Improved Structure Selection for Nonlinear Models Based on Term Clustering

L A Aguirre, and S A Billings

Department of Automatic Control and Systems Engineering
University of Sheffield
P.O. Box 600
Mappin Street
Sheffield S1 4DU
United Kingdom

Research Report No 509

March 1994

Improved Structure Selection for Nonlinear Models Based on Term Clustering

LUIS A. AGUIRRE[†] and S. A. BILLINGS

Department of Automatic Control and Systems Engineering
University of Sheffield
P.O. Box 600, Mappin Street — Sheffield S1 4DU - UK

Abstract

In this paper the concepts of *term clusters* and *cluster coefficients* are defined and used in the context of system identification. It is argued that if a certain type of term in a nonlinear model is spurious, the respective cluster coefficient is small compared to the coefficients of the other clusters represented in the model. Once the spurious clusters have been detected, the corresponding terms can be deleted from the set of candidate terms. The consequences of doing this are i) a drastic reduction in the size of the set of candidate terms and consequently a substantial gain in computation time is achieved, ii) the final estimated model is more likely to reproduce the dynamics of the original system, and iii) the final model is more robust to overparametrization. Numerical examples are included to illustrate the new procedure.

1 Introduction

The use of nonlinear models in system identification and analysis is, in many instances, essential because a number of phenomena observed in practice arise from nonlinearities in the original system. Thus the use of nonlinear models raises the hopes of reproducing and therefore analysing phenomena such as nonlinear oscillations, bifurcations and chaos. These dynamical regimes cannot be produced by linear models.

A number of nonlinear representations are currently available and these include polynomials, rational and piecewise linear models, radial basis functions and neural networks. A difficulty which appears to be common to most nonlinear representations is that the number of possible structures increases exponentially, and even for simple nonlinear systems the total number of terms often becomes impractical. Moreover, excessively complex models have two main disadvantages, namely i) the numerical problem becomes ill-conditioned, and ii) overparametrized nonlinear models tend to destroy the original dynamics.

Consequently several techniques for selecting the best model structure have been suggested in the literature (Billings et al., 1988; Billings et al., 1989; Haber and Unbehauen,

[†]e-mail:aguirre@acse.sheffield.ac.uk

1990; Kadtke et al., 1993). Most of these techniques provide a detailed way of selecting the most relevant terms usually from a large set of candidates or, in some cases, of eliminating the unnecessary terms in a trial model. However, the performance of such methods usually depends on factors such as the sampling period and measurement noise.

It is the objective of this paper to develop tools and define a new term clustering procedure which can be used in conjunction with existing techniques to enhance structure selection for nonlinear models such that overparametrization and numerical ill-conditioning are avoided.

In the present study the concepts of *term clusters* and *cluster coefficients* are defined for polynomial models. Such concepts are applicable when the data has been produced by sampling a continuous process. It is also shown how such concepts can be used as a powerful aid to structure selection. In this respect, a structure selection criterion which has been developed previously (Billings and Chen, 1989; Billings et al., 1989) is used as a basis.

The paper is organised as follows. In § 2.1 term clusters and cluster coefficients are defined. The effect of spurious clusters on the respective model is investigated in § 2.2. The question of how to assess the importance of a given cluster is addressed in § 2.3. Some related results concerning systems with integral action are developed in § 3. Section 4 provides three numerical examples to illustrate the application of the new concepts. Finally, the main points of the paper are summarised in § 5.

2 Term Clustering

In this section the concepts of *term clusters* and *cluster coefficients* are introduced and a few related features are discussed. The application of such concepts to the structure selection of nonlinear models is illustrated in § 4.

2.1 Definition

Consider the nonlinear autoregressive moving average model with exogenous inputs (NARMAX) (Billings and Leontaritis, 1981; Leontaritis and Billings, 1985a; Leontaritis and Billings, 1985b)

$$y(k) = F^{\ell}[\, y(k-1), \ldots, y(k-n_y), u(k-d), \quad \ldots \quad , u(k-d-n_u+1), e(k), \ldots \\ \ldots \quad , e(k-n_e)\,] \;, \tag{1}$$

where n_y, n_u and n_e are the maximum lags considered for the output, input and noise terms, respectively and d is the delay measured in sampling intervals, T_s. Moreover, $u(k)$ and $y(k)$ are respectively input and output time series obtained by sampling the continuous data $u(t)$ and $y(t)$ at T_s, $e(k)$ accounts for uncertainties, possible noise, unmodelled dynamics, etc.

and $F^{\ell}[\cdot]$ is some nonlinear function of $y(k)$, $u(k)$ and $e(k)$ with nonlinearity degree $\ell \in \mathbb{Z}^+$. In this paper, the map $F^{\ell}[\cdot]$ is taken to be a polynomial of degree ℓ.

The deterministic part of a NARMAX model, that is, a NARX model, can be expanded as the summation of terms with degrees of nonlinearity in the range $1 \leq m \leq \ell$. Each mth-order term can contain a pth-order factor in $y(k-n_i)$ and a $(m-p)$th-order factor in $u(k-n_i)$ and is multiplied by a coefficient $c_{p,m-p}(n_1,\ldots,n_m)$ as follows (Peyton-Jones and Billings, 1989)

$$y(k) = \sum_{m=0}^{\ell} \sum_{p=0}^{m} \sum_{n_1,n_m}^{n_y,n_u} c_{p,m-p}(n_1,\ldots,n_m) \prod_{i=1}^{p} y(k-n_i) \prod_{i=p+1}^{m} u(k-n_i) \ , \tag{2}$$

where

$$\sum_{n_1,n_m}^{n_y,n_u} \equiv \sum_{n_1=1}^{n_y} \cdots \sum_{n_m=1}^{n_u} \ , \tag{3}$$

and the upper limit is n_y if the summation refers to factors in $y(k-n_i)$ or n_u for factors in $u(k-n_i)$. For instance, expanding equation (2) up to second order, that is, $\ell = 2$ gives

$$\begin{aligned} y(k) \ = \ & c_{0,0} + \sum_{n_1=1}^{n_y} c_{1,0}(n_1) y(k-n_1) + \sum_{n_1=1}^{n_u} c_{0,1}(n_1) u(k-n_1) \\ & + \sum_{n_1}^{n_y} \sum_{n_2}^{n_y} c_{2,0}(n_1,n_2) y(k-n_1) y(k-n_2) \\ & + \sum_{n_1}^{n_y} \sum_{n_2}^{n_u} c_{1,1}(n_1,n_2) y(k-n_1) u(k-n_2) \\ & + \sum_{n_1}^{n_u} \sum_{n_2}^{n_u} c_{0,2}(n_1,n_2) u(k-n_1) u(k-n_2) \ . \end{aligned} \tag{4}$$

It should be noted that the term coefficients depend on the sampling time and should therefore be represented as $c_{p,m-p}(T_s, n_1, \ldots, n_m)$. However, for the sake of clarity, the argument T_s is dropped.

