Monograph:

Wei, H.L., Billings, S.A. and Balikhin, M.A. (2003) Wavelet Based Nonparametric NARX Models for Nonlinear Input-Output System Identification. Research Report. ACSE Research Report 831 . Department of Automatic Control and Systems Engineering

Wavelet Based Nonparametric NARX Models for Nonlinear Input-Output System Identification

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

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

Research Report No. 831

January 2003

Wavelet Based Nonparametric NARX Models for Nonlinear Input-Output System Identification

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

Abstract— Wavelet based nonparametric additive NARX models are proposed for nonlinear input-output system identification. By expanding each functional component of the nonparametric NARX model into wavelet multiresolution expansions, the nonparametric estimation problem becomes a linear-in-the-parameters problem, and least-squares-based methods such as the orthogonal forward regression (OFR) approach can be used to select the model terms and estimate the parameters. Wavelet based additive models, combined with model order determination and variable selection approaches, are capable of handling problems of high dimensionality.

Keywords—Nonlinear system identification; wavelets; NARX model; nonparametric, additive models.

1. Introduction

In the past few decades, system identification and analysis methods for nonlinear systems have been widely studied with many applications in approximation, prediction and control. Several nonlinear models have been proposed in the literature including the NARMAX (*Nonlinear Autoregressive Moving Average with eXogenous* inputs) model representation which was initially proposed by Leontaritis and Billings (1985). NARMAX can describe a wide range of nonlinear dynamic systems and includes several other linear and nonlinear model types, including the Volterra, Hammerstein, Wiener, ARMAX, and NARX models as special cases (Pearson 1995).

A general desire in data-driven modelling procedures for nonlinear systems is to develop efficient model construction procedures that overcome the curse-of-dimensionality. Several authors have studied this problem and a key idea of the methods that have been developed is to represent a multivariate function as additive superpositions of functions of fewer variables. The projection pursuit algorithm(Friedman 1981), multi-layer perceptron (MPL) architecture(Haykin 1994), and radial basis functions(Chen et al 1990, 1992) are among these representations for multivariate functions. Although Kolmogrov's theorem (Lorentz 1996), which states that any continuous function of n-variables can be completely specified by a function of a single argument, guarantees the existence of a univariate (continuous) function that completely characterises any continuous n-variable function, currently there are neither transparent methods to get a univariate function, nor numerically feasible algorithms to compute it. The existing strategies that attempt to approximate general functions in high dimensionality are based on additive functional models (Friedman 1991) , in which the additive functions are often referred to as *functional components*(Stone 1985). The functional components can be arbitrary functions with fewer arguments and with global or local properties. Kernel functions, splines, polynomials and other basis functions can all be chosen as functional components (Hastie and Tibshirani 1990).

In this paper, wavelet-based NARX models are considered. Wavelets, which have excellent approximation capabilities, are chosen as the functional components in the additive models. The wavelet analysis procedure involves adopting a wavelet prototype function, called the *mother wavelet* or simply *wavelet*. Temporal analysis is performed with a contracted, high-frequency version of the same function. Because the signal to be studied can be represented in terms of a wavelet expansion, data operations can be performed using only the corresponding wavelet coefficients. By expanding the functional components using wavelet basis functions, the additive models then become an ordinary linear-in-the-parameters problem which can be solved using least-squares of algorithms. The new wavelet-based additive routine, combined with model-order determination and variable selection approaches (Chen et al 1989, Billings et al 1990, Savit and Green 1991, He and Asada 1992), is capable of handling problems of moderately high dimensionality.

2. The general form of additive NARX models

The NARX model

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

is often used to describe the input-output relationship for nonlinear systems, where f is an unknown nonlinear mapping, $u(t)$ and $y(t)$ are the sampled input and output sequences, $\varepsilon(t)$ is an independent identically distributed random variable, and n_u, n_y are the maximum input and output lags, respectively.

The NARX model is a special case of the NARMAX model, which takes the form of the nonlinear difference equation(Leontaritis and Billings 1985):

$$y(t) = f(y(t-1),\cdots,y(t-n_y),u(t-1),\cdots,u(t-n_u),e(t-1),\cdots,e(t-n_e)) + e(t) \tag{2}$$

where the noise variable $e(t)$ with maximum lag n_e, is immeasurable but is assumed to be bounded and uncorrelated with the inputs. The model (2) relates the inputs and outputs and takes into account the combination effects of measurement noise, modelling errors and unmeasured disturbances represented by the variable $e(t)$.

