A Modified Orthogonal Forward Regression Least-Squares Algorithm for System Modelling From Noisy Regressors

Guo, L. Z. and Billings, S. A.

Department of Automatic Control and Systems Engineering
University of Sheffield
Sheffield, S1 3JD
UK

Research Report No. 911
November 2005

A modified orthogonal forward regression least-squares algorithm for system modelling from noisy regressors

Guo, L. Z. and Billings, S. A.

Department of Automatic Control and Systems Engineering
University of Sheffield
Sheffield S1 3JD, UK

Abstract

In this paper, a modified Orthogonal Forward Regression (OFR) least-squares algorithm is presented for system identification and modelling from noisy regressors. Under the assumption that the energy and signal-to-noise ratio (SNR) of the signals are known or can be estimated, it is shown that unbiased estimates of the Error Reduction Ratios (ERRs) and the parameters can be obtained in each forward regression step. Examples are provided to illustrate the proposed approach.

1 Introduction

In system identification and modelling, the Orthogonal Forward Regression (OFR) least-squares algorithm (Billings, Chen, and Korenberg 1989, Chen, Billings, and Luo 1989, and Billings, Korenberg, and Chen 1988) has proved to be an effective algorithm for determining significant model terms or the model structure and the associated parameter estimates. The OFR algorithm involves a stepwise orthogonalisation of the regressors and a forward selection of the relevant terms based on the Error Reduction Ratio (ERR) criterion (Billings, Chen, and Kronenberg 1989). In recent years, many variants of the OFR algorithm have been introduced to improve the performance of the OFR algorithm including D-optimality OFR (Hong and Harris 2001), variable pre-selection OFR (Wei, Billing, and Liu 2004), piecewise linearization (Mao and Billings 1999), minimal model structure detection (Mao and Billings 1997) etc. For the past two decades, the OFR algorithm and its variants have been successfully applied in a variety of fields in system identification and modelling (Aguirre and Billings 1995a,b; Billings, Chen, and Backhouse 1989; Billings, Fadzil, Sulley, and Johnson 1988; Coca, Zheng, Mayhew, and Billings 2000; Coca and Billings 2001; Liu, Kung, and Chao 2001; Balikhin, Zhu, and Billings 2005).

Although there have been many variations and refinements of the original OFR algorithm all these are based on the assumption that any noise only affects the measured output values. This is the classical system identification formulation where it is assumed that the input, which is often designed by the user to be persistently exciting (Leontaritis and Billings 1987), is assumed to be measured perfectly or as a noise free signal. If the outputs are corrupted by noise, which can be nonlinear and coloured, the OFR algorithm can still be applied and often noise models are fitted to ensure unbiased estimates (Billings, Chen, and Kronenberg 1989). However, there are situations where all the potential regressors, which can be made up of both input and output terms, are corrupted by noise. This arises for example in the identification of coupled map lattices (Coca and Billings 2001, Guo and Billings 2004) where the (noisy) outputs of some neighbouring nodes are considered as inputs for the identification of a model at other node locations. If the classical OFR is applied to this class of problems the ordering and the value of the ERRs can be incorrect. This means that the structure of the model cannot be determined and that the estimates of the model parameters will probably be biased. There appear to be no results in the literature relating to the applications of the OFR algorithm to this case, where there is noise on the regressors. This oversight is addressed in the present study where the effects of noise on the model regression terms are studied in detail. It is shown that if the classical OFR algorithm is applied in this case both the ERR values and the parameter estimates will be biased. The analysis of why this occurs leads to the derivation of a new modified OFR routine which overcomes these limitations. The new algorithm however requires a knowledge of the power and the SNR of the signals involved. These requirements are analysed in detail and it is shown that the ordering or ranking of the model terms should still be possible. The sensitivity of the new algorithm to the values of the power and SNR of the signals is investigated and it is shown that the new method can still work reasonably well even when the estimates of these values are not perfectly accurate.

The paper is organised as follows. Section 2 presents a brief introduction to the principle of the classical OFR algorithm. Section 3 discusses the OFR algorithm and its applications to the case where the noise is on the output only. A detailed analysis of the effects of noise on ERR is given in section 4. The modified OFR algorithm for detecting the correct terms and determining the associated parameter estimates is presented in section 5. Section 6 provides a sensitivity analysis of the proposed modified OFR algorithm to the energy. Section 7 illustrates the proposed approach using numerical simulations, and finally conclusions are given in section 8.

2 The classical OFR least-squares algorithm

Let $p_0, p_1, \cdots, p_n$ be independent variables and y the output response of a system. Assume that there is a subset I of $\{0, 1, \cdots, n\}$ such that a linear relationship

$$y = \sum_{i \in I} \theta_i p_i \tag{1}$$

exists. Given a set of observations, the system modelling problem of interest is to determine the subset I and the values of θ_i. The OFR algorithm for this problem involves three steps

- Orthogonalise the regressors to remove the correlations between these variables;
- Select significant terms using the ERR as a criterion;
- Estimate the corresponding parameters for the selected terms.

Formally, the classical OFR least-squares algorithm can be stated as follows (Billings, Korenberg, and Chen 1988).

Let $p_0(t), p_1(t), \cdots, p_n(t)$ and $y(t)$, $t = 1, 2, \cdots, N$ be the series of observations. Denote $Y = (y(1), y(2), \cdots, y(N))^T$ and $P_i = (p_i(1), p_i(2), \cdots, p_i(N))^T$, $i = 0, 1, \cdots, n$, then the following linear regression model can be formed

$$Y = P\theta + \Xi \tag{2}$$

where $P = (P_0, P_1, \cdots, P_N)$ is the regression matrix, $\theta = (\theta_1, \theta_2, \cdots, \theta_n)^T$ represents the unknown parameters to be estimated, and $\Xi = (\xi(1), \xi(2), \cdots, \xi(N))^T$ is some modelling error vector. The three steps in the OFR algorithm are

1) ORTHOGONALISATION The orthogonal decomposition $P = WA$, where A is an $(n+1) \times (n+1)$ upper triangular matrix with unity diagonal elements, of the regression matrix P provides an alternative representation of eqn. (2)

$$Y = P\theta + \Xi = WA\theta + \Xi = Wg + \Xi \tag{3}$$

where W is an $N \times (n+1)$ matrix with orthogonal columns W_i such that $W^T W = D$ in which D is an $(n+1) \times (n+1)$ diagonal matrix with elements $d_i = < W_i, W_i >$, $i = 0, 1, \cdots, n$. Note that $< \cdot, \cdot >$ denotes the inner product so that $d_i = < W_i, W_i > = \sum_{t=1}^{N} w_i(t) w_i(t)$, $i = 0, 1, \cdots, n$.