Example 2.1

The model

$$\begin{aligned} y(t) \ = \ & 2.1579\, y(t-1) - 1.3203\, y(t-2) + 0.16239\, y(t-3) \\ & + 0.22480 \times 10^{-3} y(t-3)^3 - 0.48196 \times 10^{-2} y(t-1)^3 \\ & + 0.19463 \times 10^{-2} u(t-2) + 0.34160 \times 10^{-3} u(t-1) \\ & + 0.35230 \times 10^{-2} y(t-1)^2 y(t-2) \\ & - 0.12162 \times 10^{-2} y(t-1) y(t-2) y(t-3) \end{aligned} \tag{5}$$

can be described equation (2) with

$$\left.\begin{array}{ll} c_{1,0}(1) = 2.1579 & c_{1,0}(2) = -1.3203 \\ c_{1,0}(3) = 0.16239 & c_{3,0}(3,3,3) = 0.22480 \times 10^{-3} \\ c_{3,0}(1,1,1) = -0.48196 \times 10^{-2} & c_{0,1}(2) = 0.19463 \times 10^{-2} \\ c_{0,1}(1) = 0.34160 \times 10^{-3} & c_{3,0}(1,1,2) = 0.35230 \times 10^{-2} \\ c_{3,0}(1,2,3) = -0.12162 \times 10^{-2} & \text{else } c_{p,m-p}(\cdot) = 0 \end{array}\right\} . \tag{6}$$

□

If the sampling time T_s is short enough such that

$$\left.\begin{array}{l} y(k-1) \approx y(k-2) \approx \ldots \approx y(k-n_y) \\ u(k-1) \approx u(k-2) \approx \ldots \approx u(k-n_u) \end{array}\right\} , \tag{7}$$

then equation (2) can be rewritten as

$$y(k) \approx \sum_{n_1,n_m}^{n_y,n_u} c_{p,m-p}(n_1,\ldots,n_m) \sum_{m=0}^{\ell} \sum_{p=0}^{m} y(k-1)^p u(k-1)^{m-p} . \tag{8}$$

Definition 2.1 The constants $\sum_{n_1,n_m}^{n_y,n_u} c_{p,m-p}(n_1,\ldots,n_m)$ in equation (8) are the coefficients of the *term clusters* $\Omega_{y^p u^{m-p}}$, which contain terms of the form $y(k-i)^p u(k-j)^{m-p}$ for $m=0,\ldots,\ell$ and $p=0,\ldots,m$. Such coefficients are called *cluster coefficients* and are represented as $\Sigma_{y^p u^{m-p}}$. □

The approximations in equation (7) have been made in order to point out the rationale behind term clustering and also to introduce a more formal definition of term clusters and cluster coefficients. However, such concepts are valid in most applications in which the sampling period is selected in order to enable reliable parameter estimation, as will be illustrated in § 4.

Clearly, the set of candidate terms for a NARX model is the union of all possible clusters up to degree ℓ, that is

$$\begin{aligned} \{\text{all possible terms}\} &= \bigcup_{\substack{p=0\ldots m \\ m=0\ldots \ell}} \Omega_{y^p u^{m-p}} \\ &= \text{constant} \cup \Omega_y \cup \Omega_u \cup \Omega_{y^2} \cup \Omega_{yu} \cup \Omega_{u^2} \cup \ldots \\ &\quad \ldots \cup \text{ all possible combinations up to degree } \ell . \end{aligned} \tag{9}$$

Example 2.2

The cluster coefficients of the model in equation (5) are

$$\left.\begin{array}{l} c_{1,0}(1)+c_{1,0}(2)+c_{1,0}(3)=\Sigma_y=0.99999 \\ c_{3,0}(3,3,3)+c_{3,0}(1,1,1)+c_{3,0}(1,1,2)+c_{3,0}(1,2,3)=\Sigma_{y^3}=-2.2880\times10^{-3} \\ c_{0,1}(2)+c_{0,1}(1)=\Sigma_u=2.2879\times10^{-3} \end{array}\right\} \quad (10)$$

which correspond to the term clusters Ω_y, Ω_{y^3} and Ω_u, respectively. □

Therefore a cluster, $\Omega_{y^p u^{m-p}}$, is a set of terms of the form $y(k-i)^p u(k-j)^{m-p}$ for $m=0,\ldots,\ell$ and $p=0,\ldots,m$, and the respective coefficient, $\Sigma_{y^p u^{m-p}}$, is the summation corresponding to the coefficients of all the terms in a model which are contained in such a cluster. Moreover, it is interesting to note that in the limit when $T_s \to 0$ all the terms in a cluster become indistinguishable and all the cluster coefficients vanish except Σ_y which in the limit equals unity, that is

$$\lim_{T_s\to 0}\Sigma_y=1, \qquad \lim_{T_s\to 0}\Sigma_{y^p u^{m-p}}=0 \ \text{ for all other clusters .} \quad (11)$$

2.2 The effect of spurious clusters

In this subsection it will be argued that the dynamical effect on the residuals of terms taken from spurious clusters is negligible. In order to show this a rather intuitive approach will be followed.

Assume that a set of data has been recorded by sampling the output and possibly the input of a continuous dynamical system. If the model

$$y(k)=\Psi^{\mathrm{T}}(k-1)\hat{\Theta}+\xi(k) \quad (12)$$

is fitted to such data and the parameter estimates are unbiased, it is known that the residuals, $\xi(k)$, should be unpredictable from all combinations of linear and nonlinear terms. This can be readily verified using nonlinear correlation functions (Billings and Voon, 1986; Billings and Tao, 1991). In other words, a model is said to be unbiased if it explains all relevant dynamics in the data. If some dynamics are left unexplained, they would appear in the residuals and would be detected by appropriate nonlinear validation tests.

Consider the following illustration. The Duffing-Ueda system (Ueda, 1980)

$$\ddot{y}(t)+0.1\,\dot{y}(t)+y^3(t)=u(t) \quad (13)$$

was simulated and 1800 data points of the input and output were sampled at $T_s=\pi/100$. Such data were used to identify a NARX model with fifteen terms which were selected

automatically using a criterion based on the *error reduction ratio* (ERR) (Billings et al., 1988). The estimated model, denoted as in equation (12), can be expressed as

$$y(k) = \Psi_E^T \hat{\Theta}_E + [\Psi_S^T \hat{\Theta}_S + \xi(k)] , \tag{14}$$

where $\Psi^T(k-1) = [\Psi_E^T \; \Psi_S^T]$ and $\hat{\Theta} = [\hat{\Theta}_E^T \; \hat{\Theta}_S^T]^T$. The subscripts E and S denote *effective* and *spurious* terms in the model.

In order to assess the dynamical effect of each term in the model, the following procedure was devised. Firstly, a correlation index was defined which gives a measure of the level of nonlinear correlation found in the residuals. The way this index is calculated is irrelevant to the following discussion and is therefore omitted. Secondly, such an index was calculated for fifteen 'sub-models' of the form $y(k) = \Psi_E^T \hat{\Theta}_E$. The first of these sub-models was composed only of the first term of the original fifteen-term model. In other words, initially $\Psi_E^T \hat{\Theta}_E$ = first term of $\{\Psi^T(k-1)\hat{\Theta}\}$ and the last sub-model was the complete fifteen-term model, hence finally $\Psi_E^T \hat{\Theta}_E = \Psi^T(k-1)\hat{\Theta}$.

