**Monograph:**

Wei, H.L., Balikhin, M. and Billings, S.A. (2003) Nonlinear Time-Varying System Identification Using the NARMAX Model and Multiresolutions Wavelet Expansions. Research Report. ACSE Research Report 829 . Department of Automatic Control and Systems Engineering

# Nonlinear Time-Varying System Identification Using the NARMAX Model and Multiresolution Wavelet Expansions

H. L. Wei, M. Balikhin, S.A. Billings

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

RERVM COGNOSCERE CAVSAS

Research Report No. 829

January 2003

# Nonlinear Time-varying System Identification Using the NARMAX Model and Multiresolution Wavelet Expansions

H. L. Wei, M. Balikhin and S.A. Billings

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

***Abstract:*** *Identification techniques for nonlinear time-varying systems are investigated based on the NARMAX model and multiresolution wavelet expansions. It is shown that a NARMAX model with time-varying coefficients can be reduced to a time-invariant linear-in-the-parameters analysis problem by expanding each coefficient as a multiresolution wavelet expansion. An orthogonal least squares algorithm is then adapted to estimate the parameters. An application to data relating to magnetic storms is used to illustrate the realistic application of the new identification technique.*
***Keywords:*** *Nonlinear time-varying systems; NARMAX model; system identification; wavelets.*

## 1. Introduction

Many types of signals encountered in the real world are nonlinear, nonstationary or time-varying. Examples include speech and image processing, seismic signal analysis, communications and control systems, econometrics, ecology, and astronomical observations. Although a great amount of work has been done on linear time-varying system identification including time-varying AR and ARMA models (Kozin and Nakajima 1980, Grenier 1983, Charbonnier et al 1987, Dembo and Zeitouni 1988, Morikawa 1990, Tsatsanis and Giannakis 1993), state space representations and Kalman filtering approaches (Ljung 1983, Young 1994), time-frequency representations, and nonparametric approaches for nonstationary spectral estimation (Martin and Flandrin 1985, Cho et al 1991, Jones and Parks 1992), few authors have studied the identification of nonlinear time-varying systems using parametric approaches. This may be because the parametric representation for such systems is quite difficult compared to linear parametric models. However, since linear models cannot capture the rich dynamic behaviour associated with nonlinear systems, it is important to investigate and develop methodologies for the parametric identification of nonlinear time-varying systems.

The identification of a nonlinear time-varying parametric model for a black-box system involves several problems including how to determine the model structure and model order, how to detect the degree of nonlinearity of the system, and how to estimate the time dependent parameters. Identification and modelling approaches for linear time-invariant systems have been extensively studied, and some methodologies for nonlinear time-invariant systems have been developed. Among the latter, the NARMAX identification and modelling methodology(Leontaritis and Billings 1985), which constitutes the foundation of the present work, has been successfully applied to many systems (Tabrizi 1990, Cooper 1991, Noshiro et al 1993, Jang and Kim 1994, Aguirre and Billings 1995, Radhakrishnan et al 1999, Glass and Franchek 1999). It is the wide range of successful applications of the NARMAX model to real systems that motivates the adoption of this methodology in the present study by combining this approach with multiresolution wavelet expansions for the identification and modelling of nonlinear time-varying systems.

Parametric identification of nonlinear time-varying systems is possible if each of 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 the problem can be reduced to a linear-in-the-parameters regression problem.

Several types of functions including polynomials, Fourier bases, splines etc. have been chosen as the basis functions to represent the time-varying coefficients. In the present study, orthogonal wavelet basis functions will be employed due to their distinct approximation properties in both the time and the frequency domain (Chui 1992) and the time-varying coefficients will be expanded as multiresolution wavelet expansions (Mallat 1989). Based on a NARMAX model structure, the *orthogonal forward regression* (OFR) algorithm (Korenberg et al 1988, Billings et al 1988,1989a,b, Chen et al 1989b) is then adapted to estimate the parameters in the expansions.

This paper is organised as follows. In section 2, the time-varying NARMAX model is introduced and the NARMAX methodology is briefly reviewed. In section 3, the time-varying problem is reduced to a linear-in-the-parameters regression problem by expanding the time-varying coefficients as multiresolution wavelet expansions, and an orthogonal least squares algorithm is adapted to estimate the parameters in the expansions. In nonlinear system identification, it is vitally important to determine the model structure or to decide which terms should be included in the model and this is discussed in section 4. In section 5, satellite data relating to magnetic storms is analysed to illustrate the effectiveness of the nonlinear time-varying modelling and identification approach. This is followed by brief conclusions, which are given in section 6.

