This is a repository copy of *Identification of Hybrid Systems Using a Class of Unified Wavelet-NARX models*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/84874/

Monograph:
Wei, H.L. and Billings, S.A. (2004) Identification of Hybrid Systems Using a Class of Unified Wavelet-NARX models. Research Report. ACSE Research Report 856 . Department of Automatic Control and Systems Engineering

Identification of Hybrid Systems Using a Class of Unified Wavelet-NARX Models

H.L. Wei and S.A. Billings

Research Report No.856

Identification of Hybrid Systems Using a Class of Unified Wavelet-NARX Models

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

A novel approach is proposed for the identification of hybrid systems based on a unified wavelet-based modelling framework, where the system input-output relationship is connected using mixed multiresolution wavelet decompositions with respect to different types of wavelet and associated scaling functions. No a priori information on the system dynamics is required to estimate a wavelet-NARX model but only a given input-output data set with an assumption that the system input and the output are bounded in a finite interval. Several examples, which involve piecewise affine (PWA) systems, open-loop threshold autoregressive systems (TARSO), piecewise rational systems, and hinging hyperplane ARX (HHARX) systems, are provided to demonstrate the effectiveness and applicability of the new identification procedures.

Keywords: Hybrid systems; identification; multiresolution analysis; NARX models; wavelets

1. Introduction

Hybrid systems, are a wide class of dynamical systems incorporating both continuous-time and discrete-time or discrete-event dynamics, and have attracted great interest from various areas (Morse et al., 1999). The continuous dynamics in such a system are usually associated with first principles, while the discrete-time or discrete-event behaviour is determined by discrete logic rules or sequences. Hybrid systems exist widely in many contexts, in both engineering and non-engineering areas. One of the simplest examples of such systems is the commonly used electrical radiator with a thermostat in a room. Several types of models have been proposed to represent hybrid systems (Heemels et al., 2001), including differential automata (Tavernini, 1987), Brockett's model (Brockett, 1993), hybrid Petri-nets (David and Alla, 1994), hybrid automata (Branicky et al., 1998), complementarity models (Van der Schaft and Schumacher, 1998; Heemels et al., 2000; Camlibel et al., 2003), integrator hybrid system model (Tittus and Egardt,1998), mixed logical dynamic (MLD) models (Bemporad and Morari, 1999), and piecewise affine (PWA) models (Bemporad et al., 2000, 2003 and the references therein). In a broad sense, some linear multimodels such as the open-loop threshold autoregressive models (Tong, 1983, 1990) and piecewise autoregressive moving average models (Johansen and Foss, 1993) can also be categorized into a special case of hybrid system representations.

As pointed out in Roll et al. (2004), most studies on hybrid systems in the literature are concerned with system analysis, stability analysis, control, verification and fault detection. Different analysis tools are closely related to different concrete model structures. Whilst these forward problems in hybrid systems have been extensively studied in recent years, the inverse problem, which is concerned with obtaining a model from experimental data sets, that is, the identification problem, has received relatively little attention and relatively few results have been achieved. Recent contributions concerning identification of hybrid systems and related results include the papers

by Skeppstedt et al. (1992), Breiman (1993), Johansen and Foss (1995), Bemporad et al. (2003) and Ferrari-Trecate et al. (2003). For a comprehensive discussion and evaluation on identification methods of hybrid systems, see the recent papers by Ferrari-Trecate et al. (2003) and Roll et al. (2004) and the references therein.

PWA models are a class of attractive and popular representations for hybrid systems. Several existing approaches can result in or at least are related to piecewise affine schemes, these include Chua's canonical representation (Chua and Deng, 1988), neural networks with piecewise affine perceptrons (Batruni, 1991), threshold autoregressive models (Tong, 1983, 1990) and linear multimodels (Johansen and Foss, 1993), and hinging hyperplanes (Breiman, 1993). The main difficulty for most piecewise approaches is the problem of partitioning the system operating region, where the estimation of the piecewise models cannot be easily separated from the task of finding the domain for each submodel. One of the main advantages of piecewise affine models is that they can easily be recast using state-space representations and this enables system analysis and control to be easily facilitated. The common drawback of these approaches is that the system behaviour must be continuous in the whole operating regime and the system should vary slowly while entering the domain of one submodel from another to guarantee the global validity of the resulting models. Roll et al. (2004) proposed an identification approach for PWA systems based on mixed integer linear or quadratic programming (MILP/MIQP), where the optimality of the estimate can be proved, but it was found (Ferrari-Trecate et al., 2003; Roll, 2003) that the number of the integer variables involved in the MILP/MIQP increases dramatically with the length of the data set. For example, the number of feasible integer variables involved in a MIQP is usually an astronomical figure, and the total number of a δ combination in the MIQP is by much larger. Therefore, this approach can only be applied to solve a problem with a very short data set (Ferrari-Trecate et al., 2003). Ferrari-Trecate et al. (2003) introduced an efficient method for identifying discrete-time PWA systems, even systems with possibly discontinuities by combining some clustering, linear identification and pattern recognition techniques. But, this approach requires some a priori information on the system under study, for example, it is assumed that the estimation data set is generated from a PWA model and the number of submodels is known, and so on. Therefore applications of this identification method may be restricted to a limited range.