2) TERM SELECTION The orthogonal least squares solution to g is given by

$$\hat{g}_i = \frac{< Y, W_i >}{< W_i, W_i >} = \frac{W_i^T Y}{W_i^T W_i}, i = 0, 1, \cdots, n \tag{4}$$

The fraction of variance not explained by a regression of Y on Wg is

$$\frac{< E, E >}{< Y, Y >} = \frac{< Y - Wg, Y - Wg >}{< Y, Y >} = \frac{< Y, Y > - < Wg, Wg >}{< Y, Y >} \tag{5}$$

Thus the error reduction ratio (ERR) caused by term i, $i = 0, 1, \cdots, n$ is defined as

$$ERR_i = \frac{< W_i g_i, W_i g_i >}{< Y, Y >} \tag{6}$$

The OFR least-squares algorithm selects the subset I, that is a subset of regressors in a forward-regression manner by maximising the contribution of a regressor to the explained desired response variance, that is its ERR.

3) PARAMETER ESTIMATION Once the parameters g_i, $i \in I$ have been estimated using (4) the parameters θ_i, $i \in I$ in the regression equation (1)can be calculated as

$$\hat{\theta} = A^{-1}\hat{g} \tag{7}$$

From the definition of ERR (6), it can be observed that the OFR is equivalent to maximising the product moment correlation coefficient. In fact, the product moment correlation coefficient ρ_i of term i satisfies

$$\rho_i^2 = \frac{< Y, W_i >^2}{< Y, Y >< W_i, W_i >} = \frac{\frac{< Y, W_i >^2}{< W_i, W_i >^2} < W_i, W_i >}{< Y, Y >} = \frac{< W_i g_i, W_i g_i >}{< Y, Y >} = ERR_i \tag{8}$$

3 The OFR algorithm with noise on the output measurements

There are two distinct cases when applying the OFR algorithm to system identification problems. The first is the classical case where only the output measurements are contaminated by noise, and the second is the case where both input and output, that is (some of) the regressors, are corrupted by noise. In this section, the applications of the OFR algorithm to the first case are presented. For brevity, all discussions will be carried out on single-input single-output (SISO) systems throughout the paper, however, the results can be readily extended to the MIMO case.

According to Leontaritis and Billings (1985), under some mild assumptions a discrete-time SISO nonlinear stochastic dynamic system where $y(t)$ is a random variable representing the system output and $u(t)$ is a deterministic variable representing the system input, can be described by the NARMAX model

$$y(t) = f(y(t-1), \cdots, y(t-n_y); u(t-1), \cdots, u(t-n_u); e(t-1), \cdots, e(t-n_e)) + e(t) \tag{9}$$

where $e(t) = y(t) - E[y(t)|y(t-1), y(t-2), \cdots, y(1); u(t), u(t-1), \cdots, u(1)]$ is the innovation sequence, $E[\cdot]$ denotes the expectation; n_y, n_u, and n_e are the maximum lags in the output, input, and innovation variable. $f(\cdot)$ is some nonlinear function representing the input-output relationship of the underlying stochastic dynamic system. This represents case one noted above where the possible correlated noise, $e(t)$ is accommodated within the model by the lagged noise terms $e(t-1), \cdots, e(t-n_e)$. This is the classical system identification formulation where the input $u(t)$, which is often generated as a persistently exciting test input, is assumed to be noise free.

The objective of system identification is to find a suitable model to approximate the underlying relationship $f(\cdot)$ using a set of input and output observations. For the purpose of implementing an identification algorithm, the system (9) can generally be parameterised using some basis functions, such as radial basis functions, wavelets, and polynomials, into a linear-in-the-parameters model structure. The generic form of a linear-in-the-parameters model is as follows

$$y(t) = \sum_{i=1}^{n} \theta_i p_i(t) + \xi(t), t = 1, 2, \cdots, N \tag{10}$$

where N is the data length, $p_i(\cdot)$ are model terms which are formed by combining some of the lagged values of the input, output, and innovation variables, n is the number of all the distinct terms, $\xi(t)$ is the modelling error, and θ_i are the parameters to be estimated. A regression matrix form corresponding to (10) is

$$Y = P\theta + \Xi \tag{11}$$

where $Y = (y(1), y(2), \cdots, y(N))^T$, $P = (P_0, P_1, \cdots, P_N)$ with $P_i = (p_i(1), p_i(2), \cdots, p_i(N))^T$, $i = 0, 1, \cdots, n$, $\theta = (\theta_1, \theta_2, \cdots, \theta_n)^T$, and $\Xi = (\xi(1), \xi(2), \cdots, \xi(N))^T$. Note that the form of eqn. (11) is exactly the same as eqn. (2) so that it seems that the OFR algorithm can be applied directly in the exact manner as described in Section 2. Unfortunately, the innovation or prediction error $e(t)$ can not be measured or calculated *a priori*. Then the following question naturally arises, 'what will happen if the OFR algorithm is applied to eqn. (11) directly?'

To answer the question, consider an orthogonal decomposition of (11) as (3)

$$Y = P\theta + \Xi = WA\theta + \Xi = Wg + \Xi \tag{12}$$

where W is an $N \times (n+1)$ matrix with orthogonal columns W_i such that $W^T W = D$ in which D is an $(n+1) \times (n+1)$ diagonal matrix with element $d_i =< W_i, W_i >$, $i = 0, 1, \cdots, n$. The orthogonal matrix W can be divided into $W = (W_p | W_n)$ where the submatrix W_p only contains the process data such as y and u, and W_n contains the data involving the innovations e, following which eqn. (12) can be rewritten as

$$Y = P\theta + \Xi = W_p g_p + W_n g_n + \Xi \tag{13}$$

Next, the orthogonal least squares solution to g is given by (4) and the fraction of variance not explained by a regression of Y on Wg is (5). Substituting $W = (W_p, W_n)$ into (5) yields

$$\frac{<E,E>}{<Y,Y>} = \frac{<Y,Y> - <W_p g_p + W_n g_n, W_p g_p + W_n g_n>}{<Y,Y>} \tag{14}$$
$$= \frac{<Y,Y> - <W_p g_p, W_p g_p> -2 <W_p g_p W_n g_n> - <W_n g_n, W_n g_n>}{<Y,Y>}$$

considering the orthogonality of W, (14) becomes

$$\frac{<E,E>}{<Y,Y>} = \frac{<Y,Y> - <W_p g_p, W_p g_p> - <W_n g_n, W_n g_n>}{<Y,Y>} \tag{15}$$

Therefore, the error reduction ratio (ERR) caused by term i, $i = 0, 1, \cdots, n$ is still

$$ERR_i = \frac{<W_i g_i, W_i g_i>}{<Y,Y>} \tag{16}$$

irrespective of whether the term is in W_p or in W_n. This implies that the ERR values for the process terms can be calculated independently of the noise terms and the process and noise terms can be determined separately. It follows that the OFR algorithm is able to detect the correct process terms without any knowledge of the prediction error $e(t)$. To detect the noise terms and provide a noise model, the innovation sequence $e(t)$ must be estimated. In order to solve this problem, the OFR algorithm is used in an iterative manner. Initially, the OFR algorithm is employed to determine the significant process terms, then an estimate to the prediction error sequence $\{e(t)\}$ can be generated using these initially determined terms. The whole procedure involves applying the OFR algorithm iteratively with the input and output data together with the estimated $\{e(t)\}$ at each iterative step until convergence. For more details see Billings, Chen, and Kronenberg (1989).