## 2. System representation: the time-varying NARMAX model

Under some mild assumption, a discrete-time multivariable system with $m$ outputs and $r$ inputs can be descried by the NARMAX (*Nonlinear AutoRegressive Moving Average with eXogenous* inputs) model (Leontaritis and Billings 1985)

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

where

$$y(t) = [y_1(t) \;\; y_2(t) \;\; \cdots \;\; y_m(t)]^T$$
$$u(t) = [u_1(t) \;\; u_2(t) \;\; \cdots \;\; u_r(t)]^T$$
$$e(t) = [e_1(t) \;\; e_2(t) \;\; \cdots \;\; e_m(t)]^T$$

are the system output, input and noise, respectively; $f(\cdot)$ is a nonlinear mapping vector; $n_y$, $n_u$ and $n_e$ are the maximum lags in the output, input and noise; the noise variable $\{e(t)\}$ is a zero mean independent sequence which accommodates the effects of measurement noise, modelling errors and unmeasured disturbances; $e(t)$ is sometimes called the prediction error which is defined as

$$e(t) = y(t) - \hat{y}(t) \tag{2}$$

where

$$\hat{y}(t) = E[y(t) | y^{t-1}, u^{t-1}] \tag{3}$$
$$y^{t-1} = [y(1) \;\; y(2) \;\; \cdots \;\; y(t-1)]^T \tag{4}$$
$$u^{t-1} = [u(1) \;\; u(2) \;\; \cdots \;\; u(t-1)]^T \tag{5}$$
$$E[e(t) | y^{t-1}, u^{t-1}] = 0 \tag{6}$$

A special case of the general NARMAX model (1) is the NARX (*Nonlinear AutoRegressive with eXogenous* inputs) model

$$y(t) = f(y(t-1),\cdots,y(t-n_y),u(t-1),\cdots,u(t-n_u)) + e(t) \tag{7}$$

In most cases the form of the nonlinear function $f(\cdot)$ in (1) and (7) will be unknown. The unknown function $f(\cdot)$ can however be arbitrarily well approximated by polynomial models (Chen and Billings 1989a) or by other functional expressions (Sjoberg et al 1995). Taking the case of a time-varying SISO system as an example and expanding model (1) by defining the function $f(\cdot)$ to be a polynomial of degree $\ell$ gives the representation

$$y(t) = \theta_0(t) + \sum_{i_1=1}^{n}\theta_{i_1}(t)x_{i_1}(t) + \sum_{i_1=1}^{n}\sum_{i_2=i_1}^{n}\theta_{i_1 i_2}(t)x_{i_1}(t)x_{i_2}(t) + \cdots$$
$$+ \sum_{i_1=1}^{n}\cdots\sum_{i_\ell=i_{\ell-1}}^{n}\theta_{i_1 i_2\cdots i_\ell}(t)x_{i_1}(t)x_{i_2}(t)\cdots x_{i_\ell}(t) + e(t) \tag{8}$$

where

$$n = n_y + n_u + n_e$$

and

$$x_\tau(t) = \begin{cases} y(t-\tau) & 1 \le \tau \le n_y \\ u(t-(\tau-n_y)) & n_y + 1 \le \tau \le n_y + n_u \\ e(t-(\tau-n_y-n_u)) & n_y + n_u + 1 \le \tau \le n_y + n_u + n_e \end{cases}$$

Model (8) belongs to the linear-in-the-parameters time-varying regression class of models

$$y(t) = \theta_0(t) + \sum_{m=1}^{M}\theta_m(t)x_m(t) + e(t) \tag{9}$$

where

$\theta_0(t)$ is a function to indicate the intrinsic varying trend of $y(t)$, usually $\theta_0(t) = const$,

$\theta_m(t)$ is a time-varying parameter,

$$x_m(t) = y(t-n_{y1})\cdots y(t-n_{y,N_y})u(t-n_{u1})\cdots u(t-n_{u,N_u})\ e(t-n_{e1})\cdots e(t-n_{e,N_e}) \tag{10}$$

$m = 1,2,\cdots,M$,

$1 \le n_{y1}, n_{y2},\cdots,n_{y,N_y} \le n_y$,

$1 \le n_{u1}, n_{u2},\cdots,n_{u,N_u} \le n_u$,

$1 \le n_{e1}, n_{e2},\cdots,n_{e,N_e} \le n_e$, $N_y, N_u, N_e \ge 0$,

and

$N_y = 0$ indicates that $x_m(t)$ contains no $y(\cdot)$ terms,

$N_u = 0$ indicates that $x_m(t)$ contains no $u(\cdot)$ terms,

$N_e = 0$ indicates that $x_m(t)$ contains no $e(\cdot)$ terms.

The degree of a multivariate polynomial is defined as the highest order of all the terms, for example, the degree of the polynomial $h(x,y,z) = (x-a)^4 + (y-b)(z-c) + (x-a)^2(y-b)(z-c)^2$ is 2+1+2=5. Similarly, a NARMAX model with polynomial degree $\ell$ means that the order of each term in the model is not higher than $\ell$.

If all the coefficients $\theta_m(t)$ are not dependent on the index $t$, that is, $\theta_m(t) = \theta_m$=*const*, $m = 0,1,\cdots,M$, then model (1) reduces to the conventional NARMAX model. As a general and natural representation for a wide class of linear and nonlinear systems, model (8) includes, as special cases, several traditional 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 1999).

## 3. Expanding the time-varying coefficients into multiresolution wavelet expansions

In this section a routine, which translates the time-varying model Eq (9) into a time-invariant regresson model, is explored by expanding each time-varying coefficient as a multiresolution wavelet expansion. Problems relating to the selection of the model structure and terms, and how to determine the model orders $n_y$, $n_u$ and $n_e$, as well as other problems, will be discussed in the next section.

