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

Extended model set, global data and threshold model identification of severely non-linear systems

S. A. Billings [a]; S. Chen [a]
[a] Department of Control Engineering, University of Sheffield, Sheffield, SI, England, U.K.

Online Publication Date: 01 November 1989
To cite this Article: Billings, S. A. and Chen, S. (1989) 'Extended model set, global data and threshold model identification of severely non-linear systems', International Journal of Control, 50:5, 1897 - 1923
To link to this article: DOI: 10.1080/00207178908953473
URL: http://dx.doi.org/10.1080/00207178908953473

INT. J. CONTROL, 1989, VOL. 50, NO. 5, 1897–1923

Extended model set, global data and threshold model identification of severely non-linear systems

S. A. BILLINGS† and S. CHEN†

New parameter estimation algorithms, based on an extended model set, a global data model and a threshold model formulation, are derived for identifying severely non-linear systems. It is shown that in each case an integrated structure determination and parameter estimation algorithm based on an orthogonal decomposition of the regression matrix can be derived to provide procedures for identifying parsimonious models of unknown systems with complex structure. Simulation studies are included to illustrate the techniques discussed.

1. Introduction

If the response of a system is dominated by non-linear characteristics it is often necessary to use a non-linear model and this immediately raises the problem of what class of models to use. The non-linear autoregressive moving average with exogenous inputs (NARMAX) model which was first introduced by Billings and Leontaritis (1981) and rigorously derived by Leontaritis and Billings (1985) provides a unified representation for a wide class of discrete-time non-linear stochastic systems.

Model structure determination is often vital for identification of non-linear systems. Even if attention is restricted to polynomial expansions of the NARMAX model, a difficulty quickly arises because the number of terms can increase rapidly leading to an excessively complex model and numerical ill-conditioning in the identification procedure. In order to find an adequate model that uses only a few terms it is necessary to select only the significant terms from a large set of candidate terms. An orthogonal algorithm (Korenberg *et al.* 1988, Billings *et al.* 1988) has proved to be very efficient in determining the significant terms and providing corresponding parameter estimates. Chen *et al.* (1989) have shown that this estimator is in fact an orthogonal least-squares algorithm based on the classical Gram–Schmidt method and have derived several equivalent estimators by using different orthogonal decomposition techniques such as the modified Gram–Schmidt and Householder transformation algorithms. Numerous applications (for example, Billings *et al.* 1989 a, b, Liu *et al.* 1987) have demonstrated that these estimators provide very powerful procedures for identifying parsimonious models of systems with non-linear structure.

In the present study, the algorithms described above for the polynomial NARMAX model are modified and extended using different model formulations so that they can be applied to severely non-linear systems. By allowing functions such as absolute value, exponential, logarithmic, sgn (·) to be terms within a linear-in-the-parameters model, an extended model set is created. This broadens the base for modelling non-linear systems and it is shown that the orthogonal estimator can be

Received 10 February 1989.

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

0020-7179/89 $3.00

employed to provide structure detection and parameter estimation for this class of models. A new intelligent structure detection strategy is then introduced that learns the threshold for term selection, this provides a fully automatic term-selection algorithm.

Global data fitting is investigated as a means of estimating the parameters in a single global model using several data sets, each of which covers a local region of plant operation. This type of analysis is appropriate for systems where it is not possible to excite the system over the total operating range in one experiment but where several small perturbation experiments about different operating points are permitted.

If a system is severely non-linear, however, it may not be possible to find one model that provides an adequate representation; the estimation of threshold models is investigated as a solution to this problem.

Several simulation examples are included to demonstrate the effectiveness of the new algorithms. For notational simplicity, only the single-input single-output case is considered throughout although the results are valid for multi-input multi-output systems.

2. System representation

Under some mild assumptions a discrete-time non-linear stochastic control system can be described by the NARMAX model (Leontaritis and Billings 1985)

$$y(t) = f(y(t-1), \ldots, y(t-n_y), u(t-1), \ldots, u(t-n_u), \varepsilon(t-1), \ldots, \varepsilon(t-n_\varepsilon)) + \varepsilon(t) \quad (1)$$

where $y(t)$, $u(t)$ and $\varepsilon(t)$ are the output, input and prediction error, respectively, and $f(\cdot)$ is some non-linear function. A special case of the model (1) that may be referred to as the non-linear autoregressive with exogenous inputs (NARX) model is

$$y(t) = f(y(t-1), \ldots, y(t-n_y), u(t-1), \ldots, u(t-n_u)) + \varepsilon(t) \quad (2)$$

The model (1) is about as far as one can go in terms of specifying a general finite-dimensional input–output non-linear system. Various possibilities of parametrizing the function $f(\cdot)$ exist (Chen and Billings 1989), one of which is a polynomial approximation of $f(\cdot)$.

3. Orthogonal least-squares estimators

Solving least-squares problems by an orthogonal decomposition of the regression matrix is a well-developed technique. Consider a linear regression description

$$y(t) = \sum_{i=1}^{n} \phi_i(t)\theta_i + \varepsilon(t) \quad (3)$$

and assume that the data length is N. Collecting (3) from $t = 1$ to N together yields the following matrix equation

$$\mathbf{y} = \mathbf{\Phi\Theta} + \mathbf{\Xi} \quad (4)$$

where

$$\mathbf{y} = \begin{bmatrix} y(1) \\ \vdots \\ y(N) \end{bmatrix}, \quad \mathbf{\Phi} = [\mathbf{\Phi}_1 \quad \ldots \quad \mathbf{\Phi}_n], \quad \mathbf{\Theta} = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_n \end{bmatrix}, \quad \mathbf{\Xi} = \begin{bmatrix} \varepsilon(1) \\ \vdots \\ \varepsilon(N) \end{bmatrix} \quad (5)$$

and

$$\mathbf{\Phi}_i = [\phi_i(1) \quad \ldots \quad \phi_i(N)]^{\mathrm{T}} \quad i = 1, \ldots, n \tag{6}$$

3.1. *Orthogonal decomposition of the regression matrix*

The matrix $\mathbf{\Phi}$ is often referred to as the regression matrix. An orthogonal decomposition of $\mathbf{\Phi}$ is given as

$$\mathbf{\Phi} = \mathbf{WA} \tag{7}$$

Here

$$\mathbf{A} = \begin{bmatrix} 1 & \alpha_{12} & \alpha_{13} & \ldots & \alpha_{1n} \\ & 1 & \alpha_{23} & \ldots & \alpha_{2n} \\ & & \ddots & & \vdots \\ & & & 1 & \alpha_{n-1\,n} \\ & & & & 1 \end{bmatrix} \tag{8}$$

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

$$\mathbf{W} = [\mathbf{w}_1 \quad \ldots \quad \mathbf{w}_n] \tag{9}$$

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

$$\mathbf{W}^{\mathrm{T}}\mathbf{W} = \mathbf{D} \tag{10}$$

and $\mathbf{D}$ is a positive diagonal matrix

$$\mathbf{D} = \mathrm{diag}\,\{d_1 \quad \ldots \quad d_n\} \tag{11}$$

with

$$d_i = \langle \mathbf{w}_i, \mathbf{w}_i \rangle \tag{12}$$

where $\langle \cdot \rangle$ denotes the inner product, that is

$$\langle \mathbf{w}_i, \mathbf{w}_j \rangle = \mathbf{w}_i^T \mathbf{w}_j = \sum_{t=1}^{N} w_i(t) w_j(t) \tag{13}$$

Equation (4) can now be rewritten as

$$\mathbf{y} = [\mathbf{\Phi A}^{-1}][\mathbf{A\Theta}] + \mathbf{\Xi} = \mathbf{Wg} + \mathbf{\Xi} \tag{14}$$

where

$$\mathbf{A\Theta} = \mathbf{g} \tag{15}$$

It is straightforward to show that

$$\mathbf{g} = \mathbf{D}^{-1}\mathbf{W}^{\mathrm{T}}\mathbf{y} \tag{16}$$

or

$$g_i = \frac{\langle \mathbf{w}_i, \mathbf{y} \rangle}{\langle \mathbf{w}_i, \mathbf{w}_i \rangle} \quad i = 1, \ldots, n \tag{17}$$

Alternatively, by normalizing the columns of $\mathbf{W}$ and augmenting the resulting matrix with $N - n$ further orthonormal columns to make up a full set of N

orthonormal vectors for N-dimensional euclidean space, a decomposition equivalent to (7) is obtained

$$\mathbf{\Phi} = \bar{\mathbf{W}} \begin{bmatrix} \mathbf{R} \\ 0 \end{bmatrix} \tag{18}$$

where $\mathbf{R}$ is an $n \times n$ upper triangular matrix with positive diagonal elements and

$$\bar{\mathbf{W}} = [\bar{\mathbf{w}}_1 \quad \dots \quad \bar{\mathbf{w}}_N] \tag{19}$$

is an $N \times N$ orthogonal matrix, i.e.

$$\langle \bar{\mathbf{w}}_i, \bar{\mathbf{w}}_j \rangle = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \tag{20}$$

Using $\bar{\mathbf{W}}^{\mathrm{T}}$ to transfer $\mathbf{y}$ results in

$$\bar{\mathbf{W}}^{\mathrm{T}}\mathbf{y} = [\bar{y}_1 \quad \dots \quad \bar{y}_n \quad \bar{y}_{n+1} \quad \dots \quad \bar{y}_N]^{\mathrm{T}} \tag{21}$$

It is readily seen that the parameter estimate $\hat{\mathbf{\Theta}}$ satisfies

$$\mathbf{R}\hat{\mathbf{\Theta}} = [\bar{y}_1 \quad \dots \quad \bar{y}_n]^{\mathrm{T}} \tag{22}$$

3.2. *Structure determination*

Orthogonal least-squares methods were originally used as effective numerical techniques to obtain the triangular system (15) or (22) and, consequently, to solve for the parameter estimate $\hat{\mathbf{\Theta}}$. These include the classical Gram–Schmidt and modified Gram–Schmidt methods (Björck 1967), the Householder transformation method (Golub 1965) and the Givens method (Gentleman 1973). By modifying and augmenting these orthogonal least-squares estimators, efficient algorithms for combining structure determination (or determining which terms to include in the model) and parameter estimation can be derived (Chen *et al.* 1989).

From (14), the sum of squares of the output is

$$\langle \mathbf{y}, \mathbf{y} \rangle = \sum_{i=1}^{n} g_i^2 \langle \mathbf{w}_i, \mathbf{w}_i \rangle + \langle \mathbf{\Xi}, \mathbf{\Xi} \rangle \tag{23}$$

The error-reduction ratio due to $\mathbf{w}_i$ is thus defined as the proportion of the output variance explained by $\mathbf{w}_i$

$$[\mathrm{err}]_i = \frac{g_i^2 \langle \mathbf{w}_i, \mathbf{w}_i \rangle}{\langle \mathbf{y}, \mathbf{y} \rangle} \quad 1 \leqslant i \leqslant n \tag{24}$$

If the decomposition (18) is employed, a similar definition of error-reduction ratio is given as

$$[\mathrm{err}]_i = \frac{(\bar{y}_i)^2}{\langle \mathbf{y}, \mathbf{y} \rangle} \quad 1 \leqslant i \leqslant n \tag{25}$$

The error-reduction ratio offers a simple and effective means of selecting a subset of significant terms from a large number of candidate terms in a forward-regression manner. At each step, a term is selected if it produces the largest value of $[\mathrm{err}]_i$ among the rest of the candidate terms. The selection procedure is terminated when

$$1 - \sum_{i=1}^{n_s} [\mathrm{err}]_i < \rho \tag{26}$$

where ρ ($0 < \rho < 1$) is a desired tolerance, and this leads to a subset model of n_s ($n_s < n$) terms. It is immediately seen that $1 - \sum [\text{err}]_i$ is the proportion of the unexplained output variance. Details of several such forward-regression orthogonal algorithms have been given by Chen *et al.* (1989) and will not therefore be repeated here. The general procedure is summarized below.