The present study aims to provide a unified modelling framework for hybrid system identification based on a wavelet-NARX representation, which can represent a wide class of hybrid systems, even systems with discontinuities. In the wavelet-NARX representation, a hybrid system is initially expressed as a superimposition of a number of low-dimensional submodels (Wei et al., 2002; Billings and Wei, 2003), which can be described using mixed wavelet multiresolution decompositions (Mallat,1989). By expanding each submodel using truncated wavelet decompositions, a high-dimensional wavelet-NARX model can then be converted into a linear-in-the-parameters problem, which can be solved using least-squares type methods. An efficient model term detection approach based on a forward orthogonal least squares (OLS) algorithm, along with the error reduction ratio (ERR) criterion (Billings *et al.*, 1989; Chen *et al.*, 1989) is applied to solve the linear-in-the-parameters problem in the present study.

The rest of the paper is organised as follows. In Section 2, wavelet decompositions are briefly reviewed. In Section 3, a unified wavelet-NARX representation based on mixed wavelet multiresolution decompositions is introduced. Section 4 provides some case studies, where several illustrative examples are given.

2. Wavelet Multiresolution Decompositions

This section briefly reviews results on wavelet decompositions. For more details about these results, see for example the work of Mallat (1989), Chui (1992), Daubechies (1992) and Meyer (1993). In the following, it is assumed that the independent variable x of a function $f \in L^2(\mathbb{R})$ is defined in the unit interval [0,1]. In addition, for the sake of simplicity, one-dimensional wavelets are considered as an example to illustrate related concepts.

2.1 One-dimensional wavelet multiresolution decompositions

From wavelet theory (Mallat, 1989; Chui, 1992; Daubechies, 1992; Meyer, 1993), an orthogonal wavelet system can be constructed using *multiresolution analysis* (MRA) under some assumptions and considerations. Assume that the wavelet ψ and associated scaling function ϕ constitute an orthogonal wavelet system, then any function $f \in L^2(R)$ can be expressed as a *multiresolution wavelet decomposition*

$$f(x) = \sum_k \alpha_{j_0,k} \phi_{j_0,k}(x) + \sum_{j \geq j_0} \sum_k \beta_{j,k} \psi_{j,k}(x) \tag{1}$$

where $\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k)$, $\phi_{j,k}(x) = 2^{j/2}\phi(2^j x - k)$, $j,k \in \mathbb{Z}$, $\alpha_{j_0,k}$ and $\beta_{j,k}$ are the wavelet approximation and detail coefficients, and j_0 is an arbitrary integer representing the coarsest resolution or scale level.

2.2 Extending to high-dimensional space

The results for one-dimensional case can easily be extended to high-dimensions. One commonly used approach is to a generate separable function by the tensor products of several one-dimensional functions. For example, an n-dimensional function $\Phi : \mathbb{R}^d \mapsto \mathbb{R}$ can be constructed using a scalar function ψ as follows

$$\Phi(x) = \Phi(x_1, x_2, \cdots, x_d) = \prod_{i=1}^d \psi(x_i) \tag{2}$$

Another popular scheme is to use radial functions. For example, the n-dimensional Gaussian type functions can be constructed as

$$\Phi(x) = \Phi(x_1, x_2, \cdots, x_d) = x_1 x_2 \cdots x_d e^{-\frac{1}{2}\|x\|^2} \tag{3}$$

where $\|x\|^2 = x^T x = \sum_{i=1}^d x_i^2$. Similarly, the n-dimensional Mexican hat (also called the Marr) function can be expressed as $\Phi(x) = (n - \|x\|^2) \exp(-\|x\|^2 / 2)$.

