linear transformation which maps the input space onto a new space, and the output layer implements a linear combiner on this new space. This is basically how an RBF network operates. If the centres are not predetermined and are regarded as adjustable parameters, the structure represented in Fig. 1 then becomes a two-layered feed-forward neural network (Lippmann 1987, Cybenko 1989). Each term $\phi(\|\mathbf{x} - \mathbf{c}_i\|)$ forms a basic component of the network called a hidden neuron. The relationship between RBF and two-layered neural networks was discussed in detail by Broomhead and Lowe (1988).

Because the centres are adjustable parameters in a two-layered neural network, there is more freedom for designing the network to suit given application tasks. Owing to this flexibility, neural networks prove to be highly successful in dealing with complex data. There is, however, a heavy price to pay for this. The structure of the neural network is highly non-linear in the parameters. By carefully pre-fixing centres, it is hoped that RBF networks can closely match the performance of two-layered neural networks and yet have a linear-in-the-parameter formulation. It is important to emphasize that the performance of RBF networks depends critically upon the given centres.

3.2. *Capabilities of radial basis functions*

This subsection briefly discusses how the RBF expansion (6) can be used to realize or to approximate some unknown and complicated nonlinear mapping $f: \mathbb{R}^n \rightarrow \mathbb{R}$. Although the approximation capabilities of RBF have been established by researchers working in multi-dimensional interpolation techniques (e.g. Powell 1987), the strong connections between RBF and neural networks offer a heuristic explanation.

Assume for the time being that the centres are adjustable parameters and consider the two-layered neural network represented in Fig. 1. The surface generated by the gaussian neuron (8) looks like a 'bump' as illustrated in Fig. 2 (*a*) in the two-dimen-

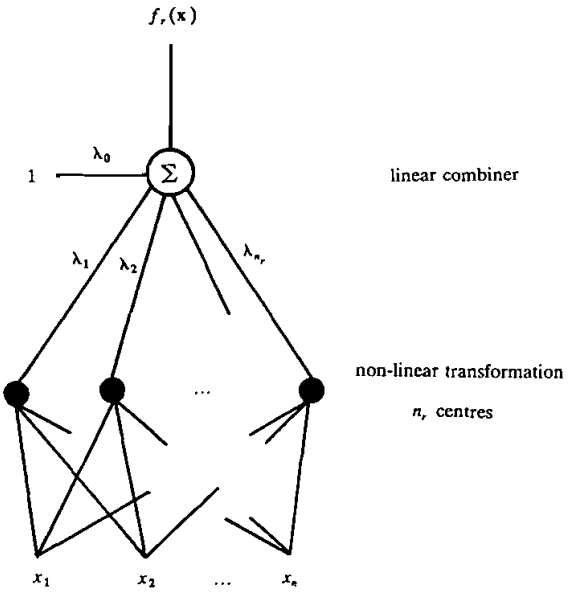


Figure 1. RBF expansion.

sional case. If the first layer of the network consists of enough neurons, sufficient 'bumps' can be generated in appropriate positions by adjusting the parameters of the first layer. A variety of continuous surfaces from $\mathbb{R}^n \to \mathbb{R}$ can thus be well approximated by adding up a series of such 'bumps'. Lapedes and Farber (1988) used these ideas to explain how feedforward neural networks are able to approximate a large class of functions. In an RBF network the centres are all fixed. If, however, a sufficient number of centres are used and they are suitably distributed in the input domain, it is possible that the much simpler RBF model will also provide a reasonable representation for various functions. Obviously a similar conclusion can be reached for the inverse multiquadric non-linearity (10).

The approximation capabilities of the RBF expansion (6) are highly related to its localization properties (Powell 1987). The weights λ_i are generally computed based on observed data $f(\mathbf{x})$. Let $\mathbf{D} \subset \mathbb{R}^n$ be the domain of the approximation and let $\mathbf{D}_0$ be any part of $\mathbf{D}$. $f_r(\mathbf{x})$ is said to have good localization properties if the contribution to $f_r(\mathbf{x})$ from $\{f(\tilde{\mathbf{x}}): \tilde{\mathbf{x}} \in \mathbf{D}_0\}$ is small for $\mathbf{x} \in \mathbf{D}$ that are far away from $\mathbf{D}_0$. Notice that for non-linearities (8) and (10), $\phi(v) \to 0$ as $v \to \infty$. For these two choices of non-linearity the RBF approximation (6) has good localization properties. The surface generated by the thin-plate-spline function (7) in the two-dimensional case is illustrated in Fig. 2 (*b*). It is apparent that for this non-linearity and for the multiquadric non-linearity (9), $\phi(v) \to \infty$ as $v \to \infty$. The surprising results of Powell (1987) indicate that these kinds of RBF approximations can have good localization properties and, in fact, the success of approximation is easier to achieve if $\phi(v) \to \infty$ as $v \to \infty$. These results suggest that the choice of the non-linearity $\phi(\cdot)$ is not crucial to the performance.

Cybenko (1989) has rigorously proved that the two-layered feedforward neural network can uniformly approximate any continuous function. By choosing centres appropriately, it is possible that the simpler RBF model can offer similar approxi-

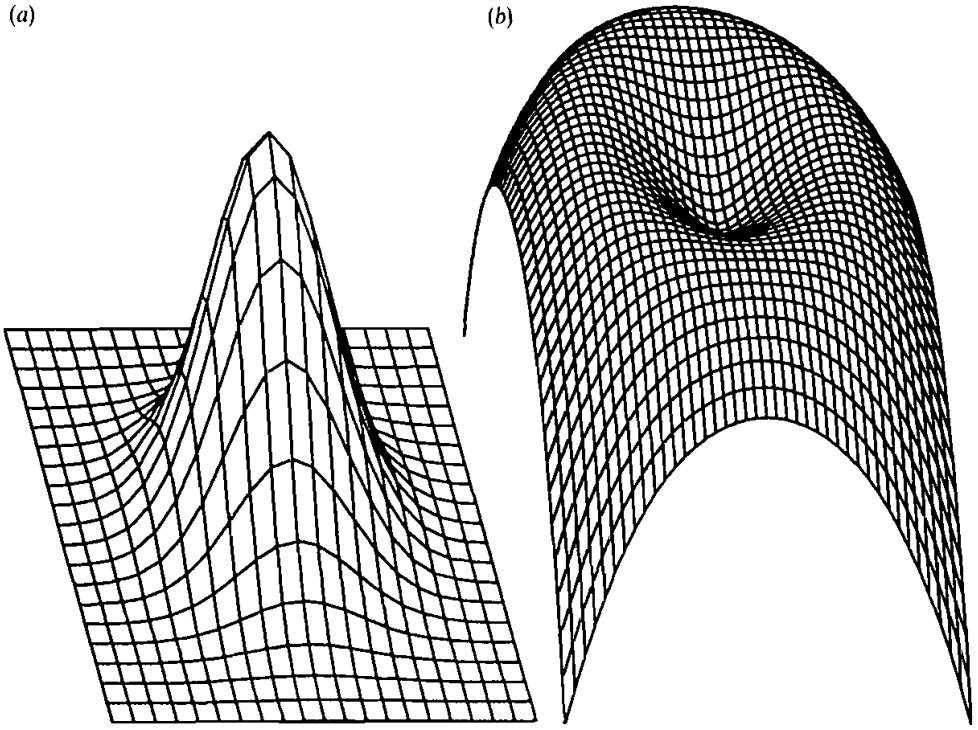


Figure 2. Surface generated by $\phi(\|\mathbf{x}-\mathbf{c}\|)$ in 2-D case: (*a*) gaussian $\phi(\cdot)$; (*b*) thin-plate-spline $-\phi(\cdot)$.

mation capabilities. The crucial problem is then how to select centres appropriately. In practice the centres are normally chosen from the data points. In a real problem the number of data are often large, say, a few hundred. To include all the data as centres would give rise to an extremely complex and possibly useless model. On the other hand, arbitrarily choosing some data points as centres may not always satisfy the requirement that centres should suitably sample the input domain. Fortunately it can readily be shown that the OFR algorithm (Chen *et al.* 1989) can be employed to select centres so that adequate and parsimonious RBF models can be obtained. In order to apply this technique to fit RBF models in the NARMAX expansion, an iterative procedure is required and this is discussed next.

