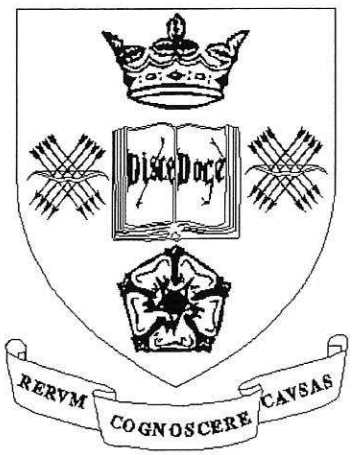
Identification and Reconstruction of Chaotic Systems Using Multiresolution Wavelet Decompositions

H. L. Wei S. A. Billings

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



Research Report No. 845

Sep. 2003

Identification and Reconstruction of Chaotic Systems Using Multiresolution Wavelet Decompositions

H.L. Wei, S.A. Billings
Department of Automatic Control and Systems Engineering, University of Sheffield
Mappin Street, Sheffield, S1 3JD, UK

A new modelling framework for identifying and reconstructing chaotic systems is developed based on multiresolution wavelet decompositions. Qualitative model validation is used to compare the multiresolution wavelet models and it is shown that the dynamical features of chaotic systems can be captured by the identified models providing the wavelet basis functions are properly selected. Two basis selection algorithms, orthogonal least squares (OLS) and a new matching pursuit orthogonal least squares (MPOLS), are considered and compared. Several examples are included to illustrate the results.

1. Introduction

Intensive studies on nonlinear dynamics over the past few decades have shown that chaotic phenomena flourish in nature and engineering systems. Chaos, which occurs widely in biology, chemistry, ecology, engineering, medicine, meteorology, physics, and the social sciences [Parker & Chua, 1989], has attracted extensive interest over many years with many analytical and experimental studies. Chaotic systems have been described both quantitatively and qualitatively, and are often characterized by a number of dynamical invariants such as Lyapunov exponents [Wolf et al., 1985; Eckmann et al., 1986; Brown et al., 1991], correlation dimensions [Grassberger & Procaccia, 1983], the geometry of the attractors [Lorenz, 1963; Packard et al., 1980; Eckmann & Ruelle, 1985; Milnor, 1985], Poincare maps and bifurcation diagrams [Guckenheimer & Holmes, 1983; Parker & Chua, 1989].

Modelling plays a major role in analyzing, understanding and controlling chaotic systems. Since large noise-free data sets are often required to accurately compute the dynamical invariants, models can be simulated to generate the required data. Although a model is not always required to control a chaotic system, conventional model-based control of chaos has been very successful [Chen & Dong, 1993].

Nonlinear system identification plays an important role in the modelling of chaotic systems. Most existing nonlinear identification approaches can easily be applied to chaotic systems. Following the pioneering work by Packard et al. [1980], Takens [1981], Crutchfield & McNamara [1987] and Farmer & Sidorowich [1987], prediction and reconstruction of chaotic systems have become important topics in nonlinear dynamics. Several approaches have been proposed for predicting and reconstructing chaotic systems. These include local, global, or semi-local modelling methods, see, for example, the papers by Casdagli [1989], Abarbanel et al. [1989,1990], Gouesbet [1991,1992], Linsay [1991], Albano et al. [1992], Principe et al. [1992], Smith [1992], Mees [1993], Aguirre & Billings [1994,1995a], Cao et al. [1995,1997], Mendes & Billings [1997], Allingham et al. [1998], Bagarinao et al. [1999], Billings & Coca [1999], Menard et al. [2000] and the references therein.

Chaotic systems often exhibit extremely sensitive dependence on initial conditions, that is, initially nearby trajectories diverge and eventually become uncorrelated. This property imposes limits on the ability to make long term predictions and also makes it difficult to determine whether the identified model is faithful to the

chaotic system under study. It follows that a simple comparison of predictors is not a good indicator of model quality. To complement the statistical validation techniques [Billings & Voon, 1986], new model validation techniques involving the calculation and reconstruction of dynamical invariants were proposed by Principe et al. [1992]. The issue of dynamic modelling and dynamical equivalence was discussed in detail as an integral part of system identification in Haynes and Billings [1994], Aguirre and Billings [1995a], and a number of studies confirmed the importance of qualitative model validation techniques [Aguirre & Billings, 1994; Aguirre & Billings, 1995b; Mendes & Billings, 1998; Billings & Coca, 1999; Billings & Zheng, 1999; Zheng & Billings, 1999].

An efficient nonlinear system identification methodology studied over the past two decades is the NARMAX method which is based on an input-output model initially proposed by Billings and Leontaritis [1982], Leontaritis and Billings [1985] and which has been successfully applied to many systems including chaotic systems [Haynes & Billings, 1994; Aguirre & Billings, 1995a; Aguirre & Mendes, 1996; Billings & Coca, 1999; Correa et al., 2000; Billings and Yang, 2003].

In this paper, a new modelling framework for identifying and reconstructing chaotic systems is developed based on multiresolution wavelet decompositions. The multiresolution wavelet model, which is capable of reconstructing dynamical invariants such as the attractors and Poincare maps of chaotic systems, provides an applicable and effective representation for chaotic systems provided that the wavelet basis functions are properly selected. Wavelets have been successfully applied to model chaotic systems by previous authors including Cao et al.[1995, 1997], Coca and Billings [1997], Allingham et al. [1998] and Billings and Coca [1999]. The new wavelet model proposed here is however similar to that adopted by Billings and Coca [1999] but different from the approaches suggested by Cao et al. [1995] and Allingham et al. [1998].

A common and important problem in wavelet based modelling is how to select the significant wavelet basis functions from a given over-complete wavelet dictionary in order to avoid over-parameterization and ensure a parsimonious dynamical reconstruction. It follows that the basis functions selected from the same wavelet dictionary based on different basis selection algorithms might be greatly different from each other. In addition, to achieve the same approximation accuracy, the number of model terms selected by different algorithms may also be different [Billings & Wei, 2003]. In this paper, two basis selection algorithms, the orthogonal least squares (OLS) and a new matching pursuit orthogonal least squares (MPOLS), are considered and compared.

The outline of the paper is as follows. In Sec. 2, the multiresolution wavelet model for nonlinear system identification is introduced. This includes brief descriptions on multiresolution wavelet decompositions, the NARAMX representations for nonlinear systems, and the mutiresolution wavelet-based model structure. The basis selection problem is discussed in Sec. 3. Three examples, the Henon map, the modified Van der Pol oscillator and a one-dimensional cellular automata, are presented in Sec. 4. Finally, the main points of the work are summarized in Sec. 5.

2. The Multiresolution Wavelet Models for Nonlinear System Identification

In this section, multiresolution wavelet decompositions are briefly reviewed. The intrinsic nonlinearity of a system is then expressed by means of the wavelet-NARX representation and multivariable multiresoluttion wavelet decompositions.

2.1 Multiresolution wavelet decompositions

It is known that for identification problems it is useful to have a basis of orthogonal (semi-orthogonal or bi-orthogonal) functions whose support can be made as small as required and which provides a universal approximation to any $L^2(R)$ function with arbitrary desired accuracy. Under some assumptions and considerations, an orthogonal wavelet system can be constructed using *multiresolution analysis* (MRA) [Mallat, 1989; Chui 1992]. Assume that the wavelet φ and associated scaling function ϕ constitute an orthogonal wavelet system, then any function $f \in L^2(R)$ can be expressed as a *multiresolution wavelet decomposition*

$$f(x) = \sum_k a_{j_0,k}\phi_{j_0,k}(x) + \sum_{j\geq j_0}\sum_k d_{j,k}\varphi_{j,k}(x) \tag{1}$$

where $\varphi_{j,k}(x) = 2^{j/2}\varphi(2^j x - k)$ and $\phi_{j,k}(x) = 2^{j/2}\phi(2^j x - k)$, $j,k \in Z$,and the wavelet approximation coefficient $a_{j_0,k}$ and the wavelet detail coefficient $d_{j,k}$ can be calculated in theory by the inner products:

$$a_{j_0,k} =< f, \phi_{j_0,k} >= \int f(x)\overline{\phi_{j_0,k}(x)}dx \tag{2}$$

$$d_{j,k} =< f, \varphi_{j,k} >= \int f(x)\overline{\varphi_{j,k}(x)}dx \tag{3}$$

The over-bar above the functions $\varphi(\cdot)$ and $\phi(\cdot)$ in (2) and (3) indicates complex conjugate, and j_0 is an arbitrary integer representing the lowest resolution or scaling level.