To calculate the correlation index for the sub-model $y(k) = \Psi_E^T \hat{\Theta}_E$ is effectively to measure the whiteness of the 'residuals' which for such a sub-model are given by the terms in square brackets in equation (14). Hence, if the correlation of such 'residuals' and of $\xi(k)$ is negligible, it seems appropriate to infer that the dynamical contribution of $\Psi_S^T \hat{\Theta}_S$ must also be negligible . It is stressed that in the procedure outlined above, the parameters were only estimated once for the model. The sub-models were then obtained by truncating $\Psi_S^T \hat{\Theta}_S$ off the original fifteen-term model.

The normalised correlation index for the fifteen sub-models is plotted as a function of the number of terms in $\Psi_E^T \hat{\Theta}_E$ in figure 1. Two distinct phases are evident in the results. In the first phase, not all the *effective terms* in the model are included in $\Psi_E^T \hat{\Theta}_E$ and consequently the effect of such terms appears in the 'residuals' corresponding to that particular sub-model. That is $\Psi_S^T \hat{\Theta}_S + \xi(k)$ are coloured and were detected by the correlation tests. This is indicated by a large value of the correlation index. Conversely, when all the effective terms had been included in the sub-model the correlation index produced much lower values.

An important point to note is that the value of the correlation index when $\Psi_E^T \hat{\Theta}_E$ comprised seven terms was as low as for the full model. In other words, the whiteness of $\Psi_S^T \hat{\Theta}_S + \xi(k)$ for the seven-term sub-model is comparable to the whitness of $\xi(k)$ for the full model. Therefore, it seems reasonable to assume that the remaining eight terms in $\Psi_S^T \hat{\Theta}_S$ do not contribute to the residuals and are not explaining any dynamics.

It should be realised that such a conclusion could only be arrived at because parameters were *not* reestimated and therefore the whole eight-term part $\Psi_S^T \hat{\Theta}_S$ should be regarded as ineffective. Consequently, if $\Psi_S^T \hat{\Theta}_S = 0$ is substituted in the original model and the correlation index is calculated, the value shown in figure 1 for the sub-model with seven terms would be obtained. Moreover, such a value is as good as for the complete model and therefore the aforementioned substitution seems justifiable.

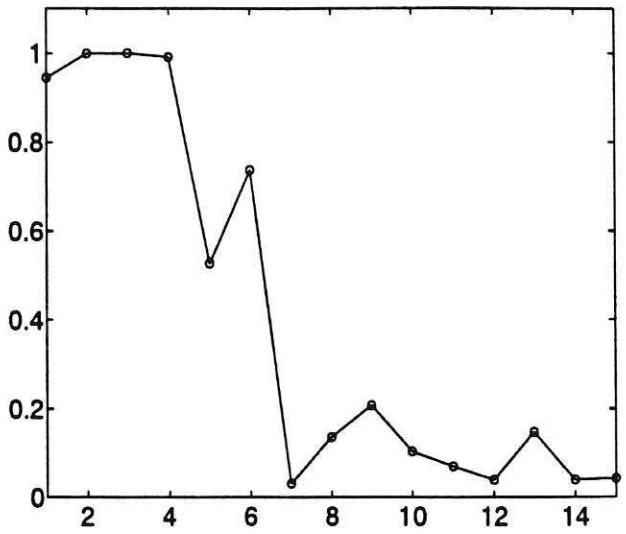

Figure 1: Correlation index as a function of the number of terms in $\Psi_E^T \hat{\Theta}_E$.

The model which was estimated in the above illustration, based on the ERR criterion, produced terms from the following clusters Ω_y, Ω_u, Ω_{y^3}, Ω_{u^3} and Ω_{yu^2}. Interestingly, the first seven terms belong to the first three clusters, whereas the eight terms in $\Psi_S^T \hat{\Theta}_S$ were selected from the two latter clusters. This strongly suggests that the clusters Ω_{u^3} and Ω_{yu^2} are spurious and consequently only terms from Ω_y, Ω_u, Ω_{y^3} should be considered for inclusion in the final model. In fact, if the original system of equation (13) is discretised by Euler's rule, the final model contains the following terms $y(k-1)$, $y(k-2)$, $y(k-1)^3$ and $u(k-1)$ confirming our conclusiong.

These effects seem to be analogous to the pole-zero cancellation effect in linear estimation. Overfitting linear models tends to provide cancelling pole-zero pairs. The additional parameters which correspond to these pairs can be deleted from the model without seriously affecting the whitness of the residuals. The term clustering ideas appear to be the equivalent phenomenon for nonlinear models.

2.3 Verifying the importance of a cluster

In the preceding section it has been argued that the effect of spurious clusters is negligible as far as residuals are concerned. This conclusion has been arrived at by calculating a correlation index for sub-models obtained by truncating the full estimated model. In this section it will be shown that the importance of each cluster is revealed by the respective cluster coefficient. Thus if a certain cluster is spurious, this will be revealed by a negligible cluster coefficient. This is because the dynamical effect of $\Psi_S^T \hat{\Theta}_S$ on the residuals is canceled within each cluster represented in the model. This result is stated in the following lemma.

Lemma 2.1 *Assuming that the contribution of $\Psi_S^T \hat{\Theta}_S$ to the residuals $\xi(k)$ of the model in equation (14) is negligible such that $\Psi_S^T \hat{\Theta}_S \to 0$, then the cluster coefficients of the term clusters represented in $\Psi_S^T \hat{\Theta}_S$ are null.*

Proof.

Grouping the terms of $\Psi_S^T \hat{\Theta}_S$ into clusters and assuming that the sampling period is short enough such that the terms in a cluster can be approximately represented by a single term, the following can be written

$$\begin{aligned} 0 &= \Psi_S^T \hat{\Theta}_S \\ &= \sum_{m=0}^{\ell} \sum_{p=0}^{m} \Sigma_{y^p u^{m-p}}\, y(k-1)^p u(k-1)^{m-p} \\ &\qquad \forall\, m,p \mid y(k-i)^p u(k-j)^{m-p} \in \cup\Omega_S, \forall\, i,j \ , \end{aligned} \tag{15}$$

where $\cup\Omega_S$ is the union of all the term clusters represented in $\Psi_S^T \hat{\Theta}_S$. A set of N_e equations can be written using equation (15) and taking the values of $y(k-1)^p u(k-1)^{m-p}$ from the data records for various values of k. Such a set of equations can be expressed in matrix form as

$$\mathbf{0} = \Phi \vec{\Sigma} \ , \tag{16}$$

where $\mathbf{0} \in \mathbb{R}^{N_e \times 1}$, $\Phi \in \mathbb{R}^{N_e \times N_{cs}}$ and $\vec{\Sigma} \in \mathbb{R}^{N_{cs} \times 1}$, where N_{cs} is the number of clusters in $\Psi_S^T \hat{\Theta}_S$ and $N_e > N_{cs}$.

