Monograph:
Wei, H.L. and Billings, S.A. (2001) Identification of Time-Varying Systems Using Multiresolution Wavelet Models. Research Report. ACSE Research Report 801 . Department of Automatic Control and Systems Engineering

IDENTIFICATION OF TIME-VARYING SYSTEMS USING MULTIRESOLUTION WAVELET MODELS

H.L. WEI, S.A. BILLINGS

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

Research Report No.801

September 2001

IDENTIFICATION OF TIME-VARYING SYSTEMS USING MULTIRESOLUTION WAVELET MODELS

H.L. WEI , S.A. BILLINGS

Abstract: Identification of linear and nonlinear time-varying systems is investigated and a new wavelet model identification algorithm is introduced. By expanding each time-varying coefficient using a multiresolution wavelet expansion, the time-varying problem is reduced to a time invariant problem and the identification reduces to regressor selection and parameter estimation. Several examples are included to illustrate the application of the new algorithm.

Keywords: Nonlinear systems; time-varying systems; identification; wavelets

1. INTRODUCTION

There are many cases where the signals encountered in applications, such as in speech processing and seismic analysis, fail to satisfy the stationary assumption. This has led to a growing interest in nonstationary signal processing including time-frequency representations [Jones & Parks, 1992; Sattar & Salomonsson 1999; Potamianos & Maragos, 2001], time-varying spectral analysis [Cho et al, 1991; Cakrak & Loughlin, 2001], and time-varying parametric methods [Kozin & Nakajima, 1980; Grenier, 1983; Niedawiecki, 1988; Tsatsais & Giannkis, 1993; Young, 1994]. In contrast with most nonparametric methods including narrow-band filtering, complex demodulation, short-time Fourier transforms and several transformations leading to time-frequency representations which are relatively well established, alternative parsimonious descriptions can be employed in cases where the signal can be described by a time-varying parametric model.

Several approaches have been adopted to deal with time-varying modelling problems. One of the most popular approaches to identify a time-varying system is to employ an adaptive algorithm under the assumption that the time-variations are slow so that the system trajectory can be tracked. In order to guarantee that an adaptive algorithm can track time variation of the system, several assumptions are needed and more explicit modelling of the variation of the coefficients is required. One approach is to use a stochastic model structure where the coefficient trajectories are regarded as random processes, the coefficients can then be estimated using Kalman filtering. The problem with this approach is the need to determine an appropriate model for the coefficient trajectories and how to estimate the parameters.

Parametric identification of linear and nonlinear time-varying systems is possible if the time-varying coefficients can be expanded as a finite set of basis functions. The problem then becomes time-invariant with respect to the parameters in the expansions and is hence reduced to regression selection. The two main problems, which are encountered when this approach is applied to general time-varying systems, include how to choose the basis functions, and how to select the significant ones from the family of the basis functions. If these problems can be solved, the final model can be expressed using these "significant" basis functions.

Several classes of functions have been proposed, as a solution to the first problem, including Legendre polynomials and Fourier bases (sine/cosine functions) [Niedzwiecki, 1988]. The normal solution to the second problem is to truncate the function expansions at an appropriate order, and to select significant terms according to some practical rule.

An alternative approach is to use wavelets as the basis functions. Wavelets have excellent approximation properties, they outperform many other approximation schemes and are well-suited for approximating general nonlinear signals, even those with sharp discontinuities. Wavelets have found many applications in system identification including the works of Tsatsanis & Giannakis [1993]; Coca & Billings[1997]; Billings & Coca [1999]; Sureshbabu & Farrell [1999].

Tsatsanis & Giannakis [1993] introduced a wavelet basis for time-varying system identification for linear systems by expanding the time-varying coefficients as the combination of multiresolution dyadic *perfect reconstruction filter banks* (*PRFB*s). The *F*-test and *AIC* method were then used to select the significant terms. In the present paper nonlinear time-varying systems are studied and an alternative approach is introduced. This consists of expressing the time-varying coefficients as multiresolution wavelet series expansions and using the *orthogonal least squares* (*OLS*) algorithm and the *error reduction ratio* (*ERR*) [Korenberg & Billings et al,1988; Billings et al,1989] to replace the perfect reconstruction filter banks and the *F*-test and *AIC* method which were adopted by Tsatsanis & Giannakis [1993] in the linear model case.

The paper is organised as follows. Section 2 introduces the input-output representation for nonlinear systems. In section 3, wavelet theory is briefly reviewed to provide the basis of multiresolution expansions for arbitrary functions. Although wavelets have been widely used in many fields, not much work has been done on applying them in time-varying system identification. Based on a multiresolution wavelet expansion, we propose a new approach for time-dependent parameter estimation, and this is introduced in section 4. Examples are provided in section 5, and conclusions are given in section 6.

2. PROBLEM REPRESENTATION

2.1 *Modelling nonlinear systems*

A generic model for nonlinear systems, the NARMAX, which was introduced by Leontaritis & Billings [1985], has been developed in several papers [see, for example, Chen & Billings, 1989; Billings & Chen, 1989]. NARMAX can describe a wide range of nonlinear dynamic systems and includes several other linear and nonlinear model types, including the Volterra, Hammerstein, Wiener, AR, ARMA, ARMAX, and bilinear models as special cases.

The NARMAX (*Nonlinear Autoregressive Moving Average with eXogenous* inputs) model takes the form of nonlinear difference equation

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

where f is a nonlinear mapping, $u(t)$ and $y(t)$ are the input and output vector, n_u and n_y are the maximum input and output lags, respectively. The noise variable $e(t)$, with maximum lag n_e, accommodates the effects of measurement noise, modelling errors and unmeasured disturbances. A rigorous derivation of the NARMAX model and many applications have been proposed in the literature, some examples are Leontaritis & Billings [1985]; Billings & Chen, 1989; Chen & Billings, 1989; Tabrizi, 1990; Cooper, 1991; Noshiro et al,1993; Jang & Kim, 1994; Aguirre & Billings,1995; Billings and Coca,1999; Tabrizi, 1998; Radhakrishnan et al, 1999; Glass & Franchek, 1999].

2.2 *Input-output representation of time-varying systems*

Consider the time-varying NARX model as an example to illustrate the expansion of the coefficients as a finite set of basis functions. Expanding (1) by defining the function $f(\cdot)$ to be a polynomial of degree M gives the representation [Korenberg & Billings et al, 1988]

$$y(t) = \sum_{i=0}^{M} a_i(t)p_i(t) + e(t) \tag{2}$$

where

$p_0(t) = 1$, $a_0(t) = a_0 = const$

$a_i(t)$ is time-varying parameter

$p_i(t) = y(t-n_1)\cdots y(t-n_j)u(t-m_1)\cdots u(t-m_k)$, $i = 1,2,\cdots,M$

$1 \le n_1, n_2, \cdots, n_j \le n_y$, $\quad 0 \le m_1, m_2, \cdots, m_k \le n_u$, $\quad k, j \ge 0$.

and

$j = 0$ indicates that $p_i(t)$ contains no $y(\cdot)$ terms

$k = 0$ indicates that $p_i(t)$ contains no $u(\cdot)$ terms

If each coefficient $a_i(t)$ can be approximated by a linear combination of some basis functions, $\xi_\ell(t)$, say, $\ell = 1,2,\cdots,L$

$$a_i(t) = \sum_{\ell=1}^{L} \theta_\ell^{(i)} \xi_\ell(t) \tag{3}$$

then the identification can be implemented by estimating the time-invariant coefficients $\{\theta_\ell^{(i)}\}_{\ell=1,L}^{i=1,M}$. Substituting (3) into (2), gives a set of linear equations, which can be solved by several methods in the least-squares class of algorithms providing the significant terms can be selected.

In the approach proposed below, multiresolution wavelets and scaling functions, which will be discussed in the next section, are used as the basis functions to express the time dependent coefficients in time-varying models.

3. WAVELET TRANSFORMS AND WAVELET SERIES

Among almost all the functions used for approximating arbitrary signals or functions, none has had such an impact and spurred so much interest as *wavelets*. Multiresolution wavelet expansions outperform many other approximation schemes and offer a flexible capability for approximating arbitrary functions. Wavelet basis functions have the property of localization in both time and frequency. Due to this inherent property, wavelet approximations provide the foundation for representing arbitrary functions economically, using just a small number of basis functions. Wavelet algorithms process data at different scales or resolutions.

Wavelet analysis is based on a wavelet prototype function, called the *analysing wavelet*, *mother wavelet*, or simply *wavelet*. Temporal analysis is performed using a contracted, high-frequency version of the same function. Because the signal or function to be studied can be represented in terms of a wavelet expansion, data operations can also be performed using the corresponding wavelet coefficients.