Using the concept of *tensor products*, the multiresolution decomposition (1) can be immediately generalised to the muti-dimensional case, where a multiresolution wavelet decomposition can be defined by taking the *tensor product* of the one-dimensional scaling and wavelet functions [Mallat,1989]. Let $f \in L^2(R^d)$, then $f(x)$ can be represented by the *multiresolution wavelet decomposition* as

$$f(x_1,\cdots,x_d) = \sum_k \alpha_{j_0,k}\Phi_{j_0,k}(x_1,\cdots,x_d) + \sum_{j\geq j_0}\sum_k\sum_{l=1}^{2^d-1}\beta_{j,k}^{(l)}\Psi_{j,k}^{(l)}(x_1,\cdots,x_d) \tag{4}$$

where $k = (k_1,k_2,\cdots,k_d) \in Z^d$ and

$$\Phi_{j_0,k}(x_1,\cdots,x_d) = 2^{j_0 d/2}\prod_{i=1}^{d}\phi(2^{j_0}x_i - k_i) \tag{5}$$

$$\Psi_{j,k}^{(l)}(x_1,\cdots,x_d) = 2^{jd/2}\prod_{i=1}^{d}\eta^{(i)}(2^j x_i - k_i) \tag{6}$$

with $\eta^{(i)} = \phi$ or φ (scalar scaling function or the mother wavelet) but at least one $\eta^{(i)} = \varphi$. In the two-dimensional case, the multiresolution approximation can be generated, for example, in terms of the dilation and translation of the two-dimensional scaling and wavelet functions

$$\begin{cases} \Phi_{j,k_1,k_2}(x_1,x_2) = \phi_{j,k_1}(x_1)\phi_{j,k_2}(x_2) \\ \Psi_{j,k_1,k_2}^{(1)}(x_1,x_2) = \phi_{j,k_1}(x_1)\varphi_{j,k_2}(x_2) \\ \Psi_{j,k_1,k_2}^{(2)}(x_1,x_2) = \varphi_{j,k_1}(x_1)\phi_{j,k_2}(x_2) \\ \Psi_{j,k_1,k_2}^{(3)}(x_1,x_2) = \varphi_{j,k_1}(x_1)\varphi_{j,k_2}(x_2) \end{cases} \tag{7}$$

Notice that if $j_0 \to -\infty$, the approximation representation (1) can be expressed using only the scaling function ϕ, that is, there exists a sufficiently large integer J, such that

$$f(x_1,\cdots,x_d) \approx \sum_k \alpha_{J,k}\Phi_{J,k}(x_1,\cdots,x_d) = \sum_{k_1,k_2,\cdots,k_d} 2^{Jd/2}\prod_{i=1}^{d}\phi(2^J x_i - k_i) \tag{8}$$

2.2 The NARMAX representation for nonlinear systems

The NARMAX model representation, which was initially proposed by Billings and Leontaritis [Billings & Leontaritis 1982, Leontaritis & Billings 1985], takes the form of the following nonlinear difference equation:

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

where f is an unknown nonlinear mapping, $u(t)$ and $y(t)$ are the sampled input and output sequences, n_u and n_y are the maximum input and output lags, respectively. The noise variable $e(t)$ with maximum lag n_e, is unobservable but is assumed to be bounded and uncorrelated with the inputs and the past outputs. The model (9) relates the inputs and outputs and takes into account the combined effects of measurement noise, modelling errors and unmeasured disturbances represented by the noise variable $e(t)$. As a general and natural representation for a wide class of linear and nonlinear systems, model (9) includes, as special cases, several model types, including the Volterra and Wiener representations, time-invariant and time-varying AR(X), NARX and ARMA(X) structures, output-affine and rational models, and the bilinear model [Pearson, 1995; 1999].

One of the popular representations for the NARMAX model (9) is the polynomial representation which takes the function $f(\cdot)$ as a polynomial of degree ℓ and gives the form as

$$\begin{aligned} y(t) = \theta_0 &+ \sum_{i_i=1}^{n} f_{i_1}(x_{i_1}(t)) + \sum_{i_1=1}^{n}\sum_{i_2=i_1}^{n} f_{i_1 i_2}(x_{i_1}(t),x_{i_2}(t)) + \cdots \\ &+ \sum_{i_1=1}^{n}\cdots\sum_{i_\ell=i_{\ell-1}}^{n} f_{i_1 i_2\cdots i_\ell}(x_{i_1}(t),x_{i_2}(t),\cdots,x_{i_\ell}(t)) + e(t) \end{aligned} \tag{10}$$

where $\theta_{i_1 i_2\cdots i_m}$ are parameters, $n = n_y + n_u + n_e$ and

$$\begin{aligned} &f_{i_1 i_2\cdots i_m}(x_{i_1}(t),x_{i_2}(t),\cdots,x_{i_m}(t)) = \theta_{i_1 i_2\cdots i_m}\prod_{k=1}^{m} x_{i_k}(t),\ 1\le m\le \ell, \\ &x_k(t) = \begin{cases} y(t-k) & 1\le k\le n_y \\ u(t-(k-n_y)) & n_y+1\le k\le n_y+n_u \\ e(t-(k-n_y-n_u)) & n_y+n_u+1\le k\le n_y+n_u+n_e \end{cases} \end{aligned} \tag{11}$$

The degree of a multivariate polynomial is defined as the highest order among the terms. For example, the degree of the polynomial $h(x_1,x_2,x_3) = a_1x_1^4 + a_2x_2x_3 + a_3x_1^2x_2x_3^2$ is $\ell = 2+1+2=5$. Similarly, a NARMAX model with polynomial degree ℓ means that the order of each term in the model is not higher than ℓ.

The NARX model is a special case of the NARMAX model and takes the form

$$y(t) = f(y(t-1),\cdots,y(t-n_y),u(t-1),\cdots,u(t-n_u)) + e(t) \tag{12}$$

Similar to (11), (12) can be described using a polynomial representation with

$$x_k(t) = \begin{cases} y(t-k), & 1 \le k \le n_y \\ u(t-k+n_y), & n_y+1 \le k \le n = n_y + n_u \end{cases} \tag{13}$$

2.3 The wavelet-based quasi-ANOVA expansions

Generally, a multivariate nonlinear function can often be decomposed into a superposition of a number of functional components similar to the well known functional analysis of variance (ANOVA) expansions as below

$$\begin{aligned} y(t) &= f(x_1(t), x_2(t), \cdots, x_n(t)) \\ &= f_0 + \sum_{i=1}^{n} f_i(x_i(t)) + \sum_{1 \le i < j \le n} f_{ij}(x_i(t), x_j(t)) + \sum_{1 \le i < j < k \le n} f_{ijk}(x_i, x_j, x_k) + \cdots \\ &\quad + \sum_{1 \le i_1 < \cdots < i_m \le n} f_{i_1 i_2 \cdots i_m}(x_{i_1}(t), x_{i_2}(t), \cdots, x_{i_m}(t)) + \cdots + f_{12\cdots n}(x_1(t), x_2(t), \cdots, x_n(t)) + e(t) \end{aligned} \tag{14}$$

where the first functional component f_0 is a constant to indicate the intrinsic varying trend; f_i, f_{ij}, $\cdots$, are univariate, bivariate, etc., functional components. The univariate functional components $f_i(x_i)$ represent the independent contribution to the system output that arises from the action of the ith variable x_i alone; the bivariate functional components $f_{ij}(x_i, x_j)$ represent the interacting contribution to the system output from the input variables x_i and x_j, etc. Let $x_k(t)$ (k=1,2,...,n) be defined as (11) or (13), the quasi-ANOVA expansion (14) can then be viewed as a special form of the NARMAX or NARX models for dynamic input and output systems.

The expansion (14) can be referred to as the quasi-ANOVA decomposition of the NARAMX or NARX models. Although the quasi-ANOVA expansion (14) involves up to 2^n different functional components, experience shows that a truncated representation containing the components up to the bivariate functional terms is often sufficient

$$y(t) = f_0 + \sum_{p=1}^{n} f_p(x_p(t)) + \sum_{p=1}^{n} \sum_{q=p+1}^{n} f_{pq}(x_p(t), x_q(t)) + e(t) \tag{15}$$

This can often provide a satisfactory description of $y(t)$ for many high dimensional problems providing that the input variables are properly selected. The presence of only low order functional components does not necessarily imply that the high order variable interactions are not significant, nor does it mean the nature of the nonlinearity of the system is less severe. An exhaustive search for all the possible submodel structures of (14) is demanding and can be prohibitive because of the curse-of-dimensionality. A truncated representation is advantageous and practical if the higher order terms can be ignored. In practice, the constant term f_0 can often be omitted since it can be combined into other functional components.

In practice, many types of functions, such as kernel functions, splines, polynomials and other basis functions can be chosen to express the functional components in model (14) and (15). In the present study, however, mutiresolution wavelet decompositions will be chosen to describe the functional components. The functional

components $f_p(x_p(t))$ (p=1,2,…,n) and $f_{pq}(x_p(t),x_q(t))$ ($1 \le p < q \le n$) can be expressed using the multiresolution wavelet decompositions (1) and (4) or (8) as

$$f_p(x_p(t)) = \sum_k \alpha^{(p)}_{j_1,k}\phi_{j_1,k}(x_p(t)) + \sum_{j\ge j_1}\sum_k \beta^{(p)}_{j,k}\varphi_{j,k}(x_p(t)),\ p=1,2,\cdots,n, \tag{16}$$

$$\begin{aligned} f_{pq}(x_p(t),x_q(t)) &= \sum_{k_1}\sum_{k_2}\alpha^{(pq)(1)}_{j_2;k_1,k_2}\phi_{j_2,k_1}(x_p(t))\phi_{j_2,k_2}(x_q(t)) \\ &+ \sum_{j\ge j_2}\sum_{k_1}\sum_{k_2}\beta^{(pq)(1)}_{j;k_1,k_2}\phi_{j,k_1}(x_p(t))\varphi_{j,k_2}(x_q(t)) \\ &+ \sum_{j\ge j_2}\sum_{k_1}\sum_{k_2}\beta^{(pq)(2)}_{j;k_1,k_2}\varphi_{j,k_1}(x_p(t))\phi_{j,k_2}(x_q(t)) \\ &+ \sum_{j\ge j_2}\sum_{k_1}\sum_{k_2}\beta^{(pq)(3)}_{j;k_1,k_2}\varphi_{j,k_1}(x_p(t))\varphi_{j,k_2}(x_q(t)),\ 1\le p<q\le n. \end{aligned} \tag{17}$$

$$f_{pq}(x_p(t),x_q(t)) = \sum_{k_1}\sum_{k_2}\alpha_{J;k_1,k_2}\phi_{J,k_1}(x_p(t))\phi_{J,k_2}(x_q(t)) \tag{18}$$

2.4 The wavelet-NARX model

Consider the NARX model (12) and assume that the nonlinear mapping f can be decomposed into a number of functional components up to the rth-variate functional terms using the quasi-ANOVA expansion (14), then the NARX model (14) can be expressed as

$$\begin{aligned} y(t) &= f(x_1(t),x_2(t),\cdots,x_n(t)) + e(t) \\ &= F_1(x(t)) + F_2(x(t)) + \cdots + F_r(x(t)) + e(t) \end{aligned} \tag{19}$$

where $x(t) = [x_1(t),x_2(t),\cdots,x_n(t)]^T$ and

$$x_k(t) = \begin{cases} y(t-k), & 1\le k\le n_y \\ u(t-k+n_y), & n_y+1\le k\le n = n_y+n_u \end{cases} \tag{19a}$$

$$F_1(x(t)) = \sum_{i=1}^{n} f_i(x_i(t)) \tag{19c}$$

$$F_2(x(t)) = \sum_{i=1}^{n}\sum_{j=i+1}^{n} f_{ij}(x_i(t),x_j(t)) \tag{19d}$$

$$F_r(x(t)) = \sum_{1\le i_1<i_2<\cdots<i_r\le n} f_{i_1i_2\cdots i_r}(x_{i_1}(t),x_{i_2}(t),\cdots,x_{i_r}(t)),\quad 2<r\le n, \tag{19e}$$

Each functional component $f_{i_1i_2\cdots i_r}(x_{i_1}(t),x_{i_2}(t),\cdots,x_{i_r}(t))$ can be approximated using the multiresolution wavelet decompositions (1), (4) or (8). The multiresolution wavelet model (19) will be referred to as the wavelet-NARX model, or WANARX [Wei & Billings, 2003], and will be used for identifying and reconstructing chaotic systems. Some implementation issues on mutiresolution wavelet models such as data normalization, highest resolution level determination, translation parameter selection and wavelet dictionary determination are discussed in detail in [Billings & Wei, 2003; Wei & Billing, 2003].