Procedure for the linear regression model

Initially the user specifies the full model set of n candidate terms and the value of ρ. At the n_sth stage:

(i) compute the values of the error-reduction ratio for each of the $n - n_s + 1$ remaining candidate terms by assuming that each is the n_sth term in the selected model, and perform the corresponding orthogonalization;

(ii) the term that gives the largest value of error-reduction ratio is then selected. If condition (26) is satisfied, go to (iii). Otherwise set $n_s := n_s + 1$ and go to (i).

(iii) the final model contains n_s terms and, depending on which orthogonal decomposition technique has been used, its parameter estimate $\hat{\boldsymbol{\Theta}}_s$ is computed from

$$\mathbf{A}_s\boldsymbol{\Theta}_s = \mathbf{g}_s \quad \text{or} \quad \mathbf{R}_s\boldsymbol{\Theta}_s = [\bar{y}_1 \quad \dots \quad \bar{y}_{n_s}]^{\mathrm{T}} \tag{27}$$

where $\mathbf{A}_s$ and $\mathbf{R}_s$ are $n_s \times n_s$ unit or positive upper-triangular matrices respectively.

4. Extended-model-set representation

In order to use model (1) or (2) as a basis for identification, a means of parametrization is required. In the present study linear-in-the-parameters models are considered, and such models take the following general form

$$y(t) = \sum_{i=1}^{n} \phi_i(y(t-1), \dots, y(t-n_y), u(t-1), \dots, u(t-n_u), \varepsilon(t-1), \dots, \varepsilon(t-n_\varepsilon))\theta_i + \varepsilon(t) \tag{28}$$

Equation (28) is a pseudo-linear regression model where the 'regressors' are transformations of lagged input, output and prediction-error values. The analogous linear regression model that corresponds to the NARX model (2) can be written as

$$y(t) = \sum_{i=1}^{n} \phi_i(y(t-1), \dots, y(t-n_y), u(t-1), \dots, u(t-n_u))\theta_i + \varepsilon(t) \tag{29}$$

where the regressors are some non-linear functions of lagged inputs and outputs only. Of course, before using model (28) or (29), the functions $\phi_i(\cdot)$ must be specified. Examples of $\phi_i(\cdot)$ are now given.

Polynomial model

If the function $\phi_i(\cdot)$ are chosen as monomials of lagged input, output and/or prediction-error values, a polynomial model is obtained. If the non-linear function $f(\cdot)$ in (1) is continuous it can always be arbitrarily well approximated by polynomial models (Chen and Billings 1989). Thus the polynomial model offers a very general representation for non-linear systems. Practical identification of many real non-linear systems has been based on the polynomial model (Billings *et al.* 1989 a, b, Liu *et al.*

1987). When the system is severely non-linear it would be advantageous if other choices of $\phi_i(\cdot)$ could be considered; this is investigated below.

Exponential time-series model

The exponential time-series model of Ozaki (1985) takes the form

$$y(t) = \sum_{i=1}^{m} [\alpha_i + \beta_i \exp(-y^2(t-1))]y(t-i)\theta_i + \varepsilon(t) \tag{30}$$

where α_i and β_i are parameters. Under very weak conditions the stochastic process defined by (30) is stationary and the model (30) possesses rich dynamic behaviour such as limit cycles associated with non-linear systems, which a linear model cannot capture. The choice of $\exp(-y^2(t-1))$ is, however, rather arbitrary and there is no reason against using the more general form

$$y(t) = \sum_{i=1}^{m} [\alpha_i + \beta_i \exp(-y^2(t-k_i))]y(t-i)\theta_i + \varepsilon(t) \tag{31}$$

It is highly desirable that an identification package can search through a set of terms of the form $\exp(-y^2(t-k_i))y(t-i)$ and determine a suitable set of k_i values that will fit the observed time-series values better.

Other examples

Consider the discretization of a damped and forced non-linear oscillator governed by, for example, the differential equation

$$\frac{d^2z(\tau)}{d\tau^2} + \alpha_1\frac{dz(\tau)}{d\tau} + \alpha_2 z(\tau) + \alpha_3\frac{dz(\tau)}{d\tau}\left|\frac{dz(\tau)}{d\tau}\right| + \alpha_4 z^3(\tau) + \alpha_5 \sin(z(\tau)) = u(\tau) \tag{32}$$

where

$$\frac{dz(\tau)}{d\tau}\left|\frac{dz(\tau)}{d\tau}\right|$$

is known as quadratic damping and $z^3(\tau)$ is called cubic stiffness. A simple pendulum undergoing arbitrarily large oscillations, for example, would introduce the term $\sin(z(\tau))$. Assume that a zero-order-hold (ZOH) device is used, that is $u(\tau) = u(\tau_t)$, $\tau_t \leqslant \tau < \tau_{t+1}$, a fixed sample rate is employed with a sampling period h and t is used to replace τ_t. Then an approximate difference equation model is obtained as

$$\begin{aligned} z(t) = {} & \theta_1 z(t-1) + \theta_2 z(t-2) + \theta_3 z(t-1)|z(t-1) - z(t-2)| \\ & + \theta_4 z(t-2)|z(t-1) - z(t-2)| + \theta_5 z^3(t-2) \\ & + \theta_6 \sin(z(t-2)) + \theta_7 u(t-2) \end{aligned} \tag{33}$$

using the forward difference scheme. If the output is corrupted by additive noise

$$y(t) = z(t) + \varepsilon(t) \tag{34}$$

the absolute-valued and trigonometric functions of lagged outputs and prediction errors will appear in the model. It is possible to use a polynomial approximation to model this sampled non-linear oscillator. This would, however, require a large number of terms to achieve an accurate approximation because, for example, the

polynomial expansion of the sinusoidal function $\sin(x)$ is

$$\sin(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \ldots \tag{35}$$

Clearly, the possibility of selecting the $\phi_i(\cdot)$ functions in (28) to match the specific non-linear terms $|z(t-i) - z(t-j)|$, $\sin(\cdot)$ in (33) would be preferable.

The swing equation for a class of synchronous generators can be described as (Polchai and Hsu 1985)

$$H\frac{d^2\delta(\tau)}{d\tau^2} = P_m - P_e(\delta(\tau)) - C(\delta(\tau))\frac{d\delta(\tau)}{d\tau} \tag{36}$$

where $\delta(\tau)$ is the electrical load angle, H the inertia constant, P_m a mechanical input, $C(\delta(\tau))$ the variable damping and $P_e(\delta(\tau))$ the electrical power output considering transient saliency. The last two quantities are given by

$$\left.\begin{aligned} C(\delta(\tau)) &= b_1 \sin^2(\delta(\tau)) + b_2 \cos^2(\delta(\tau)) \\ P_e(\delta(\tau)) &= b_3 \sin(\delta(\tau)) + b_4 \sin(2\delta(\tau)) \end{aligned}\right\} \tag{37}$$

Applying the same scheme as that for (32) produces a discrete-time model that is linear-in-the-parameters with some of $\phi_i(\cdot)$ taken as trigonometric functions.

Some other common examples of $\phi_i(\cdot)$ that arise in non-linear dynamics are the hyperbolic functions $\sinh(\cdot)$, $\cosh(\cdot)$ and $\tanh(\cdot)$, the inverse trigonometric function $\operatorname{atan}(\cdot)$, the Coulomb friction $\operatorname{sgn}(\cdot)$, and the saturation.

It is clear from the above discussion that monomials of lagged input, output and prediction-error values are primary choices for $\phi_i(\cdot)$, and these terms form a basic model set from which a parsimonious model can be selected. This basic model set can be extended to include terms such as $\sin(\cdot)$, $\cos(\cdot)$, $\exp(\cdot)$, $\sinh(\cdot)$, $\tanh(\cdot)$, which commonly exist in non-linear systems. Such an extended model set gives a much richer description to non-linear systems and more effective modelling can be achieved. In practice, physical knowledge of the system to be identified can often be used to narrow down the choice of $\phi_i(\cdot)$ and consequently to reduce the size of the model set.

When identifying non-linear systems with unknown structure, it is important to avoid losing significant terms that must be included in the final model; consequently the experimenter is forced to start with a large model set. Even if only monomials are used to form the model set, the number of terms can be huge. For example, if $n_y = n_u = n_e = 10$ and the polynomial degree is $l = 3$, the number of terms for the full model set is $n = 5456$. If other functions with different lagged inputs, outputs and prediction errors are included in the model set, the size will become even larger. The selection of a few significant terms from such a large number of terms is by no means an easy task. Fortunately the orthogonal least-squares estimator can be adopted (Billings *et al.* 1989 b, Chen *et al.* 1989) to provide an efficient solution.

4.1. *Estimation and structure detection for the extended model set*

Since the model (29) is a linear regression model, the procedure summarized in § 3.2 can be applied. For the general pseudo-linear regression model (28), however, delayed prediction errors are involved in $\phi_i(t)$, and $\varepsilon(t)$ itself can only be computed if

the values of the parameters are given. An iterative procedure is therefore required. Partition Φ into

$$\Phi = [\Phi_p : \Phi_r] \tag{38}$$

where each of the n_p columns in Φ_p is a known function of the measured input–output data only, each of the n_r columns in Φ_r is a known function of the prediction errors and the input–output data, and $n_p + n_r = n$. Φ_p may be referred to as the full process model set and Φ_r as the full noise model set. A subset model Φ_s of Φ is similarly represented as

$$\Phi_s = [\Phi_{p_s} : \Phi_{r_s}] \tag{39}$$

Procedure for the pseudo-linear regression model

Initially the algorithm selects columns of Φ_{p_s} from Φ_p in an identical way to that indicated in the procedure for the linear regression model. The selection is terminated when the condition

$$1 - \sum_{i=1}^{n_{p_s}} [\text{err}]_i < \rho_p \tag{40}$$

is satisfied where ρ_p is the tolerance for the process model. The initial subset model parameter estimate $\hat{\Theta}_s^{(0)}$ (having n_{p_s} elements) can be computed and this in turn allows the first prediction error or residual sequence $\{\varepsilon^{(1)}(t)\}$ to be generated.