4 The OFR algorithm with noise on the regressors

In this section, the effects of noise on the regressors upon the OFR algorithm is analysed. The problem can be formalised as follows.

Let $p_0, p_1, \cdots, p_n$ be independent variables and y the output response. Assume that there is a subset I of $\{0, 1, \cdots, n\}$ such that a linear relationship

$$y = \sum_{i \in I} \theta_i p_i \tag{17}$$

exists. Let $\bar{p}_0(t), \bar{p}_1(t), \cdots, \bar{p}_n(t)$ and $\bar{y}(t)$, $t = 1, 2, \cdots, N$ be the series of noisy observations, where $\bar{p}_i(t) = p_i(t) + \epsilon_i(t)$, $i = 0, 1, \cdots, n$, and $\bar{y}(t) = y(t) + \varepsilon(t)$. Note that this is distinctly different from the case considered in section 3 because here the noise is assumed to be on all the potential regressors in eqn. (17) including both input and output terms. The objective is to apply the OFR algorithm to estimate the relationship (17) from these noisy observations. Following the definition of Y and P in (2), gives

$$\bar{Y} = \begin{pmatrix} y(1) + \varepsilon(1) \\ y(2) + \varepsilon(2) \\ \vdots \\ y(N) + \varepsilon(N) \end{pmatrix} = Y + \varepsilon \tag{18}$$

$$\bar{P}_i = \begin{pmatrix} p_i(1) + \epsilon_i(1) \\ p_i(2) + \epsilon_i(2) \\ \vdots \\ p_i(N) + \epsilon_i(N) \end{pmatrix} = P_i + \epsilon_i \tag{19}$$

In order to analyse the effects of noise on the regressors in the OFR algorithm, the following assumptions are made

- $\epsilon_i(t)$, $i = 0, 1, \cdots, n$, is an independent noise sequence with zero-mean and finite variance $\sigma^2_{\epsilon_i}$.
- ε is an independent noise sequence with zero-mean and finite variance σ^2_{ε}.
- For any $i \neq j$, ϵ_i and ϵ_j are mutually independent.
- For any i, ϵ_i and ε are mutually independent.
- Both ϵ_i, $i = 0, 1, \cdots, n$, and ε are independent of any of the true signals (p_j, $j = 0, 1, \cdots, n$, and y).

Applying the OFR algorithm to the noisy data $\bar{Y}$ and $\bar{P}_i, i = 1, \cdots, n$, yields

$$\bar{W}_i = \bar{P}_i - \sum_{j=0}^{i-1} \bar{\alpha}_{j,i} \bar{W}_j \tag{20}$$

where

$$\bar{\alpha}_{j,i} = \frac{< \bar{P}_i, \bar{W}_j >}{< \bar{W}_j, \bar{W}_j >}, j = 0, 1, \cdots, i-1 \tag{21}$$

and $\bar{g}_i =< \bar{Y}, \bar{W}_i > / < \bar{W}_i, \bar{W}_i >$.

Note that the results obtained by applying the OFR algorithm to the noise-free data are

$$W_i = P_i - \sum_{j=0}^{i-1} \alpha_{j,i} W_j, i = 1, 2, \cdots, n \tag{22}$$

where

$$\alpha_{j,i} = \frac{< P_i, W_j >}{< W_j, W_j >}, j = 0, 1, \cdots, i-1 \tag{23}$$

and $g_i =< Y, W_i > / < W_i, W_i >$. Define $\Delta W_i = \bar{W}_i - W_i$, then

$$\begin{aligned} \Delta W_i &= (\bar{P}_i - P_i) - (\sum_{j=0}^{i-1} \bar{\alpha}_{j,i} \bar{W}_j - \sum_{j=0}^{i-1} \alpha_{j,i} W_j) \\ &= \epsilon_i - (\sum_{j=0}^{i-1} \bar{\alpha}_{j,i} \bar{W}_j - \sum_{j=0}^{i-1} \alpha_{j,i} W_j) \end{aligned} \tag{24}$$

From the definition of ERR

$$\overline{ERR}_i = \bar{\rho}_i^2 = \frac{< \bar{Y}, \bar{W}_i >^2}{< \bar{Y}, \bar{Y} >< \bar{W}_i, \bar{W}_i >} \tag{25}$$

From the assumptions on the noise, it follows that

$$\begin{aligned} \bar{\rho}_i &= \frac{< \bar{Y}, \bar{W}_i >}{\sqrt{< \bar{Y}, \bar{Y} >}\sqrt{< \bar{W}_i, \bar{W}_i >}} \\ &= \frac{< Y + \varepsilon, W_i + \Delta W_i >}{\sqrt{< \bar{Y}, \bar{Y} >}\sqrt{< \bar{W}_i, \bar{W}_i >}} \\ &= \frac{< Y, W_i > + < Y, \Delta W_i >}{\sqrt{< \bar{Y}, \bar{Y} >}\sqrt{< \bar{W}_i, \bar{W}_i >}} \end{aligned} \tag{26}$$

The first term on the right hand side of (26) is

$$\begin{aligned} \frac{< Y, W_i >}{\sqrt{< \bar{Y}, \bar{Y} >}\sqrt{< \bar{W}_i, \bar{W}_i >}} &= \frac{< Y, W_i >}{\sqrt{< Y, Y >}\sqrt{< W_i, W_i >}} \frac{\sqrt{< Y, Y >}\sqrt{< W_i, W_i >}}{\sqrt{< \bar{Y}, \bar{Y} >}\sqrt{< \bar{W}_i, \bar{W}_i >}} \\ &= \frac{\eta_Y}{\sqrt{\eta_Y^2 + 1}} \sqrt{\frac{< W_i, W_i >}{< \bar{W}_i, \bar{W}_i >}} \rho_i \end{aligned} \tag{27}$$

where $\eta_Y^2 = < Y, Y > /\sigma_\varepsilon^2$ can be considered as the signal-to-noise ratio of Y.