It is noted that the vector $\vec{\Sigma}$ contains the cluster coefficients of $\Psi_S^T \hat{\Theta}_S$. The columns of Φ are formed by terms of the form $y(k-1)^p u(k-1)^{m-p}$ and consequently Φ has full column rank, that is $\text{rank}(\Phi) = N_{cs}$. Hence from the dimension theorem

$$\text{rank}\{\Phi\} + \dim[\ker\{\Phi\}] = N_{cs} \ , \tag{17}$$

where $\dim[\cdot]$ and $\ker\{\cdot\}$ denote the dimension and kernel, respectively. Consequently, $\dim[\ker\{\Phi\}]=0$, or in other words the null space of Φ has dimension zero and the only vector in such a subspace is the null vector. Thus equation (16) only has the trivial solution

$$\Sigma_{y^p u^{m-p}} = 0, \quad \forall\, m,p \mid y(k-i)^p u(k-j)^{m-p} \in \cup\Omega_S \ . \tag{18}$$

This completes the proof. □

It should be noted that if the terms in $\Psi_S^T \hat{\Theta}_S$ were not clustered to compose Φ, then in general such a matrix would probably be rank deficient. Consequently, many other solutions would exist besides the trivial one. Moreover, in practice the sampling will be short but probably not sufficiently short as to ensure that all the terms in a cluster can be adequately

approximated by a single term. Consequently, the coefficients of spurious clusters will not be exactly null. However, such coefficients should be sufficiently small to indicate that the respective clusters are indeed spurious. This is illustrated in the following example.

Example 2.3

For the fifteen-term model estimated in § 2.2 the cluster coefficients are

$$\begin{array}{lll} \Sigma_y = 0.99986 & \Sigma_u = 1.64564 \times 10^{-3} & \Sigma_{y^3} = -1.64349 \times 10^{-3} \\ \Sigma_{u^3} = 5.68100 \times 10^{-6} & \Sigma_{yu^2} = -7.3000 \times 10^{-9} \ . & \end{array} \tag{19}$$

This suggests that Ω_{u^3} and Ω_{yu^2} are spurious clusters. It is stressed that besides the magnitude of the cluster coefficients, other aspects also suggest which clusters are spurious. This will be illustrated in example 4.1. □

An important thing to note is that although the dynamical effect of $\Psi_S^T \hat{\Theta}_S$ on the residuals may be negligible, the inclusion of such terms in the final model can drastically influence the overall dynamics (Aguirre and Billings, 1994c). Therefore, every effort in recognising spurious clusters and subsequently deleting unnecessary terms from the model is very worthwhile

3 Results for Systems with Integral Action

The results in this section are presented for the sake of completion and also because systems with integral action are not uncommon in practice.

3.1 The linear case

Consider the linear system

$$\begin{aligned} y(k) &= \sum_{i=1}^{n} a_i\, y(k-i) + \sum_{i=1}^{m} b_i\, u(k-i) \\ A(q^{-1})y(k) &= B(q^{-1})u(k) \ , \end{aligned} \tag{20}$$

where q^{-1} is the backward time-shift operator such that $y(k)q^{-1} = y(k-1)$. Note that for a linear system, there are only two term clusters, namely Ω_y and Ω_u. Taking the Z-transform of equation (20) and expressing the polynomials in terms of the variable z yields $y(k) = B(z)u(k)/A(z)$. Then the following result holds.

Lemma 3.1 *Given the model in equation (20) with $A(z)$ monic, then $\Sigma_y = 1$ iff $A(z)$ has a zero at $z = 1$.*

Proof.

First it will be proved that if $A(z)$ has a zero at $z = 1$ then $\Sigma_y = 1$. Consider the following polynomials and the summation of their coefficients up to the $(n-1)$th power

$$\begin{aligned}
(z+\alpha_1) &\rightarrow \Sigma_\alpha^1 = [\alpha_1] \ , \\
(z+\alpha_1)(z+\alpha_2) &\rightarrow \Sigma_\alpha^2 = <\alpha_1> + [\alpha_2] + \{\alpha_1\alpha_2\} \ , \\
(z+\alpha_1)(z+\alpha_2)(z+\alpha_3) &\rightarrow \Sigma_\alpha^3 = <\alpha_1> + <\alpha_2> + <\alpha_1\alpha_2> + [\alpha_3] \ , \\
&\quad + \{\alpha_1\alpha_3\} + \{\alpha_2\alpha_3\} + \{\alpha_1\alpha_2\alpha_3\} \ , \\
\vdots \quad \vdots & \qquad \vdots \qquad \vdots
\end{aligned}$$

where the superscripts indicate the polynomial orders.

The following pattern is clear i) the terms in $< >$ are copied from the previous step, ii) the terms in [ ] are the new terms added at each step, and iii) the terms in { } are the product of the terms in $< >$ and the term in [ ] of the current step. Therefore, the algorithm for generating Σ_α^n is

$$\Sigma_\alpha^n = <\Sigma_\alpha^{n-1}> \ + [\alpha_n] \ + \ \{\Sigma_\alpha^{n-1}\alpha_n\} \ . \tag{21}$$

It is clear that $-\Sigma_\alpha^n$ is the coefficient of Ω_y for the model in equation (20) with

$$A(z) = (z+\alpha_1)(z+\alpha_2)\ldots(z+\alpha_n) \ . \tag{22}$$

Thus $\Sigma_\alpha^n = -\Sigma_y$. Equation (21) can then be rewritten as

$$-\Sigma_y = [1+\alpha_n]\,\Sigma_\alpha^{n-1} \ + \ \alpha_n \ . \tag{23}$$

If $A(z)$ has a zero at $z = 1$, then $\alpha_i = -1$ for $i \in [1, 2, \ldots, n]$. Without loss of generality it is assumed that $\alpha_n = -1$, and from equation (23) it follows that $\Sigma_y = 1$.

Finally, it will be proved that if $\Sigma_y = 1$ then $A(z)$ will have a zero at $z = 1$. Thus if $\Sigma_y = 1$ equation (23) can be written as follows

$$\alpha_n = -\frac{\Sigma_\alpha^{n-1} + 1}{\Sigma_\alpha^{n-1} + 1} \ . \tag{24}$$

Clearly, if $\Sigma_\alpha^{n-1} \neq -1$ then $\alpha_n = -1$ which means that the last factor in equation (22) corresponds to a root of $A(z)$ at $z = 1$. If $\alpha_n \neq -1$ this necessarily requires that $\Sigma_\alpha^{n-1} = -1$. Note that Σ_α^{n-1} is the coefficient of Ω_y for an $(n-1)$th-order polynomial analogous to $A(z)$. Therefore

$$\alpha_{n-1} = -\frac{\Sigma_\alpha^{n-2} + 1}{\Sigma_\alpha^{n-2} + 1} \tag{25}$$

and the proof is completed by following the same reasoning as before. □

3.2 The nonlinear case

The extension of lemma 3.1 to a NARX model is based on an nth-order Volterra model and the multi-dimensional Fourier transform of such a model. The independent variable of this model is a vector of frequencies $\omega = \omega_1, \ldots, \omega_n$ or $\mathbf{z} = z_1, \ldots, z_n$ if the model is represented using the Z-transform.

Lemma 3.2 *The total nonlinear frequency response of a NARX model has a pole at* $\mathbf{z}=1$, *that is,* $\{z_i\}_{i=1}^{n} = 1$ *iff* $\Sigma_y = 1$.

Proof.