4. Identification in the presence of correlated noise

The objective of the present study is to use the RBF expansion to model non-linear systems described by (2). Define the centre dimension as

$$n = n_y + n_u + n_e \tag{11}$$

and the data vector at sample t as

$$\mathbf{x}(t) = [y(t-1) \quad \ldots \quad y(t-n_y) \quad u(t-1) \quad \ldots \quad u(t-n_u) \quad e(t-1) \quad \ldots \quad e(t-n_e)]^{\mathrm{T}} \tag{12}$$

The RBF expansion $f_r(\mathbf{x}(t))$ is then a one-step-ahead prediction for $y(t)$. The identification thus involves determining centres $\mathbf{c}_i$ and values of λ_i based on the input–

output observations $\{u(t), y(t)\}_{t=1}^{N}$, where N is the data length. Because the system noise $e(t)$ is generally unobserved, it can only be replaced by the prediction error or residual $\varepsilon(t)$ and $\mathbf{x}(t)$ must be modified accordingly

$$\mathbf{x}(t) = [y(t-1) \quad \ldots \quad y(t-n_y) \quad u(t-1) \quad \ldots \quad u(t-n_u) \quad \varepsilon(t-1) \quad \ldots \quad \varepsilon(t-n_e)]^{\mathrm{T}} \tag{13}$$

First a general linear regression model which includes the RBF expansion as a special case is introduced:

$$y(t) = \sum_{i=1}^{M} p_i(t)\,\theta_i + \varepsilon(t) \tag{14}$$

where $\theta_i(t)$ are known as the regressors, which are some fixed non-linear functions of lagged outputs, inputs and prediction errors. That is

$$p_i(t) = p_i(\mathbf{x}(t)) \tag{15}$$

with $\mathbf{x}(t)$ defined in (13). A constant term is included in (14) by setting the corresponding $p_i(t) = 1$. $y(t)$ is also called the dependent variable. $\varepsilon(t)$ is assumed to be uncorrelated with the regressors $p_i(t)$. The set that contains a variety of $p_i(t)$ is referred to as the extended model set (Billings and Chen 1989 b). The model (14), although linear-in-the-parameters, provides a very rich description of non-linear systems. Effective modelling can be achieved by selecting a parsimonious subset model from the extended model set.

It is obvious that a given centre $\mathbf{c}_i$ with a fixed nonlinearity $\phi(\cdot)$ corresponds to a regressor $p_i(t)$ in (14). The RBF expansion is therefore a particular choice of the extended model set representation. The problem of how to select a subset of centres from a large number of candidate centres can thus be regarded as an example of how to select a subset of significant regressors from a given extended model set. An efficient technique for selecting a subset model from (14) has previously been developed (Chen *et al.* 1989) and is briefly described below.

4.1. *Orthogonal forward regression*

Equation (14) for $t = 1, \ldots, N$ can be collected together in the matrix form

$$\mathbf{y} = \mathbf{P}\Theta + E \tag{16}$$

where

$$\left.\begin{aligned} \mathbf{y} &= \begin{bmatrix} y(1) \\ \vdots \\ y(N) \end{bmatrix}, \quad \mathbf{P} = [\mathbf{p}_1 \quad \ldots \quad \mathbf{p}_M] \\ \Theta &= \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_M \end{bmatrix}, \quad E = \begin{bmatrix} \varepsilon(1) \\ \vdots \\ \varepsilon(N) \end{bmatrix} \end{aligned}\right\} \tag{17}$$

$$\mathbf{p}_i = [p_i(1) \quad \ldots \quad p_i(N)]^{\mathrm{T}}, \quad 1 \leqslant i \leqslant M \tag{18}$$

$\mathbf{P}$, known as the regression matrix, can be decomposed into

$$\mathbf{P} = \mathbf{WA} \tag{19}$$

where

$$\mathbf{A} = \begin{bmatrix} 1 & \alpha_{12} & \alpha_{13} & \cdots & \alpha_{1M} \\ & 1 & \alpha_{23} & \cdots & \alpha_{2M} \\ & & \ddots & \ddots & \vdots \\ & & & 1 & \alpha_{M-1M} \\ & & & & 1 \end{bmatrix} \tag{20}$$

is an $M \times M$ unit upper triangular matrix and

$$\mathbf{W} = [\mathbf{w}_1 \quad \cdots \quad \mathbf{w}_M] \tag{21}$$

is an $N \times M$ matrix with orthogonal columns that satisfy

$$\mathbf{W}^T\mathbf{W} = \mathbf{H} \tag{22}$$

Here $\mathbf{H}$ is a positive diagonal matrix

$$\mathbf{H} = \begin{bmatrix} h_1 & & 0 \\ & \ddots & \\ 0 & & h_M \end{bmatrix} \tag{23}$$

with

$$h_i = \mathbf{w}_i^T\mathbf{w}_i = \sum_{t=1}^{N} w_i(t)w_i(t), \quad 1 \leqslant i \leqslant M \tag{24}$$

Equation (16) can be rearranged as

$$\mathbf{y} = (\mathbf{P}\mathbf{A}^{-1})(\mathbf{A}\Theta) + E = \mathbf{W}\mathbf{g} + E \tag{25}$$

where

$$\mathbf{A}\Theta = \mathbf{g} \tag{26}$$

Because $\varepsilon(t)$ is uncorrelated with the $p_i(t)$, it is easy to show that

$$\mathbf{g} = \mathbf{H}^{-1}\mathbf{W}^T\mathbf{y} \tag{27}$$

or

$$g_i = \frac{\mathbf{w}_i^T\mathbf{y}}{\mathbf{w}_i^T\mathbf{w}_i}, \quad 1 \leqslant i \leqslant M \tag{28}$$

Techniques for solving for the parameter estimate $\hat{\Theta}$ from the triangular system (26) are commonly known as orthogonal least-squares methods. The classical Gram-Schmidt and modified Gram-Schmidt methods (Björck 1967) can be used to derive the triangular system (26). A similar orthogonal decomposition of $\mathbf{P}$ can be obtained using the Householder transformation method (Golub 1965).

The number M of all the candidate regressors can be very large, as for, example, in the case of the RBF expansion. Adequate modelling may only require $M_s(\ll M)$ significant regressors. These significant regressors can be identified using the OFR algorithm (Chen *et al.* 1989) derived by modifying and augmenting the orthogonal

least-squares methods. From (25), the sum of squares of $y(t)$ is

$$\mathbf{y}^{\mathrm{T}}\mathbf{y} = \sum_{i=1}^{M} g_i^2 \mathbf{w}_i^{\mathrm{T}}\mathbf{w}_i + E^{\mathrm{T}}E \tag{29}$$

It is seen that each $\mathbf{w}_i$ explains a proportion of the dependent variable variance. Therefore the error reduction ratio due to $\mathbf{w}_i$ is defined as

$$[err]_i = \frac{g_i^2 \mathbf{w}_i^{\mathrm{T}}\mathbf{w}_i}{\mathbf{y}^{\mathrm{T}}\mathbf{y}}, \quad 1 \leqslant i \leqslant M \tag{30}$$

The error reduction ratio offers a simple and effective means of selecting a subset of significant regressors from a large number of candidates in a forward-regression manner. At the jth step, a regressor is selected if it produces the largest value of $[err]_j$ from among the rest of the candidates. The selection procedure is terminated when

$$1 - \sum_{j=1}^{M_s} [\mathrm{err}]_j < \rho \tag{31}$$

where $0 < \rho < 1$ is a chosen tolerance. The parameter estimate $\hat{\Theta}_s$ for the resulting subset model is then computed from

$$\mathbf{A}_s\Theta_s = \mathbf{g}_s \tag{32}$$

where $\mathbf{A}_s$ is the $M_s \times M_s$ unit upper triangular matrix. The detailed selection procedure is described by Chen *et al.* (1989) and Billings and Chen (1989 b).

The desired tolerance ρ can actually be learnt during the forward-regression procedure so that the regressor selection becomes an automatic procedure. This aspect is discussed by Billings and Chen (1989 b). Alternatively, the forward-regression procedure can be terminated using Akaike's information criterion and the detailed implementation of this is also given by Chen *et al.* (1989) and Billings and Chen (1989 b).

4.2. *Model validity tests*

If the modelling is adequate, the residual $\varepsilon(t)$ should be unpredictable from (uncorrelated with) all linear and nonlinear combinations of past inputs and outputs, and this can be tested by means of the following correlation functions (Billings and Voon 1986, Billings and Chen 1989 a):

$$\left.\begin{aligned} \Psi_{\varepsilon\varepsilon}(k) &= \delta(t) \text{ an impulse function} \\ \Psi_{u\varepsilon}(k) &= 0, \quad \forall\, k \\ \Psi_{\varepsilon(\varepsilon u)}(k) &= 0 \quad k \geqslant 0 \\ \Psi_{u^{2\prime}\varepsilon}(k) &= 0, \quad \forall\, k \\ \Psi_{u^{2\prime}\varepsilon^2}(k) &= 0, \quad \forall\, k \end{aligned}\right\} \tag{33}$$

where $\Psi_{xz}(k)$ indicates the cross-correlation function between $x(t)$ and $z(t)$, $\varepsilon u(t) = \varepsilon(t+1)\,u(t+1)$,

$$u^{2\prime}(t) = u^2(t) - \overline{u^2(t)}$$

and $\overline{u^2(t)}$ represents the time average or mean value of $u^2(t)$.