The second term on the right hand side of (26) is

$$\begin{aligned} \frac{< Y, \Delta W_i >}{\sqrt{< \bar{Y}, \bar{Y} >}\sqrt{< \bar{W}_i, \bar{W}_i >}} &= \frac{1}{\sqrt{< \bar{Y}, \bar{Y} >}\sqrt{< \bar{W}_i, \bar{W}_i >}} < Y, \epsilon_i - (\sum_{j=0}^{i-1} \bar{\alpha}_{j,i}\bar{W}_j - \sum_{j=0}^{i-1} \alpha_{j,i} W_j) > \qquad (28) \\ &= \frac{1}{\sqrt{< \bar{Y}, \bar{Y} >}\sqrt{< \bar{W}_i, \bar{W}_i >}} (\sum_{j=0}^{i-1} \frac{< P_i, W_j >}{< W_j, W_j >} < Y, W_j > - \sum_{j=0}^{i-1} \frac{< \bar{P}_i, \bar{W}_j >}{\bar{W}_j, \bar{W}_j} < Y, \bar{W}_j >) \\ &= \sum_{j=0}^{i-1} \frac{< P_i, W_j >}{< W_j, W_j >} \frac{\eta_Y}{\sqrt{\eta_Y^2 + 1}} \sqrt{\frac{< W_j, W_j >}{< \bar{W}_i, \bar{W}_i >}} \rho_j - \sum_{j=0}^{i-1} \frac{< \bar{P}_i, \bar{W}_j >}{< \bar{W}_j, \bar{W}_j >} \sqrt{\frac{< \bar{W}_j, \bar{W}_j >}{< \bar{W}_i, \bar{W}_i >}} \bar{\rho}_j \end{aligned}$$

Substituting (27) and (28) into (26) yields

$$\bar{\rho}_i = \frac{\eta_Y}{\sqrt{\eta_Y^2 + 1}} \sqrt{\frac{< W_i, W_i >}{< \bar{W}_i, \bar{W}_i >}} \rho_i + \sum_{j=0}^{i-1} \frac{< P_i, W_j >}{< W_j, W_j >} \frac{\eta_Y}{\sqrt{\eta_Y^2 + 1}} \sqrt{\frac{< W_j, W_j >}{< \bar{W}_i, \bar{W}_i >}} \rho_j - \sum_{j=0}^{i-1} \frac{< \bar{P}_i, \bar{W}_j >}{\bar{W}_j, \bar{W}_j} \sqrt{\frac{< \bar{W}_j, \bar{W}_j >}{< \bar{W}_i, \bar{W}_i >}} \bar{\rho}_j \qquad (29)$$

From eqn. (29), it can be observed that $\bar{\rho}_i$ is generally not equal to ρ_i because of the presence of the noise on the regressors. This difference may produce critical effects on the results when applying the OFR algorithm with noise on the regressors because at each step the OFR algorithm may not be able to pick the correct term (the term with maximal ρ^2) and consequently may not produce the correct parameter estimates. In the following sections, a new algorithm will be derived to overcome these problems.

5 A modified OFR least-squares algorithm for the case where there is noise on the regressors

A new OFR algorithm for the case where there is noise on the regressors is derived below.

5.1 Correction for the biased ERR values

In this section, a compensation approach is proposed to correct the bias in the ERR values when there is noise on the regressors. From the definition (6) of ERR, it follows that

$$ERR_i = \frac{< W_i g_i, W_i g_i >}{< Y, Y >} \qquad (30)$$

$$= \frac{<Y, W_i>^2}{<Y, Y><W_i, W_i>}$$

where for $i = 1, 2, \cdots, n$

$$\begin{aligned}
<Y, W_i> &= <\bar{Y} - \varepsilon, \bar{W}_i - \Delta W_i> \qquad (31) \\
&= <\bar{Y}, \bar{W}_i> - <\bar{Y}, \Delta W_i> \\
&= <\bar{Y}, \bar{W}_i> - <\bar{Y}, \epsilon_i + \sum_{j=0}^{i-1} \frac{<P_i, W_j>}{<W_j, W_j>} W_j - \sum_{j=0}^{i-1} \frac{<\bar{P}_i, \bar{W}_j>}{<\bar{W}_j, \bar{W}_j>} \bar{W}_j> \\
&= <\bar{Y}, \bar{W}_i> - \sum_{j=0}^{i-1} \frac{<P_i, W_j>}{<W_j, W_j>} <\bar{Y}, W_j> + \sum_{j=0}^{i-1} \frac{<\bar{P}_i, \bar{W}_j>}{<\bar{W}_j, \bar{W}_j>} <\bar{Y}, \bar{W}_j> \\
&= <\bar{Y}, \bar{W}_i> - \sum_{j=0}^{i-1} \frac{<P_i, W_j>}{<W_j, W_j>} <Y, W_j> + \sum_{j=0}^{i-1} \frac{<\bar{P}_i, \bar{W}_j>}{<\bar{W}_j, \bar{W}_j>} <\bar{Y}, \bar{W}_j>
\end{aligned} \qquad (32)$$

and

$$<Y, W_0> = <\bar{Y}, \bar{W}_0> \qquad (33)$$

Inspection of (31) and (33) shows that unbiased estimates of ERR_i can be determined if $<P_i, W_j>$ and $<W_j, W_j>$, $j = 0, 1, \cdots, i$, and $<Y, Y>$ are known. In order to estimate $<P_i, W_j>$, $j = 0, 1, \cdots, i-1$ and $<W_i, W_i>$, it is assumed that the energy and SNRs of the true signals are known, that is $<Y, Y>$ and $<P_i, P_i>$, $i = 0, 1, \cdots, n$ are all known. In practice, this is not an unrealistic assumption because the energy of signals and the measurement noise can often be estimated. In what follows an estimation of $<P_i, W_j>$ and $<W_i, W_i>$ will be derived using the known quantities $<Y, Y>$ and $<P_i, P_i>$, $i = 0, 1, \cdots, n$. In fact for any $i = 1, 2, \cdots, n$

$$\begin{aligned}
<P_i, W_0> &= <P_i, \bar{P}_0 - \epsilon_0> \qquad (34) \\
&= <P_i, \bar{P}_0> \\
&= <\bar{P}_i, \bar{P}_0> \\
<W_0, W_0> &= <P_0, P_0>
\end{aligned}$$

and for $j = 1, 2, \cdots, i-1$ and $i = 2, 3, \cdots, n$

$$<P_i, W_j> = <\bar{P}_i, \bar{W}_j> - \sum_{k=0}^{j-1} \frac{<P_j, W_k>}{<W_k, W_k>} <P_i, W_k> + \sum_{k=0}^{j-1} \frac{<\bar{P}_j, \bar{W}_k>}{<\bar{W}_k, \bar{W}_k>} <\bar{P}_i, \bar{W}_k> \qquad (35)$$

$$< W_{i-1}, W_{i-1} > \; = \; < P_{i-1}, P_{i-1} > - \sum_{j=0}^{i-2} \frac{< P_{i-1}, W_j >^2}{< W_j, W_j >}$$

Note that eqn. (35) can be solved iteratively with initial condition (34).