Assume that M bases (scalar mother wavelet or scaling functions or multiplication of some scalar wavelet and scaling functions) are required to expand the NARX model (19), and for convenience of representation also

assume that the M wavelet bases are ordered according to a single index m, that is, the wavelet dictionary $D = \{p_m\}_{m=1}^{M}$, then (19) can be expressed as a linear-in-the-parameters form as below:

$$y(t) = \sum_{m=1}^{M} \theta_m p_m(t) + e(t) \tag{20}$$

which can be solved using linear regression techniques. Note that for large n_y and n_u, the model (20) might involve a great number of model terms or regressors. Experience shows that often many of the model terms are redundant and therefore are insignificant to the system output and can be removed from the model. An efficient algorithm is required to determine which terms should be included in the model. The significant model term selection problem is discussed in the next section.

3. Model Term Selection

The selection of which terms should be included in the model is vital if a parsimonious representation of the system is to be identified. For a selected basic wavelet and associated scaling function, once the initial resolution scale level is given, simply increasing the orders n_y and n_u of the dynamic terms and the highest resolutions in the multiresolution wavelet model will in general result in an excessively over parameterised complex model. Fortunately, experience has shown that only a small subset of these model terms is significant and the remainder can be discarded with little deterioration in prediction accuracy. Several possible algorithms can be used to determine which terms are significant and should be included in the model, including the orthogonal least squares (OLS) algorithm [Billings et al. 1988, 1989; Korenberg et al. 1988; Chen et al. 1989] and the matching pursuit method [Mallat & Zhang, 1993].

Consider the regression equation (20). Define

$$P^{(m)} = \{p_{i_k} : 1 \le i_k \le M;\ \ k = 1,2,\cdots,m\},\ \ m=1,2,\ldots,M, \tag{21}$$

The model term selection procedure is in fact an iterative process which searches through a nested term set in the sense that

$$P^{(1)} \subset P^{(2)} \subset \cdots \subset P^{(m)} \subset \cdots \tag{22}$$

This makes both the complexity and the accuracy of the representation based on these term sets increase until a suitable term set is found, that is, there exists an integer M_0 (generally $M_0 << M$), such that the model

$$y(t) = \sum_{k=1}^{M_0} \theta_{i_k} p_{i_k}(t) + e(t) \tag{23}$$

provides a satisfactory representation for the dynamics under test.

3.1 The forward orthogonal least squares (OLS) algorithm

A fast and efficient model structure determination approach can be implemented using the forward orthogonal least squares (OLS) algorithm and the error reduction ratio (ERR) criterion, which was originally introduced to determine which terms should be included in nonlinear models [Billings et al., 1988, 1989; Korenberg et al., 1988; Chen et al., 1989]. This approach has been extensively studied and widely applied in nonlinear system

identification, see, for example, the papers by Chen et al. [1991], Wang & Mendel [1992], Zhu & Billings [1996], Zhang [1997], Hong & Harris [2001] and the references therein. The forward OLS algorithm involves a stepwise orthogonalization of the regressors and a forward selection of the relevant terms in (20) based on the error reduction ratio (ERR) [Billings et al. 1988, 1989]. The procedure can be briefly summarised as follows.

A compact matrix form corresponding to (20) is

$$Y = P\Theta + \Xi \tag{24}$$

where $Y=[y(1),y(2),\cdots,y(N)]^T$, $P=[p_1,p_2,\cdots,p_M]$, $p_i=[p_i(1),p_i(2),\cdots,p_i(N)]^T$, $\Theta=[\theta_1,\theta_2,\cdots,\theta_M]^T$, $\Xi=[e(1),e(2),\cdots,e(N)]^T$. Assume that the regression matrix P can be orthogonally decomposed as

$$P = WA \tag{25}$$

where A is an $M\times M$ unit upper triangular matrix and W is an $N\times M$ matrix with orthogonal columns $w_1,w_2,\cdots,w_M$ in the sense that $W^TW=D=diag[d_1,d_2,\cdots,d_M]$. Equation (24) can be expressed as

$$Y = (PA^{-1})(A\Theta) + \Xi = WG + \Xi \tag{26}$$

where $G=[g_1,g_2,\cdots,g_M]^T$ is an auxiliary parameter vector, which can be calculated directly from Y and W by means of the property of orthogonality as

$$g_i = \frac{Y^T w_i}{w_i^T w_i}, \qquad i=1,2,\cdots,M \tag{27}$$

The parameter vector Θ is related to G by the equation $A\Theta = G$, and this can be solved using either a classical or modified Gram-Schmidt algorithm [Chen et al. 1989].

The number M of all the candidate terms in model (20) is often very large. Some of these terms may be redundant and should be removed to give a parsimonious model with only M_0 terms ($M_0 << M$). Assume that the residual signal $e(t)$ in the model (20) is uncorrelated with the past outputs of the system, then the output variance can be expressed as

$$\frac{1}{N}Y^TY = \frac{1}{N}\sum_{i=1}^{M} g_i^2 w_i^T w_i + \frac{1}{N}\Xi^T\Xi \tag{28}$$

Note that the output variance consists of two parts, one is the desired output, $(1/N)\sum_{i=1}^{M} g_i^2 w_i^T w_i$, which can be explained by the regressors, and the other part, $(1/N)\Xi^T\Xi$, represents the unexplained variance. Thus $(1/N)\sum_{i=1}^{M} g_i^2 w_i^T w_i$ is the increment to the explained desired output variance brought by w_i, and the ith error reduction ratio, ERR_i, introduced by w_i, can be defined as

$$ERR_i = \frac{g_i^2 w_i^T w_i}{Y^TY}\times 100\% = \frac{(Y^T w_i)^2}{(Y^TY)(w_i^T w_i)}\times 100\%, \qquad i=1,2,\cdots,M, \tag{29}$$

This ratio provides a simple but effective means for seeking a subset of significant regressors. The significant terms can be selected in a forward-regression manner according to the value of ERR_i. Several orthogonalization procedures, such as Gram-Schmidt, modified Gram-Schmidt and Householder transformation [Chen et al., 1989] can be applied to implement the orthogonal decomposition. See [Billings et al., 1988, 1989; Korenberg et al., 1988; Chen et al., 1989] for details.

3.2 Matching pursuit orthogonal least squares (MPOLS) algorithm

Note that in the forward Gram-Schmidt OLS algorithm, at each step all the unselected regressors are made to orthogonalize with the previously selected regressors, and most of the computational cost is based on these orthogonalization transforms. An iterated orthogonal projection algorithm, the matching pursuit method, proposed by Mallat and Zhang [1993] is a simple regressor selection algorithm which is relatively computationally efficient. But the matching pursuit algorithm is less efficient than OLS, since the number of regressors selected by the matching pursuit algorithm is almost always larger than that selected by OLS for the same given threshold value of approximation accuracy. A trade-off between the efficiency and the computational cost is considered here by combining the advantages of the forward OLS with the matching pursuit algorithm to create a new algorithm called the matching pursuit orthogonal least squares (MPOLS) algorithm [Billings & Wei, 2003]. The algorithm is described below.

For the output vector $Y=[y(1),y(2),\cdots,y(N)]^T$ in (20) or (24), find a vector p_{ℓ_1} from the candidate regressor family $\{p_1,p_2,\cdots,p_M\}$, so that p_{ℓ_1} is the "best" matching regressor to Y, i.e., p_{ℓ_1} makes the mean squared error of the following linear regression

$$y(t)=c_m p_m(t)+\xi_m(t) \tag{30}$$

achieve a minimum in the sense that

$$\frac{1}{N}\sum_{t=1}^{N}\xi_{\ell_1}^2(t)=\frac{1}{N}\sum_{t=1}^{N}(y(t)-c_{\ell_1}p_{\ell_1}(t))^2=\min_m\left\{\frac{1}{N}\sum_{t=1}^{N}[y(t)-c_m p_m(t)]^2\right\} \tag{31}$$

The "best" matching regressor p_{ℓ_1} can be found by means of a geometrical approach, see Figure 1. From Figure 1,

$$\cos\alpha=\frac{Y^T p_m}{\sqrt{Y^TY}\sqrt{p_m^T p_m}} \tag{32}$$

$$\left\|p_m^1\right\|=\|Y\|\cos\alpha=\frac{Y^T p_m}{\sqrt{p_m^T p_m}} \tag{33}$$

Thus

$$\sum_{t=1}^{N}\xi_m^2(t)=\|\xi_m\|^2=\|Y\|^2-\left\|p_m^1\right\|^2=Y^TY-\frac{(Y^T p_m)^2}{p_m^T p_m} \tag{34}$$

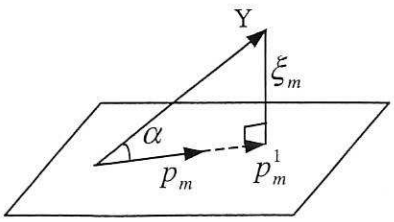


Figure 1 Diagram of the least squares algorithm

Therefore,

$$\ell_1 = \arg\max_m \left\{ \frac{(Y^T p_m)^2}{p_m^T p_m}, 1 \le m \le M \right\} \tag{35}$$

Set $q_1(t) = p_{\ell_1}(t)$, $w_1(t) = q_1(t)$, $g_1 = (Y^T w_1)/(w_1^T w_1)$, $ERR_1 = g_1^2 (w_1^T w_1)/(Y^T Y)$, and $\eta_1(t) = y(t) - g_1 w_1(t)$.