At the kth iteration ($k \geqslant 1$), $\Phi_r^{(k)}$ can be formed from the kth residual sequence $\{\varepsilon^{(k)}(t)\}$ and the input–output data, and this enables the algorithm to select columns of $\Phi_{r_s}^{(k)}$. Assume that after $n_{r_s}^{(k)}$ columns have been added to Φ_{p_s} the condition

$$1 - \sum_{i=1}^{n_{p_s} + n_{r_s}^{(k)}} [\text{err}]_i < \rho_r \tag{41}$$

is satisfied, where ρ_r ($< \rho_p$) is the tolerance for the noise model, then the selection is stopped. The kth subset model parameter estimate $\hat{\Theta}_s^{(k)}$ (having $n_{p_s} + n_{r_s}^{(k)}$ elements) can thus be computed and this allows the residual sequence to be updated. $\Phi_r^{(k+1)}$ can then be formed and the subset noise model re-selected. Experience has shown that typically 4 to 6 iterations are usually sufficient.

5. Intelligent structure detection

Consider first the linear regression model (29). It is clear that ρ, the stopping criterion (26), is an important instrument that affects both the prediction accuracy and complexity of the final model. If the value of ρ is chosen to be too large, the model will be inadequate. If, on the other hand, ρ is chosen to be too small the model will become unnecessarily complex. From (23) it is seen that ideally ρ should be larger than but very close to the ratio $\sigma_\varepsilon^2/\sigma_y^2$, where σ_ε^2 is the variance of the residuals and σ_y^2 the variance of the measured process output. Since σ_ε^2 is not known *a priori*, an appropriate value of ρ may need to be found by trial and error. Fortunately, a simple learning strategy can often be implemented. An initial guess is assigned to ρ. Once a model is selected, an estimate $\hat{\sigma}_\varepsilon^2$ for σ_ε^2 can be computed and since σ_y^2 is known from the measured data, the direction for improving ρ is given (choosing ρ close to $\hat{\sigma}_\varepsilon^2/\sigma_y^2$). If model validity tests (Billings and Voon 1986, Billings and Chen 1989) are coupled with the algorithm, the correct model structure can often be found very quickly.

The criterion (26) emphasises only the performance of the model (variance of residuals). Because a more accurate performance is often achieved at the expense of using a more complex model, a trade-off between the performance and complexity of the model is often appropriate. Akaike-type criteria which compromise between the performance and the number of parameters take the form

$$\mathrm{AIC}(\eta) = N \log(\sigma_\varepsilon^2) + n_s \eta \tag{42}$$

where η is the critical value of the chi-squared distribution with one degree of freedom for a given level of significance. $\eta = 4$ corresponds to the significance level of 0·0456 and is often a suitable choice (Leontaritis and Billings 1987 a). Other statistical criteria (Söderström 1977, Bozdogan 1988) can also be employed to terminate the selection.

For the pseudo-linear regression model (28) it follows from the discussion in § 4.1 that two tolerances ρ_p and ρ_r are required. It can be seen that the best value for ρ_r is closely related to the ratio $\sigma_\varepsilon^2/\sigma_y^2$. There is, however, no simple way to determine the best value for ρ_p. Notice that the procedure given in § 4.1 decouples the selection of the process and noise-model parameters and in this way it is convenient for updating $\{\varepsilon(t)\}$. However it also poses the problem of selecting a suitable value for ρ_p that defines the threshold at the initial stage of selecting process-model terms. Ideally the process- and noise-model terms should be treated equivalently if $\{\varepsilon(t)\}$ is given, and this suggests a slightly different procedure that alleviates the difficulty of selecting ρ_p.

The procedure begins by assigning an initial guess to ρ_p. Once a process model is selected based on this ρ_p, $\{\varepsilon^{(1)}(t)\}$ can be computed and an estimate for ρ_r can be found. The process- and noise-model terms are then re-selected together and the selection is terminated using ρ_r. As this iterative procedure progresses, the estimate of σ_ε^2 is improved and hence so is ρ_r. This strategy can also be combined with Akaike's information criterion (AIC). First use $\mathrm{AIC}(\eta)$ to terminate the process-model regression and to produce the initial residuals. At the kth iteration, the process- and noise-model terms are reselected together and the procedure is stopped using $\mathrm{AIC}(\eta)$.

6. Global-model fitting

Most processes encountered in the real world are non-linear. A non-linear global model is often desirable in order to infer effectively the dynamic behaviour of the system over a large range of operation and to design a global control law that is valid for the whole operating range. A quick review of piecewise linearization and its drawbacks provides a useful starting point for discussion.

Piecewise linear modelling identifies a series of locally linear models that approximate the non-linear system under study over some defined operating range. There are several possible ways in which non-linear systems can be approximated by locally linear models (e.g. Billings and Voon 1987). The spatial piecewise linear model will, providing the non-linearities are smooth, provide an adequate representation of a non-linear system. This may, however, only be achieved at the expense of fitting a very large number of linear models, each valid in a small region of operation. Alternatively, if the non-linearity is produced by a measurable system variable, a series of signal-dependent linear models can be estimated and these models can then be patched together to yield a non-linear description of the system. The resulting non-linear model will be valid only for relatively slow-moving inputs. That is, the model will produce an excellent prediction of the system output when perturbed by the input

used for the identification or by an input that causes the system to traverse slowly from one operating region to another. If, however, the non-linearity is not perfectly dependent upon one signal or the model is perturbed by inputs that cause rapid transient changes in the operating point, the model output will be considerably different from the system output. This means that the identified model does not provide an adequate representation of the non-linear system but is valid only for a small class of input signals. It may be better in these situations to fit a non-linear model to the system rather than attempting to approximate it by a series of linear models.

The design of inputs for non-linear system identification is a very complex problem and some useful results have been given by Leontaritis and Billings (1987 b). Roughly speaking, whenever possible the input signal chosen for an identification experiment should be persistently exciting. This means that the input should excite all the frequencies of interest in the system and should also excite the process over the whole amplitude range of operation. It is this last requirement that is often difficult to meet in practice. Normal operation of an industrial plant is often concerned with controlling the plant close to some operating points. Perturbing signals that the experimenter injects into the plant can usually only have a small amplitude in order not to cause large disturbances to the operation of the plant. If normal operation of the plant includes several operating levels, the experimenter may be able to superimpose on these levels a uniformly distributed or similar signal having a bandwidth sufficient to excite all the dynamic modes within the plant without violating the amplitude constraints for normal operation, as illustrated in Fig. 1. Instead of fitting several linear models to different data segments, a non-linear model can be fitted simultaneously to the data segments (A), (B), (C) and (D). It is reasonable to believe that the resulting non-linear model (called a global data model) will provide an excellent global representation of the plant since typical operating regions have all been taken into account. In Fig. 1, the data segments (A), (B), (C) and (D) are shown as continguous. This will not usually be the case. Each of the experiments may have been performed on quite separate occasions, possibly weeks apart, and the objective is to process the data to estimate a global model.

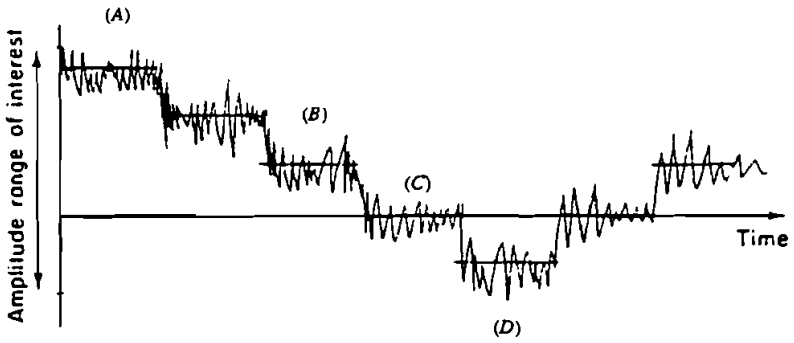


Figure 1. Input design for non-linear systems.

Generally, assume that M separate sets of input–output data have been recorded from a system of unknown structure and that each set corresponds to a different operating region of the system under study. Ideally, the regions should have some degree of overlap. As discussed in § 4, (28) represents a wide class of non-linear

systems and is used to model the system to be identified. This leads to M equations

$$\mathbf{y}^{[k]} = \mathbf{\Phi}^{[k]}\mathbf{\Theta} + \mathbf{\Xi}^{[k]} \quad k = 1, \ldots, M \tag{43}$$

where the superscript $[k]$ denotes each individual experimental condition. Even if the experiments were not contiguous, the data could be collected together in the model

$$\begin{bmatrix} \mathbf{y}^{[1]} \\ \vdots \\ \mathbf{y}^{[M]} \end{bmatrix} = \begin{bmatrix} \mathbf{\Phi}^{[1]} \\ \vdots \\ \mathbf{\Phi}^{[M]} \end{bmatrix} \mathbf{\Theta} + \begin{bmatrix} \mathbf{\Xi}^{[1]} \\ \vdots \\ \mathbf{\Xi}^{[M]} \end{bmatrix} \tag{44}$$

and the orthogonal least-squares algorithms of § 4.1 can readily be applied to (44). Unlike the approach of piecewise linear modelling and the approach of Dang Van Mien and Normand-Cyrot (1984), a single non-linear global data model is directly identified. This technique provides a simple and practical way to obtain global descriptions of non-linear systems.

An alternative application of the global data model (44) would be where a large data set is available that cannot be processed in one pass because of limitations on the size of vectors within the computer. For example, ten thousand data samples may be available while only 1000 can be processed at any one time. This scenario can be accommodated within the framework of (44) by selecting subsets of the full data set corresponding to each superscript in (43). Optimal search procedures that select appropriate subsets according to preset criteria could be designed as part of such an algorithm.

7. Threshold-model fitting

Using a global non-linear model to represent a non-linear system obviously has many advantages. It all depends, however, upon whether the system under investigation can be represented by a single model. Some severely non-linear systems may require more than one model to capture different dynamic behaviour over different operating regions. The threshold NARMAX (TNARMAX) model may be particularly useful for modelling such systems. A general TNARMAX model can be described as

$$\begin{aligned} y(t) = f^{(i)}(&y(t-1), \ldots, y(t-n_y), u(t-1), \ldots, u(t-n_u), \varepsilon(t-1), \ldots, \varepsilon(t-n_\varepsilon)) \\ &+ \varepsilon(t) \quad \text{if } \mathbf{x}_t \in \mathbb{R}^{(i)};\ i = 1, \ldots, p \end{aligned} \tag{45}$$

where

$$\mathbf{x}_t = (y(t-1), \ldots, y(t-n_y), u(t-1), \ldots, u(t-n_u)) \tag{46}$$

the $\mathbb{R}^{(i)}$ are given regions of the $(n_y + n_u)$-dimensional euclidean space, and the $f^{(i)}(\cdot)$ are some non-linear functions. The lags n_y, n_u and n_ε for different sub-models $f^{(i)}(\cdot)$ can take different values. Here, for the convenience of description, the same notation is used for all the sub-models. Notice that $\mathbf{x}_t$ does not include $\varepsilon(\cdot)$ since $\varepsilon(t)$ represents the system noise and cannot be measured directly. Again our attention is focused on the linear-in-the-parameters TNARMAX model because effective algorithms for structure determination and parameter estimation discussed in the previous sections can be applied. Thus (45) can be expressed as

$$y(t) = \sum_{j=1}^{n^{(i)}} \phi_j^{(i)}(t)\theta_j^{(i)} + \varepsilon(t), \quad \text{if } \mathbf{x}_t \in \mathbb{R}^{(i)};\ i = 1, \ldots, p \tag{47}$$

where $\phi_j^{(i)}(t)$ is shorthand notation for $\phi_j^{(i)}(y(t-1),\ldots,y(t-n_y),u(t-1),\ldots,u(t-n_u),\varepsilon(t-1),\ldots,\varepsilon(t-n_\varepsilon))$. If none of the $\phi_j^{(i)}(t)$ involve $\varepsilon(\cdot)$, model (47) becomes a linear-in-the-parameters TNARX model. The TNARX and TNARMAX models can be regarded as generalizations of the TNAR and TNARMA time-series models (Ozaki 1981, Lai and Hsieh 1988), which have been widely used for modelling non-linear random vibrations. Two cases of threshold model fitting are now considered.