If each coefficient $\theta_m(t), m = 0,1,\cdots,M$, in Eq (9) can be approximated by a linear combination of the basis functions $\xi_l(t)$, $l = 1,2,\cdots,L$

$$\theta_m(t) = \sum_{l=1}^{L} c_l^{(m)} \xi_l(t), \quad m = 0,1,\cdots,M, \tag{11}$$

then the identification can be implemented by estimating the time-invariant coefficients $\{c_l^{(m)}\}_{l=1,L}^{m=1,M}$. Substituting (11) into (9), gives a set of equations that are linear-in-the parameters and which can be solved by several methods in the least-squares class of algorithms.

In the approach proposed here, multiresolution wavelets and scaling functions (Mallat 1989) are chosen as the basis functions for the time dependent coefficients $\theta_m(t)$. With this choice, each coefficient $\theta_m(t)$ can be expressed as the resolution-limited representation

$$\theta_m(t) = \sum_{k=k_0}^{K_{j_0}} \alpha_{j_0,k}^{(m)} \phi_{j_0,k}(t) + \sum_{j=j_0}^{J} \sum_{k=k_0}^{K_j} \beta_{j,k}^{(m)} \varphi_{j,k}(t) \tag{12}$$

where $\varphi(t)$ and $\phi(t)$ are the basic mother wavelet and the basic scaling function, respectively, and $J$ is the highest resolution level . The function $g_{j,k} \in \{\phi_{j,k}(t), \varphi_{j,k}(t), j,k \in Z\}$ is defined as

$$g_{j,k}(t) = 2^{\frac{j}{2}} g(2^j t - k) \tag{13}$$

Substituting (12) into (9), yields

$$y(t) = \sum_{m=0}^{M} \sum_{k=k_0}^{K_{j0}} \alpha_{j_0,k}^{(m)} \phi_{j_0,k}(t) x_m(t) + \sum_{m=0}^{M} \sum_{j=j_0}^{J} \sum_{k=k_0}^{K_j} \beta_{j,k}^{(m)} \varphi_{j,k}(t) x_m(t) + \varepsilon(t) \tag{14}$$

Where $\varepsilon(t)$ is the modelling error, and $x_0(t) = 1$. Eq.(14) is a time-invariant equation with respect to the parameters of the wavelet coefficients $\{\alpha_{j_0,k}^{(m)}\}$ and $\{\beta_{j,k}^{(m)}\}$, $m = 0,1,\cdots,M$.

Introduce the following multiresolution wavelet expansion matrices

(*M1*) $P(t) = [x_0(t), \ x_1(t), \ \cdots, x_M(t)]$

(*M2*) $\Gamma(t) = [\phi_{j_0,k_0}(t), \ \phi_{j_0,k_0+1}(t), \ \cdots, \phi_{j_0,K_{j_0}}(t)]$

(*M3*) $A(t) = P(t) \otimes \Gamma(t)$

(*M4*) $\Lambda_j(t) = [\varphi_{j,k_0}(t), \ \varphi_{j,k_0+1}(t), \ \cdots, \varphi_{j,K_j}(t)]$ , $j = j_0, j_0+1, \cdots, J$

(*M5*) $B_j(t) = P(t) \otimes \Lambda_j(t)$, $j = j_0, j_0+1, \cdots, J$

(*M6*) $B(t) = [B_{j_0}^T(t), B_{j_0+1}^T(t), \cdots, B_J^T(t)]$

(*M7*) $\alpha^T = [\alpha^{(0)}_{j_0,k_0}, \alpha^{(0)}_{j_0,k_0+1}, \cdots, \alpha^{(0)}_{j_0,K_{j_0}} \vdots \alpha^{(1)}_{j_0,k_0}, \alpha^{(1)}_{j_0,k_0+1}, \cdots, \alpha^{(1)}_{j_0,K_{j_0}} \vdots \cdots \vdots \alpha^{(M)}_{j_0,k_0}, \alpha^{(M)}_{j_0,k_0+1}, \cdots, \alpha^{(M)}_{j_0,K_{j_0}}]$

(*M8*) $\beta_j^T = [\beta^{(0)}_{j,k_0}, \beta^{(0)}_{j,k_0+1}, \cdots, \beta^{(0)}_{j,K_j} \vdots \beta^{(1)}_{j,k_0}, \beta^{(1)}_{j,k_0+1}, \cdots, \beta^{(1)}_{j,K_j} \vdots \cdots \vdots \beta^{(M)}_{j,k_0}, \beta^{(M)}_{j,k_0+1}, \cdots, \alpha^{(M)}_{j,K_j}]$,

$j = j_0, j_0+1, \cdots, J$,

(*M9*) $\beta^T = [\beta^T_{j_0}, \beta^T_{j_0+1}, \cdots, \beta^T_J]$

where the symbol "$\otimes$" denotes the *Kronecker product*.

