Monograph:

Wei, H.L., Billings, S.A. and Balikhin, M.A. (2002) Wavelet Based Nonparametric Additive Models for Nonlinear System Identification and Prediction. Research Report. ACSE Research Report 826 . Department of Automatic Control and Systems Engineering

Wavelet Based Nonparametric Additive Models for Nonlinear System Identification and Prediction

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

RERVM COGNOSCERE CAVSAS

Research Report No. 826

July 2002

Wavelet Based Nonparametric Additive Models for Nonlinear System Identification and Prediction

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
w.hualiang@sheffield.ac.uk, s.billings@sheffield.ac.uk, m.balikhin@sheffield.ac.uk

Abstract— Wavelet based nonparametric additive models are considered for nonlinear system identification. Additive functional component representations are an important class of models for describing nonlinear input-output relationships, and wavelets, which have excellent approximation capabilities, can be chosen as the functional components in the additive models. Wavelet based additive models, combined with model order determination and variable selection, are capable of handling problems of high dimensionality. Examples are given to demonstrate the efficiency of this new modelling approach.

Keywords—Nonparametric additive models; wavelets; nonlinear system identification; NARMAX model; prediction.

1. Introduction

In the past few decades, model representations and analysis for nonlinear systems have gained much attention for the purpose of approximation, prediction and control. Many nonlinear models have been proposed in the literature and many applications have been reported. Among the existing approaches a particular nonlinear modelling methodology, which has been extensively studied over the past two decades, is the NARMAX (*Nonlinear Autoregressive Moving Average with eXogenous* inputs) method founded on the NARMAX input-output model representation initially proposed in [1]. The NARMAX model is the nonlinear extension of the ARMAX model used in linear modelling and system identification. 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, NARX (*Nonlinear AutoRegressive with eXogenous* inputs), as special cases [2].

A general desire in data-driven modelling procedures for nonlinear systems is the ability to obtain the system input-output mapping by means of finite samples in an n-dimensional space. The model output will be a function of n variables taking values over the n-dimensional space. For the problems of high dimensionality (with large n), several approaches have been proposed to overcome the difficulty of the curse-of-dimensionality. The main idea of these approaches is to represent a multivariate function as additive superpositions of functions of fewer variables. The projection pursuit algorithm [3], multi-layer perceptron (MPL) architecture [4], and radial basis functions [5]-[7] are among these representations for multivariate functions. Although Kolmogrov's theorem [8], which states that any continuous function of n-variables can be completely specified by a function of a single argument, guarantees the existence of the 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 [9], in which the additive functions are often referred to as *functional components* [10]. The functional components can be arbitrary functions with fewer arguments and global or local properties. Kernel functions, splines, polynomials and other basis functions can all be chosen as functional components [11].

In this paper, wavelet-based additive models are considered. That is, wavelets, notable for having 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 or function 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 wavelet-based functional components as the combination of the corresponding wavelet basis functions, the additive models then become ordinary linear regression representations. This wavelet-based additive routine, combined with model-order determination and variable selection approaches [12]-[15], is capable of handling problems of moderately high dimensionality.

2. Additive Model Formulations

For noise free nonlinear systems, the following NARX model [1]

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

is often used to describe the input-output relationship. Where, f is an unknown nonlinear mapping, $u(t)$ and $y(t)$ are input and output sequences, $\varepsilon(t)$ is modelling error, n_u and n_y are the maximum input and output lags, respectively.

The experimental data is normally corrupted by measurement noise. The input-output representation should, therefore, include additional stochastic variables. This leads to the NARMAX model, which takes the form of following nonlinear difference equation [1]:

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

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 $e(t)$.

2.1 The General Form of Additive Models

Consider the NARX model (1) and assume that the nonlinear mapping f can be expressed as a finite set of hierarchical function supersitions in terms of the input variables, such that

$$y(t) = f_0 + \sum_{i=1}^{n} f_i(x_i(t)) + \sum_{1\le i\le j\le n} f_{ij}(x_i(t),x_j(t)) + \sum_{1\le i\le j\le k\le n} f_{ijk}(x_i(t),x_j(t),x_k(t)) + \cdots +$$

$$\sum_{1\le i_1\le\cdots\le i_\upsilon\le n} f_{i_1 i_2\cdots i_\upsilon}(x_{i_1}(t),x_{i_2}(t),\cdots,x_{i_\upsilon}(t)) + \cdots + \sum_{1\le i_1\le\cdots\le i_n\le n} f_{i_1 i_2\cdots i_n}(x_{i_1}(t),x_{i_2}(t),\cdots,x_{i_n}(t)) + \varepsilon(t) \qquad (3)$$

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 zeroth 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.

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 [16].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.

2.2 Expanding the Additive Models Using Wavelets

According to wavelet theory [17], 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 dilating and shifting factors. A special case is to restrict the double index to a regular grid to form a wavelet frame defined as follows

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

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 $\alpha = 2$, $\beta = 1$.