The first situation is a direct extension of piecewise linear modelling where p sets of data records have been obtained and each set corresponds to a different operating region of the system. Assume that a global non-linear model has been fitted to the data sets and that model validity tests (§ 8) have revealed that the model does not provide an adequate description of the data. This warrants an investigation using the TNARMAX model (47), where each threshold model corresponds to a particular operating condition; thus from (43)

$$\mathbf{y}^{[k]} = \mathbf{\Phi}^{[k]}\mathbf{\Theta}^{(k} + \mathbf{\Xi}^{[k]}, \quad \text{if system is in operating region } k;\ k = 1,\ldots,p \tag{48}$$

Each of the p sub-models can then be identified separately using the orthogonal estimator. If the sub-models for some operating regions turn out to be very similar, these models can be patched together or a single model can be re-identified to represent these operating domains or combinations of them.

The second possibility corresponds to the situation where only a single data record from one experiment that covers the whole domain of operation is available. If the estimation of a global model produces a poor fit, the data should be split by defining a threshold to yield two subsets of data and hence two models. The procedure could be continued until an acceptable fit is obtained. This is almost the converse of the global data model idea. The kth sub-model in such a procedure would be defined by

$$\mathbf{y}^{(k} = \mathbf{\Phi}^{(k}\mathbf{\Theta}^{(k} + \mathbf{\Xi}^{(k}, \quad \text{if } \mathbf{x}_t \in \mathbb{R}^{(k};\ k = 1,\ldots,p \tag{49}$$

The algorithm is similar to that developed for threshold time-series models (for example, Tong and Lim 1980). As a simple illustration, consider the following TNARMAX model:

$$y(t) = \begin{cases} \sum_{j=1}^{n^{(1}} \phi_j^{(1}(t)\theta_j^{(1} + \varepsilon(t), & \text{if } y(t-d) \leqslant T \\ \sum_{j=1}^{n^{(2}} \phi_j^{(2}(t)\theta_j^{(2} + \varepsilon(t), & \text{if } y(t-d) > T \end{cases} \tag{50}$$

or more concisely

$$\left.\begin{aligned} \mathbf{y}^{(1} &= \mathbf{\Phi}^{(1}\mathbf{\Theta}^{(1} + \mathbf{\Xi}^{(1}, \quad \text{if } y(t-d) \leqslant T \\ \mathbf{y}^{(2} &= \mathbf{\Phi}^{(2}\mathbf{\Theta}^{(2} + \mathbf{\Xi}^{(2}, \quad \text{if } y(t-d) > T \end{aligned}\right\} \tag{51}$$

For given values of the time lag d and the threshold T, two subset models can quickly be identified from (51) using the orthogonal estimator. The optimal values of d and T can be found, based on AIC(η), by searching through a set of d and T (Tong and Lim 1980).

8. Model validation

If the model structure and parameter values are correct, $\varepsilon(t)$ will be unpredictable from (uncorrelated with) all linear and non-linear combinations of past inputs and

outputs; this can be tested by means of the following correlation functions (Billings and Voon 1986, Billings and Chen 1989)

$$\left.\begin{aligned} \Psi_{\varepsilon\varepsilon}(k) &= 0 \quad k \neq 0 \\ \Psi_{u\varepsilon}(k) &= 0 \quad \forall k \\ \Psi_{\varepsilon(\varepsilon u)}(k) &= 0 \quad k \geqslant 0 \\ \Psi_{u^{2'}\varepsilon}(k) &= 0 \quad \forall k \\ \Psi_{u^{2'}\varepsilon^2}(k) &= 0 \quad \forall k \end{aligned}\right\} \tag{52}$$

where $\varepsilon u(t) = \varepsilon(t+1)u(t+1)$, $u^{2'}(t) = u^2(t) - \overline{u^2(t)}$ and the bar indicates the time average. If these correlation functions fall within the 95 % confidence interval $\pm 1{\cdot}96/\sqrt{N}$, the model is regarded as adequate.

Alternatively a statistical test known as the chi-squared test (Bohlin 1978, Leontaritis and Billings 1987 a) can be used to validate an estimated model. Define an s-dimensional vector-valued function

$$\boldsymbol{\Omega}(t) = \boldsymbol{\Omega}(y(1), \ldots, y(t-1), u(1), \ldots, u(t-1), \varepsilon(1), \ldots, \varepsilon(t-1)) \tag{53}$$

and

$$\boldsymbol{\Gamma}^{\mathrm{T}}\boldsymbol{\Gamma} = \frac{1}{N}\sum_{t=1}^{N} \boldsymbol{\Omega}(t)\boldsymbol{\Omega}^{\mathrm{T}}(t) \tag{54}$$

Then the chi-squared statistic is calculated using the formula

$$\zeta = N\boldsymbol{\mu}^{\mathrm{T}}(\boldsymbol{\Gamma}^{\mathrm{T}}\boldsymbol{\Gamma})^{-1}\boldsymbol{\mu} \tag{55}$$

where

$$\boldsymbol{\mu} = \frac{1}{N}\sum_{t=1}^{N} \boldsymbol{\Omega}(t)\varepsilon(t)/\sigma_{\varepsilon} \tag{56}$$

and σ_{ε}^2 is the variance of the residuals. 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.

In previous applications (Leontaritis and Billings 1987 a, Billings and Chen 1989) the following vector-valued function $\boldsymbol{\Omega}(t)$ has been proposed

$$\boldsymbol{\Omega}(t) = [\omega(t) \quad \omega(t-1) \quad \ldots \quad \omega(t-s+1)]^{\mathrm{T}} \tag{57}$$

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

$$\zeta < \chi_s^2(\alpha) \tag{58}$$

the model can be regarded as adequate, where $\chi^2(\alpha)$ is the critical value of the chi-squared distribution with s degrees of freedom for a given level of significance α (0·05). For the current extended-model set the choice of $\omega(t)$ as monomials of past inputs, outputs and prediction errors is, of course, still valid. It is, however, very useful to include commonly existing non-linear functions such as exp (·), sin (·), and atan (·) as additional choices for $\omega(t)$.

9. Simulation study

Three simulated examples are used to illustrate some of the techniques described in the present study.

Example 1

The data were generated by

$$y(t) = 0{\cdot}5y(t-1) + u(t-2) + 0{\cdot}1u^2(t-1) + 0{\cdot}5e(t-1) + 0{\cdot}2u(t-1)e(t-2) + e(t)$$

where the system noise $e(t)$ was a gaussian white-noise sequence with mean zero and variance 0·04 and the system input $u(t)$ was a uniformly distributed independent sequence with mean zero and variance 1·0. The inputs and outputs of the system are shown in Fig. 2. This example was used in Chen *et al.* (1989) and the current purpose is to demonstrate how the intelligent structure-detection strategy of § 5 works.

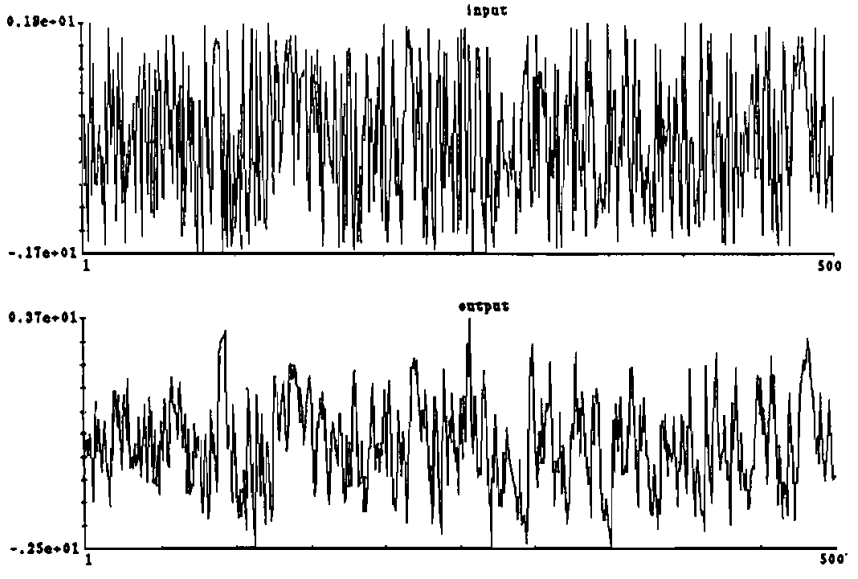


Figure 2. Inputs and outputs of Example 1.

A monomial model set with $n_y = n_u = n_e = 3$ and non-linear degree $l = 3$ was used to fit the data. The full model set contained 220 terms. Assume that the system is a black box and the experimenter simply assigns an initial value 0·032 to ρ_p. The results obtained using the structure detection procedure of § 5 are shown in Table 1. It is seen that the appropriate range of ρ_r for this example is

$$0{\cdot}025 < \rho_r \leqslant 0{\cdot}026$$

and the iterative procedure found an adequate value for ρ_r through learning. Notice how the structure of the system (or the terms within the estimated model set) changes as the algorithm learns ρ_r, finally converging to both the correct model structure and parameter estimates.