The tests (33) were developed based on the fact that $\Psi_{\varepsilon\varepsilon}(k)$ and $\Psi_{u\varepsilon}(k)$ alone are

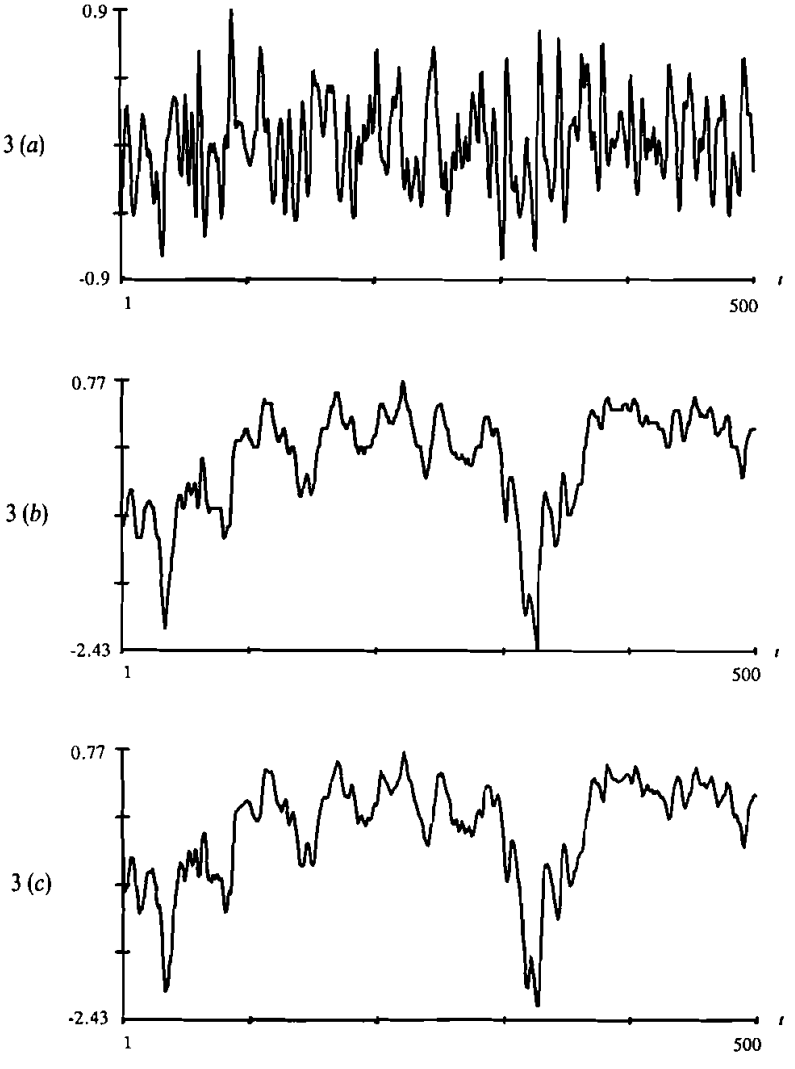
not sufficient to validate non-linear models and may even give misleading information regarding the adequacy of the fitted model. The higher-order correlation tests are thus included. In general, if the correlation functions in (33) are within the 95% confidence intervals, $\pm 1{\cdot}96/\sqrt{N}$, the model is regarded as adequate.

Alternatively, a chi-squared statistical test (Leontaritis and Billings 1987) can be employed to validate the identified model. Define $\Omega(t)$ as an s-dimensional vector-valued function of the past inputs, outputs and prediction errors, and let

$$\Gamma^{\mathrm{T}}\Gamma = \frac{1}{N}\sum_{t=1}^{N} \Omega(t)\,\Omega^{\mathrm{T}}(t) \tag{34}$$

Then the chi-squared statistic is calculated from the formula

$$\zeta = N\mu^{\mathrm{T}}(\Gamma^{\mathrm{T}}\Gamma)^{-1}\mu \tag{35}$$



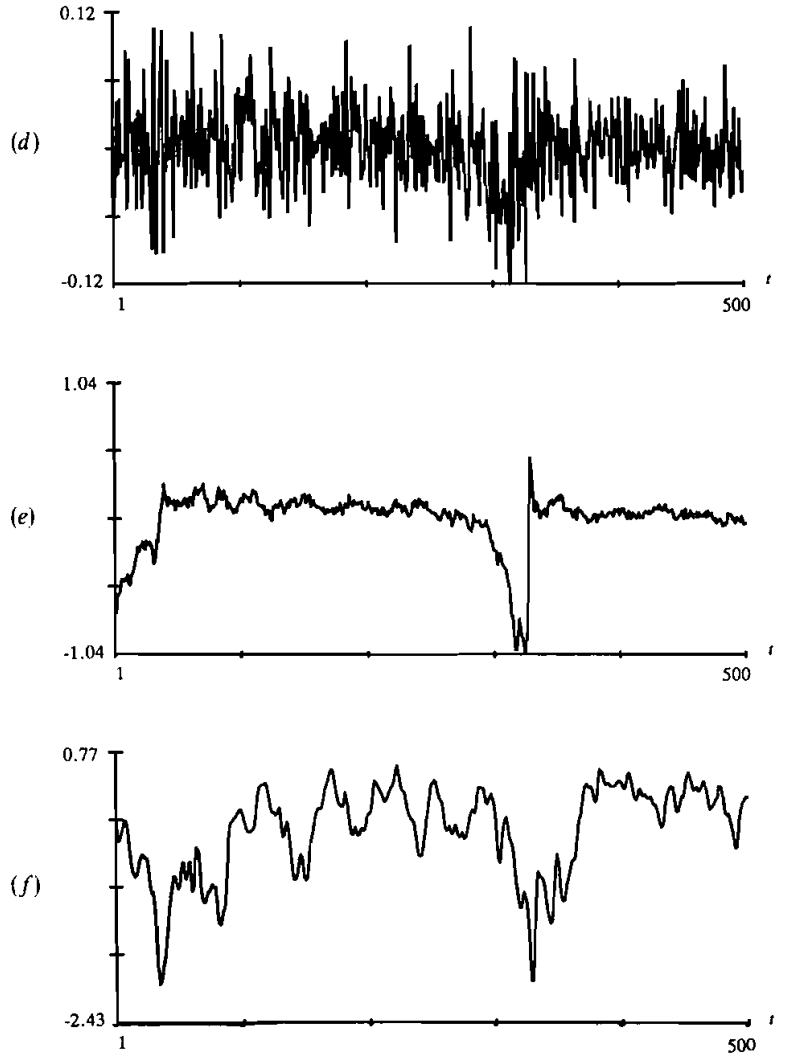


Figure 3. Data set and RBF model response (Example 1); (*a*) $u(t)$, (*b*) $y(t)$, (*c*) $\hat{y}(t)$, (*d*) $\varepsilon(t)$, (*e*) $\varepsilon_d(t)$, (*f*) $\hat{y}_d(t)$.

where

$$\mu = \frac{1}{N} \sum_{t=1}^{N} \Omega(t)\, \varepsilon(t)/\sigma_\varepsilon \tag{36}$$

and σ_ε^2 is the variance of the residual $\varepsilon(t)$. Under the null hypothesis that the data are generated by the model, the statistic ζ is asymptotically chi-squared distributed with s degrees of freedom. A convenient choice for $\Omega(t)$ is

$$\Omega(t) = [\omega(t) \quad \omega(t-1) \quad \ldots \quad \omega(t-s+1)]^{\mathrm{T}} \tag{37}$$

where $\omega(t)$ is some chosen function of the past inputs, outputs and prediction errors. If the values of ζ for several different choices of $\omega(t)$ are within the 95% acceptance

i	Weight estimate	Error reduction ratio
1	−3·59759e−01	7·21759e−01
2	3·15711e−02	2·43152e−01
3	−7·56932e−03	1·22606e−02
4	8·71656e−02	2·60421e−03
5	−6·28845e−01	3·40850e−03
6	−2·61171e−02	1·90152e−03
7	1·09946e−01	1·92118e−03
8	−2·93926e−01	1·36604e−03
9	−1·44274e−01	3·50552e−04
10	−1·11876e−01	5·20098e−04
11	−5·62868e−01	3·18867e−04
12	−2·31590e+00	1·39608e−04
13	1·94153e+00	1·75716e−04
14	−1·31791e+00	1·91326e−04
15	2·24206e−01	1·15201e−04
16	3·41577e−01	9·53731e−05
17	6·15907e−01	1·35794e−04
0	1·63431e+00	1·87789e−04
18	−1·11642e−01	2·48672e−04
19	2·01416e−01	7·19119e−04
20	4·74203e−01	2·61420e−04
21	−5·93372e−02	1·91899e−04
22	7·04118e−01	1·12029e−04
23	1·27583e−01	1·15415e−04
24	−8·54994e−02	1·37082e−04
25	−1·26371e+00	8·72300e−05
26	1·07860e+00	1·51230e−04
27	2·77538e−01	2·34625e−04
28	1·62097e−01	4·73117e−05
29	2·39621e−01	6·19713e−05
30	9·95672e−01	5·88282e−05
31	−1·47061e−01	5·05323e−05
32	−7·23539e−02	8·14058e−05
33	1·19353e−01	7·36802e−05
σ_e^2	2·56156e−03	

Table 1. Subset RBF model of Example 1.

region, that is

$$\zeta < \chi_s^2(\alpha) \tag{38}$$

the model is regarded as adequate, where $\chi_s^2(\alpha)$ is the critical value of the chi-squared distribution with s degrees of freedom for the given significance level α (0·05). The choices of $\omega(t)$ should generally include some non-linear functions.