Multiresolution wavelet expansions can also be adopted to express a given function $g(x) \in L^2(R^d)$ by representing $g(x)$ as the multiresolution wavelet series as

$$g(x_1, \cdots, x_d) = \sum_k \alpha_{j_0,k} \varphi_{j_0,k}(x_1, \cdots, x_d) + \sum_{j \geq j_0} \sum_k \beta_{jk} \phi_{j,k}(x_1, \cdots, x_d) \tag{8}$$

where $k = (k_1, k_2, \cdots, k_d) \in Z^d$, $\phi(\cdot)$ and $\varphi(\cdot)$ are the matched scaling and wavelet basis functions. However, multiresoluttion expansions are usually studied with orthonormal (or biorthonormal) wavelet bases and thus are often restricted to problems of low dimension.

Expanding each functional component in model (3) or (4) into the wavelet expansions (5) or multiresolution wavelet expansions (8), an ordinary linear regression equation can be obtained. This can be solved using least squares type algorithms. The orthogonal forward regression (OFR) approach [12][13], which has been widely applied in system identification and parameter estimation, is recommended and will be adopted in this paper. It can be seen from these expansions that, the wavelet networks advocated in [18][19] can be considered as special cases of the wavelet based nonparametric additive models considered here.

3. Examples and Applications

In this section, two examples are provided to demonstrate the effectiveness of the wavelet based additive modelling approach. In both examples, the orthogonal forward regression (OFR) algorithm [12][13] was adopted for selecting model terms and the model validation methods proposed in [20][21] were used to verify the fitness of the models.

3.1 A Numerical Example: Duffing-Ueda Oscillator Model

Consider the Duffing-Ueda oscillator model [22]

$$\ddot{y}(t) + k\dot{y}(t) + y^3 = u(t) \tag{9}$$

where $k = 0.1$ and the input $u(t) = A\cos(t)$ with $A = 6.0$. Simulation data were obtained using the fourth-order Runge-Kutta integral method with a step length of $\Delta t = \pi / 200$ sec. To make the simulation data more realistic, white noise with a standard deviation $\sigma_\varepsilon = 0.02$ was added to the final output. The data were then sampled with a sampling interval $T_s = \pi / 50$ sec to generate 2000 input-output data points for the purpose of identification.

Initially, the significant variables $[y(t-1), y(t-2), u(t-1)]$ were chosen using our variable selection procedures. The next step involves fitting an additive model based on the following structure

$$y(t) = \sum_{i=1}^{3} f_i(x_i(t)) + \sum_{i=1}^{3}\sum_{j=i}^{3} f_{ij}(x_i(t), x_j(t)) \tag{10}$$

where $[x_1(t), x_2(t), x_3(t)] = [y(t-1), y(t-2), u(t-1)]$. All the three first order functions f_i were chosen to be the fourth-order B-spline scaling functions [17] and then expanded at scale level 5, i.e., $\alpha = 2$, $\beta = 1$ and j=5 in (7); all the second order functions f_{ij} were chosen to be Gaussian wavelet functions $\varphi(x) = x_1 x_2 \exp[-(x_1^2 + x_2^2)/2]$ and expanded at a fixed scale level 4, i.e., $\alpha = 2$, $\beta = 1$ and j=4 in (7). The first 1000 points were used for parameter estimation, and the latter 1000 points were used to evaluate the efficiency of the resulting model. The original noise-free data and the model predicted output are shown in Fig. 1. This shows that the model works very well, even for the data range [1001,2000].

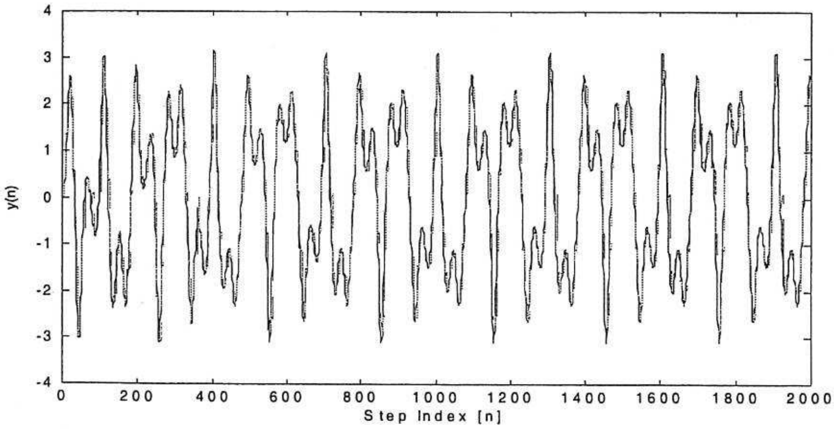

Fig 1. Duffing-Ueda oscillator output and the model predicted output. (The solid line indicates the original noise-free data; The dashed line indicates the model predicted output)