Example 2

The data were generated by

$$y(t) = 0{\cdot}5y(t-2) + u(t-1) + 0{\cdot}4[\tanh(u(t-2))]^2 + 0{\cdot}5e(t-1) + 0{\cdot}2\sin(y(t-1))e(t-2) + e(t)$$

	Term	Estimate	$[\text{err}]_i$	Standard deviation
Initial stage $\rho_p = 0{\cdot}032$	$u(t-2)$	0·10110e + 01	0·67104e + 00	0·10109e − 01
	$y(t-1)$	0·63448e + 00	0·28703e + 00	0·21079e − 01
	$u^2(t-1)$	0·86768e − 01	0·85520e − 02	0·75799e − 02
	$y(t-3)u(t-2)$	−0·20542e − 01	0·63062e − 03	0·78179e − 02
	$u^3(t-3)$	−0·64772e − 01	0·55407e − 03	0·11322e − 01
	$y(t-2)$	−0·68662e − 01	0·15841e − 02	0·13537e − 01
	$\hat{\sigma}_\varepsilon^2$	0·50169e − 01		
	$\hat{\sigma}_\varepsilon^2/\sigma_y^2$	0·30603e − 01		
1st iteration $\rho_r = 0{\cdot}03$	$u(t-2)$	0·10073e + 01	0·67104e + 00	0·93644e − 02
	$y(t-1)$	0·50464e + 00	0·28703e + 00	0·77004e − 02
	$u^2(t-1)$	0·92469e − 01	0·85520e − 02	0·71586e − 02
	$\varepsilon(t-1)$	0·40841e + 00	0·49355e − 02	0·43167e − 01
	$\hat{\sigma}_\varepsilon^2$	0·44924e − 01		
	$\hat{\sigma}_\varepsilon^2/\sigma_y^2$	0·27404e − 01		
2nd iteration $\rho_r = 0{\cdot}027$	$u(t-2)$	0·10052e + 01	0·67103e + 00	0·89177e − 02
	$y(t-1)$	0·50226e + 00	0·28703e + 00	0·73259e − 02
	$u^2(t-1)$	0·90645e − 01	0·85520e − 02	0·68288e − 02
	$\varepsilon(t-1)$	0·49454e + 00	0·61062e − 02	0·43467e − 01
	$u(t-1)\varepsilon(t-2)$	0·22055e + 00	0·14793e − 02	0·40539e − 01
	$\hat{\sigma}_\varepsilon^2$	0·40719e − 01		
	$\hat{\sigma}_\varepsilon^2/\sigma_y^2$	0·24839e − 01		
3rd iteration $\rho_r = 0{\cdot}025$	$u(t-2)$	0·10045e + 01	0·67104e + 00	0·88643e − 02
	$y(t-1)$	0·50171e + 00	0·28703e + 00	0·72757e − 02
	$u^2(t-1)$	0·90395e − 01	0·85520e − 02	0·67828e − 02
	$\varepsilon(t-1)$	0·54986e + 00	0·68936e − 02	0·45330e − 01
	$u(t-1)\varepsilon(t-2)$	0·25074e + 00	0·17351e − 02	0·42299e − 01
	$\hat{\sigma}_\varepsilon^2$	0·40230e − 01		
	$\hat{\sigma}_\varepsilon^2/\sigma_y^2$	0·24540e − 01		
4th iteration $\rho_r = 0{\cdot}025$	$u(t-2)$	0·10033e + 01	0·67047e + 00	0·89478e − 02
	$y(t-1)$	0·50289e + 00	0·28735e + 00	0·73364e − 02
	$u^2(t-1)$	0·90966e − 01	0·85640e − 02	0·68505e − 02
	$\varepsilon(t-1)$	0·54796e + 00	0·69244e − 02	0·45523e − 01
	$u(t-1)\varepsilon(t-2)$	0·23848e + 00	0·16174e − 02	0·42076e − 01
	$\hat{\sigma}_\varepsilon^2$	0·41020e − 01		
	$\hat{\sigma}_\varepsilon^2/\sigma_y^2$	0·25073e − 01		
5th iteration $\rho_r = 0{\cdot}0251$	$u(t-2)$	0·10032e + 01	0·67047e + 00	0·89479e − 02
	$y(t-1)$	0·50281e + 00	0·28735e + 00	0·73374e − 02
	$u^2(t-1)$	0·91097e − 01	0·85640e − 02	0·68514e − 02
	$\varepsilon(t-1)$	0·54843e + 00	0·69317e − 02	0·45539e − 01
	$u(t-1)\varepsilon(t-2)$	0·23785e + 00	0·16098e − 02	0·42062e − 01
	$\hat{\sigma}_\varepsilon^2$	0·41019e − 01		
	$\hat{\sigma}_\varepsilon^2/\sigma_y^2$	0·25072e − 01		

Table 1. Iterative procedure for Example 1.

where $e(t)$ was a gaussian white-noise sequence with zero mean and variance 1·0 and $u(t)$ was an independent sequence of uniform distribution with zero mean and variance 2·0. The inputs and outputs of the system are plotted in Fig. 3.

The correct model structure was identified using the procedure of § 5 from an

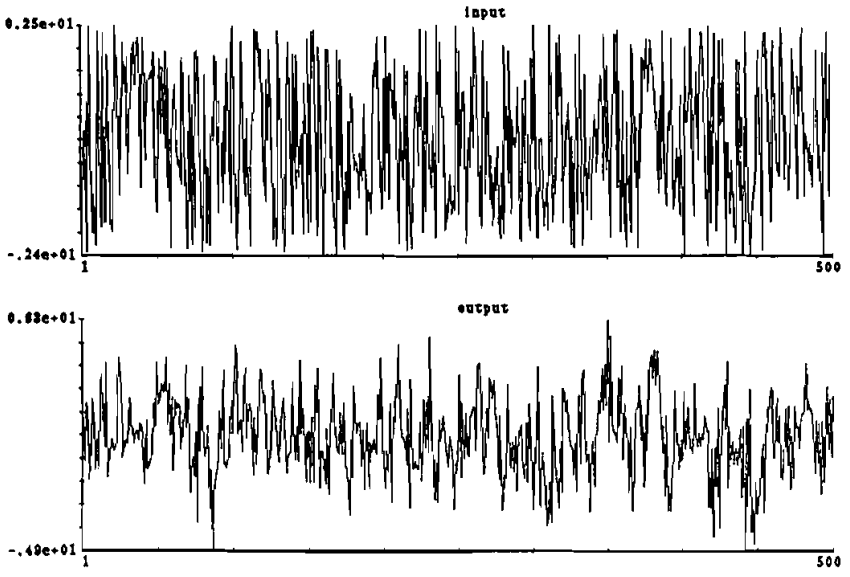
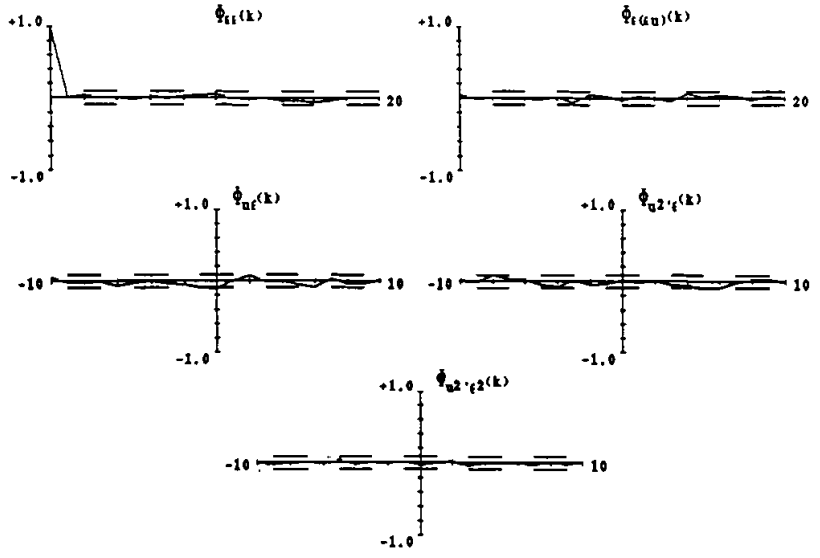


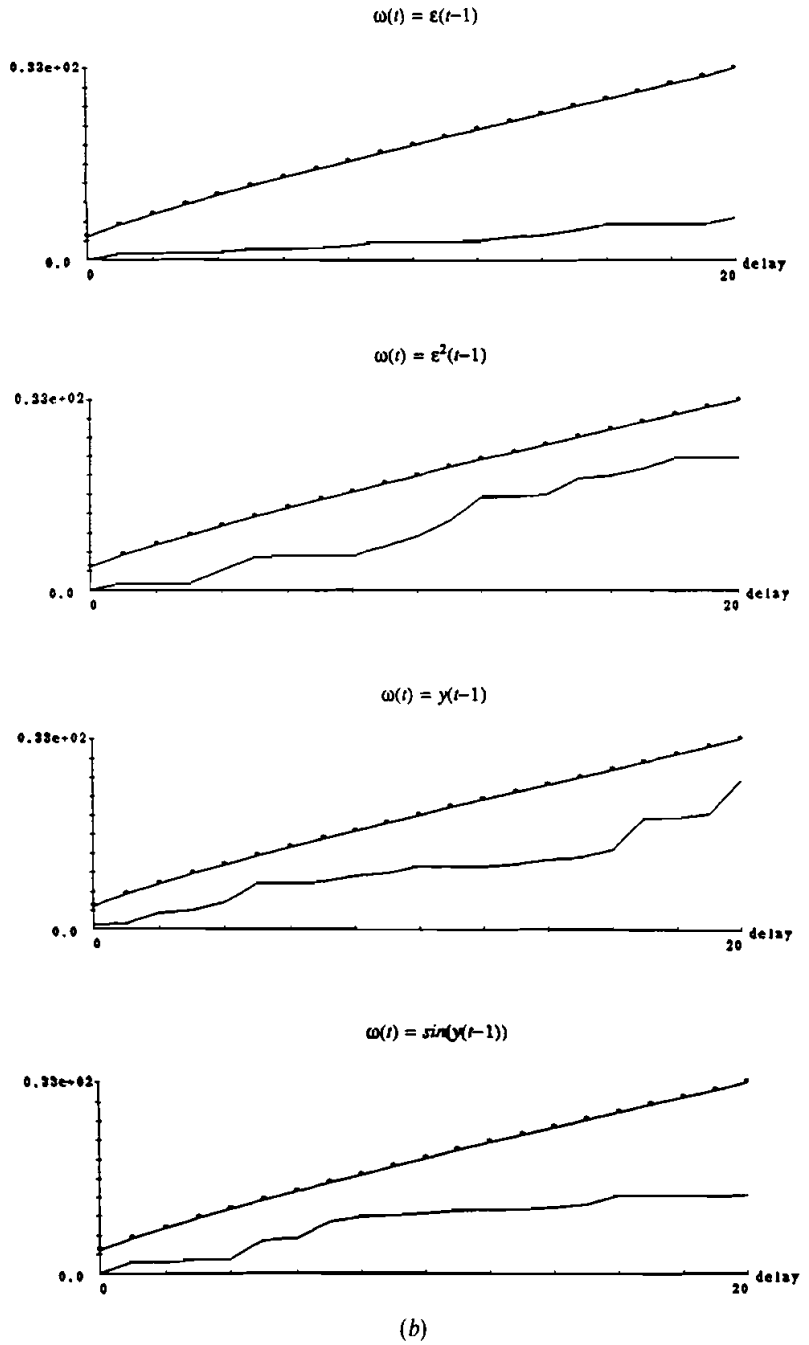
Figure 3. Inputs and outputs of Example 2.

Term	Estimate	$[\text{err}]_i$	Standard deviation
$u(t-1)$	0·98346e + 00	0·41375e + 00	0·31111e − 01
$y(t-2)$	0·45094e + 00	0·25516e + 00	0·22825e − 01
$\varepsilon(t-1)$	0·55903e + 00	0·69506e − 01	0·45110e − 01
$[\tanh(u(t-2))]^2$	0·45231e + 00	0·21649e − 01	0·66539e − 01
$\varepsilon(t-2)\sin(y(t-1))$	0·17601e + 00	0·36649e − 02	0·63326e − 01
$\hat{\sigma}_\varepsilon^2$	0·99213e + 00		
$\hat{\sigma}_\varepsilon^2/\sigma_y^2$	0·23627e + 00		

Table 2. Extended-model-set identification of Example 2.