At the second step, find a vector p_{ℓ_2} from the candidate regressor family $\{p_m : 1 \le m \le M, m \ne \ell_1\}$, so that p_{ℓ_2} is the "best" matching regrssor to η_1. Following the approach in (31) and (35), ℓ_2 should be chosen as

$$\ell_2 = \arg\max_m \left\{ \frac{(\eta_1^T p_m)^2}{p_m^T p_m}, 1 \le m \le M, m \ne \ell_1 \right\} \tag{36}$$

Set $q_2(t) = p_{\ell_2}(t)$. Orthogonalize q_2 with w_1 as below

$$w_2 = q_2 - \frac{w_1^T q_2}{w_1^T w_1} w_1 \tag{37}$$

And set $g_2 = (Y^T w_2)/(w_2^T w_2)$, $ERR_2 = g_2^2 (w_2^T w_2)/(Y^T Y)$, and $\eta_2(t) = \eta_1(t) - g_2 w_2(t)$.

Generally, at step k, select

$$\ell_k = \arg\max_m \left\{ \frac{(\eta_{k-1}^T p_m)^2}{p_m^T p_m}, 1 \le m \le M, m \ne \ell_1, m \ne \ell_2, \cdots, m \ne \ell_{k-1} \right\} \tag{38}$$

Set $q_k(t) = p_{\ell_k}(t)$ and orthogonalize q_k with $w_1, w_2, \cdots, w_{k-1}$ as below

$$w_k = q_k - \frac{w_1^T q_k}{w_1^T w_1} w_1 - \frac{w_2^T q_k}{w_2^T w_2} w_2 - \cdots - \frac{w_{k-1}^T q_k}{w_{k-1}^T w_{k-1}} w_{k-1} \tag{39}$$

Calculate $g_k = (Y^T w_k)/(w_k^T w_k)$, $ERR_k = g_k^2 (w_k^T w_k)/(Y^T Y)$, and set $\eta_k(t) = \eta_{k-1}(t) - g_k w_k(t)$.

A similar algorithm has been used for basis selection in wavelet neural networks [Xu, 2002]. Note that in the MPOLS algorithm, only the most recently selected regressor $q_j = p_{\ell_j}$ at step j is made to be orthogonal with the previous selected regressors $q_k = p_{\ell_k}$ (k=1,2,...,j-1). Therefore, the computational load of the orthogonalization procedure in OLS, which involves making all the unselected regressors orthogonal with the previously selected regressors, is significantly reduced in the MPOLS algorithm. The computational cost of the MPOLS algorithm is much less than that of the OLS algorithm, and this makes the MPOLS algorithm much faster than most existing OLS and fast OLS algorithms. Also notice that, for the same problem, MPOLS may select different model terms (regressors) and different numbers of model terms compared with OLS even for the same threshold value of termination. It is nearly always true that the MPOLS selects more model terms than that of OLS. The comparison of computational efficiency and the performance efficiency of the MPOLS algorithm compared with OLS will be illustrated via three examples provided in the next section.

4. Numerical Examples

In this section, three numerical examples are given to demonstrate the modelling procedure proposed. The comparisons of the performance and computational efficiency of the OLS and MPOLS algorithms are also illustrated.

4.1 Example 1: Modelling the Henon map

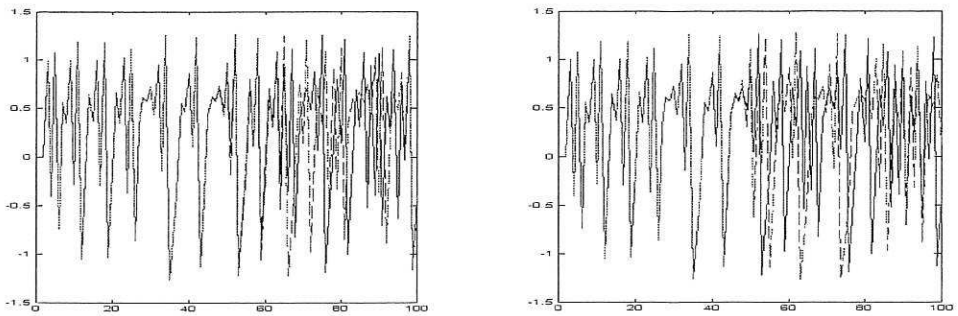
The Henon map [Henon, 1976] is described by the equation

$$\begin{cases} x(t) = 1 - ax^2(t-1) + y(t-1) \\ y(t) = bx(t-1) \end{cases} \tag{40}$$

where a=1.4, b=0.3. This is a popular example of two-dimensional maps, which is extremely sensitive with respect to not only the initial conditions but also the parameters a and b, see Fig. 2. These properties make it difficult to obtain long term predictions. The objective here is to identify the system using a multiresolution wavelet model without any prior knowledge about the system but only with an observational data set of $x(t)$.

Three thousand data points were simulated using the Henon map (40) and noise with a variance of σ_e^2 =0.0001was added to the data. Denote by $\tilde{x}(t)$ (t=1,2,...,3000) the simulated data with the additive noise. Only the first 250 points, which were taken from the time series after transients had died out, were used for model estimation after normalization into [0, 1]. Denote by $x(t)$(t=1,2,...,250) the 250 normalized data points.

The initial wavelet model for the Henon map was chosen as



(a) (b)

Fig. 2 Sensitivity dependence of the Henon map on (a) initial conditions (solid line is for $x(0)$=0, $y(0)$=0 and dashed line is for $x(0)$=0, $y(0)$=10^{-10}) and (b) parameters (solid line is for a=1.4, b=0.3 and dashed line is for a=1.4, b=0.3+10^{-10}).

$$x(t) = f(x(t-1), x(t-2), \cdots, x(t-5)) + e(t)$$

$$= \sum_{p=1}^{5} f_p(x(t-p)) + \sum_{p=1}^{4} \sum_{q=p+1}^{5} f_{pq}(x(t-p), x(t-q)) + e(t) \tag{41}$$

where each univariate function component $f_p(x(t-p))$ was expanded using the wavelet decompositions (16) with the starting resolution scale $j_1 = 0$ and the highest resolution scale $j_{\max} = 4$, and each bi-variate function component $f_{pq}(x(t-p), x(t-q))$ was expanded using the wavelet decompositions (18) with the resolution scale J=2. The 4th order B-spline wavelet and scaling functions [Chui & Wang, 1992] were used in these wavelet decompositions. After the initial model (41) was expanded using these wavelet decompositions, 815 candidate model terms (basis functions) were involved. Both the OLS and MPOLS algorithms were used to select the significant model terms, and finally two parsimonious models were obtained with the form

$$x(t) = \sum_{k=1}^{M_0} \theta_k p_k(t) + e(t) \tag{42}$$

where the basis functions $p_k(t)$ and the parameters θ_k identified using both the OLS and MPOLS algorithms are listed in Table 1 and Table 2, respectively. In Table 1 and Table 2, $\varphi_{j,k}(x) = 2^{j/2}\varphi(2^j x - k)$ and $\phi_{j,k}(x) = 2^{j/2}\phi(2^j x - k)$ are the 4th-order B-splne mother wavelet and scaling functions [Chui 1992]. Fig. 3 shows the attractor (the first return map) computed from the noisy data set, the original Henon map, and the attractors recovered from the OLS and MPOLS identified models. From Table 1 and Table 2 , it is clear that to achieve the same approximation accuracy (corresponding to the shreshold value ρ for termination), MPOLS selects many basis functions (19 terms) using much less time (0.99 sec) than the OLS algorithm (8 model terms were selected over 3.91sec). The variables selected by OLS are $x(t\text{-}1)$ and $x(t\text{-}2)$, these are the primary dependent variables for the original system, clearly OLS has selected these correctly and has discarded other redundant variables. The variables selected by MPOLS are $x(t\text{-}1)$, $x(t\text{-}2)$, $x(t\text{-}3)$, $x(t\text{-}4)$, $x(t\text{-}5)$, and this means that the model is dimension over-parameterized. Fig. 3 clearly shows that the OLS identified model is more adequate than the MPOLS identified model. It should be pointed out that further improvement can be achieved by using more data points and optimizing the threshold value for termination of the basis selection algorithms, but the objective in this example was to show that the Henon map can be estimated fairly accurately from a very short noise corrupted time series.

Table 1 Basis functions, parameters and the corresponding ERRs estimated using OLS for the Henon map (40)

Number k	Terms $p_k(t)$	Parameters θ_k	$ERR_k \times 100\%$
1	$\phi_{0,-1}(x(t-2))$	8.69648098e-001	7.91252048e+001
2	$\varphi_{0,-2}(x(t-1))$	-1.10046556e+000	1.92972751e+001
3	$\phi_{0,-3}(x(t-2))$	1.19680558e+000	8.35830889e-001
4	$\varphi_{1,-1}(x(t-1))$	2.84914870e-003	4.42820859e-001
5	$\varphi_{0,-4}(x(t-1))$	1.76493082e+001	1.54667941e-001
6	$\varphi_{0,0}(x(t-1))$	1.82935775e+001	1.11990690e-001
7	$\varphi_{0,-4}(x(t-2))$	3.38424997e+000	3.06518989e-002
8	$\varphi_{1,-1}(x(t-2))$	1.10947761e-002	1.44695585e-003

Note: 1) The threshold value $\rho = 10^{-5}$, $\sum_{k=1}^{8} ERR_k \geq 1-\rho$.