Using the concept of *tensor products*, the multiresolution decomposition (1) can be immediately generalised to the muti-dimensional case, where a multiresolution wavelet decomposition can be defined by taking the *tensor product* of the one-dimensional scaling and wavelet functions (Mallat, 1989; Zhang and Benveniste, 1992). Let $f \in L^2(R^d)$, then $f(x)$ can be represented by the *multiresolution wavelet decomposition* as

3

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

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

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

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

with $\eta^{(i)} = \phi$ or ψ (scalar scaling function or the mother wavelet) but at least one $\eta^{(i)} = \psi$.

Notice that if j_0 is large enough, the approximation representation (4) can be expressed using only the scaling function ϕ, that is, there exists a sufficiently large integer J, such that

$$f(x_1,\cdots,x_d) = \sum_k \alpha_{J,k} \Phi_{J,k}(x_1,\cdots,x_d) = \sum_{k_1,k_2,\cdots,k_d} 2^{Jd/2} \prod_{i=1}^{d} \phi(2^J x_i - k_i) \tag{7}$$

Although many functions can be chosen as scaling and/or wavelet functions, most of these are not suitable in system identification applications, especially in the case of multidimensional and multiresolution expansions. An implementation, which has been tested with very good results, involves B-spline wavelets in multiresolution wavelet decompositions (Billings and Coca, 1999; Coca and Billings, 2001; Billings et al., 2003), and these type wavelets will be used for hybrid system identification in the present study.

2.3   B-spline wavelets

B-splines are piece-wise polynomial functions with good local properties, and were originally introduced by Chui and Wang (1992) to define a class of semi-orthogonal wavelets for representing a signal using multiroselution decompositions. The mth order B-spline function is defined as

$$N_m(x) = \frac{1}{(m-1)!} \sum_{j=0}^{m} C_j^m (-1)^j (x-j)_+^{m-1}, \ m \geq 2 \tag{8}$$

where $C_k^m = m(m-1)\cdots(m-k+1)/k!$, and $x_+^n = x^n$ for $x \geq 0$ and $x_+^n = 0$ for $x < 0$. The mth order B-spline N_m can be calculated by the following recursive formula:

$$N_m(x) = \frac{x}{m-1} N_{m-1}(x) + \frac{m-x}{m-1} N_{m-1}(x-1), \ m \geq 2 \tag{9}$$

with

$$N_1(x) = I_{[0,1)}(x) = \begin{cases} 1 & if \ x \in [0,1) \\ 0 & otherwise \end{cases} \tag{10}$$

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

$$\phi(x) = \sum_{k=0}^{m} c_k N_m(2x - k) \tag{11}$$

4

$$\varphi(x) = \sum_{k=0}^{3m-2} d_k N_m (2x - k) \tag{12}$$

with the coefficients given by

$$c_k = \frac{1}{2^{m-1}} C_k^m \tag{13}$$

$$d_k = \frac{(-1)^k}{2^{m-1}} \sum_{j=0}^{m} C_j^m N_{2m}(k - j + 1), \quad k = 0,1, \cdots, 3m - 2 \tag{14}$$

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

$$\begin{cases} \operatorname{supp} \phi = \operatorname{supp} N_m = [0, m] \\ \operatorname{supp} \varphi = [0, 2m - 1] \end{cases} \tag{15}$$

Both the B-spline wavelets and the associated scaling functions are symmetric or anti-symmetric within the support. The most commonly used B-splines are those of orders 1 to 4, which can be explicitly expressed as Table 1.

Table 1   The B-splines of order 1 to 4

	$N_1(x)$	$N_2(x)$	$2 N_3(x)$	$6 N_4(x)$
$0 \leq x < 1$	1	x	x^2	x^3
$1 \leq x < 2$	0	$2 - x$	$-2x^2 + 6x - 3$	$-3x^3 + 12x^2 - 12x + 4$
$2 \leq x < 3$	0	0	$(x - 3)^2$	$3x^3 - 24x^2 + 60x - 44$
$3 \leq x < 4$	0	0	0	$-x^3 + 12x^2 - 48x + 64$
elsewhere	0	0	0	0

3.   The Unified Wavelet-NARX Representation

A unified modelling framework is introduced to identify hybrid systems based on the wavelet-NARX representation, which is initially expressed as a superimposition of a number of low-dimensional submodels. These submodels can then be described using mixed wavelet multiresolution decompositions.