4(*a*)



(*b*)

Figure 4. Model validation for extended-model-set identification of Example 2; (*a*) correlation tests; (*b*) chi-squared tests (—▲—▲—▲— 95 % confidence limit).

extended model set containing 28 monomial terms ($l = n_y = n_u = n_\varepsilon = 2$), and 11 other terms: $\tanh(u(t-i))$, $\tanh(u(t-i))\tanh(u(t-j))$, $\sin(y(t-i))$, $\sin(y(t-i))e(t-j)$ $(i, j = 1, 2)$. The final results are gived in Table 2. The model validity tests of Fig. 4 confirm that this model is adequate.

As a comparison, a polynomial model was also identified from a monomial set of 210 terms ($l = 4$ and $n_y = n_u = n_\varepsilon = 2$). The results are shown in Table 3. Although this polynomial model is adequate, as can be seen from Fig. 5, it requires more terms than does the model that was obtained on the basis of the extended model set.

Term	Estimate	$[\text{err}]_i$	Standard deviation
$u(t-1)$	0·75166e + 00	0·41375e + 00	0·81280e − 01
$y(t-2)$	0·45999e + 00	0·25516e + 00	0·23016e − 01
$\varepsilon(t-1)$	0·55759e + 00	0·66522e − 01	0·45248e − 01
$u^2(t-2)$	0·24676e + 00	0·17627e − 01	0·58362e − 01
$u^4(t-2)$	−0·32453e − 01	0·42702e − 02	0·12790e − 01
$u^3(t-1)$	0·49579e − 01	0·29811e − 02	0·19628e − 01
$y^2(t-2)u(t-1)$	0·11789e − 01	0·23400e − 02	0·52121e − 02
$y^3(t-1)\varepsilon(t-2)$	−0·93298e − 02	0·16952e − 02	0·28514e − 02
$y(t-1)\varepsilon(t-2)$	0·10523e + 00	0·32981e − 02	0·39565e − 01
$\hat{\sigma}_\varepsilon^2$	0·97497e + 00		
$\hat{\sigma}_\varepsilon^2/\sigma_y^2$	0·23218e + 00		

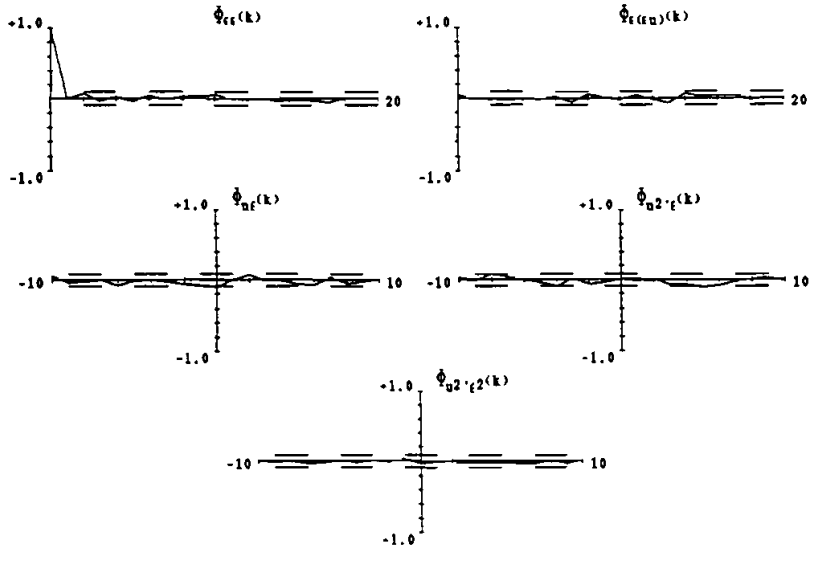
Table 3. Polynomial identification of Example 2.

Example 3

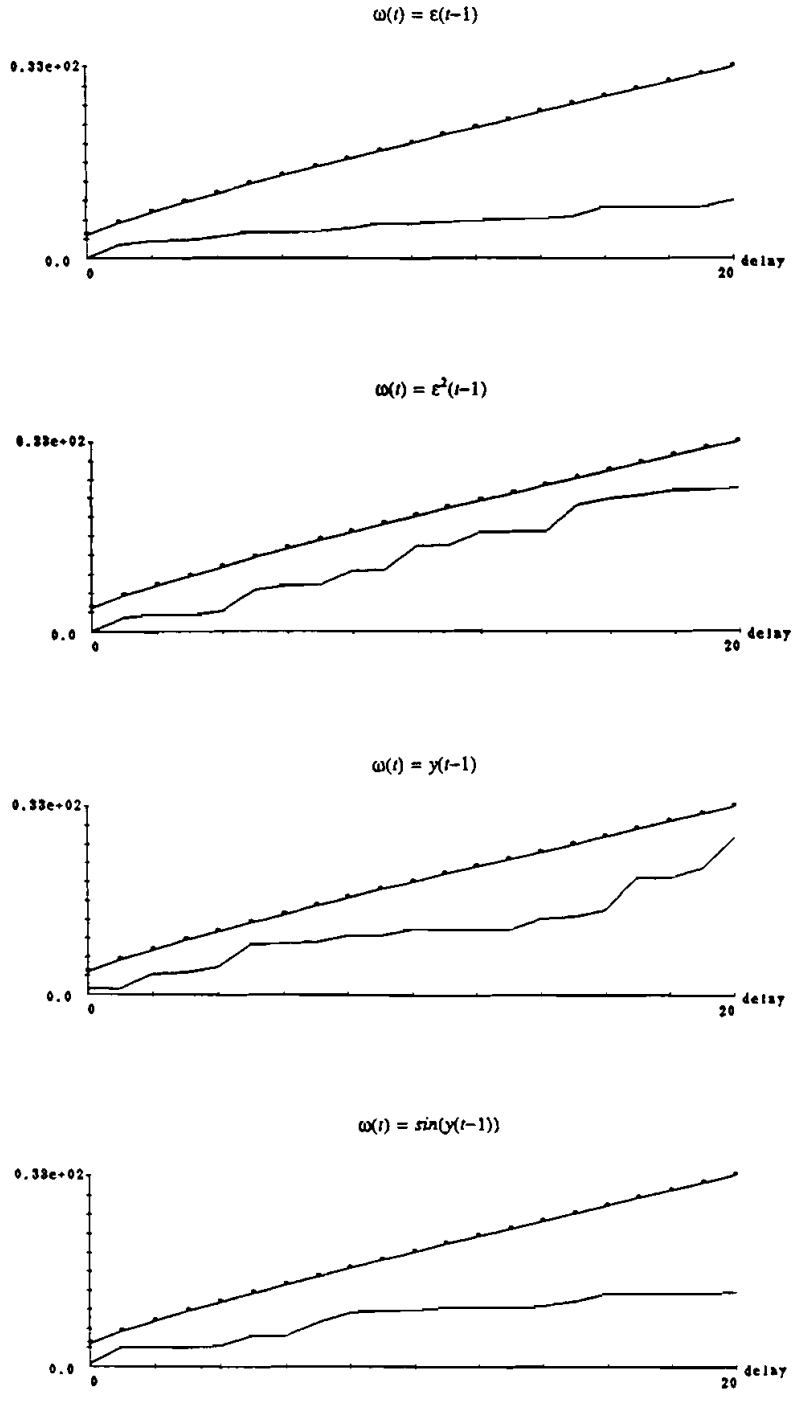
Two sets of data were generated from

$$y(t) = [0{\cdot}4 - 0{\cdot}3\exp(-u^2(t-2))]y(t-1) + [0{\cdot}5 + 0{\cdot}2\exp(-u^2(t-1))]y(t-2) + u(t-1) + 0{\cdot}2u(t-2) + e(t)$$

where $e(t)$ was a gaussian white-noise sequence with zero mean and variance 1·0, $u(t) = u_m + w(t)$, $w(t)$ was an independent sequence of uniform distribution with zero



5(*a*)



(*b*)

Figure 5. Model validation for polynomial identification of Example 2; (*a*) correlation tests; (*b*) chi-square tests (—•—•— 95% confidence limit).

mean and variance σ_w^2 and

$$u_m = 0{\cdot}0 \quad \text{and} \quad \sigma_w^2 = 1{\cdot}0 \qquad \text{for the first data set}$$
$$u_m = 10{\cdot}0 \quad \text{and} \quad \sigma_w^2 = 0{\cdot}04 \qquad \text{for the second data set}$$

These two sets of inputs and outputs are plotted in Fig. 6. This example was designed to represent a situation where two experiments had been performed on a plant, each about a different operating level. The object was to analyse the data as described in § 6 and to estimate a global data model.

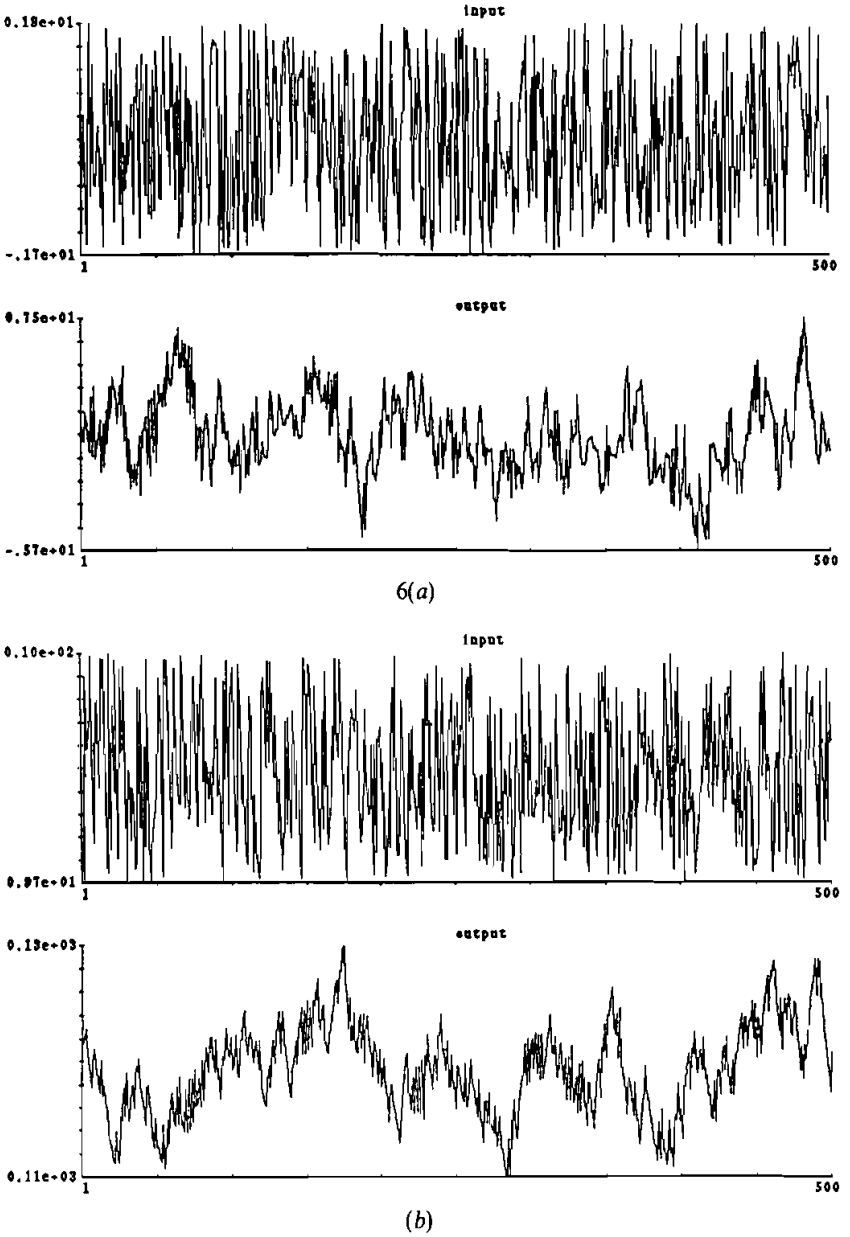
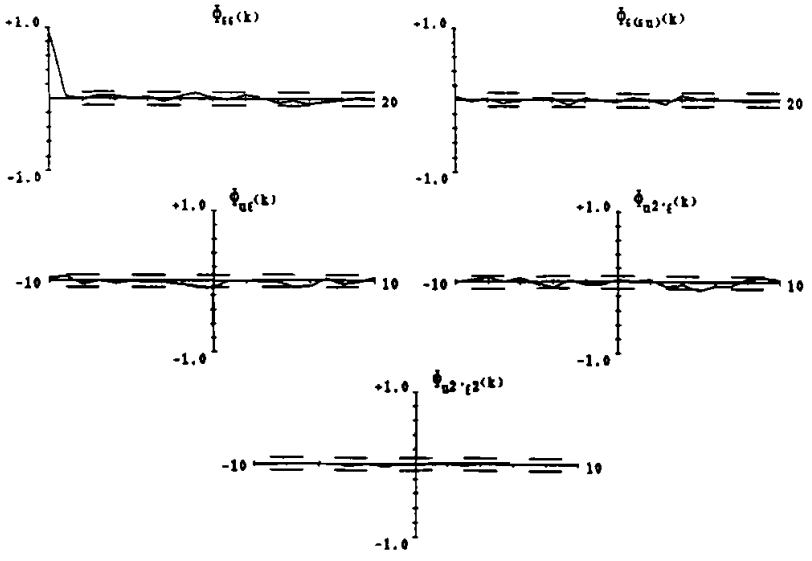