2) The CPU time spent on selecting these model terms from all the candidate model term set is 3.91s.

Table 2 Basis functions, parameters and the corresponding ERRs estimated using MPOLS for the Henon map (40)

Number k	Terms $p_k(t)$	Parameters θ_k	$ERR_k \times 100\%$
1	$\phi_{0,-1}(x(t-2))$	8.82406739e-001	7.91252048e+001
2	$\varphi_{0,-2}(x(t-1))$	-1.67466536e+000	1.92972751e+001
3	$\phi_{0,-3}(x(t-2))$	1.29634040e+000	8.35830889e-001
4	$\varphi_{1,1}(x(t-1))$	5.94558765e+000	4.34775993e-001
5	$\varphi_{3,5}(x(t-2))$	-3.69937585e-002	1.63703393e-001
6	$\varphi_{2,1}(x(t-1))$	2.28567090e-002	6.02958808e-002
7	$\varphi_{4,-3}(x(t-1))$	-2.53436201e-002	3.77007050e-002
8	$\varphi_{4,13}(x(t-3))$	9.78029037e-003	2.10365316e-002
9	$\phi_{2,-1}(x(t-2))\phi_{2,-2}(x(t-3))$	2.52914399e-002	9.42034816e-003
10	$\varphi_{2,0}(x(t-5))$	8.63856183e-003	4.76126788e-003
11	$\varphi_{4,-2}(x(t-1))$	4.49091634e-003	1.87852057e-003
12	$\varphi_{1,-4}(x(t-3))$	7.30993491e-001	2.18241588e-003
13	$\varphi_{3,-1}(x(t-5))$	-3.79557000e-003	1.21101178e-003
14	$\varphi_{2,0}(x(t-1))$	9.83784306e-003	1.32754274e-003
15	$\varphi_{4,11}(x(t-3))$	1.83792597e-003	5.40838336e-004
16	$\varphi_{2,-2}(x(t-2))$	1.02289921e-002	8.55385566e-004
17	$\varphi_{4,11}(x(t-2))$	1.54114623e-003	4.68976872e-004
18	$\phi_{2,-1}(x(t-3))\phi_{2,-1}(x(t-4))$	6.76125684e-003	4.56686214e-004
19	$\varphi_{4,4}(x(t-5))$	-1.86863493e-003	3.09554532e-004

Note: 1) The threshold value $\rho = 10^{-5}$, $\sum_{k=1}^{19} ERR_k \geq 1-\rho$.

2) The CPU time spent on selecting these model terms from all the candidate model term set is 0.99s.

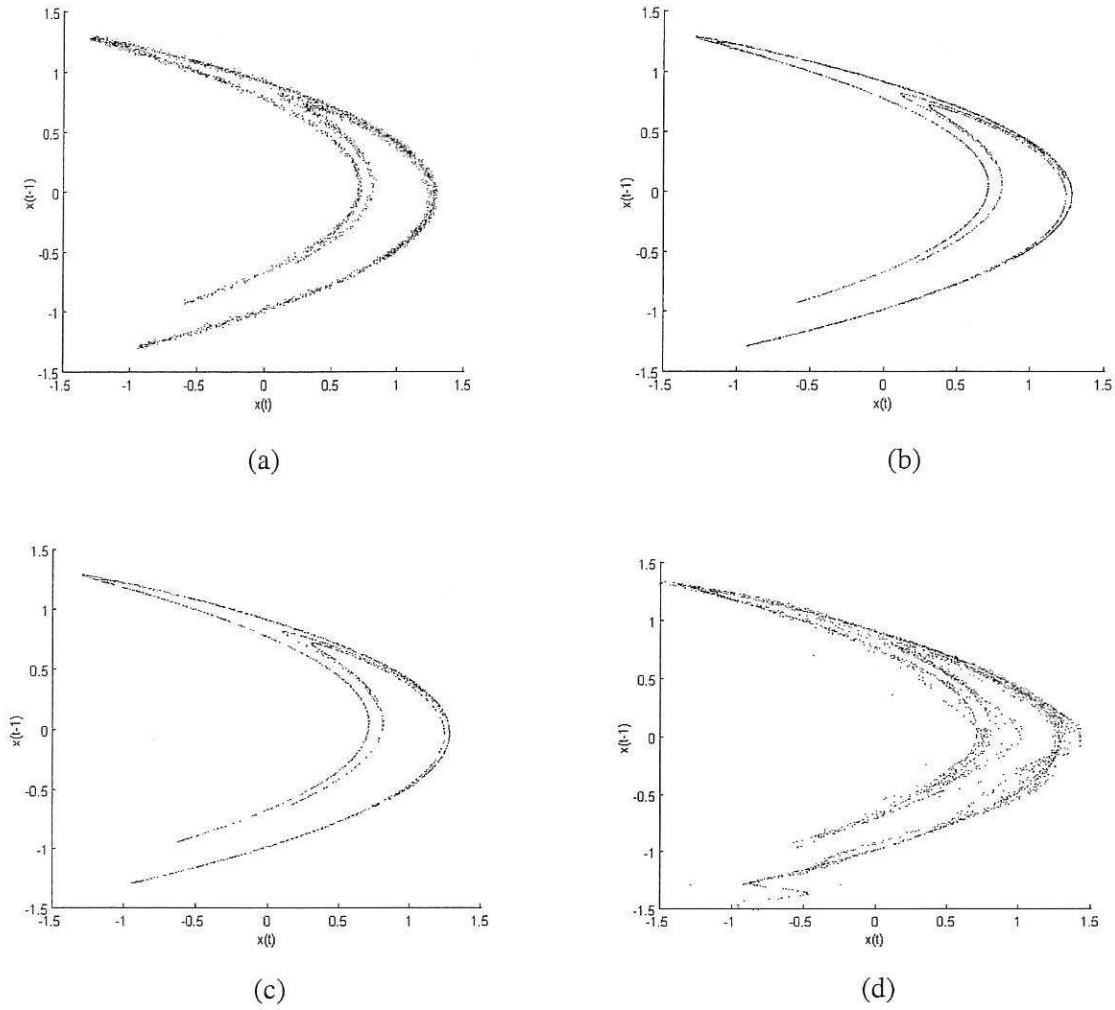


Fig. 3 First return maps for (a) Henon map with additive noise on the measurements, (b) original noise-free Henon map, (c) reconstructed from the OLS identified model and (d) reconstructed from the MPOLS identified model.

4.2 Example 2: Reconstructing the Poincare map for the modified Van der Pol oscillator

Consider the equation governing the dynamics of a forced oscillator with negative resistance [Ueda & Akamatsu, 1981]

$$\ddot{y} + \mu(y^2 - 1)\dot{y} + y^3 = u(t) \tag{43}$$

where μ=0.2, $u(t) = A\cos(\omega t)$ with A=17 and ω=4rad/s. This is known as the modified Van der Pol oscillator [Moon, 1987], which settles to a strange attractor. This system was simulated and the data were sampled at $T_s = \pi / 80$. Noise with $\sigma_e^2 = 0.001$ was added to 628 simulated data points, which were normalized to [0, 1] before model estimation. The normalized data points will be denoted by $y(t)$ (t=1,2,…,628) and were used for model estimation.

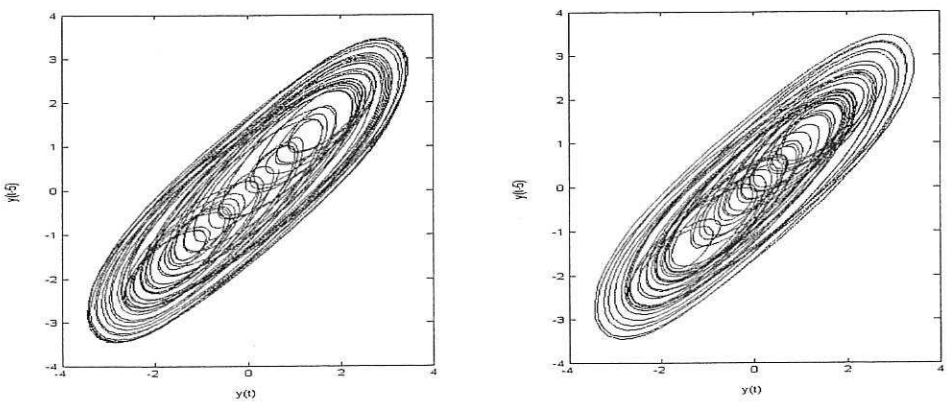
The initial wavelet model for the modified Van der Pol oscillator was chosen as

$$y(t) = f(x(t-1), x(t-2), \cdots, x(t-10)) + e(t)$$

$$= \sum_{p=1}^{10} f_p(x(t-p)) + \sum_{p=1}^{4}\sum_{q=p+1}^{5} f_{pq}(x(t-p), x(t-q))$$