Now, (9) can be expressed as

$$y(t) = A(t)\alpha + B(t)\beta + \varepsilon(t) \tag{15}$$

If $N$ measurements of the input and output signals are available, (15) can be written in the compact matrix form

$$Y = H\vartheta + \varepsilon \tag{16}$$

where

$$Y^T = [y(1) \;\; y(2) \;\; \cdots \;\; y(N)]$$
$$\varepsilon = [\varepsilon(1), \varepsilon(2), \cdots, \varepsilon(N)]$$
$$H = \begin{bmatrix} A(1) & B(1) \\ A(2) & B(2) \\ \vdots & \vdots \\ A(N) & B(N) \end{bmatrix}$$
$$\vartheta^T = [\alpha^T \;\; \beta^T] = [\vartheta_0, \vartheta_1, \cdots, \vartheta_{\Gamma(\alpha,\beta)}] \tag{17}$$

The symbol $\Gamma(\alpha, \beta)$ in (17) is used to indicate the number of unknown parameters in (14) or (15). This will be dependent on the selection of the basis functions and the truncation in the wavelet expansions. If the mother wavelet $\varphi(t)$ and the basic scaling function $\phi(t)$ are compactly supported as in the case of the Haar and the B-spline wavelets and scaling functions (Chui 1992), then the number of unknown parameters can be easily determined from (14) or (*M7*)-(*M9*).

The parameter vector $\vartheta$ in (15) can now be estimated using a least-squares-based algorithm. The *orthogonal forward regression* (OFR) algorithm (Billings et al 1989a) is recommended. The advantage of using this orthogonal algorithm is that the contributions of candidate terms are decoupled and consequently the significance of each term in the regression model can be measured based on the corresponding *error reduction ratio* (ERR)(Chen et al 1989b).

## 4. Model sturcture determination, variable selection and model validation

In contrast with linear system identification, several fundamental problems arise in the identification of time-varying nonlinear systems, namely, *i*) how to determine the model order? *ii*) how to determine the model structure and select the model terms? *iii*) how to measure model quality?

### 4.1 *Determining the model order and selecting significant variables*

In linear system identification, the model order determination problem has been extensively studied and successfully resolved. The most widely accepted information theoretic criteria are Akaike's FPE, AIC and BIC criteria (Akaike 1969,1974,1978), which have recently been applied to time-varying linear system identification (Tsatsanis and Giannakis 1993, Zheng et al 2001).

For nonlinear system identification based on the structure (1), however, the model order determination problem is much more complex than the linear model case. The NARMAX methodology provides one solution to this problem based on the *orthogonal least squares* (OLS) algorithm (Billings et al 1989a,b, Hong and Harris 2001) and the *error reduction ratio* (ERR) (Chen et al 1989b), combined with model validity tests which will be discussed in the next subsection. Other methods including conditional probability analysis (Savit and Green 1991) and sensitivity analysis (He and Asada 1992) are also available to determine model orders and select significant variables.

## 4.2 *Determining the model structure and selecting model terms*

It is vital to determine the model structure or which terms should be included in the final model, whatever basis functions are chosen to implement the model (1). For example for the polynomial NARMAX model (8), the maximum number of candidate terms is $n_{term} = \begin{pmatrix} m+\ell \\ \ell \end{pmatrix} = (m+\ell)!/[m!\ell!]$, where $m = n_y + n_u + n_e$.

This means that the candidate terms will dramatically increase as the system model orders ($n_y$, $n_u$ and $n_e$) and the polynomial degree $\ell$ increase. Similarly the number of potential unknown parameters $\{\alpha_{j_0,k}^{(m)}\}$ and $\{\beta_{j,k}^{(m)}\}$ in Eq. (16) can become large.

Several approaches have been developed to determine which terms are significant and should be included in the nonlinear model (see, for example, Billings et al 1986, 1988, 1989a,b, Mendes and Billings 2001). These approaches were derived for conventional NARMAX models but are still valid for the time-varying NARMAX model (8) and can be used for detecting the model structure and selecting the model terms in time-varying nonlinear system identification.

## 4.3 *Model validation*

Several methods of model validation have been proposed for nonlinear system identification (Billings et al 1986,1994,1995). Let $\hat{f}(\cdot)$ represent an estimate model for the system $f(\cdot)$ and let the residuals $\varepsilon(\cdot)$ be given by

$$\varepsilon(t) = y(t) - \hat{f}(y(t-1),\cdots,y(t-n_y),u(t-1),\cdots,u(t-n_u),\varepsilon(t-1),\cdots,\varepsilon(t-n_e)) \tag{18}$$

If the model structure and parameter values are correct, $\varepsilon(\cdot)$ will be unpredictable from all linear and nonlinear combinations of past inputs and outputs. For nonlinear SISO systems, this can be tested by computing the following correlation functions (Billings and Voon 1986)

$$\begin{cases} \gamma_{\varepsilon\varepsilon}(\tau) = \delta(\tau), & \forall \tau \\ \gamma_{u\varepsilon}(\tau) = 0, & \forall \tau \\ \gamma_{(u\varepsilon)\varepsilon}(\tau) = 0, & \tau \geq 1 \\ \gamma_{\overline{u}^2\varepsilon^2}(\tau) = 0, & \forall \tau \\ \gamma_{\overline{u}^2\varepsilon}(\tau) = 0, & \forall \tau \end{cases} \tag{19}$$

where $\overline{u}^2(t) = u^2(t) - \overline{u^2(t)}$, the bar indicates time averaging; the correlation function $\gamma_{\xi\zeta}(\cdot)$ is defined as

$$\gamma_{\xi\varsigma}(\tau)=\frac{\sum_{t=1}^{N-\tau}\xi(t)\varsigma(t+\tau)}{\sqrt{\left(\sum_{t=1}^{N-\tau}\xi^2(t)\right)\left(\sum_{t=1}^{N-\tau}\varsigma^2(t)\right)}} \tag{20}$$

The first two conditions in (19) form the traditional test used in linear system identification. The remaining three conditions involve cross correlation tests between the input and residuals, by which all possible omitted nonlinear terms can be detected. In practice, if these correlation functions fall within the confidence intervals at a given level $\alpha(0<\alpha<1)$, say $\alpha=0.05$ or $\alpha=0.1$, the model is viewed as adequate and acceptable.