5.2 Parameter estimation

Following the estimation of $< P_i, W_j >$ and $< W_i, W_i >$, $i = 1, 2, \cdots, n$, $j = 0, 1, \cdots, i-1$ in (35), an unbiased estimates of the noise-free parameters can be obtained. This can be done using the following equations.

$$\theta_i = \sum_{j=i}^{n} \hat{g}_j v_j \tag{36}$$

where

$$\hat{g}_i = \frac{< Y, W_i >}{< W_i, W_i >} = \frac{\sum_{t=1}^{N} y(t) w_i(t)}{\sum_{t=1}^{N} w_i^2(t)}, i = 0, 1, \cdots, n \tag{37}$$

and

$$\begin{aligned} v_i &= 1 \qquad (38) \\ v_j &= -\sum_{k=i}^{j-1} \alpha_{k,j} v_k, j = i+1, \cdots, n \end{aligned}$$

in which

$$\alpha_{k,j} = \frac{< P_j, W_k >}{< W_k, W_k >} = \frac{\sum_{t=1}^{N} p_j(t) w_k(t)}{\sum_{t=1}^{N} w_k^2(t)}, k = 0, 1, \cdots, j-1 \tag{39}$$

The modified OFR algorithm, for the case where there is noise on the regressors, can now be summarised as follows

Step 1 Apply the OFR algorithm to the noisy data to obtain $< \bar{P}_i, \bar{W}_j >$, and $< \bar{W}_j, \bar{W}_j >$, $j = 0, 1, \cdots, i-1$, $i = 1, 2, \cdots, n$ using (20), that is

$$\begin{aligned} \bar{W}_0 &= \bar{P}_0 \\ \bar{W}_i &= \bar{P}_i - \sum_{j=0}^{i-1} \bar{\alpha}_{j,i} \bar{W}_j \end{aligned}$$

where

$$\bar{\alpha}_{j,i} = \frac{<\bar{P}_i, \bar{W}_j>}{<\bar{W}_j, \bar{W}_j>} = \frac{\sum_{t=1}^{N} \bar{p}_i(t)\bar{w}_j(t)}{\sum_{t=1}^{N} \bar{w}_j^2(t)}, j = 0, 1, \cdots, i-1$$

Step 2 Estimate $<P_i, W_j>$ and $<W_i, W_i>$, $j = 0, 1, \cdots, i-1$ $i = 1, 2, \cdots, n$ using (34) and (35), that is

$$\begin{aligned} <P_i, W_0> &= <\bar{P}_i, \bar{P}_0> = \sum_{t=1}^{N} \bar{p}_i(t)\bar{p}_0(t) \\ <W_0, W_0> &= <P_0, P_0> = \sum_{t=1}^{N} p_0^2(t) \end{aligned}$$

and for $j = 1, 2, \cdots, i-1$ and $i = 2, 3, \cdots, n$

$$\begin{aligned} <P_i, W_j> &= <\bar{P}_i, \bar{W}_j> - \sum_{k=0}^{j-1} \frac{<P_j, W_k>}{<W_k, W_k>} <P_i, W_k> + \sum_{k=0}^{j-1} \frac{<\bar{P}_j, \bar{W}_k>}{<\bar{W}_k, \bar{W}_k>} <\bar{P}_i, \bar{W}_k> \\ &= \sum_{t=1}^{N} \bar{p}_i(t)\bar{w}_j(t) - \sum_{k=0}^{j-1} \frac{\sum_{t=1}^{N} p_j(t)w_k(t)}{\sum_{t=1}^{N} w_k^2(t)} \sum_{t=1}^{N} p_i(t)w_k(t) + \sum_{k=0}^{j-1} \frac{\sum_{t=1}^{N} \bar{p}_j(t)\bar{w}_k(t)}{\sum_{t=1}^{N} \bar{w}_k^2(t)} \sum_{t=1}^{N} \bar{p}_i(t)\bar{w}_k(t) \\ <W_{i-1}, W_{i-1}> &= <P_{i-1}, P_{i-1}> - \sum_{j=0}^{i-2} \frac{<P_{i-1}, W_j>^2}{<W_j, W_j>} = \sum_{t=1}^{N} p_{i-1}^2(t) - \sum_{j=0}^{i-2} \frac{(\sum_{t=1}^{N} p_{i-1}(t)w_j(t))^2}{\sum_{t=1}^{N} w_j^2(t)} \end{aligned}$$

Step 3 Obtain ERR_i, $i = 1, 2, \cdots, n$ using (31) and (30)

$$ERR_i = \frac{(\sum_{t=1}^{N} y(t)w_i(t))^2}{\sum_{t=1}^{N} y^2(t) \sum_{t=1}^{N} w_i^2(t)}$$

Step 4 Select the significant terms according to the maximal ERR criterion. The term selection procedure should stop when an error tolerance is reached.

Step 5 Calculate the parameters θ using (39), (38), (31), (37) and (36).

Remark 1 As in the classical OFR case, the final model and parameters need to be assessed. A commonly used approach to check the validity of the identified model is to use higher order statistical correlation analysis (Billings and Voon 1986, Billings and Zhu 1994). An alternative is to check both the short and the long term predictive ability of the model or some quantitative invariants such as Lyapunov exponents and correlation dimensions etc.

Remark 2 The above modified OFR algorithm is obtained under the assumption that the noise sequences are independent. In the identification of nonlinear dynamical systems, the noise is often coloured therefore, in some situations some whitening methods may be needed before the proposed algorithm is applied.

6 An analysis of the sensitivity of the ERR values in the modified OFR algorithm to $< Y, Y >$ and $< P_i, P_i >$

The proposed modified OFR algorithm, for the case where there is noise on the regressors, is based on the assumptions that the energy of the signals, $< Y, Y >$ and $< P_i, P_i >$, $i = 0, 1, \cdots, n$, are known or can be estimated. In this section, an analysis of the sensitivity of the proposed algorithm to these quantities is given. Let $< Y, Y > \epsilon_Y$ and $< P_i, P_i > \epsilon_{P_i}$, $i = 0, 1, \cdots, n$ be the perturbation of $< Y, Y >$ and $< P_i, P_i >$, where ϵ_Y and ϵ_{P_i}, $i = 0, 1, \cdots, n$ are $n + 2$ positive real numbers. When ϵ_Y and ϵ_{P_i}, $i = 0, 1, \cdots, n$ vary independently, the problem is known as multi-parametric. In this paper, it is assumed that $\epsilon_Y = \epsilon_{P_i} = \epsilon$, $i = 0, 1, \cdots, n$. Throughout this section, the quantities calculated from the perturbation are denoted by the subscript d.