3.1 *The continuous wavelet transform*

For a given function $f \in L^2(R)$, the continuos wavelet transform (CWT) with respect to the *mother wavelet* φ is defined as [Chui, 1992; Daubechies, 1992].

$$(W_\varphi f)(a,b) = \int_{-\infty}^{+\infty} f(x)\varphi_{a,b}^{*}(x)dx \tag{4}$$

where $\varphi_{a,b}(x)$ is obtained by scaling and dilating the mother wavelet $\varphi(x)$ as follows:

$$\varphi_{a,b}(x) = |a|^{-1/2}\varphi(\frac{x-b}{a}),\ \ a,b \in R, a \neq 0 \tag{5}$$

Equation (4) states that the continuous wavelet transform $(W_\varphi f)(a,b)$ is the correlation of $f(x)$ with a scaling of a and a shift (translation) of b. The over-star "*" above the function $\varphi_{a,b}(x)$ indicates the complex conjugate.

The CWT (4) is invertable subject to a mild restriction imposed on the wavelet φ, in the sense that

$$f(x) = \frac{1}{C_\varphi}\int_{-\infty}^{\infty}\frac{da}{a^2}\int_{-\infty}^{\infty}[(W_\varphi f)(a,b)]\varphi_{a,b}(x)db \tag{6}$$

with

$$C_\varphi = \int_{-\infty}^{\infty}\frac{|\hat{\varphi}(\omega)|^2}{|\omega|}d\omega < \infty \tag{7}$$

where $\hat{\varphi}$ is the Fourier transform of the function φ.

The inverse transform (6) guarantees that the function $f(x)$ can be reconstructed from the CWT and it can be interpreted in at least two different ways. On the one hand, this shows how to reconstruct the function f from the wavelet transform and, on the other, the inverse transform gives a recipe showing how to write any arbitrary f as a superposition of wavelet functions $\varphi_{a,b}(x)$.

3.2 *Wavelet series*

In practical applications the CWT is often discretised in both the scaling and dilation parameters for computational efficiency. Based on this discretization, *wavelet series* can be introduced to provide an alternative basis function representation to the conventional series expansion, for instance Fourier series, for a function in $L^2(R)$.

The most popular approach to discetise the CWT is to restrict the dilation and translation parameters to a dyadic lattice as $a_j = 2^{-j}$ and $b_{j,k} = k2^{-j}$ with $j,k \in Z$. Other non-dyadic ways of discretisition are also available.

For a given *orthogonal* wavelet φ, introduce the following derivative functional family

$$\varphi_{j,k}(x) = 2^{j/2}\varphi(2^j x - k),\ \ j,k \in Z, \tag{8}$$

then for any function $f \in L^2(R)$, the CWT can be expressed as

$$c_{j,k} = (W_\varphi f)(2^{-j}, k2^{-j}) = < f, \varphi_{j,k} >,\ \ j,k \in Z \tag{9}$$

Hence the discrete wavelet transform (9) and the wavelet family (8) can be viewed as discretised versions of the CWT (4) and the inversion formula (6), and every $f \in L^2(R)$ can be uniquely described as

$$f(x) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} c_{j,k} \varphi_{j,k}(x) \tag{10}$$

where the convergence of the series in (10) is in $L^2(R)$, namely

$$\lim_{J_1,J_2,K_1,K_2 \to \infty} \left\| f(x) - \sum_{j=-J_1}^{J_2} \sum_{k=-K_1}^{K_2} c_{j,k} \varphi_{j,k}(x) \right\| = 0 \tag{11}$$

In general, however, it is not necessary to require $\{\varphi_{j,k}\}$ to be an orthogonal basis of $L^2(R)$

$$<\varphi_{j,k}, \varphi_{\ell,m}> = \delta_{j,\ell} \cdot \delta_{k,m}, \quad j,k,\ell,m \in Z \tag{12}$$

The following two conditions are sufficient to guarantee a wavelet φ will form a wavelet series [Chui, 1992]

(*i*) *The function family $\{\varphi_{j,k}\}_{j,k \in Z}$ is a Riesz basis of $L^2(R)$, in the sense that the linear span of $\varphi_{j,k}$ is dense in $L^2(R)$, and there exist positive constants A and B, with $0 < A \le B < \infty$, such that*

$$A \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} |c_{j,k}|^2 \le \left\| \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} c_{j,k} \varphi_{j,k} \right\|_2^2 \le B \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} |c_{j,k}|^2 \tag{13}$$

for all doubly bi-infinite square-summable sequences $\{c_{j,k}\}$.

(*ii*) *There is some function $\tilde{\varphi} \in L^2(R)$, such that the family $\{\tilde{\varphi}_{j,k}\}_{j,k \in Z}$ defined as (8) is a Riesz basis of $L^2(R)$ and is dual to $\{\varphi_{j,k}\}_{j,k \in Z}$ in the sense that*

$$<\tilde{\varphi}_{j,k}, \varphi_{\ell,m}> = \delta_{j,\ell} \cdot \delta_{k,m}, \quad j,k,\ell,m \in Z \tag{14}$$

If $\{\varphi_{j,k}\}$ is an orthogonal basis of $L^2(R)$, then it is clear that (14) holds with $\tilde{\varphi}_{j,k} = \varphi_{j,k}$, or $\tilde{\varphi} \equiv \varphi$. Theoretically, if the dual pair $(\tilde{\varphi}, \varphi)$ exits and the above conditions (i) and (ii) hold, then every $f \in L^2(R)$ can be uniquely written as

$$f(x) = \sum_{j,k=-\infty}^{\infty} < f, \tilde{\varphi}_{j,k} > \varphi_{j,k}(x) \tag{15}$$

and this is called a *wavelet series.* In comparison with the CWT, the wavelet series is more computationally efficient. But this is obtained at the expense of increased restrictions on the choice of the basic wavelet φ.

3.3 *Orthogonal wavelet basis and multiresolution analysis*

It is known that for solving identification problems based on the regression representation it is useful to have a basis of orthogonal functions whose support can be made as small as required and which provides a uniform approximation to any $L^2(R)$ function. One of the original objectives of wavelet theory was to construct orthogonal (biorthogonal) basis in $L^2(R)$.

The principles for constructing orthogonal wavelets are as follows:

(*i*) There exits a function ϕ, called a *scaling function* related to the mother wavelet φ, such that the elements of the family $\{\phi(t-k)\}_{k\in Z}$ are mutually orthogonal,

(*ii*) For any given $j\in Z$, the family $\{\phi_{j,k} : \phi_{j,k}(x)=2^{j/2}\phi_{j,k}(2^j x-k), k\in Z\}$ constitute an orthogonal basis for $L^2(R)$,

(*iii*) The family $\{\varphi_{j,k}\}$ constitutes an orthogonal basis for the space $L^2(R)$, and

(*iv*) The family $\{\phi_{j_0,k}, \varphi_{j,k}\}_{j\geq j_0, k\in Z}$ also forms an orthogonal basis for $L^2(R)$.

To satisfy the above aims, an orthogonal wavelet can be constructed using *multiresolution analysis*(MRA). First introduce wavelet subspaces W_j , $j\in Z$, which are defined as the closure of the linear span of the wavelets $\{\varphi_{j,k}\}_{k\in Z}$, namely

$$W_j = \overline{span}\{\varphi_{j,k}, k\in Z\} \tag{16}$$

which satisfy

$$W_i \cap W_j = \{\varnothing\}, \text{ for any } i\neq j \tag{17}$$

where the over-bar denotes closure. It follows that $L^2(R)$ can be decomposed as a direct sum of the spaces W_j:

$$L^2(R) = \cdots \oplus W_{-1} \oplus W_0 \oplus W_1 \oplus \cdots \tag{18}$$

in the sense that every function $f\in L^2(R)$ has a unique decomposition

$$f(x) = \cdots + g_{-1}(x) + g_0(x) + g_1(x) + \cdots = \sum_{j\in Z} g_j(x) \tag{19}$$

The circles around the plus signs in (18) indicate "orthogonal sums". The decomposition of (18) is usually called an *orthogonal decomposition* of $L^2(R)$.

For each $j\in Z$, consider the closed subspaces of $L^2(R)$

$$V_j = \cdots \oplus W_{j-2} \oplus W_{j-1},\ j\in Z \tag{20}$$

which have the following properties:

(*i*) $\cdots \subset V_{-1} \subset V_0 \subset V_1 \subset \cdots$,

(*ii*) $\overline{(\bigcup_{j\in Z} V_j)} = L^2(R)$ (the over-bar here indicates closure),

(*iii*) $\bigcap_{j\in Z} V_j = \{\varnothing\}$,

(*iv*) $V_{j+1} = V_j \oplus W_j$, $j\in Z$, and

(*v*) $f(x)\in V_j \Leftrightarrow f(2x)\in V_{j+1}$, $j\in Z$.