4.3. *Iterative procedure*

By combining the OFR routine and the model validity tests, an iterative procedure can be formed to identify NARMAX models using the RBF expansion. The non-linearity $\phi(\cdot)$ in the RBF expansion will be chosen as the thin-plate-spline function (7). Apart from its good global interpolating property, the thin-plate-spline function does not require the extra parameter β and, therefore, is slightly more convenient to

i	c_{1i} c_{6i}	c_{2i} c_{7i}	c_{2i} c_{8i}	c_{4i} c_{9i}	c_{5i}
0			constant		
1	7·61719e−01 5·17578e−01	7·03125e−01 3·31253e−02	5·46875e−01 1·00582e−01	5·85940e−02 −1·34814e−03	4·29688e−01
2	−8·20313e−01 0·00000e+00	0·00000e+00 −7·86912e−01	0·00000e+00 0·00000e+00	−4·68750e−01 0·00000e+00	0·00000e+00
3	−5·66406e−01 −4·98047e−01	−3·12500e−01 1·82023e−02	−9·76560e−02 −7·58298e−02	−7·51953e−01 −2·14319e−02	−7·61719e−01
4	7·03125e−01 4·19922e−01	5·46875e−01 1·00582e−01	4·88281e−01 −1·34814e−03	4·29688e−01 7·92319e−03	5·17578e−01
5	4·68750e−01 6·54297e−01	4·10156e−01 −2·48124e−04	3·32031e−01 −4·25553e−02	4·00391e−01 1·57459e−02	5·76172e−01
6	−7·42188e−01 −2·05078e−01	−6·44531e−01 7·27092e−02	−5·27344e−01 −2·12200e−02	−4·68750e−01 −2·26521e−03	−4·19922e−01
7	3·32031e−01 1·66016e−01	2·73438e−01 −3·67755e−03	2·34375e−01 −5·46013e−03	2·05078e−01 −8·74925e−03	3·12500e−01
8	−3·90625e−01 −3·22266e−01	−2·53906e−01 −5·90254e−02	−1·56250e−01 −1·16631e−02	−9·76560e−02 −3·37688e−03	−3·32031e−01
9	2·92969e−01 −1·85547e−01	1·95313e−01 6·11390e−03	1·95313e−01 4·42455e−02	6·44531e−01 2·98252e−02	2·73438e−01
10	−7·42188e−01 −1·07422e−01	−7·42188e−01 1·71060e−02	−7·42188e−01 3·38346e−02	−2·73438e−01 1·63228e−02	−8·78910e−02
11	4·88281e−01 −2·05078e−01	5·46875e−01 1·05537e−02	6·83594e−01 1·24075e−02	−1·95313e−01 2·62395e−03	−2·92969e−01
12	2·73438e−01 2·83203e−01	1·95313e−01 2·59003e−02	1·17188e−01 8·66169e−03	2·44141e−01 1·09415e−02	2·73438e−01
13	2·73438e−01 1·66016e−01	1·95313e−01 1·08582e−02	1·95313e−01 −2·29179e−02	3·32031e−01 −1·96630e−02	3·12500e−01
14	3·32031e−01 1·75781e−01	2·53906e−01 3·05088e−02	2·73438e−01 −4·73855e−02	3·22266e−01 2·13513e−02	2·24609e−01
15	5·46875e−01 6·73828e−01	5·46875e−01 −6·09297e−03	4·88281e−01 −2·70075e−02	1·36719e−01 −1·59142e−03	4·49219e−01
16	3·71094e−01 −1·36719e−01	3·32031e−01 2·84963e−02	4·10156e−01 −4·50973e−02	3·90630e−02 5·01287e−02	2·92970e−02
17	3·32031e−01 4·78516e−01	2·53906e−01 −6·77300e−03	1·36719e−01 −2·77793e−03	2·63672e−01 1·00440e−01	4·88281e−01
18	−1·46484e+00 −7·03125e−01	−1·97266e+00 5·18907e−02	−2·42188e+00 9·36738e−02	8·78910e−02 −1·48592e−01	−4·29688e−01
19	−1·01562e+00 −1·46484e−01	−1·09375e+00 −8·23599e−03	−1·09375e+00 2·16791e−02	9·76560e−02 1·00404e−02	−3·90630e−02
20	3·90625e−01 −1·07422e−01	4·88281e−01 −2·30838e−02	4·88281e−01 3·95381e−02	−3·32031e−01 −3·29906e−02	−2·44141e−01
21	−6·64063e−01 −4·58984e−01	−8·98438e−01 6·66557e−03	−7·61719e−01 −1·09135e−01	4·98047e−01 2·92522e−02	9·76560e−02
22	4·10156e−01 4·39453e−01	4·10156e−01 −2·26586e−03	3·51563e−01 3·59494e−02	1·95313e−01 2·59324e−02	3·12500e−01
23	−3·51563e−01 6·54297e−01	−3·12500e−01 3·76687e−02	−4·49219e−01 4·47573e−02	−2·44141e−01 −2·61394e−02	2·53906e−01
24	−7·81250e−02 5·76172e−01	−3·90625e−01 4·29491e−02	−6·83594e−01 6·19762e−03	8·39844e−01 1·02739e−01	8·98438e−01
25	−1·95310e−02 −8·78910e−02	−1·95310e−02 3·04733e−02	0·00000e+00 4·49002e−03	4·88280e−02 −5·26021e−03	−5·85940e−02
26	−9·76560e−02 −2·63672e−01	−9·76560e−02 1·75251e−02	−7·81250e−02 2·96281e−02	−4·88280e−02 3·46041e−02	−8·78910e−02

(*continued*)

27	5·85940e−02	5·85940e−02	5·85940e−02	1·56250e−01	1·26953e−01
	1·07422e−01	1·09151e−02	6·15147e−03	1·29895e−02	
28	−8·20313e−01	−6·64063e−01	−4·49219e−01	−3·32031e−01	−4·78516e−01
	−5·17578e−01	−3·14039e−02	−1·43116e−02	7·48491e−02	
29	−1·30849e+00	−1·52344e+00	−1·66016e+00	1·75781e−01	9·76560e−02
	5·85940e−02	7·29334e−02	4·52477e−02	2·34002e−02	
30	1·95313e−01	1·36719e−01	1·17188e−01	2·05078e−01	1·95313e−01
	2·73438e−01	3·72316e−02	−8·05099e−03	−1·27434e−02	
31	−9·57031e−01	−1·25000e+00	−1·46484e+00	7·61719e−01	5·07813e−01
	8·78910e−02	5·80296e−02	−2·95292e−02	5·18907e−02	
32	−1·87500e+00	−1·79688e+00	−1·75781e+00	−1·95313e−01	−1·17188e−01
	4·88280e−02	−3·79891e−03	−6·15757e−02	−9·60702e−02	
33	−3·90625e−01	−6·83594e−01	−9·37500e−01	8·98438e−01	5·76172e−01
	1·07422e−01	6·19762e−03	1·02739e−01	−1·28458e−02	

Table 2. Subset RBF centres of Example 1.

use in the regressor selection than the other choices (8)–(10). The identification procedure can be summarized as follows.

(i) Choose n_y, n_u and n_e. Initially the set of candidate centres are all the

$$\mathbf{x}_d(t) = [y(t-1) \quad \ldots \quad y(t-n_y) \quad u(t-1) \quad \ldots \quad u(t-n_u)]^{\mathrm{T}} \tag{39}$$

A subset RBF model $f_r^{(0)}$ is selected using the OFR algorithm and this initial model is used to generate the initial prediction error sequence $\{\varepsilon^{(0)}(t)\}$.

(ii) An iterative loop is then entered to update the model. At the kth iteration the set of candidate centres are all the

$$\mathbf{x}^{(k)}(t) = [y(t-1) \quad \ldots \quad y(t-n_y) \quad u(t-1) \quad \ldots \quad u(t-n_u) \quad \varepsilon^{(k-1)}(t-1) \quad \ldots \quad \varepsilon^{(k-1)}(t-n_e)]^{\mathrm{T}} \tag{40}$$

The subset RBF model $f_r^{(k)}$ is selected by the OFR algorithm and this gives rise to the prediction error sequence $\{\varepsilon^{(k)}(t)\}$. Typically two to four iterations are sufficient.

(iii) The model validity tests are performed to assess the model. If the model is considered adequate the procedure is terminated. Otherwise go to step (i).

Some refinements to the above procedure are also possible. Whenever there are sufficient data points, the data should be partitioned into a fitting set and a testing set. The former is used in the subset model selection and the latter is used to validate the selected model. If modelling complex data is regarded as performing a curve fitting in a high-dimensional space, the two basic concepts in neural networks, namely learning and generalization, becomes relevant to systems identifiaction. From this viewpoint, learning corresponds to producing a surface in multidimensional space that fits the set of data in some best sense, and generalization is equivalent to interpolating the test data set on this fitting surface (Broomhead and Lowe 1988). The performance of generalization is a more 'genuine' indicator for the 'goodness' of the model. As discussed by Chen *et al.* (1990), the regressor selection can be terminated using the following alternative criterion. The subset model is chosen on the fitting set and the variance of the residuals over the testing set is computed based on the selected model. The selection is terminated when this generalization error is minimized.