Although the model validation tests are normally justified on the basis of the calculation of correlations between the input and the residuals, Billings and Zhu (1994,1995) showed that the use of the outputs enhances the performance of the tests and allows the number of individual correlation tests to be reduced. When the output is introduced, only two tests are required

$$\begin{cases}\gamma_{(\overline{y\varepsilon})\overline{\varepsilon}^2}(\tau)=\lambda\delta(\tau)\\ \gamma_{(\overline{y\varepsilon})\overline{u}^2}(\tau)=0\end{cases} \quad \text{for } \forall\tau \tag{21}$$

and these can be more efficient in the cases of MIMO system identification.

## 4.4 *An algorithm for the identification of time-varying NARMAX models*

The nonlinear, time-varying system identification algorithm can now be summarized below:

(*i*) Select initial values for $n_y$, $n_u$, $n_e$ and $\ell$ in model (1). If any prior knowledge is available this can be used to prime the model or preselect the model terms, then go to (*iii*).

(*ii*) Determine the model terms using the NARMAX structure detection procedure. Denote the term set as $\{x_m(t)\}_{m=1}^{M}$ and assume that each $x_m(t)$ is expressed as (10).

(*iii*) Form a time-varying model (9) using the model terms $\{x_m(t)\}_{m=1}^{M}$.

(*iv*) Transform the time-varying model into a time-invariant form using the method proposed in section 3.

(*v*) Form the linear-in-the-parameters regression model (16).

(*vi*) Estimate the unknown parameters in the linear-in-the-parameters model (16) using the OFR algorithm.

(*vii*) Test the model validity using (19) or (21). If testing fails, then force any indicated terms into the model (9) and repeat (*iii*)—(*vi*), or go to (*i*) to reset the initial model specification and boundary values.

It should be noted that, the model obtained using the above procedure may not be unique. Different models (different terms and different numbers of terms) might be obtained under different initial conditions. Trade-offs among the models include parsimony, accuracy, and validity. Although it may sometimes be difficult to determine which model is the best, the following criteria can be used to compare the quality of the different models.

Assume $I$ models are available for a given system. Let $\{\hat{y}_t^{(i)}\}_{t=1,N}^{i=1,I}$ be the output of the $i$ th model, and $\{y_t\}_{t=1,N}$ the output measurement of the system.

*(a) Theil's inequality coefficient*

The following Theil's inequality coefficient (Theil 1966) can be used to measure the coherency between $\hat{y}_t^{(i)}$ and $y_t$

$$\rho^{(i)}(y,\hat{y}) = \frac{\sqrt{\frac{1}{N}\sum_{t=1}^{N}(y_t - \hat{y}_t^{(i)})^2}}{\sqrt{\frac{1}{N}\sum_{t=1}^{N} y_t^2} + \sqrt{\frac{1}{N}\sum_{t=1}^{N}[\hat{y}_t^{(i)}]^2}}, \quad i = 1,2,\cdots,I \tag{22}$$

where, $0 \le \rho^{(i)} \le 1$, and the smaller the value of $\rho^{(i)}$ is, the closer between $\{\hat{y}_t^{(i)}\}$ is to $\{y_t\}$. The function $\rho^{(i)}(y,\hat{y})$ also has a very clear geometric meaning (Liu and Wei 1998).

*(b) Relational coefficient*

Define the relationship coefficient between $\hat{y}_t^{(i)}$ and $y_t$ as (Wei and Li 1997)

$$\rho_t^{(i)} = \begin{cases} 1, & \text{if } \left|y_t^{(i)} - y_t\right| \equiv 0 \\ \dfrac{\min\limits_i \min\limits_t \left|y_t^{(i)} - y_t\right| + \xi^{(i)} \max\limits_i \max\limits_t \left|\hat{y}_t^{(i)} - y_t\right|}{\left|y_t^{(i)} - y_t\right| + \xi^{(i)} \max\limits_i \max\limits_t \left|\hat{y}_t^{(i)} - y_t\right|}, & \text{otherwise} \end{cases}, \quad i = 1,2,\cdots,I \tag{23}$$

and the relationship degree as

$$r^{(i)}(y,\hat{y}) \overset{\Delta}{=} \frac{1}{N}\sum_{t=1}^{N} \lambda^{(i)} \rho_t^{(i)}, \quad i = 1,2,\cdots,I \tag{24}$$

where $\xi^{(i)} > 0$ ($i = 1,2,\cdots,I$) are called resolution coefficients. The smaller the value of $\xi^{(i)}$, the higher the resolution is. Usually $\xi^{(i)}$ is set to a value between [0, 0.1] or [0, 0.5]. $\lambda^{(i)}$ are weighted coefficients which can be given different values according to the features of the original data, for example, in different time periods of the process, different values of $\lambda^{(i)}$ might be required. The larger the value of $r^{(i)}$, the closer $\{\hat{y}_t^{(i)}\}$ is to $\{y_t\}$.