The total nonlinear frequency response of an nth-order NARX model, $H_n^{\text{asym}}(\cdot)$, is given by (Peyton-Jones and Billings, 1989)

$$(1 - \sum_{n_1=1}^{n_y} c_{1,0}(n_1)\, e^{-j(\omega_1+\ldots+j\omega_n)n_1})\, H_n^{\text{asym}}(j\omega_1, \ldots, j\omega_n) = f(c_{i,j}(\cdot), H_{k,l}(\cdot)) \ , \tag{26}$$

where $H_{k,l}(\cdot)$ is used to denote the contribution to the kth-order frequency response function that is generated by the lth degree of nonlinearity in the output.

The Z-transform is obtained by replacing $e^{j\omega_i}$ with z_i (Billings and Peyton-Jones, 1990). It follows that the poles of $H_n^{\text{asym}}(\cdot)$ in the Z-domain are the roots of

$$1 - \sum_{n_1=1}^{n_y} c_{1,0}(n_1)\,(z_1 \ \ldots \ z_n)^{-n_1} = 0 \ . \tag{27}$$

Hence, if $H_n^{\text{asym}}(\cdot)$ has a pole at $\mathbf{z}=1$, that is $z_i = 1, \ i = 1, \ldots, n$, then

$$\sum_{n_1=1}^{n_y} c_{1,0}(n_1) = 1 \ , \tag{28}$$

The first part of the proof is completed by noticing that the left hand side of equation (28) is the coefficient of the cluster Ω_y, that is, $\sum_{n_1=1}^{n_y} c_{1,0}(n_1) = \Sigma_y$.

To see that if $\Sigma_y = 1$ then $H_n^{\text{asym}}(\cdot)$ has a pole at $\mathbf{z} = 1$ it is sufficient to show that the poles of $H_n^{\text{asym}}(\cdot)$ are determined by the linear output terms only (Peyton-Jones and Billings, 1989). The proof is then completed by following a reasoning similar to the one in the second part of the proof of lemma 3.1. □

It should be noted that lemmas 3.1 and 3.2 hold regardless of the sampling period. Furthermore, such lemmas also hold for time series models, that is AR and NAR models.

Lemma 3.2 is consistent with the observation that the poles of the nth-order transfer function are determined only by the linear output terms of the NARX model (Peyton-Jones and Billings, 1989). In other words, such poles are determined exclusively by the terms in

the cluster Ω_y. Moreover, the Duffing-Ueda system has integral action, that is, a 'pole at the origin' in the s-plane. This can be readily verified from the differential equation and by the fact that for the models estimated from data generated by such equation $\Sigma_y \approx 1$, see examples 2.2 and 2.3.

4 Numerical Results

This section provides four examples concerning the use of term clustering to enhance structure selection in identification problems. The first example illustrates how models which are composed of terms taken from effective clusters are usually better than models which include terms from spurious clusters. The second example investigates the influence of noise on the term clustering approach. In the third example it is shown that if the terms corresponding to spurious clusters are removed from the set of candidate terms, the estimated models are more robust to the deleterious effects of overparametrization. The fourth example uses Chua's circuit operating in two different regimes. Application of higher-order spectral methods to noise-free data show that the nonlinear interactions in each regime is different. This is also revealed by the cluster coefficients which are calculated form data with rather high noise levels.

In order to assess the influence of the model structure on the dynamic properties of nonlinear models, bifurcation diagrams are used. It has been argued that such diagrams are far more sensitive to changes in the model structure than other indices such as variance of residuals, prediction errors and correlation tests (Aguirre and Billings, 1994c). Details on the computation of such diagrams can be found elsewhere (Parker and Chua, 1989).

The examples below use two systems, namely the Duffing-Ueda oscillator (Ueda, 1980) and Chua's circuit (Chua and Hasler, 1993). These systems present an enormous variety of dynamical regimes and are known to be quite sensitive to small variations in initial conditions and parameters. Consequently such systems have become benchmarks in the study of nonlinear dynamics.

Example 4.1

The equation of the Duffing-Ueda oscillator was used to generate a set of 900 data points sampled at $T_s = \pi/30$, where π is the Nyquist rule. The input used was a square wave of growing amplitude superimposed on a Gaussian distributed zero-mean signal. Such an input yields improved estimated models for this system (Aguirre and Billings, 1994c).

Fixing the degree of nonlinearity and the maximum lags at the following values $\ell = 3$ and $n_y = n_u = 3$, a family of models with a varying number of terms was estimated. The terms were selected automatically based on the ERR criterion (Billings et al., 1989). The

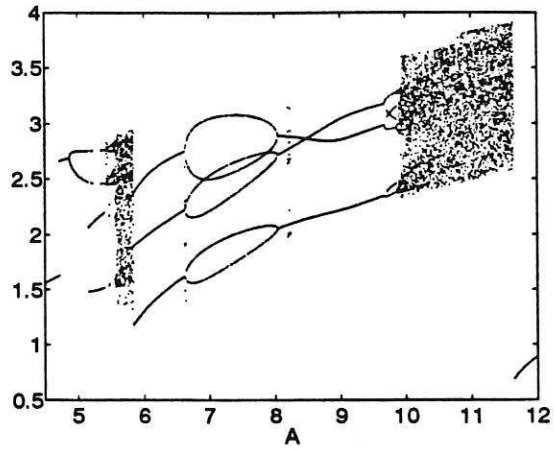

Figure 2: Bifurcation diagram of the Duffing-Ueda oscillator, equation (13).

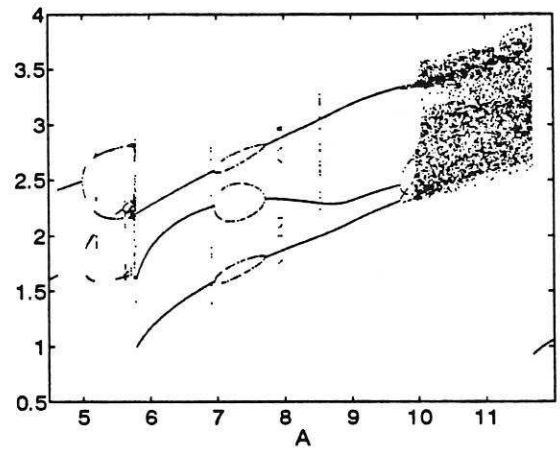

Figure 3: Bifurcation diagram of the fourteen-term estimated model for the Duffing-Ueda system, $T_s = \pi/30$. Terms selected from all possible clusters.

bifurcation diagram of the Duffing-Ueda system and the best estimated model are shown in figures 2 and 3, respectively. The estimated model has fourteen terms.

Figure 4 shows the cluster coefficients Σ_y, Σ_{y^3}, Σ_u and Σ_{u^3} for a family of models with an increasing number of terms. Three points are worth noting regarding the cluster coefficient Σ_{u^3}, i) the first term of this cluster was the tenth whereas all the other clusters were represented in models with five terms or more, ii) the value of Σ_{u^3} tends to oscillate close to zero, and iii) the magnitude of such a coefficient is far smaller than for the other clusters, in fact Σ_{u^3} reaches values as low as -6.4637$\times 10^{-8}$. Based on these observations it seems appropriate to exclude all clusters except Ω_y, Ω_{y^3} and Ω_u from the set of candidate terms.