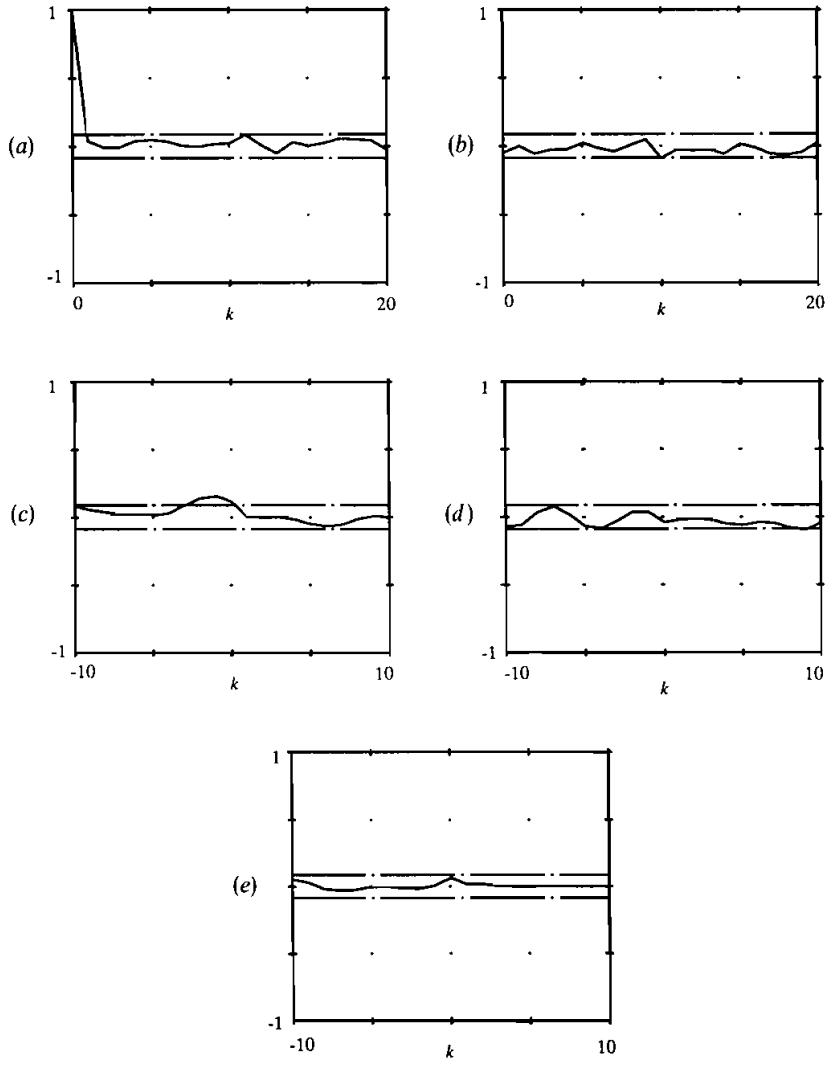


Figure 4. Correlation tests (Example 1); (*a*) $\Psi_{\varepsilon\varepsilon}(k)$, (*b*) $\Psi_{\varepsilon(\varepsilon u)}(k)$, (*c*) $\Psi_{u\varepsilon}(k)$, (*d*) $\Psi_{u^2\varepsilon}(k)$, (*e*) $\Psi_{u^2\varepsilon^2}(k)$. Dashed line 95% confidence interval.

5. Comparison with fitting polynomial NARMAX models

The iterative procedure for fitting subset RBF models is very similar to that for fitting subset polynomial models (Chen *et al.* 1989). Here we do not attempt to assess which model is superior to the other in representing non-linear systems. Rather, a brief comparison of the two models is discussed from the practical identification viewpoint.

For a polynomial expansion, the number of all the candidate regressors or monomials is determined by the formula

$$m = \sum_{i=0}^{l} m_i \tag{41}$$

i	Weight estimate	Error reduction ratio
0	4·86518e+00	9·74671e−01
1	2·89274e−03	2·43432e−02
2	−1·67923e−02	1·71149e−04
3	−8·31933e−02	3·85502e−04
4	5·60050e−01	7·91428e−05
5	−8·85746e−03	1·17319e−04
6	6·00541e−02	6·33583e−05
7	−1·64703e+00	2·06132e−05
8	1·39768e+00	2·35189e−05
9	−1·70597e−02	9·05632e−06
10	−5·41178e−02	7·50023e−06
11	−2·62244e−03	1·57312e−05
12	−7·92890e−01	1·84051e−05
13	1·61505e−01	6·41244e−06
14	7·44306e−01	7·31515e−06
15	6·42142e−02	8·62033e−06
16	−1·51422e−01	4·33155e−06
17	2·63342e−01	3·78649e−06
18	−6·78411e−01	3·35450e−06
19	2·19646e−01	2·45880e−06
20	2·27489e−01	2·18753e−06
21	−1·61403e+00	2·07859e−06
22	1·42335e+00	1·84200e−06
23	−6·57795e−01	1·07429e−06
24	−5·47704e−02	2·21430e−06
25	−9·54563e−01	9·12473e−01
26	2·03376e−01	6·40319e−07
27	4·37858e−02	1·01682e−06
28	7·30585e−01	7·67175e−07
29	−8·72148e−02	6·81383e−07
30	−2·25814e−01	9·00494e−07
31	9·23551e−01	1·06634e−06
32	5·54911e−02	1·08590e−06
σ_e^2	3·50978e−04	

Table 3. Subset RBF model of Example 2.

where l is the polynomial degree and

$$m_0 = 1, \quad m_i = m_{i-1}(n_y + n_u + n_e + i - 1)/i, \quad 1 \leqslant i \leqslant l \tag{42}$$

It is obvious that the number of monomials increases exponentially as l increases. In practice, therefore, l must be restricted, say, to 3. As an example, assume that $n_y = n_u = n_e = 7$ and $l = 3$; then $m = 2024$. Although the OFR algorithm is very efficient in subset selection, choosing a subset model from such a large set is still computationally intensive. The dimension of the candidate set for RBF modelling, on the other hand, is determined by the data length N, which is typically a few hundred. Increasing n_y, n_u or n_e does not affect the number of candidate regressors. This gives the RBF approach advantages in fitting non-linear systems with large lags in the input, output or noise.

NARMAX modelling requires the replacement of the unobserved noise $e(t)$ by the prediction error $\varepsilon(t)$. The issue of invertibility (Chen and Billings 1989 b), or

i	c_{1i} c_{6i}	c_{2i} c_{7i}	c_{3i} c_{8i}	c_{4i} c_{9i}	c_{5i}
0			constant		
1	3·11380e+00 0·00000e+00	0·00000e+00 −7·37578e−04	0·00000e+00 0·00000e+00	3·72450e+00 0·00000e+00	0·00000e+00
2	3·06030e+00 4·53940e+00	2·87700e+00 −2·50678e−02	2·77010e+00 −2·35328e−02	3·76330e+00 2·08644e−03	3·76330e+00
3	3·02980e+00 3·74390e+00	3·22840e+00 7·53651e−03	3·20540e+00 1·08585e−02	3·74390e+00 −2·16087e−02	3·74390e+00
4	3·32000e+00 4·52000e+00	3·15200e+00 −2·17232e−02	3·24360e+00 −4·45968e−02	3·74390e+00 3·73825e−02	4·52000e+00
5	3·22840e+00 3·74390e+00	3·12910e+00 3·62243e−02	3·35820e+00 −1·69932e−02	5·80060e+00 −1·63574e−02	5·80060e+00
6	3·15200e+00 3·74390e+00	3·24360e+00 −4·45968e−02	3·43460e+00 3·73825e−02	4·52000e+00 −1·83684e−02	4·52000e+00
7	3·16730e+00 3·74390e+00	3·35820e+00 3·12243e−02	3·35050e+00 −1·84543e−03	5·80060e+00 −9·11092e−02	3·74390e+00
8	3·15960e+00 3·74390e+00	3·36580e+00 9·75430e−03	3·35820e+00 1·06606e−03	5·82000e+00 −1·63189e−02	3·74390e+00
9	2·94580e+00 0·00000e+00	3·11380e+00 1·74804e−03	0·00000e+00 −7·37578e−04	4·53940e+00 0·00000e+00	3·72450e+00
10	4·36630e+00 5·82000e+00	4·16770e+00 4·02729e−02	3·93100e+00 1·31542e−02	5·04390e+00 −5·12454e−02	5·82000e+00
11	4·29760e+00 5·04390e+00	4·35100e+00 −2·79905e−02	4·36630e+00 1·00483e−02	5·04390e+00 4·02729e−02	5·04390e+00
12	3·39640e+00 4·52000e+00	3·20540e+00 7·08318e−03	3·24360e+00 9·57639e−03	3·74390e+00 1·66295e−02	4·52000e+00
13	4·61070e+00 5·82000e+00	4·63360e+00 −2·99789e−02	4·63360e+00 1·85691e−03	5·04390e+00 −3·15593e−02	5·02390e+00
14	3·08320e+00 3·74390e+00	3·09850e+00 −1·14265e−02	3·01450e+00 −2·51918e−02	4·52000e+00 2·13655e−02	4·52000e+00
15	3·45750e+00 3·74390e+00	3·70190e+00 1·58350e−02	3·93100e+00 −4·67661e−02	3·74390e+00 7·53581e−03	3·74390e+00
16	4·71000e+00 5·82000e+00	4·73290e+00 3·58209e−02	4·63360e+00 7·76617e−03	5·04390e+00 2·05549e−02	5·04390e+00
17	3·38870e+00 4·53940e+00	3·18250e+00 −8·71494e−03	3·22070e+00 −2·42880e−03	3·74390e+00 −5·96622e−03	4·53940e+00
18	3·11380e+00 3·72450e+00	3·10620e+00 1·97229e−02	3·01450e+00 −1·64108e−02	4·53940e+00 −4·70489e−04	4·53940e+00
19	3·19780e+00 3·74390e+00	3·41160e+00 −1·04871e−03	3·41930e+00 1·57596e−02	5·82000e+00 −1·22219e−02	3·74390e+00
20	4·48080e+00 5·02450e+00	4·39680e+00 2·00877e−03	4·45030e+00 9·52153e−03	5·02450e+00 −1·29715e−02	5·80060e+00
21	4·63360e+00 5·04390e+00	4·63360e+00 1·85691e−03	4·64890e+00 −3·15593e−02	5·04390e+00 −5·14224e−03	5·82000e+00
22	4·64120e+00 5·94389e+00	4·64120e+00 −1·66652e−02	4·65650e+00 −1·01000e−03	5·04390e+00 −5·10076e−03	5·82000e+00
23	4·46560e+00 5·00510e+00	4·50380e+00 2·46727e−02	4·46560e+00 1·79497e−02	5·80060e+00 1·02457e−03	5·00510e+00
24	4·34340e+00 5·94389e+00	4·25170e+00 3·65379e−02	4·28230e+00 −1·39611e−02	5·82000e+00 −2·36748e−02	5·82000e+00
25	4·63360e+00 5·82000e+00	4·64890e+00 −3·15593e−02	4·64890e+00 −5·14224e−03	5·82000e+00 −3·59283e−02	5·04390e+00
26	3·41160e+00 4·53940e+00	3·41930e+00 1·57596e−02	3·15200e+00 −1·22219e−02	3·74390e+00 −9·48673e−04	3·74390e+00