3.2 A Realistic Example: The Terrestrial Magnetosphere

Fig 2(a) and (b) show 4344 data points of the measurements of the solar wind VB_s (input) and D_{st} index (output) over the time period from January to June 1979, with a sample period of 1 hour. The output data shown in the figure were smoothed based on the raw records. In order to fit a model, 6 significant variables $D_{st}(t-1)$, $D_{st}(t-2)$, $D_{st}(t-3)$, $D_{st}(t-4)$, $VBs(t-1)$ and $VBs(t-2)$ were chosen initially using variable selection procedures. Then the nonparametric model (4) was obtained using the same wavelet basis functions as in Example 3.1. The first 3000 points were used for identification. The 10 hour-ahead prediction (ten-step-ahead prediction) based on the final model was compared with the original data and is shown in Fig 2(b) and (c). Clearly, the model predicts extremely well.

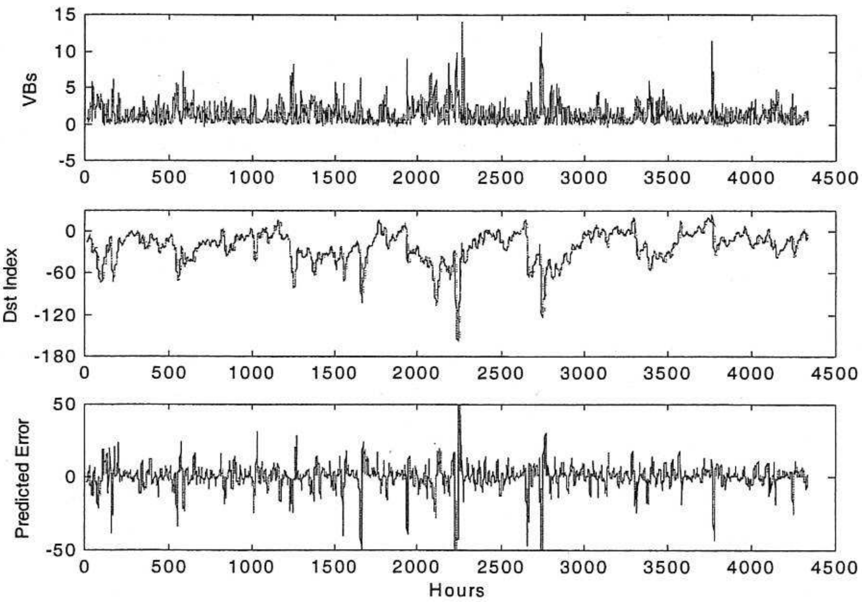

Fig 2. The identification results of the terrestrial magnetosphere process. (a) The solar wind VBs (input); (b) The comparison of ten-step-ahead prediction (the dashed line) and the original data Dst (the solid line); (c) Model predicted error (the standard deviation of the predicted error is $\sigma_e = 8.48$)

4. 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. Higher scaling levels (higher resolution) could perhaps improve the approximation accuracy but can result in over fitting of the model which will contain a greater 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 examples have demonstrated the efficiency of the modelling approach presented.

Acknowledgment

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

References

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

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

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

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

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

[6] S. Chen, S. A. Billings, C. F. N. Cowan and P. W. Grant, "Practical identification of NARMAX models using radial basis functions," *International Journal of Control*, **52(6)**, 1327-1350, Dec. 1990.

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

[8] G. G. Lorentz, M.A. Golitschek and Y. Makovoz, *Constructive approximation*. New York : Springer, 1996.

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

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

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

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

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

[14] M.B. Kennel, R. Brown and H.D. Abarbanel, "Determining embedding dimension for phase space reconstruction using a geometrical construction," *Physical Review A*, **45(6)**, 3403-3411, Mar. 1992.

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

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

[17] C. K. Chui, *An introduction to wavelets*. Boston; London : Academic Press, 1992.

[18] Q. Zhang and A. Benveniste, "Wavelet networks," *IEEE Trans. Neural Networks*, **3(6)**, 889-898, Nov. 1992.

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

[20] S. A. Billings and W. S. F. Voon, "Correlation based model validity tests for nonlinear models," *International Journal of Control*, vol. 44, no. 1, pp.235-244, Jul. 1986

[21] S. A. Billings and Q. M. Zhu, "Nonlinear model validation using correlation tests," *International Journal of Control*, vol. 60, no. 6, pp.1107-1120, Dec.1994.

[22] Y. Ueda, " Steady motions exhibited by Duffing's equations: a picture book of regular and chaotic motions, " In *NewApproaches to Nonlinear Problems in Dynamics*, Ed. P.J. Holmes, SIAM: Philadelphia, 311-322, 1980.