The same set of data was therefore used to estimate a fourteen term model with the same values of ℓ, n_y and n_u as before but where only terms from the clusters Ω_y, Ω_u and Ω_{y^3} were considered as candidate terms. As before, the terms in the final model were automatically selected using a criterion based on ERR. The bifurcation diagram of the estimated model

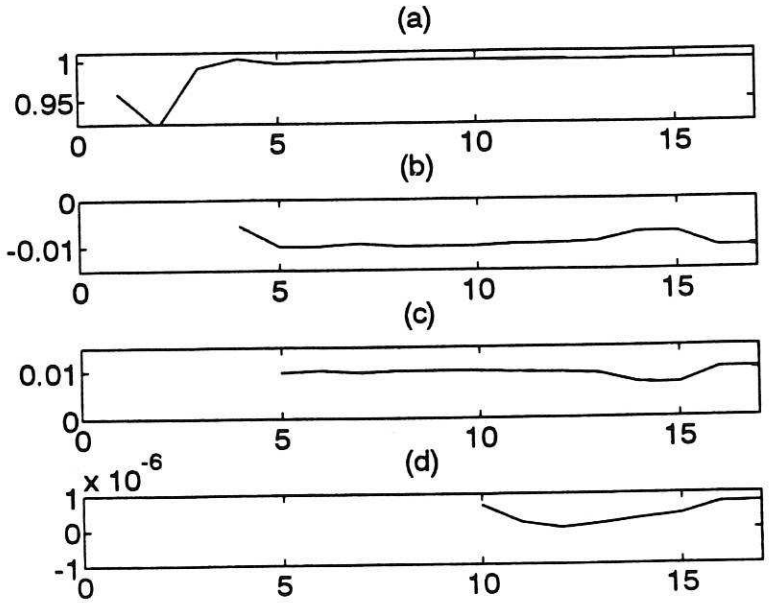

Figure 4: Cluster coefficients for the Duffing-Ueda system plotted as a function of the number of terms in the model. (a) Σ_y, (b) Σ_{y^3}, (c) Σ_u and (d) Σ_{u^3}.

is shown in figure 5. The improvement is revealed by the widening of the chaotic window at $A \approx 5.7$ and by a more accurate placement of the supercritical and subcritical pitchfork bifurcations at $A \approx 6.8$ and $A \approx 8.0$, respectively.

These results suggest that a model is more likely to reproduce faithfully the overall dynamics of the original system if the terms are chosen from effective clusters. □

Example 4.2

In this example noise was added to the output sequence used in example 4.1 such that the resulting data had a signal to noise ratio equal to 117 dB. Whilst in the noise-free case a model with $n_p = 17$ only had terms in the clusters Ω_y, Ω_u Ω_{y^3} and Ω_{yu^2}, for the noisy data, apart from such clusters terms from Ω_{yu} and Ω_{u^3} were also selected using the ERR criterion. However, the coefficients of the spurious clusters were again much smaller than the coefficients of the effective clusters. This is illustrated in Table 1 which also includes results for data with lessnoise for comparison purposes.

It is worth noticing that the coefficients of the effective clusters are not drastically changed even in the presence of moderate amounts of noise. The slight changes arise as a consequence of higher variances of the parameter estimates which is a well known consequence of the presence of noise on the data. However, the coefficients of the spurious clusters do not seem

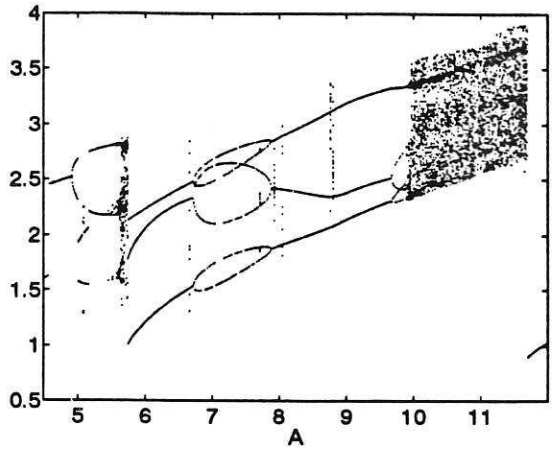

Figure 5: Bifurcation diagram of the fourteen-term estimated model for the Duffing-Ueda system, $T_s = \pi/30$. Terms selected from effective clusters only.

to follow any definite pattern as the noise in the data is increased. This is another indication that such clusters are indeed spurious.

Table 1. Cluster coefficients for a seventeen-term model of the Duffing-Ueda oscillator

Cluster	Noise-free	SNR=208 dB	SNR=117 dB
Ω_y	0.9999	0.9971	0.9994
Ω_{y^3}	-0.0103	-0.0160	-0.0196
Ω_u	0.0103	0.0165	0.0197
Ω_{yu^2}	7.5×10^{-7}	-2.4×10^{-6}	-3.2×10^{-5}
Ω_{yu}	–	–	-2.3×10^{-4}
Ω_{u^3}	–	–	-1.9×10^{-6}

□

Example 4.3

The data of example 4.1 sampled at $T_s = \pi/60$ was used in this example. A nine-term model was estimated, see equation (5), and it was shown that such a model reproduces the major dynamical invariants of the original system over a wide range of parameters (Aguirre and Billings, 1994c). When five extra terms are allowed in the estimated model, the resulting bifurcation diagram exhibits spurious dynamical regimes as shown in figure 6. In fact, the chaotic window observed for approximately $8.7 < A < 9.1$ has no counterpart in the original model and is therefore spurious. This is just one instance of the common, though sometimes

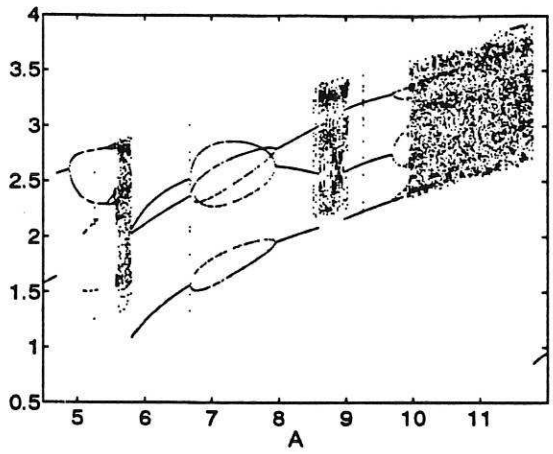

Figure 6: Bifurcation diagram of a fourteen-term estimated model for the Duffing-Ueda system. The terms were selected from every possible cluster.

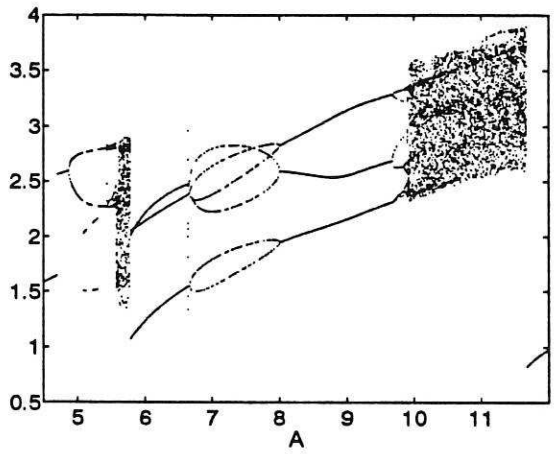

Figure 7: Bifurcation diagram of a fourteen-term estimated model for the Duffing-Ueda system. The terms were selected from effective clusters only.

overlooked, fact that overparametrizing nonlinear models usually induces ghost bifurcations and spurious dynamical regimes (Aguirre and Billings, 1994b).