(*continued*)

27	2·86940e+00 3·74390e+00	2·74720e+00 6·53485e−04	2·86940e+00 −1·25916e−02	3·74390e+00 −1·28845e−02	4·52000e+00
28	4·48850e+00 5·04390e+00	4·53430e+00 1·68135e−02	4·48850e+00 2·47397e−02	5·82000e+00 −1·46066e−02	5·04390e+00
29	3·22840e+00 3·74390e+00	3·29710e+00 −4·34802e−03	3·41930e+00 −4·91869e−02	3·74390e+00 5·99918e−02	4·52000e+00
30	3·24360e+00 4·50060e+00	3·24360e+00 8·15524e−03	3·32000e+00 2·77028e−04	3·74390e+00 4·73018e−03	3·74390e+00
31	4·60300e+00 5·80060e+00	4·61830e+00 −2·70984e−02	4·62590e+00 −6·20406e−03	5·80060e+00 −3·75853e−02	5·02450e+00
32	2·77010e+00 3·74390e+00	2·82360e+00 3·57718e−02	2·98400e+00 −1·16939e−02	4·50060e+00 −5·22487e−02	3·74390e+00

Table 4. Subset RBF centres of Example 2.

Term	Weight estimate
constant	−1·8647e+00
$y(t-1)$	2·5121e−01
$u(t-1)$	4·9787e−01
$u(t-2)$	8·3088e−01
$\varepsilon(t-1)$	2·6157e−01
$\varepsilon(t-2)$	−3·9234e−01
$\varepsilon(t-3)$	7·9580e−01
$\varepsilon(t-4)$	5·8926e−01
$\varepsilon(t-5)$	9·7346e−01
$y^2(t-1)$	5·5328e−02
$y(t-1)\,u(t-1)$	5·4122e−02
$y(t-1)\,u(t-2)$	−5·0123e−02
$y(t-1)\,\varepsilon(t-2)$	5·0442e−01
$y(t-1)\,\varepsilon(t-3)$	6·1282e−01
$y(t-1)\,\varepsilon(t-4)$	6·0015e−01
$u^2(t-1)$	−5·8763e−02
$u(t-1)\,\varepsilon(t-2)$	−2·9549e−01
$u(t-1)\,\varepsilon(t-3)$	−3·3093e−01
$u(t-1)\,\varepsilon(t-4)$	−2·1994e−01
$u(t-1)\,\varepsilon(t-5)$	−1·5602e−01
$u^2(t-2)$	−4·8272e−02
$u(t-2)\,\varepsilon(t-3)$	−2·5665e−01
$u(t-2)\,\varepsilon(t-4)$	−3·3229e−01
σ_ε^2	6·32261e−04

Table 5. Polynomial subset model of Example 2.

whether it is possible to compute $\varepsilon(t)$ using the model and given system inputs and outputs, becomes crucial. The authors' experience suggests that the RBF expansion is less likely to produce explosive $\varepsilon(t)$ sequences than the polynomial expansion.

The polynomial approach has its advantages. Polynomial modelling often requires fewer terms or is more parsimonious than RBF modelling. A subset polynomial model with less than 20 terms is usually sufficient to capture the dynamics of highly non-linear processes, while a subset RBF model typically requires 30–40 terms. Given the same number of terms or parameters, the computation of the RBF mapping f_r is obviously more expensive than that of the polynomial mapping.

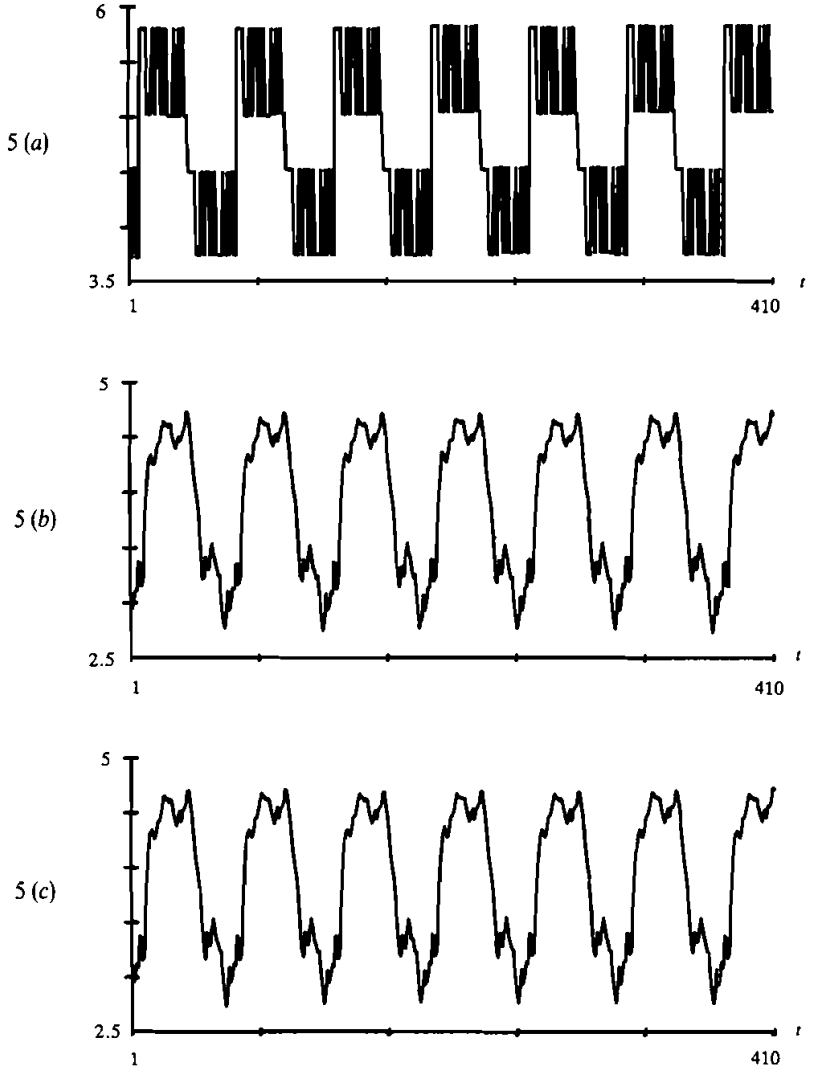
6. Identification of two real processes

6.1. *Example* 1

The process considered is a liquid level system. The system consists of a DC water pump feeding a conical flask which in turn feeds a square tank. The system input is the voltage to the pump motor and the system output is the water level in the conical flask. 500 points of data generated in an experiment are shown in Fig. 3.