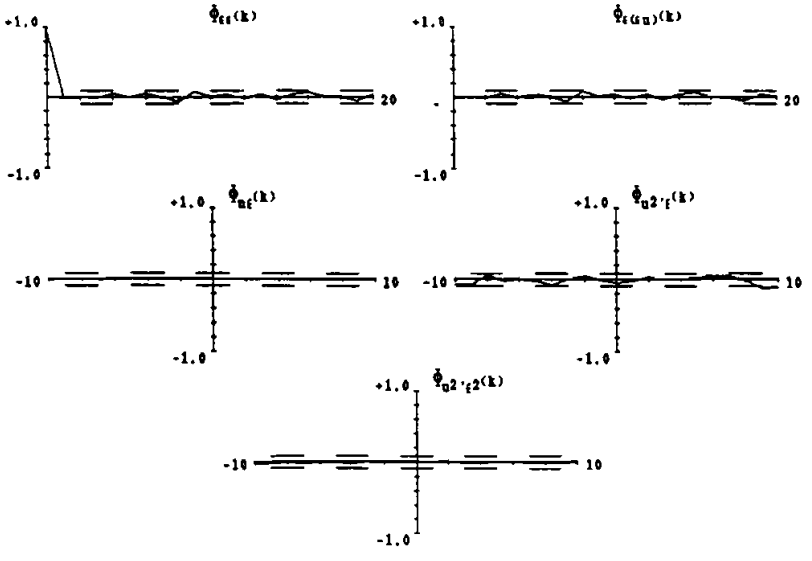
Figure 6. Inputs and outputs of Example 3; (*a*) 1st data set; (*b*) 2nd data set.

The correct global model was identified from the extended model set: 28 monomial terms ($l = 2$, $n_y = n_u = 3$ and $n_\varepsilon = 0$), and 21 other terms: $\exp(-u^2(t-i))$, $\exp(-u^2(t-i))u(t-j)$, $\exp(-u^2(t-i))y(t-j)$ $(i,j = 1, 2, 3)$. The final results are given in Table 4 where AIC(4·0) was used to terminate the process-model regression procedure and the model validity tests are shown in Fig. 7.

As a comparison, models were also fitted to each of the two data sets separately. The models were selected from the same 49-terms that defined the extended model set. The results are presented in Tables 5 and 6, respectively, where AIC(4·0) was used to terminate the model selection. These two models seem to fit the corresponding data



7(*a*)



7(*b*)

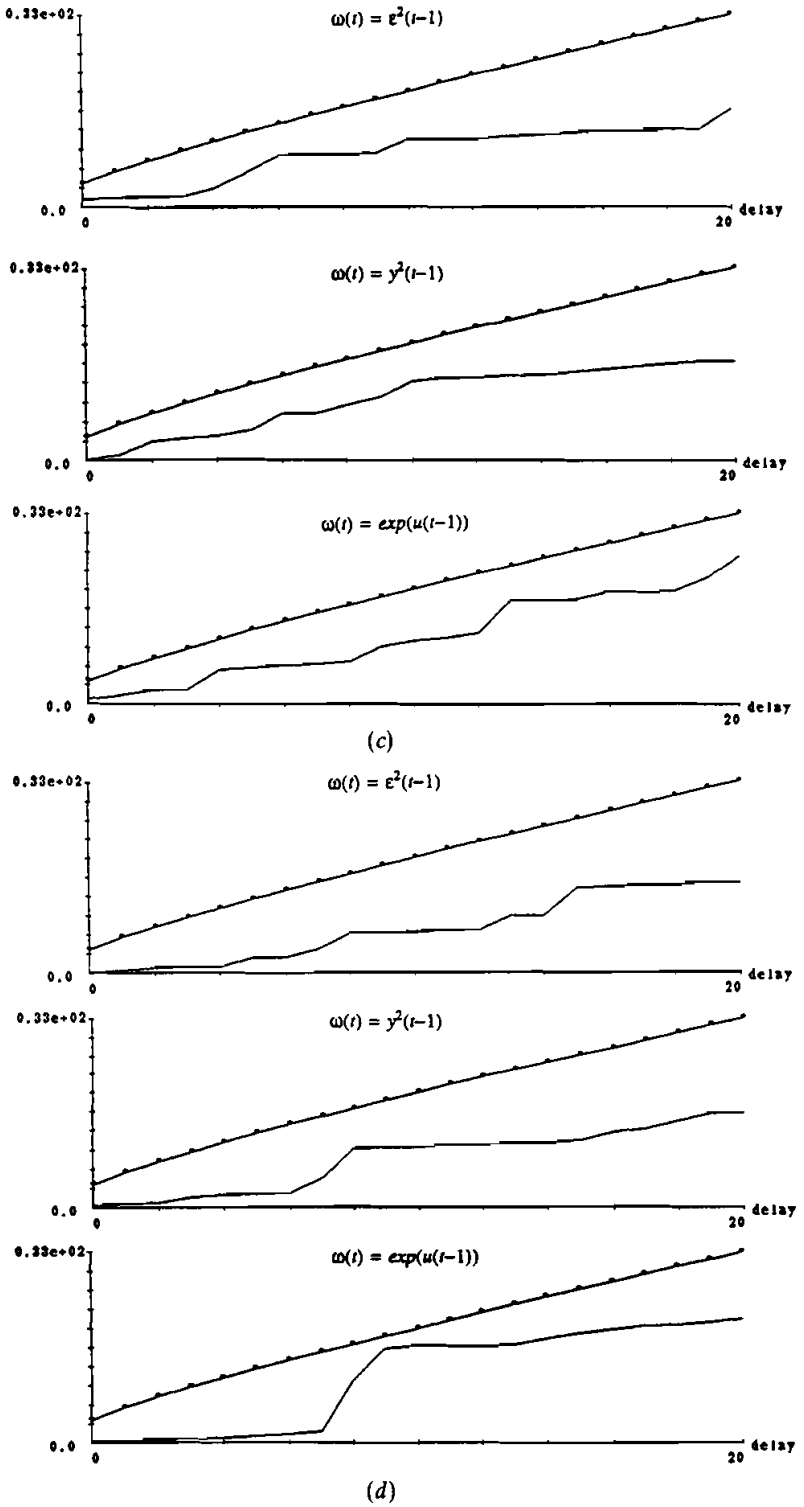
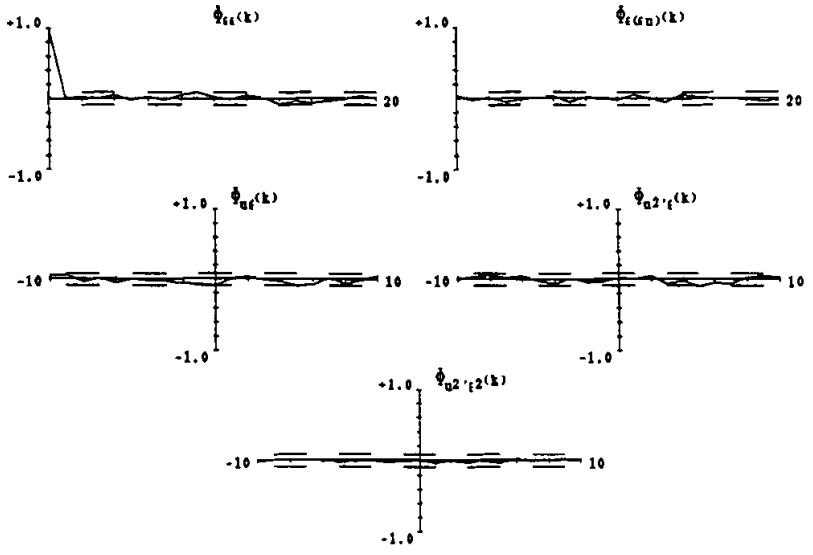
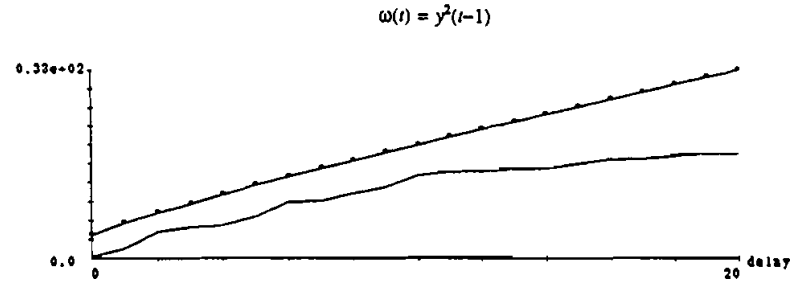


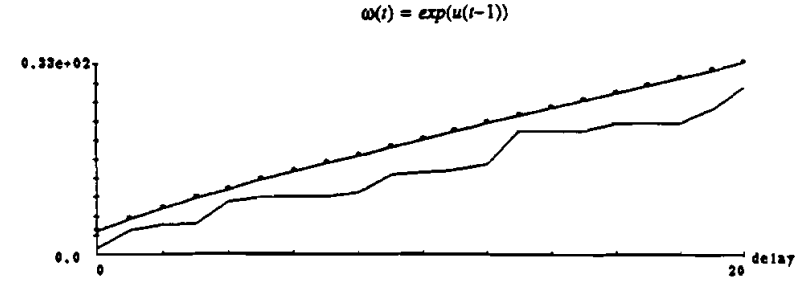
Figure 7. Model validation for global-model fitting of Example 3; (*a*) correlation tests (1st data set); (*b*) correlation tests (2nd data set); (*c*) chi-squared tests (1st data set); (*d*) chi-squared tests (2nd data set) (—•—•— 95 % confidence limit).

well, as can be seen from Figs 8 and 9. Both the models are different from the correct global model because neither of the inputs alone covers the full amplitude range of the plant; the inputs taken one at a time are not persistently exciting (Leontaritis and Billings 1987 b).