3.1   Additive representations of nonlinear dynamical systems

A wide range of nonlinear systems can be represented using the NARX (*Nonlinear AutoRegressive with eXogenous* inputs) model (Leontaritis and Billings, 1985; Pearson, 1999). Taking SISO systems as an example, this can be expressed by the following nonlinear difference equation

$$y(t) = f(y(t-1), \cdots, y(t - n_y), u(t-1), \cdots, u(t - n_u)) + e(t) \tag{16}$$

where f is an unknown nonlinear mapping, $u(t)$ and $y(t)$ are the sampled input and output sequences, n_u and n_y are the maximum input and output lags, respectively. The noise variable $e(t)$ is immeasurable but is assumed to be bounded and uncorrelated with the inputs.

Several approaches can be applied to realise the representation (16) including polynomials, neural networks, linear multimodels and other complex models, see for example Pearson (1999) for comprehensive discussions. In the present study, an additive model structure will be adopted to represent the NARX model (1). The multivariate nonlinear function f in the model (16) can be decomposed into a number of functional components via the well known functional analysis of variance (ANOVA) expansions (Friedman, 1991; Chen, 1993)

$$y(t) = f(x_1(t), x_2(t), \cdots, x_n(t))$$

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

$$+ \sum_{1 \leq i_1 < \cdots < i_m \leq n} f_{i_1 i_2 \cdots i_m}(x_{i_1}(t), x_{i_2}(t), \cdots, x_{i_m}(t)) + \cdots + f_{12 \cdots n}(x_1(t), x_2(t), \cdots, x_n(t)) + e(t) \qquad (17)$$

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

$$x_k(t) = \begin{cases} y(t-k), & 1 \leq k \leq n_y \\ u(t-k+n_y), & n_y + 1 \leq k \leq n = n_y + n_u \end{cases} \qquad (18)$$

The first functional component f_0 is a constant to indicate the intrinsic varying trend; f_i, f_{ij}, $\cdots$, are univariate, bivariate, etc., functional components. The univariate functional components $f_i(x_i)$ represent the independent contribution to the system output that arises from the action of the ith variable x_i alone; the bivariate functional components $f_{ij}(x_i, x_j)$ represent the interacting contribution to the system output from the input variables x_i and x_j, etc. The ANOVA expansion (17) can be viewed as a special form of the NARX models for input and output dynamical systems. Although the ANOVA decomposition of the NARX model (16) involves up to 2^n different functional components, experience shows that a truncated representation containing the components up to the bivariate or tri-variate functional terms is often sufficient to provide a satisfactory description of $y(t)$ for many high dimensional problems providing that the input variables are properly selected. The presence of only low order functional components does not necessarily imply that the high order variable interactions are not significant, nor does it mean the nature of the nonlinearity of the system is less severe. An exhaustive search for all the possible submodel structures of (17) is demanding and can be prohibitive because of the curse-of-dimensionality. A truncated representation is advantageous and practical if the higher order terms can be ignored. In practice, the constant term f_0 can often be omitted since it can be combined into other functional components.

In practice, many types of functions, such as kernel functions, splines, polynomials and other basis functions can be chosen to express the functional components in model (16). In the present study, however, wavelet decompositions will be chosen to describe the functional components in the additive model (17), and this was referred to as the wavelet-NARX model in Billings and Wei (2003) and Wei and Billings (2004).

6

3.2 Mixed multiresolution wavelet decompositions

In order to make wavelet multiresolution decompositions more flexible and capable of representing arbitrary hybrid system, including systems with sharp variations or possible discontinuities, the decompositions (1) and (7) are adapted and extended to mixed wavelet multiresolution decompositions using B-spline wavelets. For one-dimensional functional components in the additive model (17), the mixed decompositions can be described as

$$
f_p(x_p(t)) = \sum_k \alpha_{j_1,k}^{[1]} \phi_{j_1,k}^{[1]}(x_p(t)) + \sum_{j \geq j_1} \sum_k \beta_{j,k}^{[1]} \psi_{j,k}^{[1]}(x_p(t))
$$

$$
+ \sum_k \alpha_{j_2,k}^{[2]} \phi_{j_2,k}^{[2]}(x_p(t)) + \sum_{j \geq j_2} \sum_k \beta_{j,k}^{[2]} \psi_{j,k}^{[2]}(x_p(t))
$$