Consider the NARX model (1) and assume that the nonlinear mapping f can be expressed as a finite set of hierarchical correlated functions expanded in terms of the lagged output and input variables $y(t-i)$ and $u(t-j)$ such that

$$\begin{aligned} y(t) = f_0 &+ \sum_{i=1}^{n} f_i(x_i(t)) + \sum_{1\le i\le j\le n} f_{ij}(x_i(t),x_j(t)) + \sum_{1\le i\le j\le k\le n} f_{ijk}(x_i(t),x_j(t),x_k(t)) + \cdots + \\ &\sum_{1\le i_1\le\cdots\le i_v\le n} f_{i_1 i_2\cdots i_v}(x_{i_1}(t),x_{i_2}(t),\cdots,x_{i_v}(t)) + \cdots + \sum_{1\le i_1\le\cdots\le i_n\le n} f_{i_1 i_2\cdots i_n}(x_{i_1}(t),x_{i_2}(t),\cdots,x_{i_n}(t)) + \varepsilon(t) \end{aligned} \tag{3}$$

where $x_i(t) = y(t-i)$ for $i = 1,2,\cdots,n_y$ and $x_i(t) = u(t-i)$ for $i = n_y+1, n_y+2,\cdots,n$ with $n = n_y + n_u$. The zero-th order functional component f_0 is a constant to indicate the intrinsic varying trend of $y(t)$; the first order functional components $f_i(x_i(t))$ represents the independent contribution to $y(t)$ which arises by the action of the ith variable $x_i(t)$ alone; the second order functional components $f_{ij}(x_i(t),x_j(t))$ gives the interacting contribution to $y(t)$ by the input variables $x_i(t)$ and $x_j(t)$, etc. The last term $f_{12\cdot n}(x_1,x_2,\cdots,x_n)$ contains any residual n-th order correlated contribution of the input variables (Li et al 2001).

Experience shows that the representation of up to second order of terms in model (3)

$$y(t) = f_0 + \sum_{i=1}^{n} f_i(x_i(t)) + \sum_{i=1}^{n}\sum_{j=i}^{n} f_{ij}(x_i(t),x_j(t)) + \varepsilon(t) \tag{4}$$

can often provide a satisfactory description of $y(t)$ for many high dimensional problems providing that the input variables are properly selected (Li et al 2001).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.

In practice, many kinds of functions, such as kernel functions, splines, polynomials and other basis functions can be chosen as functional components in model (3). In the present study, however, wavelets are chosen as the functional components to express the additive model.

3. Expanding the additive NARX models using multiresolution wavelet bases

3.1 *Wavelets multiresolution expansions*

From wavelet theory (Chui 1992), any given function $g \in L^2(R^d)$ can be approximately expressed as a wavelet expansion

$$g(x_1, x_2, \cdots, x_d) = \sum_{\xi_j \in \Omega} w_j \xi_j (x_1, x_2, \cdots, x_d) \tag{5}$$

where

$$\Omega = \left\{ \xi_{(a_i, b_i)}(x) = a_i^{\frac{d}{2}} \xi\left(\frac{x - b_i}{a_i} \right) : x \in R^d, i \in Z, a_i \in R^+, b_i \in R^d \right\} \tag{6}$$

$\xi(\cdot)$ is a basic wavelet or scaling function, $a_i \in R^+$ and $b_i \in R^d$ are dilation and shift factors. Restricting the double index to a regular grid gives a special case to form a wavelet frame defined as

$$\Omega_1 = \{ \xi_{j,k}(x) = \alpha^{\frac{jd}{2}} \xi(\alpha^j x - \beta k) : j \in Z, k \in Z^d \} \tag{7}$$

where the scalar factors α and β are defined as the dilation and translation steps for discretization. The most popular choice is to restrict the dilation and translation parameters to a dyadic lattice as $\alpha = 2$, $\beta = 1$.

Multidimensional wavelet decompositions (expansions) can be defined by taking the *tensor product* of the one-dimensional scaling and wavelet functions (Mallat 1989). Let $g \in L^2(R^d)$, then $g(x)$ can be represented by the multiresolution wavelet series as

$$g(x_1, \cdots, x_d) = \sum_k \alpha_{j_0,k} \Phi_{j_0,k}(x_1, \cdots, x_d) + \sum_{j \geq j_0} \sum_k \sum_{l=1}^{2^d - 1} \beta_{j,k}^{(l)} \Psi_{j,k}^{(l)}(x_1, \cdots, x_d) \tag{8}$$