Then it is clear that every function $f\in L^2(R)$ can be approximated as closely as desirable by the projections $P_j f$ in V_j. Another important intrinsic property of these spaces is that more and more variations of $P_j f$ are removed as $j \to -\infty$. In fact, these variations are peeled off, level by level in decreasing order of the rate of variations (frequency bands) and stored in the complementary W_j, as in property (iv).

Now for every function $f \in L^2(R)$, the wavelet series expansion can therefore be expressed as

$$f(x) = \sum_k \alpha_{j_0,k}\phi_{j_0,k}(x) + \sum_{j\geq j_0}\sum_k \beta_{j,k}\varphi_{j,k}(x) \tag{21}$$

where the wavelet coefficients $\alpha_{j,k}$ and $\beta_{j,k}$ are theoretically given by the inner products:

$$\alpha_{j,k} = < f, \phi_{j,k} > = \int f(x)\varphi^*_{j,k}(x)dx \tag{22}$$

$$\beta_{j,k} = < f, \varphi_{j,k} > = \int f(x)\varphi^*_{j,k}(x)dx \tag{23}$$

However, in practice the above coefficients are usually estimated during identification. Using the concept of *tensor products*, the series expansion (21) can easily be generalised to the muti-dimensional cases, this will be discussed in a later section.

3.4 *Multiresolution B-spline wavelets*

For many applications, it is not essential for the wavelets to be orthonormal. Relaxing the condition of orthonormality results in semiothogonal, biorthogonal or other non-orthonrmal multiroselution approximations. This provides, under some conditions, a more flexible framework for function approximation.

3.4.1 *B-spline wavelets*

B-splines as piece-wise polynomial functions with good local properties, were originally introduced by Chui and Wang [1992] as wavelet and scaling functions in multiroselution expansions.

The B-spline function of m th order is defined by the following recursive formula [Chui,1992]:

$$N_m(x) = \frac{x}{m-1}N_{m-1}(x) + \frac{m-x}{m-1}N_{m-1}(x-1),\ m \geq 2 \tag{24}$$

with

$$N_1(x) = \chi_{[0,1)}(x) = \begin{cases} 1 & if\ x \in [0,1) \\ 0 & otherwise \end{cases} \tag{25}$$

Setting N_m as the scaling function, that is, $\phi(x) = N_m(x)$, then both the wavelet and the scaling function can be expressed in terms of the scaling function $N_m(x)$ as follows

$$\phi(x) = \sum_{k=0}^{m} c_k N_m(2x-k) \tag{26}$$

$$\varphi(x) = \sum_{k=0}^{3m-2} d_k N_m(2x-k) \tag{27}$$

with the coefficients given by

$$c_k = \frac{1}{2^{m-1}}\binom{m}{k} \tag{28}$$

$$d_k = \frac{(-1)^k}{2^{m-1}} \sum_{j=0}^{m} \binom{m}{j} N_{2m}(k-j+1), \ k = 0,1,\cdots,3m-2 \tag{29}$$

Clearly, the support of the m th order B-spline wavelet and the associated scaling function are

$$\begin{cases} \text{supp}\,\phi = \text{supp}N_m = [0,m] \\ \text{supp}\,\varphi = [0,2m-1] \end{cases} \tag{30}$$

Both the B-spline wavelet and the associated scaling function are symmetric in their own support. The most commonly used B-spline wavelets are the linear $(m=2)$ and cubic $(m=4)$ cases, both of which can be expressed explicitly.

The B-spline wavelets have been selected in the present study because they are particularly suitable in system identification [Billings & Coca , 1999].

3.4.2 *Mulitresolution B-spline wavelets*

A multidimensional multiresolution wavelet decomposition(expansion) can be defined by taking the *tensor product* of the one-dimensional scaling and wavelet functions. Let $f \in L^2(R^d)$, then $f(x)$ can be represented by the multiresolution wavelet series as

$$f(x_1,\cdots,x_d) = \sum_k \alpha_{j_0,k} \Phi_{j_0,k}(x_1,\cdots,x_d) + \sum_{j\ge j_0} \sum_k \sum_{l=1}^{2^d-1} \beta_{j,k}^{(l)} \Psi_{j,k}^{(l)}(x_1,\cdots,x_d) \tag{31}$$

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

$$\Phi_{j_0,k}(x_1,\cdots,x_d) = 2^{j_0 d/2} \prod_{i=1}^{d} \phi(2^{j_0} x_i - k_i) \tag{32}$$

$$\Psi_{j,k}^{(l)}(x_1,\cdots,x_d) = 2^{jd/2} \prod_{i=1}^{d} \eta^{(i)}(2^j x_i - k_i) \tag{33}$$

with $\eta^{(i)} = \phi$ or φ (scalar scaling function and the mother wavelet) but at least one $\eta^{(i)} = \varphi$. Theoretically, the wavelet coefficients are given by the inner products:

$$\alpha_{j,k} = < f, \tilde{\Phi}_{j,k} > = \int_{R^d} f(x) \tilde{\Phi}_{j,k}^*(x) dx \tag{34}$$

$$\beta_{j,k}^{(l)} = < f, \tilde{\Psi}_{j,k}^{(l)} > = \int_{R^d} f(x) \tilde{\Psi}_{j,k}^{*(l)}(x) dx \tag{35}$$

In the two-dimensional case, the multiresolution approximation can be generated, for example, in terms of the dilation and translation of a two-dimensional scaling and wavelet functions

$$\begin{cases} \Phi_{j,k_1,k_2}(x,y) = \phi_{j,k_1}(x)\phi_{j,k_2}(y) \\ \Psi_{j,k_1,k_2}^{(1)}(x,y) = \phi_{j,k_1}(x)\varphi_{j,k_2}(y) \\ \Psi_{j,k_1,k_2}^{(2)}(x,y) = \varphi_{j,k_1}(x)\phi_{j,k_2}(y) \\ \Psi_{j,k_1,k_2}^{(3)}(x,y) = \varphi_{j,k_1}(x)\varphi_{j,k_2}(y) \end{cases} \tag{36}$$

Although many functions can be chosen as scaling and wavelet functions, most of these are not suitable in system identification applications, especially in the case of multidimensional and multiresolution expansions because of the *curse-of-dimensionality*. An implementation, which has been tested with very good results, involves B-spline scaling and wavelet functions as the regressors(expansion basis) [Billings & Coca,1999]. Furthermore, not all the B-spline wavelet and scaling functions are used when modelling a dynamic system. Since the B-spline wavelet and scaling functions have local support and since the position of each basis function is determined by an integer multi-index $k=\{k_1,k_2,\cdots,k_d\}$, only a finite number of basis functions will have relevance for a particular model structure. It is obvious that only the functions whose support contains data points should be considered as candidate model terms. Thus in this case there is no need to solve the problem of positioning the centers of the basis functions, which is normally associated with the radial wavelet basis functions [Zhang, 1997]. Therefore, in practice most of the coefficients in the expansion (21) or (31) have negligible values and can be ignored, and this will lead to a very economical representation of the function $f(x)$.

4. TIME-VARYING SYSTEM IDENTIFICATION USING WAVELETS

4.1 *Expanding the time-varying coefficients into multiresolution wavelet series*

The multiresolution wavelet and scaling functions will now be used as the basis functions to describe the time-varying system models represented in section 2. Consider the model (2), and choose $\xi_\ell(t)$ in (3) as multiresolution wavelet and scaling functions. In such a case, each coefficient $a_i(t)$ can be expressed as

$$a_i(t)=\sum_{k=k_0}^{K_{j0}}\alpha_{j_0,k}^{(i)}\phi_{j_0,k}(t)+\sum_{j=j_0}^{J}\sum_{k=k_0}^{K_j}\beta_{j,k}^{(i)}\varphi_{j,k}(t) \tag{37}$$

Substituting (37) into (2), yields

$$y(t)=a_0+\sum_{i=1}^{M}\sum_{k=k_0}^{K_{j0}}\alpha_{j_0,k}^{(i)}\phi_{j_0,k}(t)p_i(t)+\sum_{i=1}^{M}\sum_{j=j_0}^{J}\sum_{k=k_0}^{K_j}\beta_{j,k}^{(i)}\varphi_{j,k}(t)p_i(t)+e(t) \tag{38}$$

This is a time-invariant equation with respect to the parameters of the wavelet coefficients $\{\alpha_{j_0,k}^{(i)}\}$ and $\{\beta_{j,k}^{(i)}\}$. Define

(*d1*) $P(t)=[p_1(t),\ p_2(t),\ \cdots,\ p_M(t)]$

(*d2*) $\Gamma(t)=[\phi_{j_0,k_0}(t),\ \phi_{j_0,k_0+1}(t),\ \cdots,\phi_{j_0,K_{j0}}(t)]$