$$
+ \sum_k \alpha_{j_4,k}^{[4]} \phi_{j_4,k}^{[4]}(x_p(t)) + \sum_{j \geq j_4} \sum_k \beta_{j,k}^{[4]} \psi_{j,k}^{[4]}(x_p(t)) \tag{19}
$$

where $\phi_{j,k}^{[m]}(x_p(t)) = 2^{j/2} \phi^{[m]}(2^j x_p(t) - k)$ and $\psi_{j,k}^{[m]}(x_p(t)) = 2^{j/2} \psi^{[m]}(2^j x_p(t) - k)$ are the mth order B-spline and wavelet functions, $\alpha_{j,k}^{[m]}$ and $\beta_{j,k}^{[m]}$ are the associated wavelet coefficients, j_m is the initial (coarsest) resolution level for the decomposition of with respect to the mth order B-spline wavelets.

Similarly, the d-dimensional ($d>1$) functional components in the additive model (17) can be expressed as

$$
f_{i_1 i_2 \cdots i_d}(x_{i_1}(t), x_{i_2}(t), \cdots, x_{i_d}(t)) = \sum_{k_1,k_2,\cdots,k_d} \alpha_{J_1;k_1,k_2,\cdots,k_d}^{[1]} \prod_{p=1}^d \phi^{[1]}(2^{J_1} x_{i_p} - k_p)
$$

$$
+ \sum_{k_1,k_2,\cdots,k_d} \alpha_{J_2;k_1,k_2,\cdots,k_d}^{[2]} \prod_{p=1}^d \phi^{[2]}(2^{J_2} x_{i_p} - k_p)
$$

$$
+ \sum_{k_1,k_2,\cdots,k_d} \alpha_{J_4;k_1,k_2,\cdots,k_d}^{[4]} \prod_{p=1}^d \phi^{[4]}(2^{J_4} x_{i_p} - k_p) \tag{20}
$$

where $\phi^{[m]}$ is the mth order B-spline function, J_m is the finest resolution scale. In the mixed wavelet multiresolution decompositions (19) and (20), the first order B-spline (the Haar) scaling and wavelet functions are piecewise constant functions, which are particularly effective to describe discontinuous dynamics of a system. The second order B-spline and wavelet functions are piecewise ramp functions, which are capable of capturing some linear trends of the dynamics including sharp varying trends. The fourth order B-spline and wavelet functions are smooth functions, which can detect the system behaviour that varies smoothly.

All the submodels in the additive model (17) can be expressed using the extended multiresolutions (19) and (20), and the resulting model consists of three parts, that is,

$$
y(t) = f(x_1(t), x_2(t), \cdots, x_n(t))
$$

$$
= f^{[1]}(x_1(t), x_2(t), \cdots, x_n(t)) + f^{[2]}(x_1(t), x_2(t), \cdots, x_n(t)) + f^{[4]}(x_1(t), x_2(t), \cdots, x_n(t)) \tag{21}
$$

where $f^{[m]}$, $m=1,2,4$ are the mixed multiresolution wavelet decomposition with respect to the mth order B-spline wavelet and scaling functions. Each $f^{[m]}$ is in the form of the additive model (17), where the submodels are

approximated using the wavelet multiresolutions (1) or (7). Model (21) will be referred to as the mixed multiresolution wavelet decomposition (MMWD).

For an identification practice, it is impossible to solve a decomposition involving infinite terms. A MMWD is therefore often truncated at a satisfactory accuracy with a chosen finest resolution scale. With the assumption that $x(t) \in [0,1]$, for a given resolution scale j, the range of the associated translation (dilation) parameter k for the wavelet $\psi_{j,k}^{[m]}(x(t))$ and scaling function $\phi_{j,k}^{[m]}(x(t))$ in the MMWD can be determined accordingly. For example, for the Haar wavelet and scaling function, $k=0,1,\ldots, 2^j -1$. For the second order B-spline function $\phi_{j,k}^{[2]}(x(t))$, $k=-1,0,1,\ldots, 2^j -1$ and for the second order B-wavelet function $\psi_{j,k}^{[2]}(x(t))$, $k=-2,-1,0,\ldots, 2^j -1$. Thus, the number of the total wavelet terms involved in a truncated MMWD can be determined beforehand. Since the unknown parameters are only the wavelet coefficients, a truncated MMWD can easily be translated into a linear-in-the-parameters problem, which can be solved using least squares type algorithms.