## 5. Applications

As an example of the algorithm described above, a nonlinear time-varying model of a dynamic process related to the terrestrial magnetosphere will be identified.

### 5.1 *Application to geomagnetic activity*

The sun is a source a continuous flow of charged particles, ions and electrons called the Solar wind. The terrestial magnetic filed shields the Earth from the solar wind, and forms a cavity in the solar wind flow that is called the terrestrial magnetosphere. The magnetopause is a boundary of the cavity and its position on the day side (sunward side) of the magnetosphere can be determined as the surface where there is a balance between the dynamic pressure of the solar wind outside the magnetosphere and the pressure of the terrestrial magnetic field inside. A complex current system exists in the magnetosphere to support the complex structure of the magnetosphere and the magnetopause. Changes in the solar wind velocity, density or magnetic field lead to changes in the shape of the magnetopause and variations in the magnetospheric current system. In addition if the solar wind magnetic field has a component directed towards the south a reconnection between the terrestrial magnetic field and the solar wind magnetic field is initiated. Such a reconnection results in a very drastic modification to the magnetospheric current system and this phenomenon is referred to as magnetic storms.

During a magnetic storm, which can last for hours, the magnetic field on the Earth's surface will change as a result of the variations of the magnetospheric current system. Changes in the magnetic field induces considerable currents in long conductors on the terrestrial surface such as power lines and pipe-lines. Unpredicted currents in power lines can lead to the blackouts of huge areas, the Ontario Blackout is just one recent example. Other undesirable effects include increased radiation to crews and passengers on long flights, and effects on communications and radio-wave propagation. Forecasting of geomagnetic storms is therefore highly desirable and can aid the prevention of such effects. The $D_{st}$ index is used to measure the disturbance of the geomagnetic field in the magnetic storm. Numerous studies of correlations between the solar wind parameters and magnetospheric disturbances show that the product of the solar wind velocity $V$ and the southward component of the magnetic field, quantified by $B_s$, represents the input that can be considered as the input to the magnetosphere. Denote the multiplied input by $VB_s$.

The input-output data consist of 4344 hours of measurements of the solar wind parameter $VB_s$ and the $D_{st}$ index over the time period from January to June 1979, with a sample period of 1 hour. This is depicted in Fig. 1.

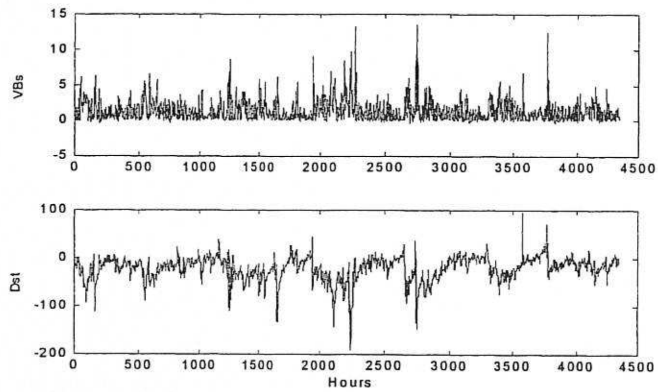


Fig. 1 The input $VB_s$ and the output $D_{st}$ of the dynamic magnetosphere process

## 5.2 *System identification*

A nonlinear, time-varying NARMAX model will be identified to represent the above dynamic process using the first 4000 points. With the initial values set to $n_y = 3, n_u = 10, n_e = 0$ and $\ell = 2$, the NARMAX procedure was applied using an intelligent search strategy (Mao and Billings 1997), and a model term set, which contained 18 terms, was obtained in Eq. (25) and ordered according to the significance of terms

$$S = \{x_m(t)\}_{m=1}^{18} = \left\{ \begin{array}{lll} y(t-1)u(t-8), & u^2(t-1), & u(t-1)u(t-6), \\ u(t-1), & y(t-1), & y(t-3)u(t-5), \\ y^2(t-2), & u^2(t-4), & u(t-2), \\ u(t-6)u(t-7), & u(t-1)u(t-7), & u(t-3), \\ u^2(t-8), & u(t-3)u(t-8), & u(t-2)u(t-5), \\ y^2(t-3), & y(t-2)u(t-2), & y(t-1)u(t-5) \end{array} \right\} \quad (25)$$

where $u$ and $y$ represent $VB_s$ and $D_{st}$ , respectively. A time-varying NARMAX model for the terrestrial magnetosphere process can therefore be expressed as

$$y(t) = \sum_{m}^{18} \theta_m(t) x_m(t) + e(t) \tag{26}$$

where $x_m(t) \in S$ and $x_m(t) \neq x_k(t)$ if $m \neq k$ .