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

$$\Phi_{j_0,k}(x_1, \cdots, x_d) = 2^{j_0 d/2} \prod_{i=1}^{d} \phi(2^{j_0} x_i - k_i) \tag{9}$$

$$\Psi_{j,k}^{(l)}(x_1, \cdots, x_d) = 2^{jd/2} \prod_{i=1}^{d} \eta^{(i)}(2^j x_i - k_i) \tag{10}$$

with $\eta^{(i)} = \phi$ or φ (scalar scaling function and 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

$$\begin{cases} \Phi_{j,k_1,k_2}(x, y) = \phi_{j,k_1}(x)\phi_{j,k_2}(y) \\ \Psi_{j,k_1,k_2}^{(1)}(x, y) = \phi_{j,k_1}(x)\varphi_{j,k_2}(y) \\ \Psi_{j,k_1,k_2}^{(2)}(x, y) = \varphi_{j,k_1}(x)\phi_{j,k_2}(y) \\ \Psi_{j,k_1,k_2}^{(3)}(x, y) = \varphi_{j,k_1}(x)\varphi_{j,k_2}(y) \end{cases} \tag{11}$$

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 (basis functions) (Billings and Coca 1999).

3.2 *Expanding the functional components using wavelets*

Expanding each functional component in model (3) or (4) into the multiresolution wavelet expansions (8), an ordinary linear-in-the-parameters equation can be obtained. Consider the model (4), this can be expanded using wavelet multiresolution expansions as follows.

$$f_p(x_p(t)) = \sum_k \alpha_{j_1,k}^{(1)} \phi_{j_1,k}(x_p(t)) + \sum_{j \geq j_1} \sum_k \beta_{j,k}^{(1)} \varphi_{j,k}(x_p(t)),\ p = 1,2,\cdots,n \tag{12}$$

$$\begin{aligned} f_{pq}(x_p(t), x_q(t)) = & \sum_{k_1} \sum_{k_2} \alpha_{j_2;k_1,k_2}^{(2,1)} \phi_{j_2,k_1}(x_p(t)) \phi_{j_2,k_2}(x_q(t)) \\ & + \sum_{j \geq j_2} \sum_{k_1} \sum_{k_2} \beta_{j;k_1,k_2}^{(2,1)} \phi_{j,k_1}(x_p(t)) \varphi_{j,k_2}(x_q(t)) \\ & + \sum_{j \geq j_2} \sum_{k_1} \sum_{k_2} \beta_{j;k_1,k_2}^{(2,2)} \varphi_{j,k_1}(x_p(t)) \phi_{j,k_2}(x_q(t)) \\ & + \sum_{j \geq j_2} \sum_{k_1} \sum_{k_2} \beta_{j;k_1,k_2}^{(2,3)} \varphi_{j,k_1}(x_p(t)) \varphi_{j,k_2}(x_q(t)),\ 1 \leq p \leq q \leq n \end{aligned} \tag{13}$$

Inserting Eqs (12) and (13) into (4) yields a linear-in-the-parameters equation with respect to the wavelet coefficients $\alpha_{j_1,k}^{(1)}$, $\beta_{j_1,k}^{(1)}$, $\alpha_{j_2;k_1,k_2}^{(1,1)}$, $\beta_{j;k_1,k_2}^{(2,i)}$ (i=1,2,3). This can be solved using least squares type algorithms. The orthogonal forward regression (OFR) approach (Billings et al 1989, Chen et al 1989), which has been widely applied in system identification and parameter estimation, is recommended and will be adopted in this paper. Notice that the wavelet networks proposed by Zhang (1997) can be considered as special cases of the wavelet based nonparametric additive model (3), where only the last functional component $f(x_1, x_2, \cdots, x_n)$ is considered and expanded using the wavelet expansion (5).

4. An example: a terrestrial geomagnetic system

Figure 1 shows 2172 data points of the measurements of the solar wind parameter VBs (input) and Dst index (output) with a sample period of 2 hours. In order to fit a model, 6 significant variables $Dst(t-1)$, $Dst(t-2)$, $VBs(t-1)$, $VBs(t-2)$, $VBs(t-3)$ and $VBs(t-5)$ were chosen initially using variable selection procedures.

Let $\tilde{x}_1(t) = Dst(t-1)$, $\tilde{x}_2(t) = Dst(t-2)$, $\tilde{x}_3(t) = VBs(t-1)$, $\tilde{x}_4(t) = VBs(t-2)$, $\tilde{x}_5(t) = VBs(t-3)$, $\tilde{x}_6(t) = VBs(t-5)$. Normalize $\tilde{x}(t)$ by setting $x_i(t) = (\tilde{x}_i(t) - Dst_{\max})/(Dst_{\max} - Dst_{\min})$ for $i = 1,2$ and $x_i(t) = (\tilde{x}_i(t) - VBs_{\min})/(VBs_{\max} - VBs_{\min})$ for i=3,4,5,6 so that $x_i(t) \in [0,1]$ over the time interval concerned, where $Dst_{\min}$ and $Dst_{\max}$ are the minimum and maximum values for the Dst index (the output), $VBs_{\min}$ and $VBs_{\max}$ are the minimum and maximum values for the solar wind parameter VBs (the input). The normalized variables $x_i(t)$ were used to identify a wavelet based additive model.