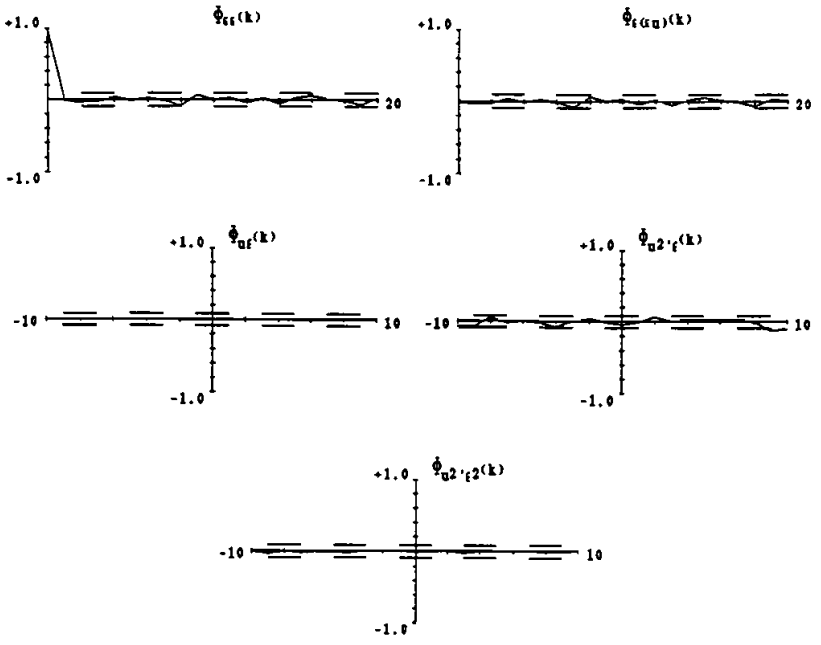
(*a*)



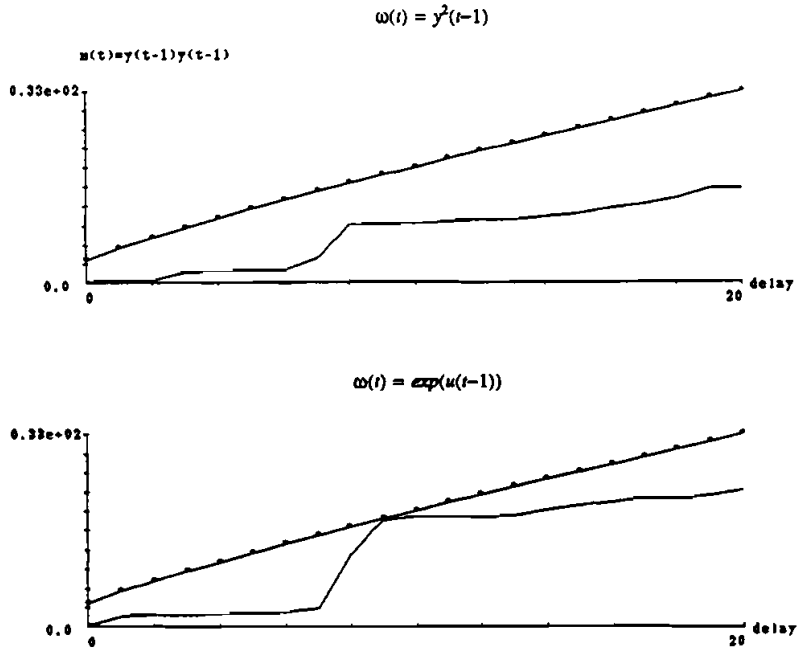


(*b*)

Figure 8. Model validation for model given in Table 5 (using 1st data set); (*a*) correlation tests; (*b*) chi-squared tests (—·—·— 95 % confidence limit).



(a)



(b)

Figure 9. Model validation for model given in Table 6 (using 2nd data set); (*a*) correlation tests; (*b*) chi-squared tests (—•—•— 95 % confidence limit).

Term	Estimate	$[\text{err}]_i$	Standard deviation
$y(t-2)$	0·48590e + 00	0·99969e + 00	0·25534e − 01
$u(t-1)$	0·10238e + 01	0·10589e − 03	0·42357e − 01
$y(t-1)$	0·40762e + 00	0·57197e − 04	0·27508e − 01
$u(t-2)$	0·25041e + 00	0·44852e − 05	0·50485e − 01
$y(t-1)\exp(-u^2(t-2))$	−0·27497e + 00	0·31337e − 05	0·43928e − 01
$y(t-2)\exp(-u^2(t-1))$	0·17738e + 00	0·23507e − 05	0·43229e − 01
combined $\hat{\sigma}_\varepsilon^2$	0·99334e + 00		
combined $\hat{\sigma}_\varepsilon^2/\sigma_y^2$	0·13878e − 03		

Table 4. Global model fitting of Example 3.

Term	Estimate	$[\text{err}]_i$	Standard deviation
$y(t-2)$	0·45036e + 00	0·52427e + 00	0·41539e − 01
$u(t-1)$	0·10363e + 01	0·19738e + 00	0·44163e − 01
$y(t-1)$	0·47644e + 00	0·77093e − 01	0·40412e − 01
$y(t-1)\exp(-u^2(t-2))$	−0·34116e + 00	0·13238e − 01	0·60462e − 01
$u(t-2)\exp(-u^2(t-2))$	0·68683e + 00	0·64432e − 02	0·16657e + 00
$y(t-2)\exp(-u^2(t-1))$	0·19363e + 00	0·36759e − 02	0·60421e − 01
$\hat{\sigma}_\varepsilon^2$	0·99745e + 00		
$\hat{\sigma}_\varepsilon^2/\sigma_y^2$	0·17789e + 00		

Table 5. Identifying the first data set of Example 3.

Term	Estimate	$[\text{err}]_i$	Standard deviation
$y(t-2)$	0·50657e + 00	0·99990e + 00	0·37121e − 01
$y(t-1)$	0·40991e + 00	0·21629e − 04	0·36902e − 01
$u(t-1)$	0·10090e + 01	0·49086e − 05	0·17320e + 00
$\hat{\sigma}_\varepsilon^2$	0·99027e + 00		
$\hat{\sigma}_\varepsilon^2/\sigma_y^2$	0·69203e − 04		

Table 6. Identifying the second data set of Example 3.

10. Conclusions

Special attention has been given to non-linear models that are linear-in-the-parameters. The concept of an extended model set has been introduced by augmenting the set of polynomial models with other functions that commonly exist in non-linear systems and this provides a unified representation for a large class of non-linear systems. Intelligent structure detection strategies have been combined with orthogonal least-squares estimators to provide efficient procedures for identifying non-linear systems, and it has been demonstrated how these algorithms can update the model structure and parameter estimates at each iteration.

It is often advantageous to identify a non-linear model that is valid over the total allowable operating regime whenever possible; and a global model fitting technique

has been proposed. Threshold modelling based on the TNARMAX model has also been discussed and shown to be complementary to the idea of global data modelling.

Simulation results have been included to illustrate the application of algorithms. Extensive applications of the extended-model-set and global-and threshold-model methods fitting real data are currently under way.

ACKNOWLEDGMENTS

This work was supported by the Science and Engineering Research Council (Grant reference GR/D/30587).

REFERENCES

BILLINGS, S. A., and CHEN, S., 1989, Identification of non-linear rational systems using a prediction-error estimation algorithm. *International Journal of Systems Science*, **20**, 467–494.

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., KORENBERG, M. J., and CHEN, S., 1988, Identification of non-linear output-affine systems using an orthogonal least-squares algorithm. *International Journal of Systems Science*, **19**, 1559–1568.

BILLINGS, S. A., and LEONTARITIS, I. J., 1981, Identification of nonlinear systems using parameter estimation techniques. *Proceedings of the I.E.E. Conference on Control and its Applications*, Warwick, England, pp. 83–187.

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; 1987, Piecewise linear identification of non-linear systems. *Ibid.*, **46**, 215–235.

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

BOHLIN, T., 1978, Maximum-power validation of models without higher-order fitting. *Automatica*, **14**, 137–146.

BOZDOGAN, H., 1988, ICOMP: a new model-selection criterion. *Classification and Related Methods of Data Analysis*, edited by H. H. Bock (Amsterdam: Elsevier), pp. 599–608.

CHEN, S., and BILLINGS, S. A., 1989, Representations of non-linear systems: the NARMAX model. *International Journal of Control*, **49**, 1013–1032.

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.

DANG VAN MIEN, H., and NORMAND-CYROT, D., 1984, Nonlinear state affine identification methods: applications to electrical power plants. *Automatica*, **20**, 175–188.

GENTLEMAN, W. M., 1973, Least squares computations by Givens transformations without square roots. *Journal of the Institute of Mathematics and its Applications*, **12**, 329–336.

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

KORENBERG, M. J., BILLINGS, S. A., LIU, Y. P., and MCILROY, P. J., 1988, Orthogonal parameter estimation algorithm for non-linear stochastic systems. *International Journal of Control*, **48**, 193–210.

LAI, H. Y., and HSIEH SU-HUA, 1988, Threshold modelling of nonlinear dynamic systems. *6th International Conference on Modal Analysis*, Hyatt Orlando, U.S.A., pp. 89–95.

LEONTARITIS, I. J., and BILLINGS, S. A., 1985, Input–output parameter 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 a, Model selection and validation methods for non-linear systems. *Ibid.*, **45**, 311–341; 1987 b, Experimental design and

identifiability for non-linear systems. *International Journal of Systems Science*, **18**, 189–202.

LIU, Y. P., KORENBERG, M. J., BILLINGS, S. A., and FADZIL, M. B., 1987, The nonlinear identification of a heat exchanger. Presented at the *26th I.E.E.E. Conference on Decision and Control*, Los Angeles, California, U.S.A.

OZAKI, T., 1981, Non-linear threshold autoregressive models for non-linear random vibrations. *Journal of Applied Probability*, **18**, 443–451; 1985, Non-linear time series models and dynamical systems. *Handbook of Statistics 5: Time Series in the Time Domain* (edited by E. J. Hannan, P. R. Krishnaiah and M. M. Rao) (Amsterdam: North-Holland), pp. 25–83.

POLCHAI, A., and HSU, C. S., 1985, Domain of stability of synchronous generators by a cell mapping approach. *International Journal of Control*, **41**, 1253–1271.

SÖDERSTRÖM, T., 1977, On model structure testing in system identification. *International Journal of Control*, **26**, 1–18.

TONG, H., and LIM, K. S., 1980, Threshold autoregression, limit cycles and cyclical data. *Journal of the Royal Statistical Society*, Pt B, **42**, 245–292.