(*d3*) $\Lambda_j(t)=[\varphi_{j,k_0}(t),\ \varphi_{j,k_0+1}(t),\ \cdots,\varphi_{j,K_j}(t)]$

(*d4*) $A(t)=P(t)\otimes\Gamma(t)$

(*d5*) $B_j(t)=P(t)\otimes\Lambda_j(t),\quad j=j_0,j_0+1,\cdots,J$

(*d6*) $\alpha^T=[\alpha_{j_0,k_0}^{(1)},\alpha_{j_0,k_0+1}^{(1)},\cdots,\alpha_{j_0,K_j}^{(1)}\ \vdots\ \alpha_{j_0,k_0}^{(2)},\alpha_{j_0,k_0+1}^{(2)},\cdots,\alpha_{j_0,K_j}^{(2)}\ \vdots\ \cdots\ \vdots\alpha_{j_0,k_0}^{(M)},\alpha_{j_0,k_0+1}^{(M)},\cdots,\alpha_{j_0,K_j}^{(M)}]$

(*d7*) $B(t)=[B_{j_0}^T(t),B_{j_0+1}^T(t),\cdots,B_J^T(t)]$

(*d8*) $\beta_j{}^T=[\beta_{j,k_0}^{(1)},\beta_{j,k_0+1}^{(1)},\cdots,\beta_{j,K_j}^{(1)}\ \vdots\ \beta_{j,k_0}^{(2)},\beta_{j,k_0+1}^{(2)},\cdots,\beta_{j,K_j}^{(2)}\ \vdots\ \cdots\vdots\ \beta_{j,k_0}^{(M)},\beta_{j,k_0+1}^{(M)},\cdots,\alpha_{j,K_j}^{(M)}]$

(*d9*) $\beta^T=[\beta_{j_0}^T,\beta_{j_0+1}^T,\cdots,\beta_J^T]$

where the symbol "$\otimes$" denotes the *Kronecker product*. Now, (38) can be recast as

$$y(t) = A(t)\alpha + B(t)\beta + e(t) \tag{39}$$

If N measurements of the input and output are available, (39) can be written in a compact matrix form as

$$Y = H\theta + \varepsilon \tag{40}$$

where

$$Y^T = [y(1) \quad y(2) \quad \cdots \quad y(N)]$$

$$\varepsilon = [e(1), e(2), \cdots, e(N)]$$

$$H = \begin{bmatrix} 1 & A(1) & B(1) \\ 1 & A(2) & B(2) \\ \vdots & \vdots & \vdots \\ 1 & A(N) & B(N) \end{bmatrix}$$

$$\theta^T = [\theta_0 \quad \alpha^T \quad \beta^T] = [\theta_0, \theta_1, \cdots, \theta_{n(\alpha,\beta)}], \quad \theta_0 = a_0$$

Here, the symbol $n(\alpha,\beta)$ is used to indicate that the number of unknown parameters in (40) dependents on the selection of the basis functions in the wavelet expansions.

The parameter vector θ in (40) can now be estimated using a least-squares-based algorithm or a prediction error routine [Billings & Voon, 1986]. Notice, however, that the number of possible terms in the model is very large and this is why detecting the model structure is a vitally important problem in nonlinear system identification.The problem is even more acute for nonlinear time varying models. The orthogonal least-squares algorithm (OLS), whose purpose is to orthogonalize all the terms in (40) by introducing an auxiliary orthogonal model, is one of the most efficient techniques that address this problem. The error reduction ratio (ERR) values can then be used as a measure of the significance of each candidate model term. Because the values of the error reduction ratios depend on the order in which candidate terms are orthogonalized into the regression equation, simply orthogonalizing candidate terms in an arbitrary order may result in incorrect information regarding the significance of terms. In order to overcome this problem, the forward regression orthogonal algorithm was introduced [Billings et al ,1989; Billings & Chen, 1989].

Consider the time-varying ARMAX model, which is as a special case of the time-varying NARMAX model, to illustrate the approach

$$y(t) = \sum_{i=1}^{p} a_i(t)y(t-i) + \sum_{j=1}^{q} b_j(t)u(t-j) + \sum_{k=1}^{s} c_k(t)e(t-k) + e(t) \tag{41}$$

where $a_i(t), b_j(t)$ and $c_k(t)$ are time dependent coefficients which can be estimated from measured values of $y(t), u(t)$ and $e(t)$ by expanding $a_i(t), b_j(t)$ and $c_k(t)$ as multiresolution wavelet series.

The values of the noise $e(t)$ are not normally available for measurement and these are usually replaced with the residuals computed as $r(t) = y(t) - \hat{y}(t|t-1)$, where $\hat{y}(t|t-1)$ is the one-step ahead predicted output. This leads to the modified model used in practice

$$y(t) = \sum_{i=1}^{p} a_i(t)y(t-i) + \sum_{i=1}^{q} b_i(t)u(t-i) + \sum_{i=1}^{s} d_i(t)r(t-i) + e(t) \tag{42}$$

Expanding the coefficients $a_i(t)$, $b_i(t)$ and $d_i(t)$ as a multiresolution wavelet series, the model (42) becomes

$$y(t) = y_1(t) + y_2(t) + y_3(t) + e(t) \tag{43}$$

where

$$y_1(t) = \sum_{i=1}^{p}\sum_{k}\alpha_{j_0,k}^{(a,i)}\phi_{j_0,k}(t)y(t-i) + \sum_{i=1}^{p}\sum_{j=j_0}^{J_1}\sum_{k}\beta_{j,k}^{(a,i)}\varphi_{j,k}(t)y(t-i) \tag{44}$$

$$y_2(t) = \sum_{i=1}^{q}\sum_{k}\alpha_{j_0,k}^{(b,i)}\phi_{j_0,k}(t)u(t-i) + \sum_{i=1}^{q}\sum_{j=j_0}^{J_2}\sum_{k}\beta_{j,k}^{(b,i)}\varphi_{j,k}(t)u(t-i) \tag{45}$$

$$y_3(t) = \sum_{i=1}^{s}\sum_{k}\alpha_{j_0,k}^{(d,i)}\phi_{j_0,k}(t)r(t-i) + \sum_{i=1}^{s}\sum_{j=j_0}^{J_3}\sum_{k}\beta_{j,k}^{(d,i)}\varphi_{j,k}(t)r(t-i) \tag{46}$$

In practice, J_1, J_2 and J_3 are often chosen to be the same values, i.e., $J_1 = J_2 = J_3 = J$.

4.2 *Selecting the multiresolution levels*

Theoretically, the multiresolution wavelet expansion contains an infinite number of terms, but in practice only a finite number of basis functions are needed to approximate a given nonlinear signal/function. However, finding or selecting the appropriate basis functions, also known as the structure selection problem, represents a very important step in constructing a parsimonious mapping f from the regression space to the output space. The multiresolution wavelet basis set $\varsigma = \{\phi_{j_0,k}, \varphi_{j,\ell} : j_0, j, k, \ell \in Z, j_0 \le j\}$ will in practice be truncated by including only the scaling functions with the given initial resolution level j_0 and the wavelet functions with the resolution levels from j_0 to a certain scale $J = j_{\max}$. In addition only those basis functions whose support contains the sampled data points will be considered. The highest resolution or scale will be such that at least one observation is within the support of the corresponding wavelet.

Assume that N-samples of observations are available from the following input-output system

$$y_i = f(x_i) + e_i, \quad i = 1,2,\cdots,N \tag{47}$$

where $x_i = [x_{1,i}, x_{2,i}, \cdots, x_{d,i}]^T \in [0,1]^d$, $f \in L^2(R)$ can be constructed from x_i and y_i using the d dimensional mutiresolution wavelet expansion (31) subject to some restrictions.If x_i is uniformly distributed on $[0,1]^d$, then the loglog type of statistical laws offer a rough interval for selecting the highest resolution scale $j_{\max}$ [Sjoberg et al.,1995]

$$\frac{N}{\ln N} \le 2^{d \cdot j_{\max}} \le \frac{2N}{\ln N} \tag{48}$$

Since most practical identification problems fail to satisfy the uniform distribution assumption, the estimate (48) provides only a rough indication for the upper scale $j_{\max}$. A practical approach is to select $j_{\max}$ such that a minimum number of observations hit the support of each basis function $\varphi_{j_{\max},k}$. Features, such as the natural frequency, for example, of the sampled signal can also be considered when determining the highest resolution scale. Assume that the maximum natural frequency of the sampled signals is $f_{\max}$, then the upper scale $j_{\max}$ can be empirically chosen as

$j_{\max} = [\log_2(Mf_{\max})]$, where M is a positive number between 10 and 20 and $[\cdot]$ denotes taking the integer value of the corresponding number.