The iterative procedure described in § 4.3 was used to identify a subset RBF model. The lags n_y, n_u and n_e were all chosen to be 3 and the number of all the candidate regressors was $M = 501$ (500 data vectors plus constant term). Three iterations were involved in the identification procedure, and an appropriate tolerance $\rho = 0{\cdot}0068$ was learnt during the identification. The selected subset model included a constant term and 33 centres. The final model thus takes the form

$$\hat{y}(t) = \hat{f}_r(\mathbf{x}(t)) = \hat{\lambda}_0 + \sum_{i=1}^{33} \hat{\lambda}_i \phi(\|\mathbf{x}(t) - \mathbf{c}_i\|)$$



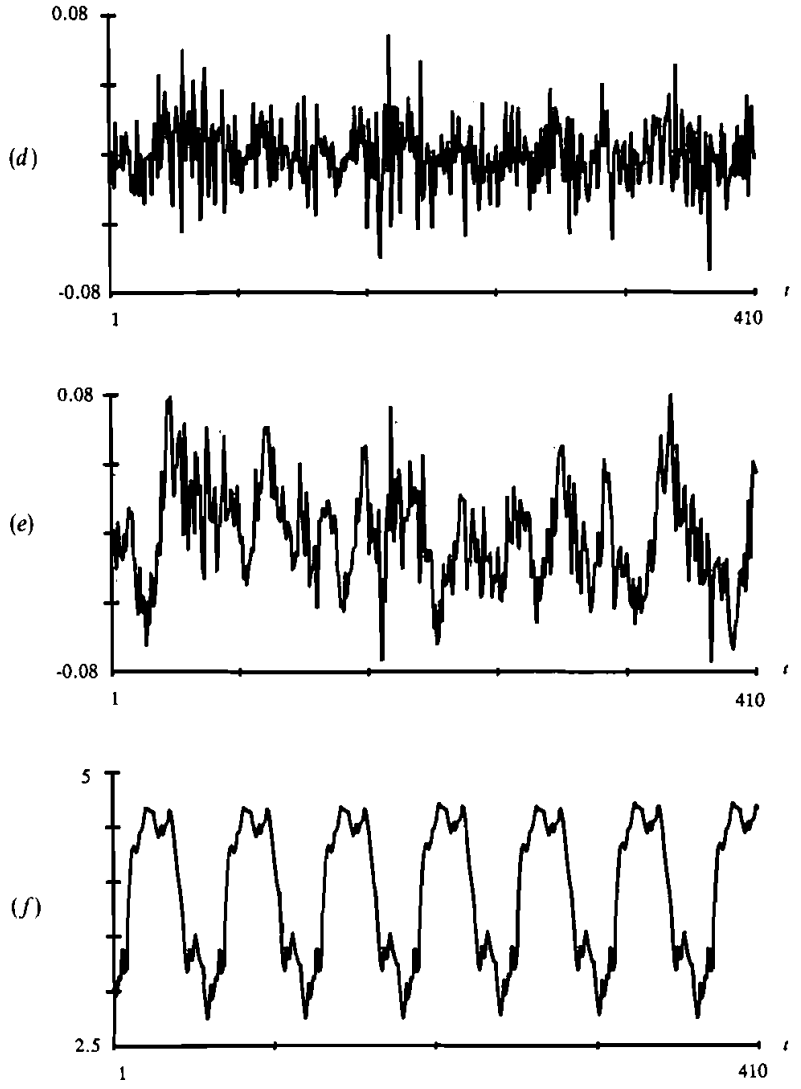


Figure 5. Data set and RBF model response (Example 2); (*a*) $u(t)$, (*b*) $y(t)$, (*c*) $\hat{y}(t)$, (*d*) $\varepsilon(t)$, (*e*) $\varepsilon_d(t)$, (*f*) $\hat{y}_d(t)$.

where $\phi(\cdot)$ is defined in (7) and

$$\|\mathbf{x}(t)-\mathbf{c}_i\| = \left[\sum_{k=1}^{3}(y(t-k)-c_{ki})^2+\sum_{k=4}^{6}(u(t-k+3)-c_{ki})^2 + \sum_{k=7}^{9}(\varepsilon(t-k+6)-c_{ki})^2\right]^{1/2}$$

for $1 \leqslant i \leqslant 33$. The coefficients $\hat{\lambda}_i$ and the centres $\mathbf{c}_i$ are listed in Tables 1 and 2, respectively.

The model responses are shown in Fig. 3, where the one-step-ahead prediction

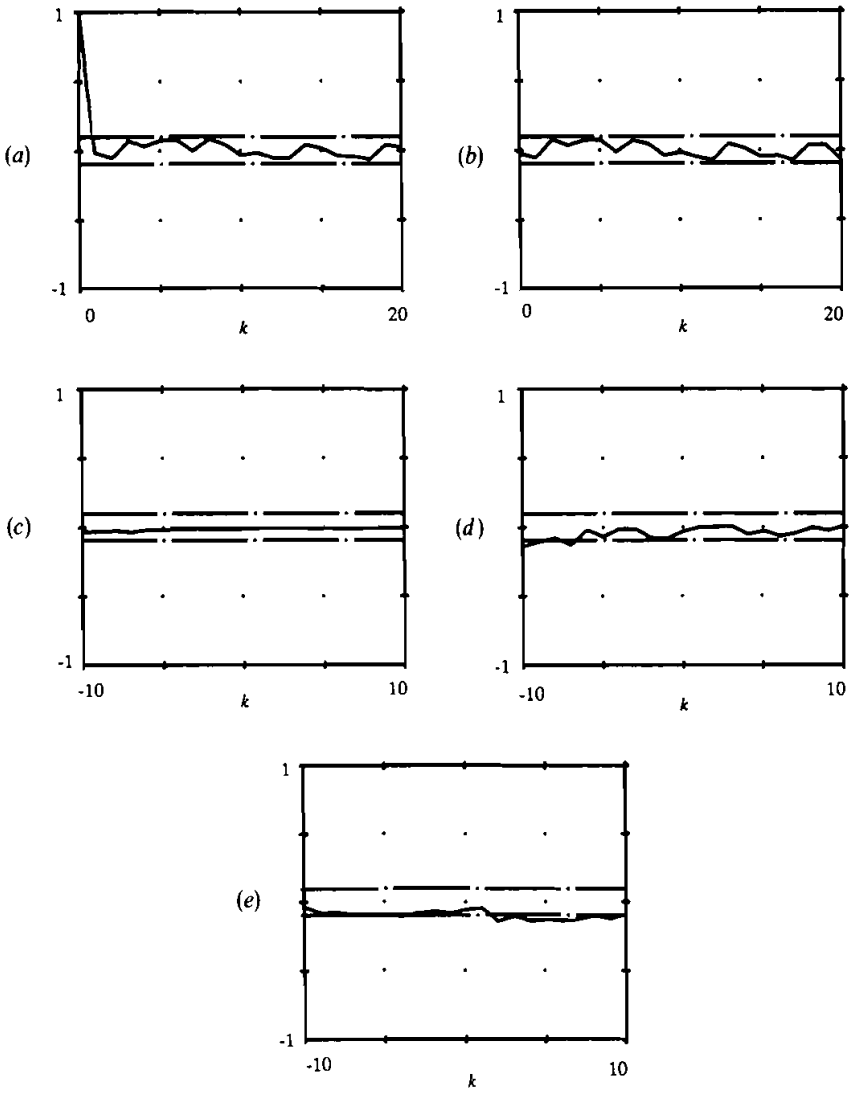


Figure 6. Correlation tests (Example 2): (*a*) $\Psi_{\varepsilon\varepsilon}(k)$; (*b*) $\Psi_{\varepsilon(\varepsilon u)}(k)$; (*c*) $\Psi_{u\varepsilon}(k)$; (*d*) $\Psi_{u^2\varepsilon}(k)$; (*e*) $\Psi_{u^2\varepsilon^2}(k)$. Dashed line 95% confidence interval.

$\hat{y}(t)$ is defined as

$$\hat{y}(t) = f_r(\mathbf{x}(t)) = f_r(y(t-1), \ldots, y(t-n_y), u(t-1), \ldots, u(t-n_u), \varepsilon(t-1), \ldots, \varepsilon(t-n_e)) \tag{43}$$

The prediction error or residual is given by

$$\varepsilon(t) = y(t) - \hat{y}(t) \tag{44}$$

The model deterministic output $\hat{y}_d(t)$ is defined as

$$\hat{y}_d(t) = f_r(\mathbf{x}_d(t)) = f_r(\hat{y}_d(t-1), \ldots, \hat{y}_d(t-n_y), u(t-1), \ldots, u(t-n_u), 0, \ldots, 0) \tag{45}$$