Each $\theta_m(t)$ was then expanded into resolution-limited representation using B-spline wavelets, with the resolution levels from $j = j_0 = 0$ to $j = j_{\max} = J(J$ =2,3,4,5,6) to give the regression models of the form of (16). Estimates of the unknown parameter vector $\hat{\vartheta}$ in (17) were then obtained using the forward regression orthogonal algorithm. Based on the criteria (29) and (30), the highest resolution scale $J$ =4 was chosen. Finally all the 18 time-varying coefficients $\theta_m(t)\,(i = 1,2,\cdots,18)$ were reconstructed using $\hat{\vartheta}$. The estimated time-varying coefficients $\hat{\theta}_m(t)$ over the range of $t \in [0,4000]$ are shown in Fig. 3(A) to Fig. 3(C).

The model validity tests of (19) for the identified model were all within the 95% confidence intervals, and the model was therefore accepted. The comparison between the one-step-ahead (OSA) prediction (the dashed line) and the original measurement data (the solid line), along with the prediction error over the range [0,4344] are shown in Fig. 4, where the value of the covariance of the prediction error is $\sigma^2$=35.11. The part over the range [4001,4344] is expanded and depicted in Fig. 5. Inspection of figure 4 and 5 shows that the model fits the data extremely well.

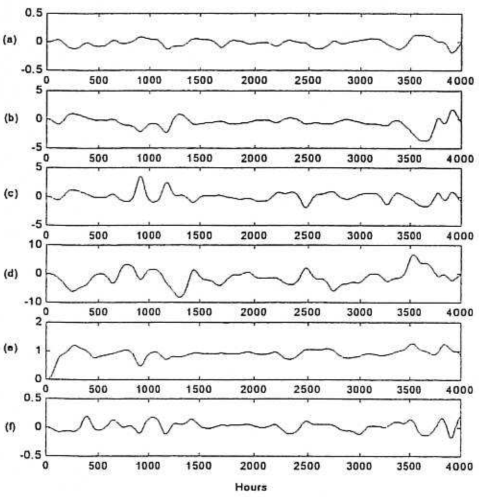


Fig. 3 (A) The estimated time-varying coefficients (a) $\hat{\theta}_1(t)$ to (f) $\hat{\theta}_6(t)$

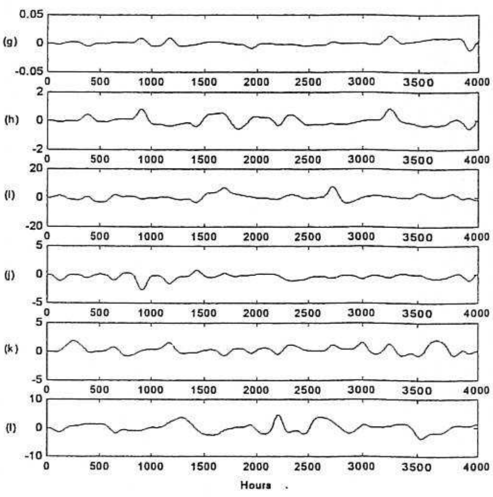


Fig. 3 (B) The estimated time-varying coefficients (g) $\hat{\theta}_7(t)$ to (i) $\hat{\theta}_{12}(t)$

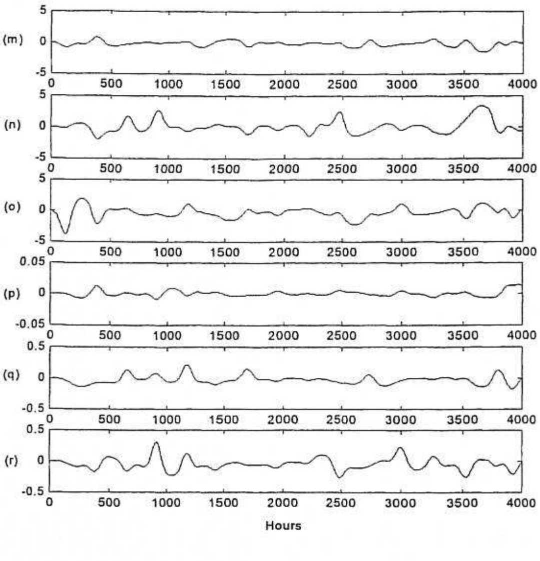


Fig. 3 (C ) The estimated time-varying coefficients (m) $\hat{\theta}_{13}(t)$ to (r ) $\hat{\theta}_{18}(t)$

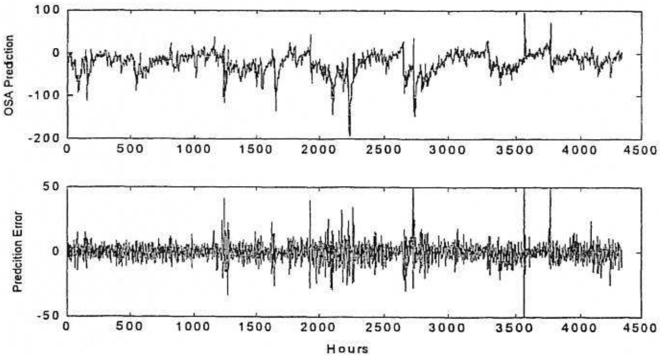


Fig. 4 One-step-ahead prediction(the dashed line) and prediction error

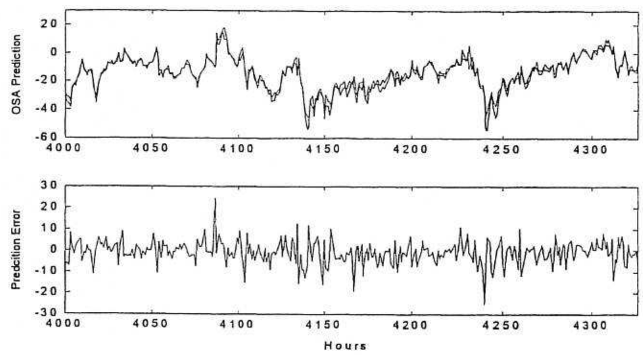


Fig. 5 One-step-ahead prediction(the dashed line) and prediction error over the range of [4001,4344]