Denote the support of the basis functions $\varphi_{j,k}$ as $S_{j,k}$

$$S_{j,k}^{(\varphi)} = \begin{cases} \{x : |\varphi_{j,k}(x)| \neq 0, j,k \in Z, x \in R^d\}, & \varphi_{j,k} \text{ is compactly supported,} \\ \{x : |\varphi_{j,k}(x)| \geq \delta \max_x |\varphi_{j,k}(x)|, j,k \in Z, x \in R^d\}, & \varphi_{j,k} \text{ is compactly supported,} \end{cases}$$

where δ is a small positive number. For each $x \in [a,b]^d$, where $[a,b]^d$ is in the problem interval, introduce the following index set

$$I_n^{(\varphi)} = \{(j,k) : x_n \in S_{j,k}^{(\varphi)}, j,k \in Z\}$$

Analogue symbols $S_{j,k}^{(\phi)}$ and $I_m^{(\phi)}$ pertaining to the scaling functions $\phi_{j,k}$ can also be introduced, such that

$$I_m^{(\phi)} = \{(j,k) : x_m \in S_{j,k}^{(\phi)}, j,k \in Z\}$$

Then the union of $I_n^{(\varphi)}, I_m^{(\phi)}, m,n = 1,2,\cdots,N$, gives the indices of the wavelets whose supports contain at least one data point. This results in a reduced set of basis functions

$$W = \left\{ g_{j,k} : (j,k) \in \left(\bigcup_{n=1}^{N} I_n^{(\varphi)} \right) \bigcup \left(\bigcup_{m=1}^{N} I_m^{(\phi)} \right), j_0 \leq j \leq j_{\max}, k \in Z \right\} \tag{49}$$

where $g_{j,k}$ denotes $\varphi_{j,k}$ or $\phi_{j,k}$. The basis functions set W in (47) forms the initial regressor set which is used to build the model structure.

Once the regressor set has been determined, the structure selection and parameter estimation procedure can be implemented by means of several algorithms. In the present study the *orthogonal forward regression* (*OFR*) algorithm [Billings & Chen, 1989; Billings et al, 1989; Billings & Coca, 1999] will be used because *OFR* can arrange the candidate model terms in the order of significance and provides an effective and upwardly extendible method of selecting the relevant model terms.

4.3 *Data pre-processing*

In some cases it is more convenient to select the starting resolution level and the range of the shift parameters if the sample data has been normalized to the unit interval [0,1], this is especially true when the Haar wavelet (the first-order B-spline wavelet) and scaling functions are chosen as the expansion basis. The Haar wavelet is very simple, but it does possess almost all the properties of multiresolution analysis. In some cases, for example if the system coefficients are known to be piece-wise constant, that is where the system exhibits discontinuous jumps, then the Haar basis may be the best choice for depicting this behaviour.

Assume that all the observations fall into the finite interval $[a,b]$, in order to deal with the end effects at both ends of the data record, the common practice of periodically extending the available data , as well as the coefficients $a_i(t)$ beyond $[a,b]$ can be followed. The original data in $[a,b]$ can now be normalized to the unit interval [0,1] by means of the following simple linear transform h

$$\zeta \in [a,b] \xrightarrow{h} \bar{\zeta} \in [0,1]$$

with

$$\bar{\zeta} = \frac{\zeta - a}{b - a} \tag{50}$$

This will result in equivalent data $\bar{y}(\bar{\zeta}) = y(\zeta)$,and equivalent coefficients $\bar{a}_i(\bar{\zeta}) = a_i(\zeta)$.

The algorithms proposed in this section can be summarised as

(i) Choose a class of wavelet ϕ and φ,

(ii) Pre-process the sampled data (where necessary),

(iii) Determine the start resolution scale j_0 and the maximum scale J, as well as the order of the model, such as p, q, s in the time varying ARMAX model (41), based on prior knowledge and computational constraints,

(iv) For $j = j_0, j_0+1, \cdots, J$, select the candidate wavelet and scaling functions from the family composed of all the possible functions $\{\phi_{j_0,k}\}_{k\in Z}$ and $\{\varphi_{j,k}\}_{k\in Z}$. The candidate functions are those, whose support contain sample data points.

(v) Perform model structure selection and parameter estimation using the *OFR* algorithm,

(vi) Validate the model using for example the tests in Billings & Voon [1983], Billings & Zhu [1994,1995].

Notice that, the proposed identification procedure is not limited to the wavelet basis case. Other bases can also be employed if there is strong evidence that they can yield a sparse expansion of the time-dependent coefficients. In addition, the proposed identification procedure has the capability to estimate the lagged model terms. For example in the simple case of an AR model of true order p_0, where p_0 is unknown. Assigning the order to be $\tilde{p}(> p_0)$ during the identification procedure, then the *OFR* algorithm [Billings & Chen, 1989] will reject all the regressors corresponding to the expansion of $a_i(t)$ for $p_0 < i < \tilde{p}$.

5. EXAMPLES

A selection of examples are described below to show the application of the new multiresolution wavelet models in the identification of time-varying systems.

5.1 *Sunspot data*

The Wolf sunspot data is a very well-known data set and records the annual sunspot index from 1700 onwards [Priestley, 1988]. The main feature of this time series is a cycle of activity varying in duration between 9 and 14 years, with an average period of approximately 11.3 years. Another feature of the series is that in each cycle the rise to the maximum tends to have a steeper gradient than that of the fall to the next minimum. This suggests that a nonlinear model might be appropriate and many different models have been fitted to this data set [Priestley, 1988]. However, few time-varying models have been fitted to this data.

In the present study the application of the new wavelet identification procedure derived above will be applied to fit a simple time-varying AR model to the sunspot data. The objective here however is not to find the definitive model for the sunspot data but rather to use this data to illustrate the new algorithm on a well known real data set. The initial candidate term set was defined by a 3-rd AR model and the wavelet expansion scales $j_0 = j_{\min} = 2$ and $J = j_{\max} = 4$. The coefficient trajectories corresponding to the model are shown in Fig. 1 together with the one-step-ahead predictions.

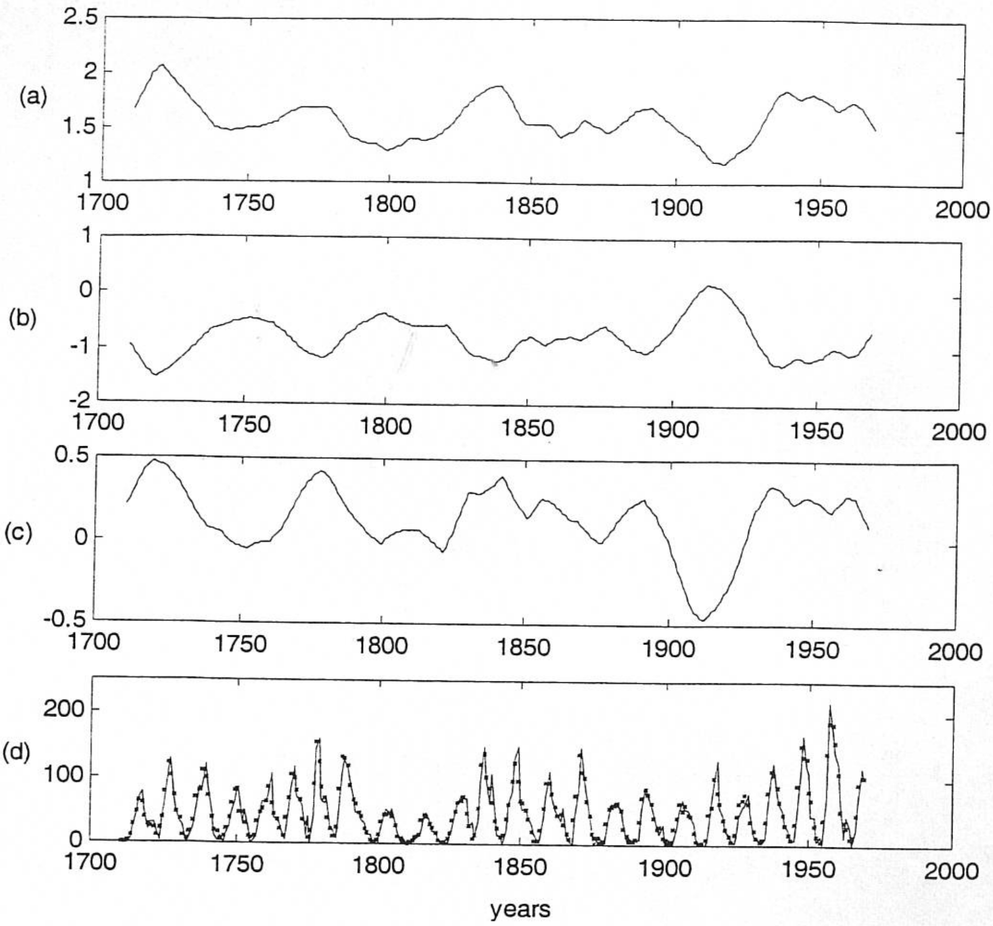

Fig. 1. Parameter trajectories for the estimated 3rd-order time-varying AR model coefficients and the one-step-ahead predictions for the sunspot data
(a) $a_1(t)$; (b) $a_2(t)$; (c) $a_3(t)$; (d) Sunspot data and one-step ahead predictions
("*" indicates true values; "—"indicates one-step-ahead predictions)