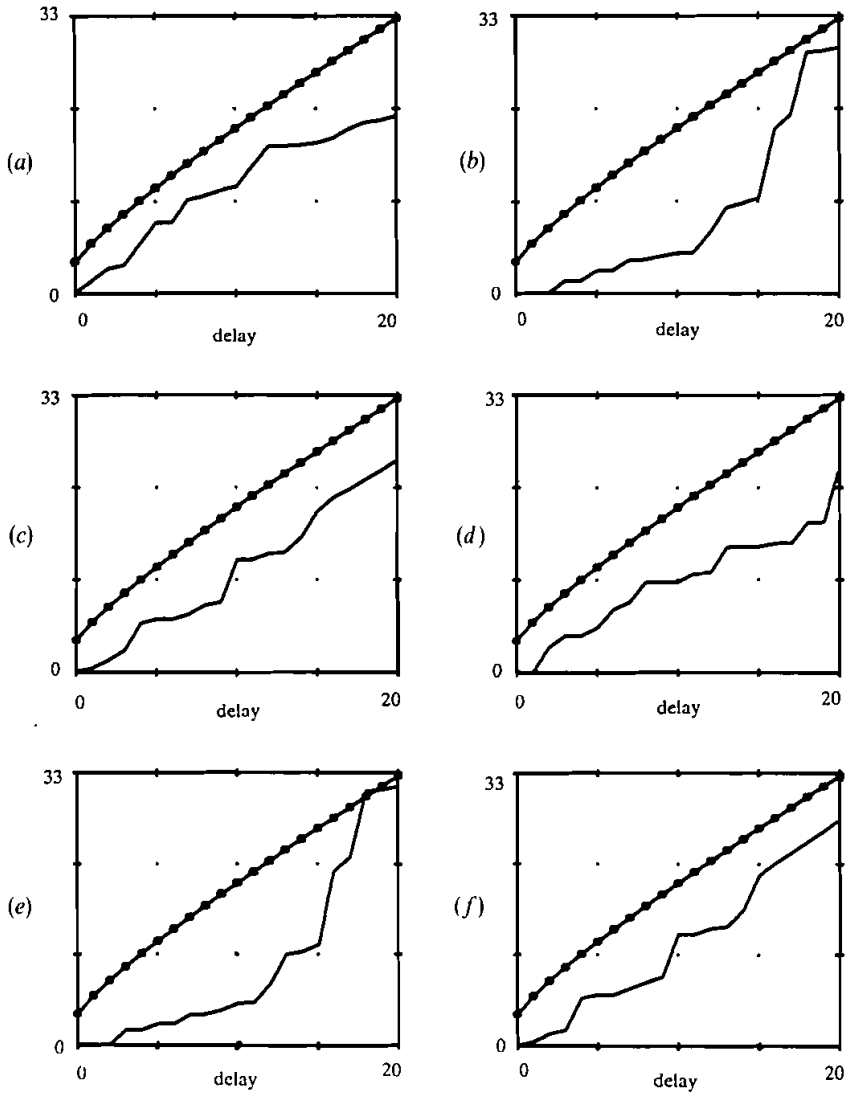


Figure 7. Chi-squared tests (Example 2); (a) $\omega(t) = \varepsilon(t-1)$, (b) $\omega(t) = u(t-1)$, (c) $\omega(t) = y(t-1)$, (d) $\omega(t) = \varepsilon^2(t-1)$, (e) $\omega(t) = u^2(t-1)$, (f) $\omega(t) = y^2(t-1)$. Dashed line 95% confidence limit.

and the deterministic error is accordingly given by

$$\varepsilon_d(t) = y(t) - \hat{y}_d(t) \tag{46}$$

The correlation tests are depicted in Fig. 4. It is observed that only at an isolated point is $\Psi_{u\varepsilon}(k)$ slightly outside the confidence bands, and this was judged a good result, considering that the data was from a real system.

6.2. *Example* 2

The data was collected from a turbocharged automotive diesel engine in a low engine speed test. The input was the fuel rack position and the output was the engine

speed. A detailed description of the experiment design was given by Billings *et al.* (1989 a). The data set contains 410 points and has been shown to be inherently non-linear. Previous results also indicated that it was very difficult to fit an adequate model to the data.

The dimension of the RBF centres was set to $n = n_y + n_u + n_e = 3 + 3 + 3$ and the dimension of the regressor set was $M = 411$. Only two iterations were involved in the identification procedure and the appropriate cutoff learnt during the identification was $\rho = 0{\cdot}000023$. The final model containing a constant term and 32 centres takes the form

$$\hat{y}(t) = \hat{f}_r(\mathbf{x}(t)) = \hat{\lambda}_0 + \sum_{i=1}^{32} \hat{\lambda}_i \phi(\|\mathbf{x}(t) - \mathbf{c}_i\|)$$

where

$$\|\mathbf{x}(t) - \mathbf{c}_i\| = \left[\sum_{k=1}^{3} (y(t-k) - c_{ki})^2 + \sum_{k=4}^{6} (u(t-k+3) - c_{ki})^2 + \sum_{k=7}^{9} (\varepsilon(t-k+6) - c_{ko})^2 \right]^{1/2}$$

for $1 \leqslant i \leqslant 32$. The coefficients $\hat{\lambda}_i$ and the centres $\mathbf{c}_i$ are listed in Tables 3 and 4, respectively. As a comparison, the subset polynomial model identified by Billings *et al.* (1989 a) is also given in Table 5. The data set and the model response are shown in Fig. 5, and the correlation tests and some chi-squared tests are illustrated in Figs 6 and 7, respectively. The model validity tests confirm that the model is adequate.

7. Conclusions

If the response of the system is dominated by non-linear characteristics, it is often necessary to use a non-linear model. The NARMAX representation gives a concise description for a large class of discrete-time non-linear systems. A practical identification algorithm for the NARMAX system has been developed based on the RBF expansion. It has been shown that the iterative identification algorithm, derived by coupling the OFR routine and the model validity tests, provides a powerful procedure for fitting parsimonious RBF models to real-world systems in the presence of correlated noise.

Acknowledgment

This work was supported by the U.K. Science and Engineering Research Council under Grant Ref. GR/E/10357.

References

Billings, S. A., and Chen, S., 1989 a, Identification of non-linear rational systems using a prediction-error estimation algorithm. *International Journal of Systems Science*, **20**, 467–494; 1989 b, Extended model set, global data and threshold model identification of severely non-linear systems. *International Journal of Control*, **50**, 1897–1923.

Billings, S. A., Chen, S., and Backhouse, R. J., 1989 a, The identification of linear and non-linear models of a turbocharged automotive diesel engine. *Mechanical Systems and Signal Processing*, **3**, 123–142.

Billings, S. A., Chen, S., and Korenberg, M. J., 1989 b, Identification of MIMO non-linear systems using a forward-regression orthogonal estimator. *International Journal of Control*, **49**, 2157–2189.

BILLINGS, S. A., and VOON, W. S. F., 1986, Correlation based model validity tests for non-linear models. *International Journal of Control*, **44**, 235–244.

BJÖRCK, A., 1967, Solving linear least squares problems by Gram-Schmidt orthogonalization. *Nordisk Tidskr. Informations-Behandling*, **7**, 1–21.

BOHLIN, T., 1987, Maximum-power validation of models without higher-order fitting. *Automatica*, **4**, 137–146.

BROOMHEAD, D. S., and LOWE, D., 1988, Multivariable functional interpolation and adaptive networks. *Complex Systems*, **2**, 321–355.

CHEN, S., and BILLINGS, S. A., 1988, Prediction-error estimation algorithm for non-linear output-affine systems. *International Journal of Control*, **47**, 309–332; 1989 a, Representation of non-linear systems: the NARMAX model. *Ibid.*, **49**, 1013–1032; 1989 b, Modelling and analysis of non-linear time series. *Ibid.*, **50**, 2151–2171.

CHEN, S., BILLINGS, S. A., and LUO, W., 1989, Orthogonal least squares methods and their application to non-linear system identification. *International Journal of Control*, **50**, 1873–1896.

CHEN, S., BILLINGS, S. A., COWAN, C. F. N., and GRANT, P. M., 1991, Non-linear systems identification using radial basis functions. *International Journal of Systems Science*, to be published.

CYBENKO, G., 1989, Approximations by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, **2**, 303–314.

GOLUB, G., 1965, Numerical methods for solving linear least squares problems. *Numerische Mathematik*, **7**, 206–216.

LAPEDES, A., and FARBER, R., 1988, How neural nets work. *Neural Information Processing Systems*, edited by D. Z. Anderson (New York: American Institute of Physics), pp. 442–456.

LEONTARITIS, I. J., and BILLINGS, S. A., 1985, Input–output parametric models for non-linear systems—Part 1: Deterministic non-linear systems; Part 2: Stochastic non-linear systems. *International Journal of Control*, **41**, 303–344; 1987, Model selection and validation methods for non-linear systems. *Ibid.*, **45**, 311–341.

LIPPMANN, R. P., 1987, An introduction to computing with neural nets. *I.E.E.E. ASSP Magazine*, **4**.

MICCHELLI, C. A., 1986, Interpolation of scattered data: distance matrices and conditionally positive definite functions. *Constructive Approximation*, **2**, 11–22.

POWELL, M. I. D., 1985, Radial basis functions for multivariable interpolation: a review. *IMA Conference on Algorithms for the Approximation of Functions and Data*, RMCS, Shrivenham; 1987, Radial basis function approximations to polynomials. *12th Biennial Numerical Analysis Conference*. Dundee, pp. 223–241.

SONTAG, E. D., 1979, *Polynomial Response Maps*, Lecture Notes in Control and Information Science, Vol. 13 (Berlin: Springer-Verlag).