From the modified OFR algorithm, the ERRs corresponding to the perturbation are

$$ERR_{d,i} = \frac{< Y, W_i >_d^2}{< Y, Y >_d < W_i, W_i >_d} \tag{40}$$

where for $i = 1, 2, \cdots, n$

$$< Y, W_i >_d = < \bar{Y}, \bar{W}_i > - \sum_{j=0}^{i-1} \frac{< P_i, W_j >_d}{< W_j, W_j >_d} < Y, W_j >_d + \sum_{j=0}^{i-1} \frac{< \bar{P}_i, \bar{W}_j >}{< \bar{W}_j, \bar{W}_j >} < \bar{Y}, \bar{W}_j > \tag{41}$$

and

$$< Y, W_0 >_d = < \bar{Y}, \bar{W}_0 > \tag{42}$$

From eqns. (34), (35), (42), and (41), the perturbed $< W_i, W_i >_d$, $< P_i, W_j >_d$, and $< Y, W_i >_d$ can be derived as

$$< W_0, W_0 >_d = < W_0, W_0 > \epsilon \tag{43}$$

$$< W_i, W_i >_d = \frac{1}{\epsilon}[1 + \frac{< P_i, P_i > (\epsilon^2 - 1)}{< W_i, W_i >} - \sum_{j=1}^{i-1} \frac{< P_i, W_j >^2}{< W_j, W_j >< W_i, W_i >}(\frac{\epsilon}{\delta_{ij}} - 1)] < W_i, W_i > \tag{44}$$

$$< P_i, W_0 >_d = < P_i, W_0 > \tag{45}$$

$$< P_i, W_j >_d = \frac{1}{\epsilon}[1 + \frac{< \bar{P}_i, \bar{W}_j > + \sum_{k=0}^{j-1} \frac{<\bar{P}_i, \bar{W}_k>}{<\bar{W}_k, \bar{W}_k>} < \bar{P}_i, \bar{W}_k >}{< P_i, W_j >}(\epsilon - 1) - \sum_{k=1}^{j-1} \frac{< P_{i-1}, W_k >< P_i, W_k >}{< W_k, W_k >< P_i, W_j >}(\frac{\epsilon}{\eta_{ij}} - 1)] < P_i, W_j > \tag{46}$$

and

$$< Y, W_0 >_d = < Y, W_0 > \tag{47}$$

$$< Y, W_i >_d = \frac{1}{\epsilon}[1 + \frac{< \bar{Y}, \bar{W}_i > + \sum_{j=0}^{i-1} \frac{<\bar{P}_i, \bar{W}_j>}{<\bar{W}_j, \bar{W}_j>} < \bar{Y}, \bar{W}_j >}{< Y, W_j >}(\epsilon - 1) - \sum_{j=1}^{i-1} \frac{< P_i, W_j >< Y, W_j >}{< W_j, W_j >< Y, W_i >}(\frac{\epsilon}{\sigma_{ij}} - 1)] < Y, W_i > \tag{48}$$

where

$$\delta_{ij} = \frac{< W_j, W_j >_d}{< W_j, W_j >} \frac{< P_i, W_j >^2}{< P_i, W_j >_d^2} \tag{49}$$

$$\eta_{ij} = \frac{< W_j, W_j >_d}{< W_j, W_j >} \frac{< P_{i-1}, W_j >}{< P_{i-1}, W_j >_d} \frac{< P_i, W_j >}{< P_i, W_j >_d} \tag{50}$$

and

$$\sigma_{ij} = \frac{< W_j, W_j >_d}{< W_j, W_j >} \frac{< P_i, W_j >}{< P_i, W_j >_d} \frac{< Y, W_j >}{< Y, W_j >_d} \tag{51}$$

Note that when ϵ approaches 1, all the parameters δ_{ij}, η_{ij}, and σ_{ij} tend to ϵ. Substituting (44), (48) and $< Y, Y >_d = < Y, Y > \epsilon$ into (40) yields

$$ERR_{d,i} = \frac{< Y, W_i >_d^2}{< Y, Y >_d < W_i, W_i >_d} = \frac{1}{\epsilon^2} ERR_i \Delta_i \tag{52}$$

where

$$\Delta_i = \frac{[1 + \frac{<\bar{Y}, \bar{W}_i> + \sum_{j=0}^{i-1} \frac{<\bar{P}_i, \bar{W}_j>}{<\bar{W}_j, \bar{W}_j>} <\bar{Y}, \bar{W}_j>}{<Y, W_j>}(\epsilon - 1) - \sum_{j=1}^{i-1} \frac{<P_i, W_j><Y, W_j>}{<W_j, W_j><Y, W_i>}(\frac{\epsilon}{\sigma_{ij}} - 1)]^2}{1 + \frac{<P_i, P_i>(\epsilon^2 - 1)}{<W_i, W_i>} - \sum_{j=1}^{i-1} \frac{<P_i, W_j>^2}{<W_j, W_j><W_i, W_i>}(\frac{\epsilon}{\delta_{ij}} - 1)} \tag{53}$$

From (53) it can be observed that if $\epsilon \approx 1$, Δ_i will be very close to 1 as well. In this case, one can conclude that

$$ERR_{d,i} \approx \frac{1}{\epsilon^2} ERR_i, i = 0, 1, \cdots, n \tag{54}$$

This implies that a small perturbation will not change the order of the ERR values. This is an important property of the new modified algorithm because it suggests that the related model terms will be ordered or ranked correctly even when there is noise on the regressors. Therefore it should be possible, using the modified OFR algorithm, to correctly select the model structure. In fact, this conclusion also extends to the parameter estimation problem. Recall the formulae for calculating the parameters θ_i (39), (38), (37) and (36). Now substituting (44), (46), and (41) into these equations yields