Note that a MMWD may involve a large number of candidate wavelet terms and might be severely redundant, especially for a high-dimensional system with several variables (large time lags for the system input and/or system output). Fortunately, this problem can be successfully resolved by employing the well-known and widely used OLS-ERR algorithm (Billings *et al.*, 1989; Chen *et al.*, 1989). A fast and efficient model structure determination approach has been implemented using the forward orthogonal least squares (OLS) algorithm and the error reduction ratio (ERR) criterion, which was originally introduced to determine which terms should be included in a model (Billings *et al.*, 1989; Chen *et al.*, 1989). This approach has been extensively studied and widely applied in nonlinear system identification, see for example Chen et al. (1991), Wang and Mendel (1992), Hong and Harris (2001), Billings and Wei (2003). The forward OLS algorithm involves a stepwise orthogonalization of the regressors (wavelet terms) and a forward selection of the relevant significant terms in all candidate model terms based on the error reduction ratio (ERR) (Chen *et al.*, 1989).

4. Case Studies

In this section, several examples are given to illustrate the new identification procedure using the MMWD models. No a priori knowledge about the system dynamics is required except the following assumption.

Assumption: The system inputs and outputs are bounded in some finite interval.

Assume that the system input $u(t)$ and output $y(t)$ are bounded in $[\underline{u}, \overline{u}]$ and $[\underline{y}, \overline{y}]$, respectively, where $-\infty < \underline{u} < \overline{u} < \infty$ and $-\infty < \underline{y} < \overline{y} < \infty$. The observational data set can easily be normalised into the unit interval [0,1], therefore the variables $x_i(t)$ described by Eq (18) satisfy $0 \le x_i(t) \le 1$. The identification procedures will be performed over the unit interval [0, 1], and finally the signals simulated from the identified model will be recovered via some appropriate inverse transform (Wei and Billings, 2004).

Example 1 Consider a simple piecewise affine system similar to that in Ferrari-Trecate et al.(2001) described by the model

$$y(t) = \begin{cases} -u(t-1) & -10 \le u(t-1) \le -5 \\ +u(t-1) & -5 < u(t-1) < 5 \\ 3u(t-1) - 2 & 5 \le u(t-1) \le 10 \end{cases} \tag{22}$$

This model was simulated by setting the input $u(t)$ as a random sequence that is uniformly distributed in [-10,10], and 5000 input-output data points with an additive noise of normal distribution $N(0, \sigma^2)$, where $\sigma = 0.5$, were collected, the first 400 of which will be used for system identification. The regression plot of $y(t)$ versus $u(t\text{-}1)$ is shown in Figure 1(a) and (b), which clearly indicates that there exists a piecewise linear relationship between the output $y(t)$ and the input $u(t\text{-}1)$, and there is a discontinuity point at $u(t-1) = \pm 5$. As shown in Figure 1(c) and (d), this PWA system possesses an interesting phase portrait which is divided into four parts. The density of each part is not the same.

Noting that the system input $u(t)$ and noisy output $y(t)$ are respectively bounded in $[\underline{u}, \overline{u}]$ and $[\underline{y}, \overline{y}]$, where $\overline{u} = -\underline{u} = 10$, $\underline{y} = -8$ and $\overline{y} = 30$, the observational data were normalised using this information. The identification procedures will be performed over the unit interval [0, 1]. The wavelet-NARX model for the PWA system was initially chosen to be
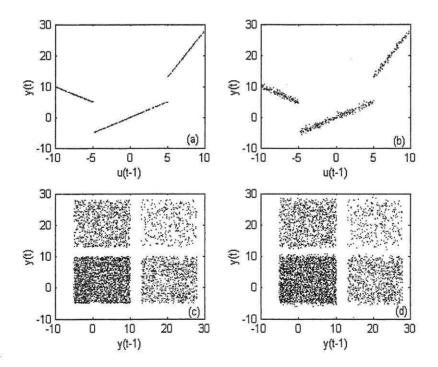

Figure 1 Phase portraits of the PWA system (22). (a) $y(t)$ vs $u(t\text{-}1)$, 400 points, noise free; (b) $y(t)$ vs $u(t\text{-}1)$ 400 points, with additive noise; (c) $y(t)$ vs $y(t\text{-}1)$, 5000 points, noise free; (d) $y(t)$ vs $u(t\text{-}1)$, 5000 points, with additive noise.