The first 1500 points were used for identification and the remaining 672 points were used to test the model. The nonparametric model (4) was adopted and the functional components in the model were expanded using the wavelet multiresolution expansion (12) and (13), where the 4-th order B-spline wavelet and scaling functions were employed. The wavelet and scaling functions were expanded at scales $j_{max} = j_1 = 4$ for the expansion (12) and $j_{max} = j_2 = 2$ for the expansion (13). The model predicted outputs based on the final model were compared with the measurements of the *Dst* index and the results are illustrated in figure 1, which shows that the model predicts very well.

5. Conclusions

In this paper, a new wavelet based additive modelling approach has been proposed. An advantage of additive models is that the dimensionality can be greatly "reduced" when dealing with problems in high dimensional spaces. The most notable property of wavelets is the excellent local approximation capability. Combining wavelets and additive models makes it possible to represent problems in high dimensionality accurately using low order functional components and enables the identification of nonlinear input-output systems even with severe nonlinearities. The number of candidate regressors in a wavelet based additive model depends on the wavelet basis (or scaling) functions and the chosen scaling levels. High scaling levels (high resolution) could perhaps improve the approximation accuracy but can result in over fitting of the model which will contain a large number of regressors, some of which may be redundant. This problem can be overcome by performing a redundant-regressor elimination procedure and significant regressor selection approach. The results from the illustrative example has demonstrated the efficiency of the modelling approach presented.

Acknowledgment

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

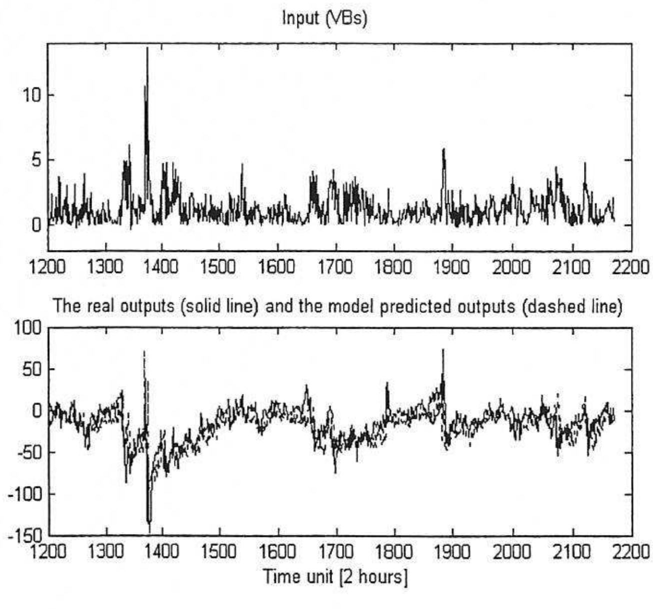

Figure 1. The identification results of the terrestrial magnetosphere process of the Dst index. (a) The solar wind VBs (input); (b) The model predicted outputs (the dashed line) and the real measurements of the Dst index (the solid line).

References

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

Billings,S. A.,Chen, S. and Korenberg, M. J.,1989,Identification of MIMO non-linear systems using 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.

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.,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.

Chen,S.,Billings, S. A. and Grant, P. W.,1992, Recursive hybrid algorithm for nonlinear system identification using radial basis function network, *International Journal of Control.*, **55(5)**,1051-1070.

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

Friedman,J.H. and Stuetzle, W., 1981, Projection pursuit regression, *Journal of the American Statistical Association*, **76(376)**, 817-823.

Friedman,J. H.,1991 Multivariate adaptive regression splines, *The Annals of Statistics*, **19(1)**, 1-67.

Haykin,S., 1994, *Neural networks: a comprehensive foundation*. New York : Macmillan; Oxford : Maxwell Macmillan International, 1994.

Hastie,T.J. and Tibshirani, R.J.,1990, *Generalized additive models* (London; Glasgow; Weinheim; New York; Tokyo; Melbourne; Madras: Chapman & Hall).

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.

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.

Li,G.,Rosenthal, C. and Rabitz,H.,2001,High dimensional model representation, *The Journal of Physical Chemistry*, **105(33)**, 7765-7777.

Lorentz,G. G.,Golitschek, M.A. and Makovoz,Y.,1996, *Constructive approximation* (New York : Springer).

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.

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

Savit,R. and Green, M.,1991, Time series and dependent variables, *Physica D*, **50(1)**, 95-116.

Stone,C. J.,1985, Additive regression and other nonparametric models, *The Annals of Statistics*, **13(2)**, 689-795.

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