5.2 *Modelling a flight vehicle simulator*

Fig. 2 shows 1000 sampled input and output data that were collected from a flight vehicle experimental simulator. The input $u(t)$, a squared voltage excitation signal with unit amplitude, was the input to the servo system. The output $y(t)$ was the spin angle of the flight vehicle simulator. The objective here is to build a simple model to describe the system input-output relationship using a time-varying ARX model. The following time-varying ARX structure was chosen according to theoretical analysis and prior knowledge

$$y(t) = a_0 + a_1(t)y(t-1) + a_2(t)y(t-2) + a_3(t)u(t-4) + e(t) \tag{51}$$

Expanding the coefficients $a_i(t)$ into a multiwavelet series, the time-varying model (51) becomes a time-invariant parameter estimation problem. Solving this problem by employing the *OFR* algorithm, the wavelet coefficients and hence the time-dependent parameters $a_i(t)$ can be identified. The identified parameter trajectories are shown in Fig. 3. The one step-ahead predicted output and the corresponding error are illustrated in Fig4, which shows that although the model (51) is very simple, it can describe the system well.

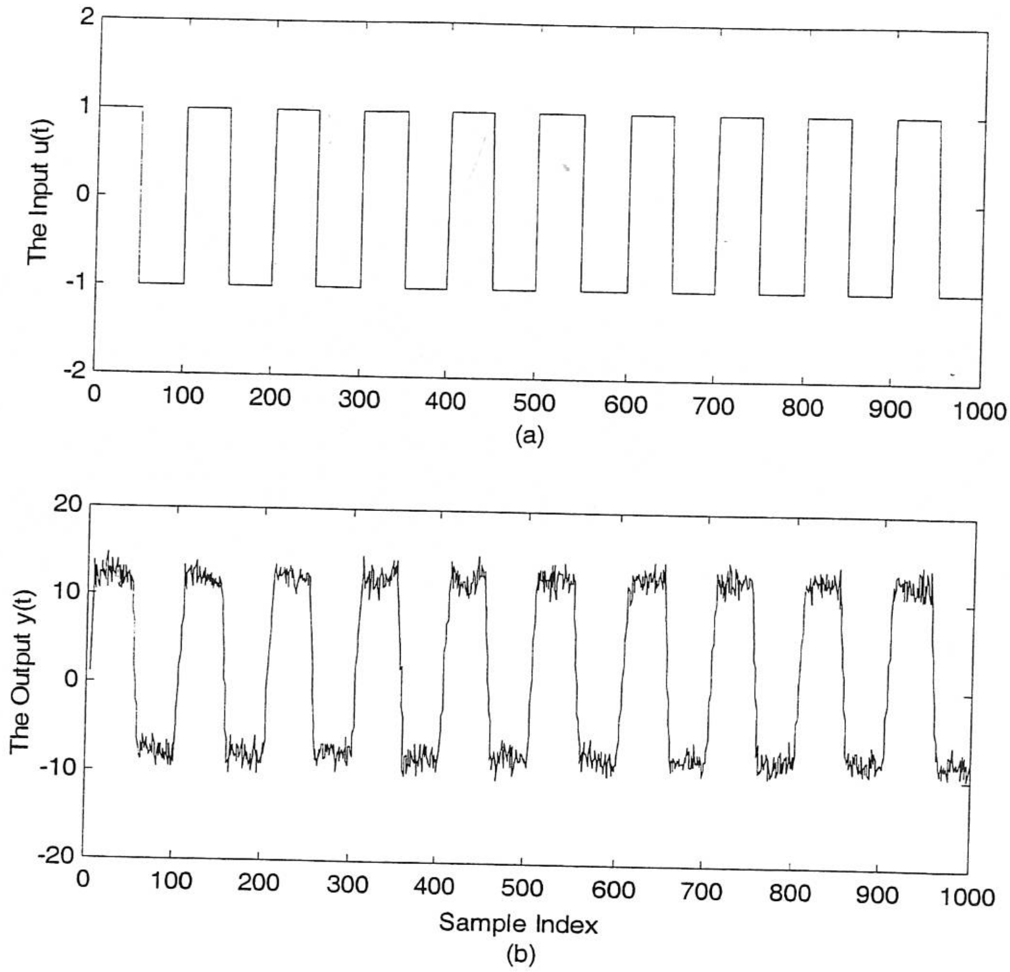

Fig. 2. System input and output for a flight vehicle experimental subsystem
(a) The input—a squared voltage excitation signal with unit amplitude; (b) The measured output