$$y(t) = f(y(t-1), u(t-1)) = f^{[1]}(y(t-1), u(t-1)) + f^{[2]}(y(t-1), u(t-1))$$

$$= f_1^{[1]}(y(t-1)) + f_2^{[1]}(u(t-1)) + f_1^{[2]}(y(t-1)) + f_2^{[2]}(u(t-1)) \tag{23}$$

where

$$f_p^{[m]}(x(t)) = \sum_k \alpha_{j_m,k}^{[m]} \phi_{j_m,k}^{[m]}(x(t)) + \sum_{j \geq j_m} \sum_k \beta_{j,k}^{[m]} \psi_{j,k}^{[m]}(x(t)) \tag{24}$$

where $f_p^{[m]}$ are wavelet decompositions with respect to the mth order B-spline and associated wavelet functions. The coarsest and finest resolution scales for the first and second-order B-spline and wavelet functions in (24) were chosen to be $j_1 = j_2 = 0$, $J_1 = 6$, and $J_2 = 5$, respectively. Although 410 candidate model terms in the initial wavelet-NARX model (23), only 8 most significant model terms were selected using an OLS-ERR algorithm (Billings *et al.*, 1989; Chen *et al.*, 1989). An parsimonious model is found to be

$$y(t) = f(y(t-1), u(t-1)) = f^{[1]}(u(t-1)) + f^{[2]}(u(t-1))$$

$$= 0.0132\psi_{0,0}^{[1]}(u(t-1)) + 0.0930\psi_{1,0}^{[1]}(u(t-1)) - 0.0744\psi_{1,1}^{[1]}(u(t-1))$$

$$+ 1.0307\phi_{0,0}^{[2]}(u(t-1)) - 0.0366\psi_{0,-2}^{[2]}(u(t-1)) - 0.0731\psi_{0,-1}^{[2]}(u(t-1))$$

$$- 0.0658\psi_{1,0}^{[2]}(u(t-1)) + 0.1974\psi_{1,1}^{[2]}(u(t-1)) \tag{25}$$

The model (25), as well as the original system (22), was simulated by setting the input $u(t)$ as a random sequence that is uniformly distributed in [-10,10]. The model simulated output from (25) was compared with that from (23), and this is shown in Figure 2, where only the data points ranging from 800 to 1000 in the data set are plotted. The discrepancy between these outputs shown in Figure 2(b) clearly indicates that the identified model (25) is excellent. Note that the model (25) is performed on normalized input and output and therefore the given input and output signals should be pre-normalized and then the output from the model can be recovered to its original amplitude by $\tilde{y}(t) = \underline{y} + (\overline{y} - \underline{y})y(t)$, where $y(t)$ is the model simulated output from (25) and $\tilde{y}(t)$ is the associated recovered signal.

To inspect how the output signal is recovered from a mixed wavelet decomposition, the model (25) was also simulated by setting the input signal $u(t) = 10\sin(2\pi t / 50)$. Note that this input induces the response in the system output a *higher-harmonic*, that is, the output oscillates at twice the input oscillation frequency, see the dashed line in Figure 3(c). The signals recovered by $f^{[1]}$, $f^{[2]}$ and $y = f^{[1]} + f^{[2]}$ in (25) are shown in Figure 3(a), (b) and (c), respectively. Clearly, the signal recovered by $f^{[1]}$ is piecewise constant, which is appropriate to describe sharp variations or discontinuous dynamics, and the signal recovered by $f^{[2]}$ is somewhat smooth. The discrepancy between the recovered signal by $f^{[1]} + f^{[2]}$ from (25) and the ideal output from (22) is shown in Figure 3(d).
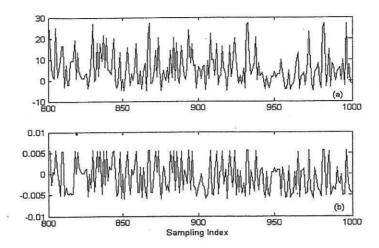
Figure 2 A comparison between the model predicted output from the model (25) and the that from the original system (22) with the input as a random sequence. (a) Overlap of the outputs from (22) and (25); (b) The discrepancy between the outputs.
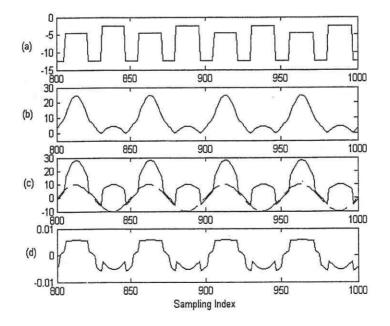

Figure 3 Signals recovered from $f^{[1]}$, $f^{[2]}$ and $f^{[1]} + f^{[2]}$ based on the model (25). (a) Recovered from $f^{[1]}$; (b) Recovered from $f^{[2]}$; (c) A comparison between the signal recovered from $f^{[1]} + f^{[2]}$ and the that from the original system (22). (d) The discrepancy between the recovered signal from (25) and that from the original system (22).