The cluster coefficients of this model are $\Sigma_y = 1.0000$, $\Sigma_{y^3} = -2.7172 \times 10^{-3}$, $\Sigma_u = 2.7110 \times 10^{-3}$ and $\Sigma_{u^3} = 2.5246 \times 10^{-8}$ and show quite clearly that the terms in the cluster Ω_{u^3} can be safely omitted from the set of candidate terms. These clusters were therefore removed and another fourteen-term model was estimated from the same set of data. With this initial model specification, only terms from the clusters Ω_y, Ω_{y^3} and Ω_u were selected based on the ERR criterion. The bifurcation diagram of this model does not present any spurious bifurcations as shown in figure 7.

These results show that the estimated model is more robust to overparametrization if the 'overparametrized terms' are taken from effective clusters. □

Example 4.4

This example uses Chua's circuit (Chua and Hasler, 1993)

$$\begin{cases} \dot{x} = \alpha(y - h(x)) \\ \dot{y} = x - y + z \\ \dot{z} = -\beta y \end{cases}, \qquad h(x) = \begin{cases} m_1 x + (m_0 - m_1) & x \geq 1 \\ m_0 x & |x| \leq 1 \\ m_1 x - (m_0 - m_1) & x \leq -1 \end{cases}, \tag{29}$$

where $m_0 = -1/7$ and $m_1 = 2/7$. Varying the parameters α and β the circuit displays several regular and chaotic regimes. The well known double scroll attractor, for instance, is obtained for $\alpha = 9$. Another attractor produced by this circuit is the spiral Chua's attractor observed for $\alpha = 8.5$ and $\beta = 100/7$.

It has been observed that in the noise-free case terms from Ω_y and Ω_{y^3} are selected using ERR to give a sixteen-term model of the double scroll attractor. Such a model reproduces the topological geometry, the largest Lyapunov exponent and the correlation dimension of the original attractor. However, when the data are on the spiral Chua's attractor terms from Ω_y, Ω_{y^2} and Ω_{y^3} are selected to form the final model (Aguirre and Billings, 1994a).

This can be explained by noticing that when the circuit is evolving on the double scroll, the x-component varies over the entire domain of the piecewise linear function, that is, x visits the three regions of $h(x)$. This would seem to lend support for approximating such a function by a smooth polynomial with a cubic nonlinearity (Khibnik et al., 1993). But when the trajectories of the system evolve on the spiral attractor, only two segments of the piecewise linear function are visited. It has then been conjectured that for such an attractor, the nonlinearity could be approximated by a quadratic polynomial (Elgar and Kennedy, 1993). These are useful theoretical observations but are of very little practical value in structure selection because they presuppose *a priori* knowledge of the system.

An explanation which does not assume any prior knowledge can be obtained by using higher-order spectral analysis. In particular, the bispectrum and trispectrum have been calculated for data on both the double scroll and the spiral strange attractors in Chua's circuit. In the former case the bispectrum did not show significant quadratic interactions whereas the trispectrum revealed strong cubic interactions between the primary peak frequency and the harmonics. In the latter case, the bispectrum and trispectrum revealed both quadratic and cubic interaction for the spiral Chua's attractor (Elgar and Chandran, 1993; Elgar and Kennedy, 1993).

In the presence of noise, however, terms from Ω_y, Ω_{y^2} and Ω_{y^3} are automatically selected using the ERR criterion in both cases, that is, from data on the double scroll and from data on the spiral attractor. To illustrate how term clustering can help to recognise the effective clusters in each case, two noisy time series, one for each attractor, with signal to noise ratios equal to 29 dB were considered. The cluster coefficients for each case are listed in Table 2

Table 2. Cluster coefficients for Chua's circuit

Attractor	Σ_y	Σ_{y^2}	Σ_{y^3}	Constant
double scroll	1.1560	-3.7029 $\times 10^{-4}$	-0.0536	–
Chua's spiral	0.9518	0.3332	0.0572	-0.6181

Table 2 clearly reveals that although terms from Ω_{y^2} had been selected for noisy data in the double scroll case, such a cluster can be confidently discarded from the set of candidate terms. For the spiral Chua's attractor, however, the value of Σ_{y^2} leaves little doubt that terms from Ω_{y^2} should be considered as candidates.

These results are in total accord with those discussed at the beginning of this example but with the difference that the results in Table 2 were obtained from noisy data which correspond to a more realistic scenario and shows how term clustering enhances the performance of higher-order spectra and the ERR criterion. □

4.1 Discussion

Although the advantages of the term clustering approach to model selection can also be assessed by verifying the variance of the residuals and the estimated parameters, bifurcation diagrams have been used because they are more sensitive to changes in the model structure.

Notice that a cluster of terms may be irrelevant to the residuals in the sense that the presence or absence of such terms in the model may not alter the 'whiteness' of the residuals. This fact was used in lemma 2.1. However, terms which are irrelevant to the residuals, which are calculated based on the one-step-ahead predictions, might induce spurious bifurcations which are characteristic of the steady state predicted output. Secondly, in spite of the estimated model being more robust to overparametrization effects if the terms are only selected from effective clusters, overparametrizing a nonlinear model should always be avoided.

The fact that *slightly* overparametrized models in which the excess terms are members of effective clusters still reproduce the original bifurcation diagrams without introducing spurious dynamical regimes should come as no surprise. This is because for reasonably short sampling periods the terms within a cluster are quite similar. Including a few more terms will have the effect of 'sharing' information between terms during parameter estimation. However, if the excess terms are taken from spurious clusters, the type of nonlinearity that such terms are apt to model is *not* present in the data and consequently such terms will induce spurious dynamics if included in the model structure.

The concept of *zeroing-and-refitting* has been suggested as one of the most powerful methods for fine tuning a model structure (Kadtke et al., 1993). Briefly, in such an approach the terms in a model which have coefficients smaller than a certain threshold are eliminated

from the original structure. Thus parameters are re-estimated for the truncated model. This can be carried out as many times as necessary.

It is pointed out that in examples 4.2 and 4.4 the use of the zeroing-and-refitting approach (Kadtke et al., 1993) would not have given a clear indication of which terms were unnecessary. In particular, for the data on the double scroll a few terms in the spurious cluster Ω_{y^2} have coefficients of the same order of magnitude as some terms in the effective clusters. However, as shown in Table 2, the cluster coefficient Σ_{y^2} is in fact much smaller.

The zeroing-and-refitting approach is indeed quite effective for low noise levels. High noise levels, however, tend to increase the value of the term coefficients and it is no longer obvious how to choose the threshold which separates the effective and spurious terms. While the zeroing-and-refitting procedure makes the unstated claim that spurious terms will have relatively small coefficients, it has been shown in lemma 2.1 that it is the composite effect of the terms in a cluster which determines if such a cluster is spurious or not. This can easily be seen by considering that two spurious but similar terms may have relatively large coefficients with opposite signs thus indicating cancellation. Thus term clustering allows for effective terms with relatively small coefficients and spurious terms with relatively large coefficients. These are not uncommon situations in practice.