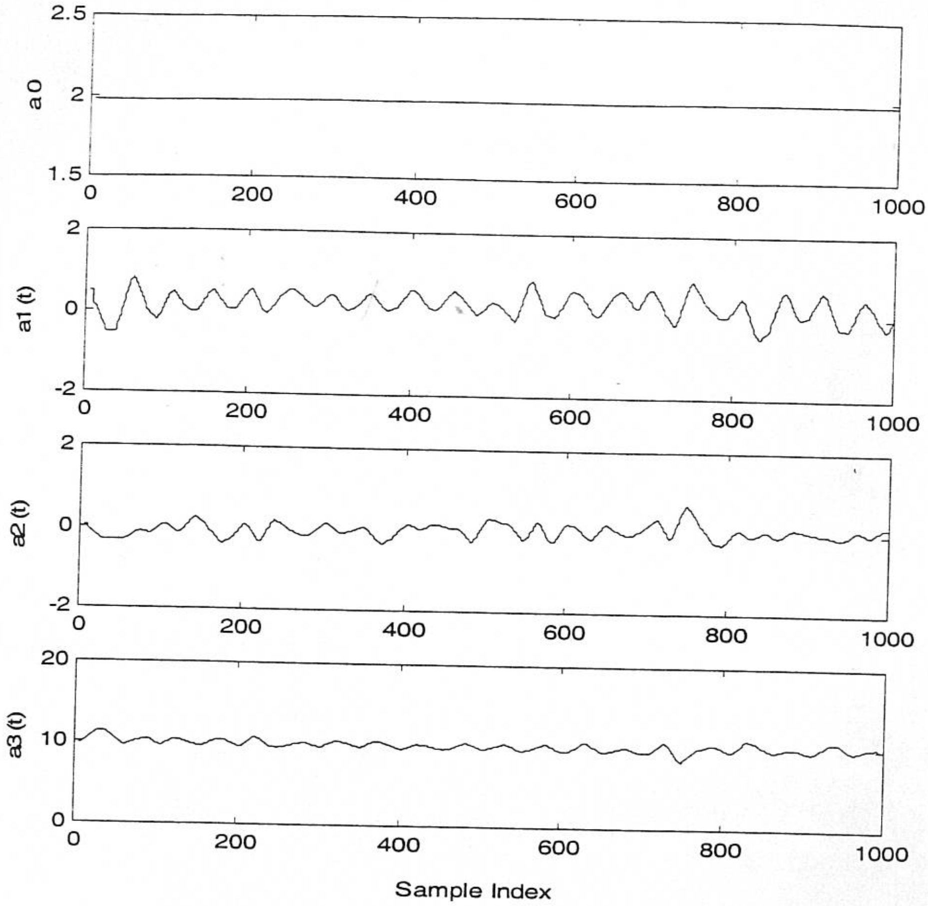

Fig. 3. Parameter trajectories of the time-varying ARX model in the section 5.2

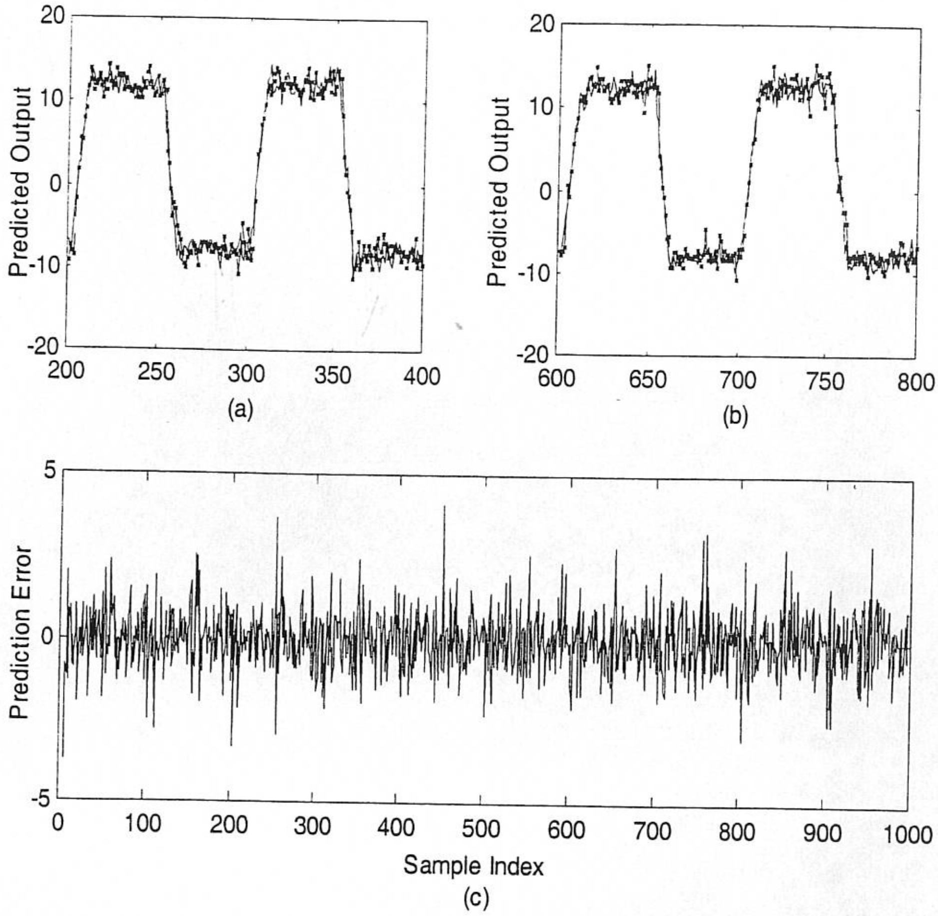

Fig. 4. One-step-ahead predicted output and prediction error for the flight vehicle experimental subsystem in section 5.2. (a) ,(b) The measurements and the one-step-ahead predicted output; (c) The one-step-ahead prediction error. ("*" indicates true values; "—" indicates one-step-ahead prediction)

5.3 *Coefficient estimation for nonlinear time-invariant continuous time systems*

Although the approach described in section 4 was proposed for time-varying systems, it can also be used for the identification of time-invariant linear and nonlinear systems. As an example, consider the *Goodwin Equation*[Coca, 1996], a nonlinear time-invariant continuous system model

$$\ddot{y}(t) + a\frac{y^2(t)-1}{y^2(t)+1}\dot{y}(t) + by(t) + cy^3(t) = -\lambda u(t) \tag{52}$$

where a, b, c and λ are time-invariant parameters.

Under the conditions $\dot{y}(0) = y(0) = 0$, $u(t) = \sin(t)$, with $a = 0.1, b = -0.5$, $c = 0.5$ and $\lambda = 37$, a 4th-order Runge-Kutta algorithm was used to simulate this model to obtain 1000 equi-spaced samples from the input and output with a sampling interval of $T = 0.01$ time units. The sampled data will be referred to subsequently as $u_k = u(kT)$, $y_k = y(kT)$.

For this continuous time model identification, the model structure is assumed to be known. If the derivatives $\dot{y}$ and $\ddot{y}$ can be observed or reconstructed, the time-invariant parameters a, b and c can be estimated using the algorithm introduced in section 4.The model (51) can be re-written as

$$a(t)\frac{y^2(t)-1}{y^2(t)+1}\dot{y}(t) + b(t)y(t) + c(t)y^3(t) = f(t) \tag{53}$$

where $f(t) = -\ddot{y}(t) - \lambda u(t)$. Expanding $a(t), b(t)$ and $c(t)$ into multiwavelet series, the time-varying model (53) becomes a time-invariant model and the wavelet coefficients can be estimated. The trajectories of $a(t), b(t)$ and $c(t)$ are shown in Fig. 5.

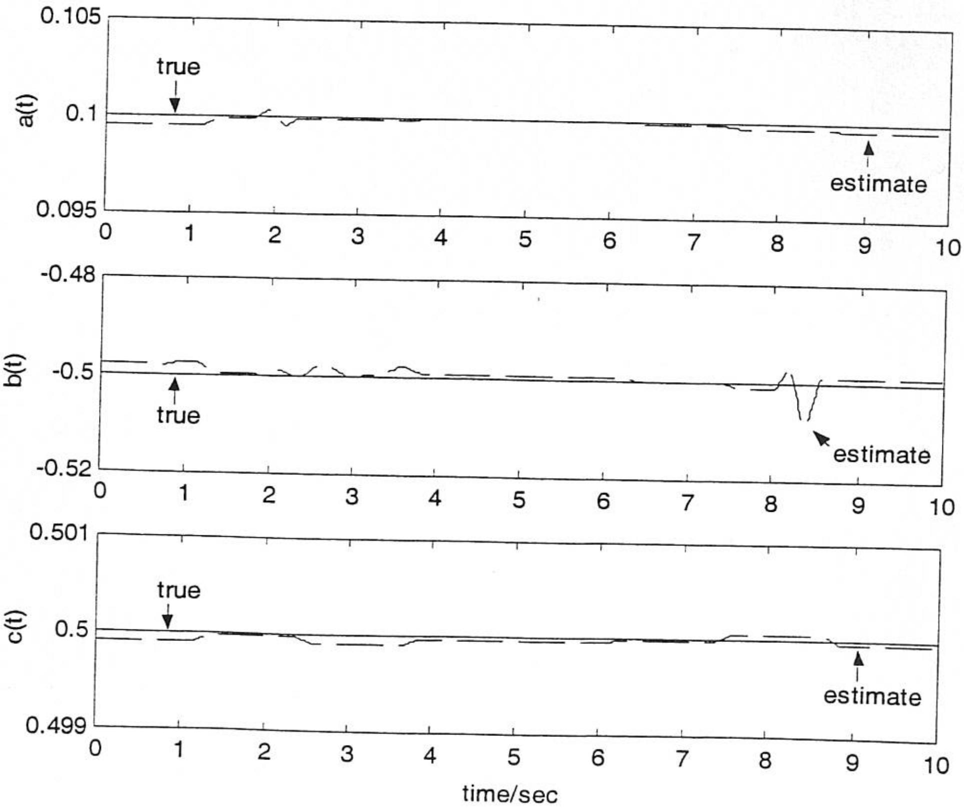

Fig. 5. The true and estimated values of the time-varying coefficients for the model in Eq. (53)

5.4 *Time-varying parameter estimation*

The following system

$$y(t) = a(t)y(t-1) + b(t)u(t-1) + v(t) \tag{54}$$

was used to generate 500 data samples of $y(t)$ (see Fig.6(a)), where $u(t)$ and $v(t)$ were independent normally distributed random sequences with zero means and variances $\sigma_u^2 = \sigma_v^2 = 1.0$. The coefficients $a(t) = 0.8$ for $0 \le t \le 500$, and $b(t)$ was a piece-wise function described as follows

$$b(t) = \begin{cases} 2 & 0 \le t \le 100, \quad 201 \le t \le 300, \quad 401 \le t \le 500 \\ 3 & 101 \le t \le 200, \quad 301 \le t \le 400 \end{cases} \tag{55}$$

Assuming that no prior knowledge about the parameters $a(t)$ and $b(t)$ were known, the aim was to identify these from the above simulated data using the algorithm introduced in section 4. Expanding the coefficients $a(t)$ and $b(t)$ into multiresolution wavelet series, the time-varying model (54) becomes a time-invariant identification problem and the wavelet coefficients can be estimated using the *OFR* algorithm. The estimated values of the parameters $a(t)$ and $b(t)$ are depicted in Fig. 6(b), which clearly shows that the wavelet expansion can track the piece-wise varying coefficient very well.