$$\hat{g_{d,i}} = \frac{< Y, W_i >_d}{< W_i, W_i >_d} = \hat{g}_i \frac{1 + \frac{<\bar{Y},\bar{W}_i> + \sum_{j=0}^{i-1} \frac{<\bar{P}_i,\bar{W}_j>}{<\bar{W}_j,\bar{W}_j>} <\bar{Y},\bar{W}_j>}{<Y,W_j>}(\epsilon - 1) - \sum_{j=1}^{i-1} \frac{<P_i,W_j><Y,W_j>}{<W_j,W_j><Y,W_i>}(\frac{\epsilon}{\sigma_{ij}} - 1)}{1 + \frac{<P_i,P_i>(\epsilon^2-1)}{<W_i,W_i>} - \sum_{j=1}^{i-1} \frac{<P_i,W_j>^2}{<W_j,W_j><W_i,W_i>}(\frac{\epsilon}{\delta_{ij}} - 1)} \tag{55}$$

$$\alpha_{d,k,j} = \frac{< P_j, W_k >_d}{< W_k, W_k >_d} = \alpha_{k,j} \frac{1 + \frac{<\bar{P}_j,\bar{W}_k> + \sum_{l=0}^{k-1} \frac{<\bar{P}_j,\bar{W}_l>}{<\bar{W}_l,\bar{W}_l>} <\bar{P}_j,\bar{W}_l>}{<P_j,W_k>}(\epsilon - 1) - \sum_{l=1}^{k-1} \frac{<P_{j-1},W_l><P_j,W_l>}{<W_l,W_l><P_j,W_k>}(\frac{\epsilon}{\eta_{ij}} - 1)}{1 + \frac{<P_k,P_k>(\epsilon^2-1)}{<W_k,W_k>} - \sum_{l=1}^{k-1} \frac{<P_k,W_l>^2}{<W_l,W_l><W_k,W_k>}(\frac{\epsilon}{\delta_{kl}} - 1)} \tag{56}$$

which tend to $\hat{g}_i$ and $\alpha_{k,j}$ when ϵ is close to 1.

Similar results should hold for the multi-parametric case when the perturbations are small enough.

7 Numerical simulations

7.1 Example 1: A linear system

Consider the following simple AR model

$$y(t) = a_1 y(t-1) + a_2 y(t-2) + a_3 y(t-3) + a_4 y(t-4) + u(t) \tag{57}$$

where $a_1 = 1.8$, $a_2 = -1.99$, $a_3 = 1.422$, and $a_4 = 0.493$. To fully excite all the modes of the system, the simulation was conducted and data was collected with $u(t)$ selected to be a white

Model Term		1	$y(t-1)$	$y(t-2)$	$y(t-3)$	$y(t-4)$	$u(t)$
Parameter Estimates	Noise-free	X	1.8000e+00	-1.9900e+00	1.4220e+00	-4.9300e-01	1.0000e+00
	OFR	2.1475e-01	8.4570e-01	-3.9986e-01	X	2.8854e-01	5.5197e-01
	Modified OFR	X	1.7476e+00	-1.8962e+00	1.3342e+00	-4.4421e-01	9.9035e-01
	Modified OFR (5%)	X	1.7088e+00	-1.8708e+00	1.2958e+00	-4.4594e-01	1.2082e+00
	Modified OFR (10%)	X	1.7515e+00	-2.0004e+00	1.4081e+00	-5.4020e-01	1.4921e+00
ERR	Noise-free	X	9.0040e-01	2.4071e-02	3.3846e-02	6.6658e-03	3.5015e-02
	OFR	1.5695e-03	8.6605e-01	1.3921e-02	X	2.9565e-02	1.5021e-02
	Modified OFR	X	9.0046e-01	2.3854e-02	3.3530e-02	5.6350e-03	3.4844e-02
	Modified OFR(5%)	X	8.8638e-01	2.6103e-02	2.8865e-02	5.7656e-03	4.0882e-02
	Modified OFR(10%)	X	8.8660e-01	2.9784e-02	2.4739e-02	7.9744e-03	4.7882e-02

Table 1: The terms and parameters of the final model for Example 1

noise sequence with mean zero and variance $\sigma^2 = 1$. The set of terms in an initial candidate model was set to be $\{1, y(t-1), y(t-2), y(t-3), y(t-4), y(t-5), y(t-6), u(t)\}$. To test the algorithm, an additive independent white noise with mean zero was added to $u(t)$ and $y(t)$ such that the SNRs of the resulting input and output signals were 6.0139dB and 17.072dB, respectively. The energies of input and output, that is $< U, U >$ and $< Y, Y >$ were calculated to be $1.6641e+05$ and $2.1241e+06$ respectively. The identified results using the original and the modified OFR algorithms with the above energy values are shown in Table (1), where only those terms selected by the algorithms are shown and X indicates the term has not been selected by a specific algorithm.

From the example it can be observed that even for a linear system where the regressros are corrupted with white noise of relatively low SNR (6.0139dB and 17.072dB), the values of ERRs calculated by the conventional OFR algorithm can be quite different from the noise-free case, for example, the ERR value for the term $y(t-1)$ changes from $9.0040e-01$ to $8.6605e-01$. This causes the conventional algorithm to incorrectly select the model terms when the noise sequences are not considered as additional regression variables whilst the proposed modified algorithm works extremely well. In order to test the robustness of the proposed algorithm with respect to the energy, the algorithm was applied with 5% and 10% errors in the above energy values and the obtained results are also shown in Table (1), which show that the algorithm can still work reasonably well.

7.2 Example 2: A nonlinear system

A nonlinear dynamical system is defined as

$$y(t) = 0.5y(t-1) + 0.1u^2(t-1) + u(t-2) \tag{58}$$

where the system input $u(t)$ was a white noise sequence of uniform distribution with mean zero and finite variance 1.0. A polynomial NARMAX model with input and output lags 2 and

Model Term		$y(t-1)$	$u^2(t-1)$	$u(t-2)$	$u^2(t-2)$	$y(t-1)u(t-1)$
Parameter Estimates	Noise-free	5.0000e-01	1.0000e-01	1.0000e+00	X	X
	OFR	5.0349e-01	1.0946e-01	1.0058e+00	-1.6782e-02	-9.4633e-03
	Modified OFR	5.0237e-01	1.0180e-01	9.9122e-01	X	X
	Modified OFR(5%)	5.2932e-01	9.9750e-02	1.0570e+00	X	X
	Modified OFR(10%)	4.9143e-01	6.2116e-02	1.0404e+00	X	X
ERR	Noise-free	8.2987e-02	7.8675e-04	9.1623e-01	X	X
	OFR	9.1499e-01	8.2803e-04	8.2706e-02	3.0542e-06	1.6453e-06
	Modified OFR	8.1779e-02	8.0391e-04	9.1810e-01	X	X
	Modified OFR(5%)	8.2933e-02	7.3987e-04	9.1548e-01	X	X
	Modified OFR(10%)	8.1619e-02	8.1643e-04	9.1708e-01	X	X