Examples 4.1–4.4 illustrate that the term clustering approach usually yields models with improved dynamics. Another major advantage is that computation time is drastically reduced. This reduction is achieved because using the concepts introduced in this paper, it is often possible to discriminate between effective and spurious clusters and therefore greatly reduce the initial set of candidate terms for the subsequent estimation of models from a particular piece of data. For the examples with the Duffing-Ueda system, the set of candidate terms considering all possible clusters was composed of 64 terms which were considered in the first run. After the first model was estimated, the effective clusters were deleted and a new set of candidate terms was composed for all subsequent experiments on that particular set of data. Thus the new set of candidate terms was composed of 16 terms since only effective clusters were considered. This corresponds to a reduction of 75% in the original size of the set of candidates. It is noted that even greater reductions can be achieved for systems of higher order.

The results in this section and several simulations using other systems suggest that the coefficients of spurious clusters are indeed much smaller than those corresponding to effective clusters. A rather intuitive proof of this has been provided in lemma 2.1. A natural question is: how much is 'much smaller'? Although a rule of thumb could be suggested such as fixing a threshold to be 10^4 times smaller than the largest cluster coefficient, this does not seem to be the best procedure. If the coefficient of a spurious cluster is not convincingly smaller than the other coefficients, it is often useful to plot the cluster coefficients as a function of the number of terms in the model such as in figure 4. Spurious clusters are usually revealed

by insignificant and/or oscillating coefficients. Moreover, if noise is added to the data, or if a different window of data is used, the coefficients of effective clusters will *not* usually change drastically. This will not be the case for spurious clusters. Because of these properties of the clustering approach to coarse structure selection, a threshold is not absolutely necessary and is hardly critical.

The term clustering approach introduced in the present study can easily be extended to MIMO models. In the multivariable case, clusters are defined in a similar way for each sub-system. Thus, for instance, the following terms are *not* members of the same cluster: $y_1(k-i)$ and $y_2(k-j)$ $\forall\, i, j$, where such terms correspond to the first and second sub-systems, respectively.

5 Conclusions

The concepts of term clustering and cluster coefficients have been introduced. The rationale behind such concepts is that information is 'shared' among terms within individual clusters during parameter estimation. Consequently if a cluster is spurious this should be revealed by a relatively low value of the respective cluster coefficient. It has been pointed out how such concepts can be used to provide important *coarse* information during structure selection for nonlinear models. Thus term clustering used in conjunction with *detailed* structure selection criteria form a consistent basis for selecting the most important terms in nonlinear modelling.

Some of the main properties of the new procedure are i) if a model is composed of terms selected from effective clusters, such a model is more likely to reproduce faithfully the dynamics of the original system, ii) the estimated models are more robust to overparametrization provided the excess terms are not taken from spurious clusters, and iii) the procedure is robust in the presence of noise.

ACKNOWLEDGMENTS

We are grateful to Eduardo Mendes for his help in the simulations and for many useful discussions. LAA gratefully acknowledges financial support from CNPq (Brazil) under grant 200597/90-6. SAB gratefully acknowledges that part of this work was funded by SERC under contract GR/H35286.

References

Aguirre, L. A. and Billings, S. A. (1994a). Discrete reconstruction of stange attractors in Chua's circuit. *Int. J. Bifurcation and Chaos*, (in press).

Aguirre, L. A. and Billings, S. A. (1994b). Dynamical effects of overparametrization in nonlinear models. (Submitted for publication).

Aguirre, L. A. and Billings, S. A. (1994c). Validating identified nonlinear models with chaotic dynamics. *Int. J. Bifurcation and Chaos*, (in press).

Billings, S. A. and Chen, S. (1989). Extended model set, global data and threshold model identification of severely nonlinear systems. *Int. J. Control*, 50(5):1897–1923.

Billings, S. A., Chen, S., and Korenberg, M. J. (1989). Identification of MIMO nonlinear systems using a forward-regression orthogonal estimator. *Int. J. Control*, 49(6):2157–2189.

Billings, S. A., Korenberg, M. J., and Chen, S. (1988). Identification of nonlinear output affine systems using an orthogonal least squares algorithm. *Int. J. Systems Sci.*, 19(8):1559–1568.

Billings, S. A. and Leontaritis, I. J. (1981). Identification of nonlinear systems using parametric estimation techniques. In *IEE Conf. Control and its applications*, pages 183–187, Warwick.

Billings, S. A. and Peyton-Jones, J. C. (1990). Mapping non-linear integro-differential equations into the frequency domain. *Int. J. Control*, 52(4):863–879.

Billings, S. A. and Tao, Q. H. (1991). Model validation tests for nonlinear signal processing applications. *Int. J. Control*, 54:157–194.

Billings, S. A. and Voon, W. S. F. (1986). Correlation based model validity tests for nonlinear models. *Int. J. Control*, 44(1):235–244.

Chua, L. O. and Hasler, M. (1993). (Guest Editors). Speciall issue on Chaos in nonlinear electronic circuits. *IEEE Trans. Circuits Syst.*, 40(10–11).

Elgar, S. and Chandran, V. (1993). Higher-order spectral analysis to detect nonlinear interactions in measured time series and an application to Chua's circuit. *Int. J. Bifurcation and Chaos*, 3(1):19–34.

Elgar, S. and Kennedy, M. P. (1993). Bispectral analysis of Chua's circuit. *J. Circuits, Systems and Computers*, 3(1):33–48.

Haber, H. and Unbehauen, H. (1990). Structure identification of nonlinear dynamic systems — A survey on input/output approaches. *Automatica*, 26(4):651–677.

Kadtke, J. B., Brush, J., and Holzfuss, J. (1993). Global dynamical equations and Lyapunov exponents from noisy chaotic time series. *Int. J. Bifurcation and Chaos*, 3(3):607–616.

Khibnik, A. I., Roose, D., and Chua, L. O. (1993). On periodic orbits and homoclinic bifurcations in Chua's circuit with a smooth nonlinearity. *Int. J. Bifurcation and Chaos*, 3(2):363–384.

Leontaritis, I. J. and Billings, S. A. (1985a). Input-output parametric models for nonlinear systems part I: deterministic nonlinear systems. *Int. J. Control*, 41(2):303–328.

Leontaritis, I. J. and Billings, S. A. (1985b). Input-output parametric models for nonlinear systems part II: stochastic nonlinear systems. *Int. J. Control*, 41(2):329–344.

Parker, T. S. and Chua, L. O. (1989). *Practical numerical algorithms for chaotic systems.* Springer Verlag, Berlin.

Peyton-Jones, J. C. and Billings, S. A. (1989). Recursive algorithm for computing the frequency response of a class of non-linear difference equation models. *Int. J. Control*, 50(5):1925–1940.

Ueda, Y. (1980). Steady motions exhibited by Duffing's equation: A picture book of regular and chaotic motions. In Holmes, P. J., editor, *New approaches to nonlinear problems in dynamics*, pages 311–322. SIAM.