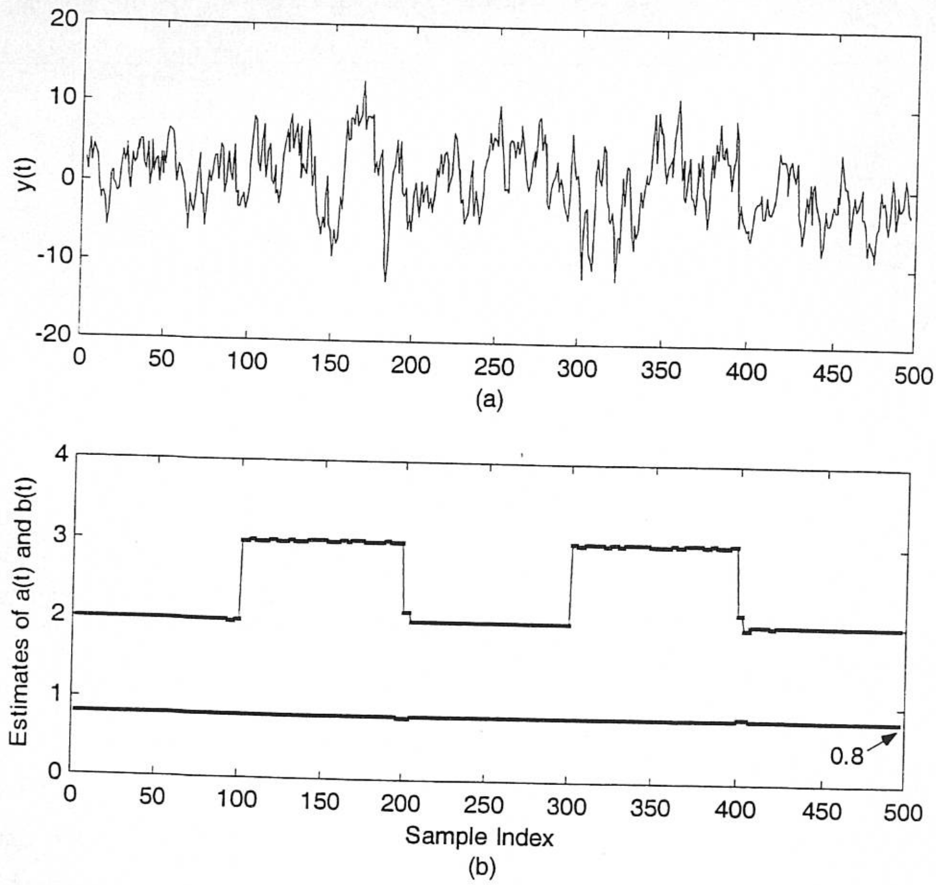

Fig. 6. System output [(a)] and the coefficient estimates[(b)] for the system in section 5.4.

6. CONCLUSIONS

Parametric identification of nonlinear time-varying systems is simplified if each time-varying coefficient can be expanded as a finite set of basis functions. The problem then becomes time-invariant with respect to the parameters in the expansions and the main problem then becomes regression selection. A multiresolutuion wavelet expansion of the time-varying model coefficients has been proposed and implemented using an orthogonal least-squares procedure as a solution to this important problem. This provides a flexible procedure that overcomes many of the limitations associated with employing the F and AIC tests when many potential candidate terms are involved. The new algorithm automatically selects the most significant model terms in the time-varying expansion to provide a parsimonious representation for both linear and nonlinear time-varying systems.

ACKNOWLEDGEMENT

SAB gratefully acknowledges that part of this work was supported by the Engineering and Physics Science Research Council UK. H L Wei is grateful to Dr J H Liu for useful discussions.

REFERENCES

Aguirre, L.A., Billings, S.A.(1995), Retrieving dynamical invariants from chaotic data using NARMAX models, *International Journal of Bifurcation and Chaos*, **5(2)**, 449-474.

Billings,S.A., and Voon,W.S.F.(1983), Structure detection and model validity tests in the identification of nonlinear systems, *Proc. Institute of Electrical Engineerings, Part D*, **130**,193-199.

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., 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., Chen,S.(1989), Extended model set, global data and threshold model identification of severely nonlinear systems, *International Journal of Control*, **50(5)**,1897-1923.

Billings,S.A. and Zhu,Q.M.(1994), Nonlinear model validation using correlation tests, *International Journal of Control*, **60(6)**,1107-1120.

Billings,S.A. and Zhu,Q.M.(1995), Model validation tests for multivariable nonlinear models including neural networks, *International Journal of Control*, **62(4)**,749-766.

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.

Bouzeghoub, M.C., Ellacott, S W., Easdown, A. and Brown, M.(2000), On the identification of non-stationary linear processes, *International Journal of Systems Science*, **31(3)**,273-286.

Cakrak,F., Loughlin,P.J.(2001), Multiple window time-varying spectral analysis, *IEEE Trans on Signal Processing*, **49(2)**,448-453.

Chen, S., Billings, S.A.(1989), Representation of non-linear systems: the NARMAX model, *International Journal of Control*, **49(3)**,1013-1032.

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

Chen, S., Billings, S.A., and Grant, P.M.(1990), Nonlinear system identification using neural networks, *International Journal of Control*, **51(6)**,1191-1214.

Cho, Y.S., Kim, S.B. and Powers E.J.(1991), Time-varying spectral estimation using AR models with variable forgetting factors, *IEEE Trans. on Signal Processing*, **39(6)**,1422-1426.

Chowdhury, F.N.(2000), Input-output modelling of nonlinear systems with time-varying linear models, *IEEE Trans. Automatic Control*, **45(7)**,1355-1358.

Chui, C. K. (1992), *An introduction to wavelets*, Academic Press.

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

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

Coca,D.(1996), A class of wavelet multiresolution decomposition for nonlinear system identification and signal processing, PhD Thesis, the University of Sheffield.

Cooper, R.A.(1991), System-identification of human-performance models, *IEEE Transactions on Systems Man and Cybernetics*, **21(1)**, 244-252.

Daubechies,I.(1992), *Ten lectures on wavelets*, Society for Industrial and Applied Mathematics, Philaelphia, Pennsylvania.

Glass,J.W., Franchek,M.A.(1999), NARMAX modelling and robust control of internal combustion engines, *International Journal of Contorl*,**72(4)**, 289-304.

Grenier,Y.(1983), Time-dependent ARMA modelling of nonstationary signals, *IEEE Trans. Signal Processing,* **ASSP 31(4)**,899-911.

Jang,H.K. and Kim,K.J.(1994), Identification of loudspeaker nonlinearities using the NARMAX modelling technique, *Journal of the Audio Engineering Society*, **42(1-2)**, 50-59.

Jones, D.L. and Parks, T.W.(1992), A resolution comparison of several time-frequency representations, *IEEE Trans. on Signal Processing*, **40(2)**,413-420.

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.

Kozin,F. and Nakajima,F.(1980), The order determination problem for linear time-varying AR models, *IEEE Trans. on Automatic Control*, **AC-25(2)**,250-257.

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

Liu, G P, Billings, S. A. and Kadirkamnathan, V.(2000), Nonilear system identification using wavelet networks, *International Journal of Systems Science*, **31(12)**,1531-1541.

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

Niedzwiecki, M.(1988), Functional series modelling approach to identification of nonstationary stochastic systems, *IEEE Trans. Automatic Control*, **33(10)**,955-961.

Noshiro, M., Furuya, M., Linkens, D., and Goode, K.(1993), Nonlinear identification of the PCO2 control-system in man, *Computer Methods and Programs In Biomedicine*, **40(3)**, 189-202.

Potamianos,A., Maragos, P.(2001), Time-frequency distributions for automatic speech recognition *IEEE Trans. on Speech and Audio Processing*, **9(3)**,196-200.

Priestley, M.B.(1988),*Non-linear and non-stationary time series analysis*. Academic Press, London.

Radhakrishnan,T.K., Sundaram, S., Chidambaram, M.(1999), Non-linear control of continuous bioreactors, *Bioprocess Engineering* ,**20(2)**, 173-178.

Sattar, F., Salomonsson, G.(1999), The use of a filter bank and the Wigner-Ville distribution for time-frequency representation, *IEEE trans. on Signal Processing*, **47(6)**,1776-1783.

Sjoberg, J., Zhang, Q.H., Ljung,L., Benveniste, A., et al.(1995), Nonlinear Black-box Modelling in system identification: a unified overview, *Automatica*, **31(12)**, 1691-1724.

Sureshbabu, N. and Farrell, J.A.(1999, Wavelet-based system identification for nonlinear control, *IEEE Trans. on Automatic Control*, **44(2)**, 412-417.

Tabrizi, M.H.N., Jamaluddin, H., Billings, S.A. and Skaggs, R.W.(1990), Use of identification techniques to develop a water-table, *Prediction Model Transactions of the ASAE*, **33 (6)**,1913-1918.

Tsatsanis, M.K., Giannakis,G.B.(1993), Time-varying system identification and model validation using wavelets, *IEEE Trans. Signal Processing,* **41(12)**,3512-3523.

Young, P.(1994), Time variable and state dependent modelling of non-stationary and nonlinear time series, in *Development in Time Series Analysis* Edited by Subba Rao,T., 375-413.

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