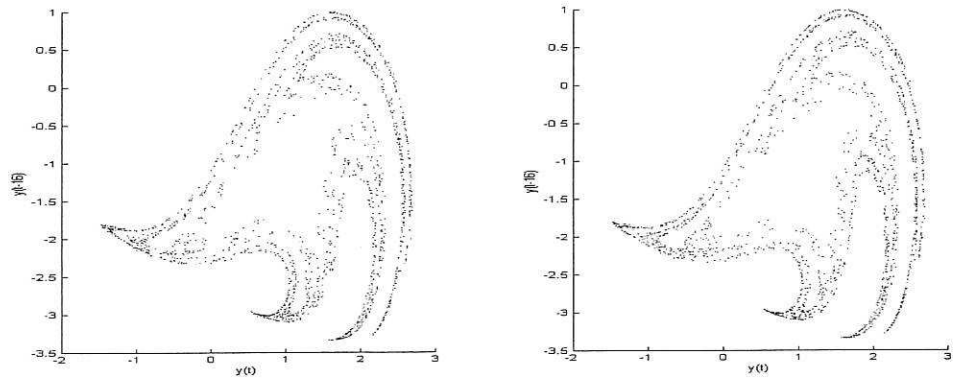
$$+ f_{123}(x(t-1), x(t-2), x(t-3)) + e(t) \tag{44}$$

where $x_i(t) = y(t-i)$ for i=1,2,3,4,5 and $x_i(t) = u(t-i+5)$ for i=6,7,8,9,10. Each univariate function component $f_p(x(t-p))$ was expanded using the wavelet decompositions (16) with the starting resolution scale $j_1 = 0$ and the highest resolution scale $j_{\max} = 4$, and each bi-variate function component $f_{pq}(x(t-p), x(t-q))$ was expanded using the wavelet decompositions (18) with the resolution scale J=2. The tri-variate function component $f_{123}(x(t-1), x(t-2), x(t-3))$ was expanded using the wavelet decomposition (8) with J=1. The 4th order B-spline wavelet and scaling functions [Chui & Wang, 1992] were used in these wavelet decompositions. After the initial model (44) was expanded using these wavelet decompositions, 1265 candidate model terms (basis functions) were involved. Both the OLS and MPOLS algorithms were used to select the significant terms, and finally two parsimonious models were obtained with the form of (42). The basis functions $p_k(t)$ and the parameters θ_k identified using the OLS algorithm are listed in Table 3. The attractor and Poincare section calculated from the OLS identified wavelet model, which is listed in Table 3, are compared with that of the noise-free original modified Van der Pol oscillator listed in Figs. 4 and 5. The results clearly show that the OLS identified wavelet model is good representation of the system which accurately reproduced the dynamics of the Van der Pol oscillator. The MPOLS algorithm was also used with the same threshold value ρ for termination and 76 basis functions were selected over 99.86s. These basis functions are not listed here due to the large number of selected terms. The MPOLS identified model could not reproduce the attractor and the Poincare map of the Van der Pol oscillator correctly. In fact, even though the short term predictions of the MPOLS model were very good, the model predicted outputs (the model free-run behaviour) were very bad and oscillated with an amplitude that was greatly different from the original Van der Pol oscillator and tended to diverge.



(a) (b)

Fig. 4 Attractors for (a) the original, noise-free modified Van der Pol oscillator (43) and (b) for the OLS identified model. $u(t) = A\cos(\omega t)$, A=17, ω=4rad/s and μ=0.2.



(a) (b)

Fig. 5 Poincare sections for (a) the original system (43) and (b) obtained from the OLS identified model.

$u(t) = A\cos(\omega t)$, A=17, ω =4rad/s and μ =0.2

Table 3 Basis functions, parameters and the corresponding ERRs estimated using OLS for the modified Van der Pol oscillator (43)

Number k	Terms $p_k(t)$	Parameters θ_k	$ERR_k \times 100\%$
1	$\phi_{0,-1}(y(t-1))$	1.35908173e+000	9.69044287e+001
2	$\phi_{0,-2}(y(t-1))$	-7.39199745e-001	2.63270355e+000
3	$\phi_{0,-2}(y(t-5))$	4.14284625e-001	3.38837349e-001
4	$\phi_{1,1}(y(t-1))\phi_{1,0}(y(t-2))\phi_{1,0}(y(t-3))$	5.22584995e-001	9.91985707e-002
5	$\phi_{1,0}(y(t-1))\phi_{1,-3}(y(t-2))\phi_{1,-3}(y(t-3))$	9.13656389e+001	1.37420686e-002
6	$\phi_{0,-2}(u(t-4))$	-5.33318600e-004	4.79798321e-003
7	$\phi_{2,1}(y(t-1))\phi_{2,0}(y(t-2))$	-1.17233604e-002	3.06706331e-003
8	$\phi_{2,-3}(y(t-2))\phi_{2,0}(y(t-5))$	-2.11908045e-002	7.32419263e-004
9	$\phi_{0,-1}(u(t-1))$	3.28552255e-002	1.13258867e-003
10	$\phi_{2,1}(y(t-2))\phi_{2,-1}(y(t-3))$	-4.40680195e-002	3.16639753e-004
11	$\varphi_{2,-4}(y(t-1))$	3.37224894e-002	2.45074831e-004
12	$\phi_{2,-2}(y(t-1))\phi_{2,-1}(y(t-4))$	-1.69085298e-002	2.40023352e-004
13	$\phi_{2,2}(y(t-2))\phi_{2,1}(y(t-3))$	1.00797779e-002	1.07146529e-004
14	$\phi_{2,1}(y(t-1))\phi_{2,-1}(y(t-5))$	1.09288211e-002	7.81006703e-005
15	$\varphi_{2,-4}(y(t-5))$	1.38377466e-003	7.06243831e-005
16	$\varphi_{0,0}(u(t-2))$	-1.12423798e-001	5.62692956e-005
17	$\phi_{2,1}(y(t-1))\phi_{2,2}(y(t-3))$	-7.35190097e-003	3.73003313e-005
18	$\phi_{1,-1}(y(t-1))\phi_{1,-1}(y(t-2))\phi_{1,-3}(y(t-3))$	6.96406971e-001	3.60984859e-005
19	$\varphi_{1,-4}(y(t-1))$	5.33474487e-001	4.97655607e-005
20	$\phi_{2,-3}(y(t-3))\phi_{2,-3}(y(t-5))$	1.01546654e-001	3.46993291e-005

Note: 1) $\varphi_{j,k}(x) = 2^{j/2}\varphi(2^j x - k)$ and $\phi_{j,k}(x) = 2^{j/2}\phi(2^j x - k)$ — the 4th-order B-splne wavelets and scaling functions;

2) The threshold value $\rho = 10^{-6}$, $\sum_{k=1}^{20} ERR_k \geq 1-\rho$.

3) The CPU time spent on selecting these model terms from all the candidate model term set is 490. 92s.

4) The MPOLS algorithm was also used with the same threshold value ρ for termination and 76 basis functions over 99.86s.

4.3 Example 3: Extracting the CA rule from a cellular automaton pattern

Cellular automata (CA) are among the simplest mathematical representations of complex systems [Ilachinski, 2001], and are characterized by states that exist only on lattices of discrete sites (cells). Each site has a finite number of possible states and site evolutions are determined by either a deterministic or a random rule, which involves some finite neighbourhood on the lattice and a finite number of previous time steps. Two types of problems, the forward problem and the inverse problem, are generally studied in cellular automata [Gutowitz, 1990]. Although the forward problem, which involves determining the natural property of a set of given rules governing the cellular automata, has been extensively studied, the inverse problem, which is concerned with finding a rule, or a set of rules, and quantitatively reproducing a given set of prescribed patterns, has received relatively little attention and relatively few results have been achieved. Modelling and identification plays a key role for solving the inverse problem in cellular automata, but few studies have been done on CA identification [Adamatzky, 1994; Billings & Yang, 2003].

Extensive empirical evidence suggests that patterns generated by all one-dimensional CA evolving from disordered initial states can be catalogued into 4 classes [Wolfram, 1983]. The CA rules in class 3 are particularly attractive since in this class a particular site value depends on the values of an ever-increasing number of initial sites and random initial values then lead to chaotic behaviour representing the essence of self—organization in cellular automata.

The objective here is to identify a rule which governs a cellular automata based on partial observations of a given CA pattern using the multiresolution wavelet model proposed in Sec. 2. Taking a one-dimensional CA as an example, this can be done by choosing the initial model as

$$s_j(t) = f(s_{j-p}(t-1), s_{j-p+1}(t-1), \cdots, s_{j-1}(t-1), s_j(t-1), s_{j+1}(t-1), \cdots, s_{j+q}(t-1)) \qquad (45)$$

where $s_j(t)$ indicates the jth cell at time t, $r_0 = \max(p,q) \geq r$, and r is the neighbourhood radius. Model (45) can be expressed using the multiresolution representation (19) and then OLS or MPOLS algorithms can be applied to obtain a parsimonious model, which can be used as the rule to reconstruct the CA pattern. The property of the Haar wavelet and scaling functions makes the multiresolution Haar wavelet models appropriate to represent binary cellular automata.

Rule22 in class 3 will be used as an example here. The 3-site one-dimensional rule, with von Neumann neighbourhood, is shown in Table 4 and the pattern of 200×200-site lattices generated by this rule with a set of random initial values is shown in Fig. 6(a), where periodic boundary conditions were considered so that the first site is a nearest neighbour of the n-th site in an n-site automata.

Table 4 The one dimensional three-site CA Rule 22--the local states and the updated central site values

$s_{j-1}(t-1)s_j(t-1)s_{j+1}(t-1)$	000	001	010	011	100	101	110	111
$s_j(t)$	0	1	1	0	1	0	0	0