Table 2: The terms and parameters of the final model for Example 2

nonlinear degree 2 was used to fit the simulated data. The full candidate model contained 15 terms, that is $\{1, y(t-1), y(t-2), u(t-1), u(t-2), y^2(t-1), y(t-1)y(t-2), y(t-1)u(t-1), y(t-1)u(t-2), y^2(t-2), y(t-2)u(t-1), y(t-2)u(t-2), u^2(t-1), u(t-1)u(t-2), u^2(t-2)\}$. To test the algorithm, an additive white noise with mean zero and finite variance was added to the data. Note that if the noise is added to input $u(t)$ and $y(t)$ directly, the regressors will have a coloured noise because of the nonlinearity which can not be dealt with by the proposed algorithm directly. Therefore, the noise in this simulation was added to the regressors so that the noisy regressors and output sequences have SNRs of around 26dB and 31.728dB, respectively. The energies for the total 15 terms above are $9.9980e+03, 1.2427e+04, 1.2424e+04, 3.2940e+03, 3.2933e+03, 2.1274e+04, 1.8713e+04, 4.0674e+03, 4.2918e+03, 2.1267e+04, 4.0821e+03, 4.0653e+03, 1.9720e+03, 1.0933e+03, 2.9571e+03$ and the energy of the output is $1.2428e+04$. The identified results using the original and modified OFR algorithms with the above energy values are shown in Table (2), where only those terms selected by the algorithms are shown.

The simulation results show that both the conventional OFR and the modified OFR can pick the correct significant terms. This is because the SNR of the signal was set to be high (26dB and 31.728dB). However, for the conventional OFR algorithm, the ERRs have been affected by the presence of noise and the order of the selected terms has been changed from $u(t-2), y(t-1), u^2(t-1)$ to $y(t-1), u(t-2)$, $u^2(t-1)$. A robustness test was also conducted by using 5% and 10% errors in the true energy values which shows the algorithm works well when the energy values are not perfectly estimated.

8 Conclusions

A modified OFR algorithm for the identification of both the model terms or structure and the unknown parameters when there is noise on the model regressors has been introduced. It has been shown that the presence of noise on the regressor terms can produce biased results for both term selection and parameter estimation. The new algorithm makes use of some energy information about the true signals to correct the ERR values and the estimated parameters. The sensitivity of the algorithm to the properties of the noise has been studied and the method has been tested on simulated data and was shown to perform very well.

9 Acknowledgement

The authors gratefully acknowledge financial support from EPSRC (UK).

References

[1] Aguirre, L. A. and Billings, S. A., (1995a) Improved structure selection for nonlinear models based on term clustering, *Int. J. Contr.*, Vol. 62, No.3, pp. 569-587.

[2] Aguirre, L. A. and Billings, S. A., (1995b) Retrieving dynamical invariants from chaotic data using narmax models, *Int. J. Bifurcation and Chaos*, Vol. 5, No. 2, pp. 449-474.

[3] Balikhin, M. A., Zhu, D., and Billings, S. A., (2005) Time domain identification of nonlinear systems: from the measurements to continuous differential equations, *The 2005 Joint Assembly Meeting of AGU, SEG, NABS and SPD/AAS*, New Orleans, Louisiana.

[4] Billings, S. A., Chen, S., and Kronenberg, M. J., (1989) Identification of MIMO nonlinear systems using a forward-regression orthogonal estimator, *Int. J. Contr.*, Vol. 49, pp. 2157-2189.

[5] Billings, S. A., Chen, S., and Backhouse, R. J., (1989), The identification of linear and nonlinear models of a turbocharged automotive diesel engine, *Mechanical Systems and Signal Processing*, Vol. 3, pp. 123-142.

[6] Billings, S. A., Fadzil, M. B., Sulley, J., and Johnson P. M., (1988), Identification of a nonlinear difference equation model of an industrial diesel generator, *Mechanical Systems and Signal Processing*, Vol. 2, pp. 59-76.

[7] Billings, S. A., Korenberg, M. J., and Chen, S., (1988) Identification of non-linear output-affine systems using an orthogonal least-squares algorithm, *Int. J. Syst. Sci.*, Vol. 19, pp. 59-76.

[8] Billings, S. A., and Voon, W. S. F. (1986) Correlation based model validity tests for nonlinear models, *International Journal of Control*, Vol. 40, No. 1, pp. 235-244.

[9] Billings, S. A. and Zhu, Q. M. (1994) Nonlinear model validation using correlation tests, *International Journal of Control*, Vol. 60, No. 6, pp. 1107-1120.

[10] 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*, Vol. 50, No. 5, pp. 1873-1896.

[11] Coca,D. and Billings, S. A., (2001) Identification of coupled map lattice models of complex spatio-temporal pattern, *Phys. Lett.*, A287, pp. 65-73.

[12] Coca, D., Zheng, Y., Mayhew, J., and Billings, S. A. (2000) Nonlinear system identification and analysis of complex dynamical behaviour in reflected light measurements of vasomotion, *International Journal of Bifurcation and Chaos*, Vol. 10, No. 2, pp. 461-476.

[13] Guo, L. Z. and Billings, S. A., (2004) Identification of coupled map lattice models of stochastic spatio-temporal dynamics using wavelets, *Dynamic Systems*, Vol. 19, No.3, pp. 265-278.

[14] Hong, X. and Harris, C. J., (2001) Nonlinear model structure design and construction using orthogonal least squares and d-optimality design, *IEEE Transactions on Neural Networks*, vol. 13, no. 5, pp. 12451250.

[15] Leontaritis, I. J., and Billings, S. A., (1985) Input-output parametric models for nonlinear systems, Part I: deterministic nonlinear Nonlinear systems; Part II: stochastic nonlinear systems, *Int. J. Control*, vol. 41, no. 2, pp. 303-344.

[16] Leontaritis, I. J., and Billings, S. A., (1987) Experimental design and identificability for nonlinear systems, *Int. J. Syst. Sci.*, vol. 18, no. 1, pp. 189-202.

[17] Liu, J. J., Kung, I. C., and Chao, H. C., (2001) Speed estimate of induction motor using a nonlinear identification technique, *Proceedings of the National Science Council*, Republic of China. Part A, vol. 25, no. 2, pp. 107-114.

[18] Mao, K. Z., and Billings, S. A., (1997) Algorithms for minimal model structure detection in nonlinear dynamic system identification, *Int. J. Control*, Vol. 68, pp. 311330.

[19] Mao, K. Z., and Billings, S. A., (1999) Variable selection in nonlinear systems modelling, *Mechanical Systems and Signal Processing*, Vol. 13, pp. 351366.

[20] Wei, H. L., Billings, S. A., and Liu, J. (2004) Term and variable selection for nonlinear system identification, *Int. J. Control*, Vol. 77, No. 1, pp. 86-110.