## 6. Conclusions

A NARMAX wavelet based identification scheme for general nonlinear time-varying systems has been introduced. Initially a NARMAX parameter estimate procedure is applied to obtain a set of model terms. A nonlinear time-varying model structure can then be formed. This model structure contains several other types of traditional linear and nonlinear models as special cases. By expanding each time-varying coefficient in the model as a wavelet expansion, the time-varying problem is transformed into a linear-in-the-parameters regression problem which can be solved using the orthogonal forward regression algorithm. The identification of a model relating to the terrestrial magnetosphere was used to illustrate the application of the new identification approach.

## Acknowledgment

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

## 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.

Akaike,H.(1969), Fitting autoregressive models for prediction, *Ann. Inst. Statist. Math.*, **21**, 243-247.

Akaike,H.(1974), A new look at the statistical model identification, *IEEE Trans. Automatic Control*, **AC-19**, 716-722.

Akaike,H.(1978), A Bayesian analysis of the minimum AIC procedure, *Ann. Inst. Statist. Math.*, **30A**, 9-14.

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, *Int. Journal of Systems Sci*, 19(8),1559-1568.

Billings,S.A., Chen,S. and Korenberg,M.J.(1989a), 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.(1989b), 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.

Charbonnier,R., Barlaud,M., Alengrin,G., Menez,J.(1987). Results on AR-modelling of nonstationary signals, *Signal Processing*, **12(2)**, 143-151.

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

Chen,S., Billings,S.A., and Luo,W.(1989b), 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., Cowan, C.F.N. and Grant, P.W.(1990), Nonlinear system identification using radial basis functions, *International Journal of Systems Sci.*, **21(12)**, 2513-2539.

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.

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

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

Dembo, A., Zeitouni, O.(1988). Maximum a posteriori estimation of time-varying ARMA processes from noisy observations, *IEEE Trans. Acoust. Speech Process*, **ASSP-36(4)**, 471-476.

Fairfield,D.H.(1992), Advances in magnetospheric and substorm research: 1989-1991, *Journal of Geophysical Research*, **97(A7)**, 10,865-10,874.

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

Gonzalez,W.D., Joselyn,J.A., Kamide,Y., Kroehl,H.W., Rostoker,G., Tsurutani,B.T., and Vasilinuas, V.M. (1994), What is a geomagnetic storm? J. *Geophysical Research*, **99(A4)**, 5771-5792.

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

He X.D. and Asada H.(1992), A new service for identifying orders of input-output models for nonlinear dynamic systems, Proceesings of the American Control Conference, San Francisco, 2520-2523.

Hong, X. and Harris, C. J. (2001), Variable selection algorithm for the construction of MIMO operating point dependent neurofuzzy networks, *IEEE Transactions On Fuzzy Systems*, **9(1)**, 88-101.

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.

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, Z.Z. and Wei, H.L. (1998), *System simulation*. Beijing: Beijing Institute of Technology Press.

Ljung, L. (1983), *Theory and practice of recursive identification.* Cambridge, Mass.; London : MIT Press.

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.

Mao,K.Z. and Billings, S.A.(1997), Algorithms for minimal model structure detection in nonlinear dynamic system identification, *International Journal of Control*, **68(2)**, 311-330.

Martin, W., Flandrin, P.(1985), Wigner-Ville spectral analysis of nonstaionary process, *IEEE Trans Acoust., Speech and Signal Processing*, **ASSP-33(6)**, 1461-1470.

Mendes, E.M.A.M., and Billings, S.A.(2001), An alternative solution to the model structure selection problem, *IEEE Trans on Sys,Man, and Cybernetics--Part A: Systems and Humans*, **31(6)**,597-608.

Morikawa, H.(1990), Adaptive estimation of time-varying model order in the ARMA speech analysis, *IEEE Trans. Signal Processing,* **38(7)**, 1073-1083.

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.

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

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.

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.

Theil, H. (1966), *Applied Economic Forecasting*, Amsterdam: North-Holland Publishing Company.

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

Wei, H.L. and Li, Z.W.(1997), Grey relational analysis and its application to the validation of computer simulation models for dynamic systems, *Chinese Journal of Systems Engineering and Electronics*, **19(2)**, 55-61.

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.

Zheng, Y., Tay, D. B. H., Lin, Z. (2001), Modelling general distributed nonstationary process and identifying time-varying autoregressive system by wavelets: theory and application, *Signal Processing*, **81(9)**, 1823-1848.