To identify the rule, the multiresolution wavelet model was initially chosen as

$$s_j(t) = f(s_{j-3}(t-1), s_{j-2}(t-1), s_{j-1}(t-1), s_j(t-1), s_{j+1}(t-1), s_{j+2}(t-1), s_{j+3}(t-1))$$

$$
\begin{aligned}
&= f(x_1(t), x_2(t), x_2(t), x_3(t), x_4(t), x_5(t), x_6(t), x_7(t)) \\
&= \sum_{i=1}^{7} f_i(x_i(t)) + \sum_{i_1=1}^{6}\sum_{i_2=i_1+1}^{7} f_{i_1 i_2}(x_{i_1}(t), x_{i_2}(t)) + \sum_{i_1=1}^{5}\sum_{i_2=i_1+1}^{6}\sum_{i_3=i_2+1}^{7} f_{i_1 i_2 i_3}(x_{i_1}(t), x_{i_2}(t), x_{i_3}(t)) \\
&+ \sum_{i_1=1}^{4}\sum_{i_2=i_1+1}^{5}\sum_{i_3=i_2+1}^{6}\sum_{i_4=i_3+1}^{7} f_{i_1 i_2 i_3 i_4}(x_{i_1}(t), x_{i_2}(t), x_{i_3}(t), x_{i_4}(t)) \\
&+ \sum_{i_1=1}^{3}\sum_{i_2=i_1+1}^{4}\sum_{i_3=i_2+1}^{5}\sum_{i_4=i_3+1}^{6}\sum_{i_5=i_4+1}^{7} f_{i_1 i_2 i_3 i_4 i_5}(x_{i_1}(t), x_{i_2}(t), \cdots, x_{i_5}(t)) \\
&+ \sum_{i_1=1}^{2}\cdots\sum_{i_6=i_5+1}^{7} f_{i_1 i_2 i_3 i_4 i_5 i_6}(x_{i_1}(t), x_{i_2}(t), \cdots, x_{i_6}(t)) \\
&+ f_{1234567}(x_1(t), x_2(t), \cdots, x_7(t))
\end{aligned} \tag{46}
$$

where $x_k(t) = s_{8-k}(t-1)$. Each univariate function component was decomposed using the Haar wavelet and scaling (the first-order B-spline wavelet and scaling) functions as

$$
f_i(x_i(t)) = a_{0,0}^{(i)}\phi_{0,0}(x_i(t)) + \sum_{j=0}^{5}\sum_{k\in K_j} d_{j,k}^{(i)}\varphi_{j,k}(x_i(t)) \tag{47}
$$

where $K_j = \{0,1,\cdots,2^j - 1\}$. Other function components were decomposed using the Haar scaling functions as

$$
f_{i_1\cdots i_m}(x_{i_1}(t), \cdots, x_{i_m}(t)) = \sum_{k_1=0}^{2^{J_m}-1}\cdots\sum_{k_m=0}^{2^{J_m}-1} a_{J_m;k_1,\cdots,k_m}^{(i_1\cdots i_m)}\prod_{p=1}^{m}\phi_{J_m,k_m}(x_{i_p}(t)) \tag{48}
$$

where J_2=3, J_3=2, J_4=J_5=1=J_6=J_7=1,and $\phi(x)$ is the Haar scaling function defined as

$$
\phi(x) = \begin{cases} 1 & \text{for } 0 \le x < 1 \\ 0 & \text{otherwise} \end{cases} \tag{49}
$$

The model (46) contains 5776 candidate basis functions (regressors) after decompositon into the multiresolution wavelet model. The MPOLS algorithm was used to select the significant model terms and detect the model structure based on 100 data points for the 7 contiguous column of sites $\{ s_{j-3}(t), \cdots, s_j(t), \cdots, s_{j+3}(t) \}$ (t=1,2,...,100). The final identified wavelet model for rule 22 was found to be

$$
\begin{aligned}
s_j^*(t) &= 0.125\phi_{3,0}(s_{j-1}(t-1))\phi_{3,0}(s_j(t-1)) + 0.125\phi_{3,0}(s_{j-1}(t-1))\phi_{3,0}(s_{j+1}(t-1)) \\
&+ 0.125\phi_{3,0}(s_j(t-1))\phi_{3,0}(s_{j+1}(t-1)) - 0.375\phi_{2,0}(s_{j-1}(t-1))\phi_{2,0}(s_j(t-1))\phi_{2,0}(s_{j+1}(t-1)) \\
&= \phi(2^3 s_{j-1}(t-1))\phi(2^3 s_j(t-1)) + \phi(2^3 s_{j-1}(t-1))\phi(2^3 s_{j+1}(t-1)) \\
&+ \phi(2^3 s_j(t-1))\phi(2^3 s_{j+1}(t-1)) - 3\phi(2^2 s_{j-1}(t-1))\phi(2^2 s_j(t-1))\phi(2^2 s_{j+1}(t-1)) \\
&= \phi(s_{j-1}(t-1))\phi(s_j(t-1)) + \phi(s_{j-1}(t-1))\phi(s_{j+1}(t-1)) + \phi(s_j(t-1))\phi(s_{j+1}(t-1)) \\
&- 3\phi(s_{j-1}(t-1))\phi(s_j(t-1))\phi(s_{j+1}(t-1))
\end{aligned} \tag{50a}
$$

where

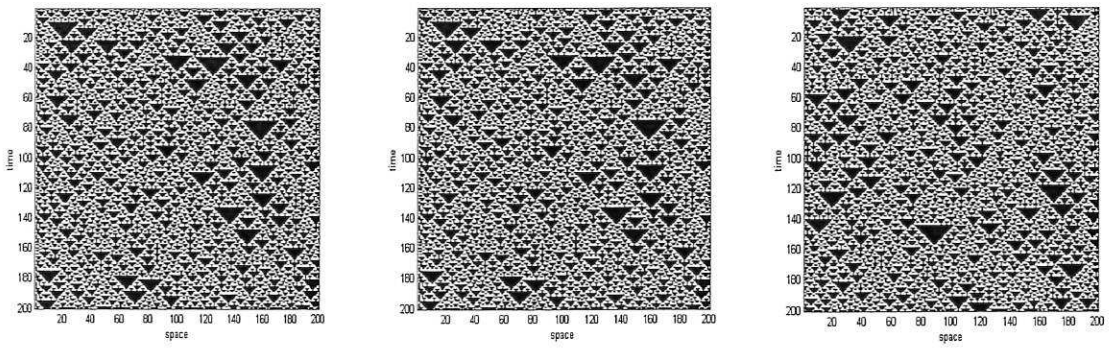
$$
s_j(t) = round(s_j^*(t)) \tag{50b}
$$

and the value of the function *round*(x) is defined as the integer that is nearest to the variable x. Clearly, the CA neighbourhood was correctly selected. The reason that the function (50b) was introduced as a filter was to

eliminate the round-off errors produced by the word-capacity limit of the computer or errors generated by the boundary effects of the wavelet functions.

Notice that only part of the observational data of the CA pattern was used for model estimation. Data points from any 7 contiguous columns of lattices $\{ s_{j-3}(t),\cdots, s_j(t),\cdots, s_{j+3}(t) \}$ (t=1,2,...,100) can be used for identification. No a priori knowledge of the system structured was assumed except that the neighbourhood radius was restricted to be not great than 3. The final model structure identified by OLS was in this example identical to that identified by the MPOLS algorithm. This is believed to be because of the special binary nature of the 1D CA problem. However, the CPU time taken by OLS to detect the model structure was much longer than for the MPOLS algorithm. In this example, the CPU time spent on detecting the model structure using the OLS and MPOLS algorithms was 68.32s and 3.82s, respectively.

The reconstructed pattern of 200×200-site lattice from the MPOLS identified model (50) with the same random initial values as in the pattern of Fig. 6(a) is shown in Fig. 6(b). The pattern generated from the MPOLS identified model (50) with a set of random initial values which are different from that in the pattern of Fig. 6(a) is shown in Fig. 6(c). Simulation results show that the reconstructed pattern in Fig. 6(b) is exactly the same as the original pattern in Fig. 6(a). The pattern in Fig 6(c) seems different from that in Fig. 6(a), but this is because the initial vales in Fig. 6(c) are different from those in Fig. 6(a).



(a) (b) (c)

Fig. 6 CA patterns for (a) the original pattern generated from Rule 22 , (b) reconstructed from the MPOLS identified model (50) and with the same initial values as that in (a), and (c) generated from the model (50) but with different initial values.

5. Conclusions

Multiresolution wavelet models have been successfully applied to identify chaotic systems including a cellular automata example. It was assumed that the structures of the models were unknown. This is equivalent to the situation which occurs in practice where the model of an unknown system is required and only input-output data and very limited information about the system is available. Consequently the initial model search can involve many model terms or regressors. Both the OLS and a new MPOLS algorithm were investigated as model selection and identification procedures to select the most significant model terms using the new multiresolution wavelet models. Qualitative model validation was used to compare the identified models. The results show that

while MPOLS is computationally more efficient compared to OLS, the latter produces parsimonious models with fewer model terms and accurately reproduces the dynamic invariants. The results suggest therefore that MPOLS tends to produce overparameterised models.

Acknowledgment

The authors gratefully acknowledge that part of this work was supported by EPSRC(UK).

References

Abarbanel, H.D.I., Brown, R., and Kadtke, J.B., 1989, Prediction and system identification in chaotic nonlinear systems: time series with broadband Fourier spectra. *Physics Letters* **A138(8)**, 401-408.

Abarbanel, H.D.I., Brown, R., and Kadtke, J.B., 1990, Prediction in chaotic nonlinear systems: methods for time series with broadband Fourier spectra. *Physical Review* **A41(4)**, 1782-1807.

Adamatzky, A., 1994, *Identification of Cellular Automata* (London: Taylors ,and Francis).

Aguirre, L.A., and Billings, S.A., 1994, Validating identified nonlinear models with chaotic dynamics. *International Journal of Bifurcation and Chaos* **5(2)**, 449-474.

Aguirre, L.A., and Billings, S.A., 1995a, Retrieving dynamical invariants from chaotic data using NARMAX Models. *International Journal of Bifurcation and Chaos* **4(1)**, 109-125.

Aguirre, L.A., and Billings, S.A.,1995b, Dynamical effects of overparameterization in nonlinear models. *Physica* **D80(1-2)**, 26-40.

Aguirre, L.A., and Mendes, E.M.A.M., 1996, Global nonlinear polynomial models: structure, term clusters and fixed points. *International Journal of Bifurcation and Chaos* **6(2)**, 279-294.

Albano, A.M., Passamante, Hediger, T., and Farrell, M.E.,1992, Using neural nets to look for chaos. *Physica* **D58(1-4)**, 1-9.

Allingham, D., West, M., and Mees, A., 1998, Wavelet reconstruction of nonlinear dynamics. *International Journal of Bifurcation and Chaos* **8(11)**, 2191-2201.

Adamatzky, A.[1994] *Identification of Cellular Automata* (London : Taylor ,and Francis).

Bagarinao, E., Pakdaman, K., Nomura, T., and Sato, S., 1999, Reconstructing bifurcation diagrams from noisy time series using nonlinear autoregressive models. *Physical Review* **E60(1)**, 1073-1076.

Billings, S.A., and Leontaritis,I.J.,1982, Parameter estimation techniques for nonlinear systems. *The 6th IFAC Symposium on Identification and Systems Parameter Estimation*, Washington, pp 427-432.

Billings, S.A., and Voon,W.S.F.,1986, Correlation based model validity tests for nonlinear models. *International Journal of Control*, **44(1)**,235-244.

Billings, S.A., Korenberg, M., and Chen, S.,1988, Identification of nonlinear output-affine systems using an orthogonal least-squares algorithm. *International Journal of Systems Science*, **19(8)**,1559-1568.

Billings, S.A., Chen, S., and Korenberg, M.J.,1989, Identification of MIMO non-linear systems suing a forward regression orthogonal estimator. *International Journal of Control*, **49(6)**,2157-2189.

Billings, S.A., and Coca, D. ,1999, Discrete wavelet models for identification and qualitative analysis of chaotic Systems. *International Journal of Bifurcation and Chaos* **9(7)**, 1263-1284.

Billings, S.A., and Zheng, G.L.,1999, Qualitative validation of radial basis function networks. *Mechanical Systems and Signal Processing* **13(2)**, 335-349.

Billings, S.A., and Wei, H.L., 2003, The wavelet-NARMAX representation: a hybrid model structure combining polynomial models with multiresolution wavelet decompositions, submitted for publication.

Billings, S.A., and Yang, Y.X.,2003, Identification of the neighbourhood and CA rules from spatio-temporal CA Patterns. *IEEE Transactons on Systems, Man and Cybernetics*, **Part B 33(2)**, 332-339.

Brown, R., Bryant, P., and Abarbanel, H.D.I., 1991, Computing Lyapunov spectrum of a dynamical system from an observed time series. *Physical Review* **A43(6)**, 2787-2806.

Cao, L.Y., Hong, Y.G., Fang, H.P., and He, G.W., 1995, Predicting chaotic time series with wavelet networks, *Physica* **D85(1-2)**, 225-238.

Cao, L.Y., Mees, A., and Judd, K., 1997, Modelling and predicting non-stationary time series. *International*

Journal of Bifurcation and Chaos **7(8)**, 1823-1831.

Casdagli, M.,1989, Nonlinear prediction of chaotic time series, *Physica* **D35(3)**, 335-356.

Chen,S., Billings, S.A., and Luo,W., 1989, Orthogonal least squares methods and their application to non-linear system identification. *International Journal of Control*, **50(5)**,1873-1896.

Chen, S, Cowan, C.F.N., and Grant, P.M., 1991, Orthogonal least-squares learning algorithm for radial basis function networks. *IEEE Trans Neural Networks*, **2 (2)**, 302-309.

Chen, G., and Dong, X., 1993, From chaos to order: perspectives and methodologies in controlling chaotic dynamical systems. *International Journal of Bifurcation and Chaos* **3(6)**, 1363-1409.

Chui,C. K.,1992, *An Introduction to Wavelets* (Boston; London : Academic Press).

Chui, C. K., and Wang, J. H.,1992, On compactly supported spline wavelets and a duality principle. *Trans. of the American Mathematical Society.* **330(2)**, 903-915.

Coca, D., and Billings, S.A.,1997, Continuous-time system identification for linear and nonlinear systems using wavelet decomposition. *International Journal of Bifurcation and Chaos*, **7(1)**, 87-96.

Correa, M.V., Aguirre, L.A., and Mendes, E.M.A.M., 2000, Modelling chaotic dynamics with discrete nonlinear rational models. *International Journal of Bifurcation and Chaos* **10(5)**, 1019-1032.

Crutchfield, J.P., and McNamara, B.S., 1987, Equations of motion from data series. *Complex Systems* **1**, 417-452.

Eckmann, J.-P., and Ruelle, D., 1985, Ergodic theory of chaos and strange attractors. *Review of Modern Physics* **57(3)**, 617-656.

Eckmann, J.-P., Kamporst, S.O., Ruelle, D., and Ciliberto, S., 1986, Lyapunov exponents from a time series. *Physical Review* **A34(6)**, 4971-4979.

Farmer, J.D., and Sidorowich, J.J., 1987, Predicting chaotic time series. *Physical Review Letters* **59(8)**, 845-848.

Grassberger, P., and Procaccia, I., 1983, Measuring the strangeness of strange attractors. *Physica* **D**, **9(1-2)**, 189-208.

Guckenheimer, J., and Holmes, P., 1983, *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields* (Springer-Verlag, New York).

Gutowitz, H.A., 1990, Introduction. *Physica* **D 45(1-3)**, vii-xiv.

Gouesbet, G.,1991, Reconstruction of standard and inverse vector fields equivalent to a Rossler system. *Physical Review* **A44(10)**, 6264-6280.

Gouesbet, G., 1992, Reconstruction of vector fields: the case of Lorenz system. *Physical Review* **A46(4)**, 1784-1796.

Hayners, B.R.,and Billings, S.A., 1994, Global analysis and model validation in nonlinear system identification. *Nonlinear Dynamics* **5(1)**, 93-130.

Henon, M., 1976, A two-dimensional map with strange attractor. *Communications in Mathematical Physics* **50**, 69-77.

Hong, X., and Harris, C. J.,2001, Nonlinear model structure detection using optimum experimental design and orthogonal least squares. *IEEE Transactions On Neural Networks*, **12(2)**, 435-439.

Ilachinski, A., 2001, *CellularAutomata: A Discrete Universe* (New Jersey : World Scientific).

Korenberg, M., Billings, S.A., Liu, Y. P., and McIlroy P.J.,1988, Orthogonal parameter estimation algorithm for non-linear stochastic systems. *International Journal of Control*, **48(1)**,193-210.

Linsay, P.S., 1991, An efficient method of forecasting chaotic time series using linear regression. *Physics Letters* **A153(6-7)**, 353-356.

Leontaritis,I.J., and Billings, S.A.,1985, Input-output parametric models for non-linear systems, (part I: deterministic non-linear systems; part II: stochastic non-linear systems). *Int. Journal of Control*, **41(2)**, 303-344.

Lorenz, E., 1963, Deterministic nonperiodic flow. *Journal of Atmospheric Sciences.* **20**, 130-141.

Mallat,S.G., 1989, A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. On Pattern analysis and machine intelligence*, **11(7)**, 674-693.

Mallat, S.G., and Zhang, Z., 1993, Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, **41(12)**, 3397-3415.

Mees, A.I., 1993, Parsimonious dynamical reconstruction. *International Journal of Bifurcation and Chaos* **3(3)**, 669-675.

Menard, O., Letellier, C., Maquet, J. ,and Gouesbet, G., 2000, Modelling maps by using rational functions. *Physical Review* **E62(5)**, 6325-6331.

Mendes, E.M.A.M., and Billings, S.A., 1997, On identifying global nonlinear discrete models from chaotic data. *International Journal of Bifurcation and Chaos* **7(11)**, 2593-2601.

Mendes, E.M.A.M., and Billings, S.A.,1998, On overparameterization of nonlinear discrete systems. *International Journal of Bifurcation and Chaos* **8(3)**, 535-556.

Milner, J., 1985, On the concept of attractor. *Communications in Mathematical Physics* **99(2)**, 177-195.

Moon, F.C., 1987, *Chaotic Vibrations: An Introduction for Applied Scientists and Engineers* (John Wiley, New York).

Packard, N.H., Crutchfield, P.J., Farmer, J.D., and Shaw, R.S., 1980, Geometry from a time series. *Physical* Review Letters **45(9)**, 712-716.

Parker, T.S., and Chua, L.O., 1989, *Practical Numerical Algorithms for Chaotic Systems* (Springer-Verlag, New York).

Pearson, R. K., 1995, Nonlinear input/output modelling. *Journal of Process Control*, **5(4)**, 197-211.

Pearson, R.K., 1999, *Descrete-time Dynamic Models* (New York; Oxford: Oxford University Press).

Principe, J.C., Rathie, A., and Kuo, J.M., 1992, Prediction of chaotic time series with neural networks and the issue of dynamic modelling. *International Journal of Bifurcation and Chaos* **2(4)**, 989-996.

Smith, L.A., 1992, Identification and prediction of low dimensional dynamics. *Physica* **D58(1-4)**, 50-76.

Takens, F., 1981, Detecting strange attractors in turbulence. In D. Rand and L.S. Young (ed.), *Lecture Notes in Mathematics* 898, (Berlin : Springer).

Ueda, Y., and Akamatsu, N., 1981, Chaotically transitional phenomena in the forced negative-resistance Oscillators. *IEEE Trans.Circuits and Systems* **28(3)**, 217-224.

Wang, L.X., and Mendel, J.M.,1992, Fuzzy basis functions, universal approximations, and orthogonal least squares learning. *IEEE Trans Neural Networks*, **3(5)**, 807-814.

Wei, H.L., and Billings, S.A., 2003, A unified wavelet-based modelling framework for nonlinear system identification: the WANARX model structure. submitted for publication.

Wolf, A., Swift, J.B., Swinney, H.L., and Vastano, J.A., 1985, Determining Lyapunov exponents from a time Series. *Physics* **D16(3)**, 285-317.

Wolfram, S., 1983, Statistical mechanics of cellular automata. *Reviews of Modern Physics* **55(3)**, 601-644.

Xu, J.H., and Ho, D.W.C., 2002, A basis seletion algorithm for wavelet neural networks. *Neurocomputing*, **48**, 681-689.

Zhang,Q., 1997, Using wavelet network in nonparametric estimation. *IEEE Trans. Neural Networks*, **8(2)**, 227- 236.

Zheng, G.L., and Billings, S.A., 1999, Qualitative validation and generalization in non-linear system Identification. *International Journal of Control*, **72(17)**, 1592-1608.

Zhu, Q.M., and Billings, S.A., 1996, Fast orthogonal identification of nonlinear stochastic models and radial basis function neural networks. *International Journal of Control*, **64(5)**, 871-886.