11

Example 2 Consider an open-loop threshold autoregressive system (TARSO) (Tong, 1990; Pearson, 1999) described by the model

$$y(t) = a_v y(t-1) + b_v y(t-2) + u(t-1) + \varepsilon(t), \ [x_1(t), x_2(t)]^T \in \Omega_v \tag{26}$$

where $x_k(t) = y(t-k)$ for $k=1,2$, $\varepsilon(t)$ is a Gaussian white noise with a standard derivation $\sigma_\varepsilon = 0.1$. $\Omega_1 = [0,\infty) \times [0,\infty)$, $\Omega_2 = [0,\infty) \times (-\infty,0)$, $\Omega_3 = (-\infty,0) \times [0,\infty)$, and $\Omega_4 = (-\infty,0) \times (-\infty,0)$. $a_1 = a_2 = 1.6$, $a_3 = a_4 = 0$, $b_1 = b_3 = -0.8$, $b_2 = b_4 = 0.8$.

Model (26) was simulated by setting the input signal as $u(t) = \sin(2\pi t / 50)$ and 400 input-output data points were used for model estimation. Using the information that $\underline{u} \leq u(t) \leq \overline{u}$ and $\underline{y} \leq y(t) \leq \overline{y}$, where $\overline{u} = -\underline{u} = 1$, $\underline{y} = -4$ and $\overline{y} = 7.5$, the observational data were normalized. Three most significant variables, $x_1(t) = y(t-1)$, $x_2(t) = y(t-2)$ and $x_3(t) = u(t-1)$ were chosen using a variable selection algorithm (Wei et al., 2004) for this system and the wavelet-NARX model with respect to the normalized variables for the TARSO system (26) was initially set to be

$$
\begin{aligned}
y(t) &= f(x_1(t), x_2(t), x_3(t)) \\[4pt]
&= f^{[1]}(x_1(t), x_2(t), x_3(t)) + f^{[2]}(x_1(t), x_2(t), x_3(t)) + f^{[4]}(x_1(t), x_2(t), x_3(t)) \\[4pt]
&= \sum_{p=1}^{3} f_p^{[1]}(x_p(t)) + \sum_{p=1}^{2}\sum_{q=2}^{3} f_{pq}^{[1]}(x_p(t), x_q(t)) \\[4pt]
&+ \sum_{p=1}^{3} f_p^{[2]}(x_p(t)) + \sum_{p=1}^{2}\sum_{q=2}^{3} f_{pq}^{[2]}(x_p(t), x_q(t)) \\[4pt]
&+ \sum_{p=1}^{3} f_p^{[4]}(x_p(t)) + \sum_{p=1}^{2}\sum_{q=2}^{3} f_{pq}^{[4]}(x_p(t), x_q(t))
\end{aligned} \tag{27}
$$

The univariate functions $f_p^{[m]}$ are wavelet decompositions with respect to the mth order B-spline and associated wavelet functions defined as in (24), where the coarsest and finest resolution scales for the first, second and fourth-order B-spline and wavelet functions were chosen to be $j_1 = j_2 = j_4 = 0$, $J_1 = 5$, $J_2 = 4$ and $J_4 = 3$. The bivariate functions $f_{pq}^{[m]}$ are wavelet decompositions with respect to the mth B-spline products as defined by (20), where the resolutions scales for the first, second and fourth-order B-splines were chosen to be $J_1 = 2$, $J_2 = 1$ and $J_4 = 0$. An identified parsimonious model was found to be

$$
\begin{aligned}
y(t) &= f(y(t-1), y(t-2), u(t-1)) \\[4pt]
&= 1.2817\phi_{0,-1}^{[4]}(y(t-1)) + 0.6484\psi_{0,-3}^{[4]}(y(t-1)) - 0.8970\psi_{0,-3}^{[4]}(y(t-2)) \\[4pt]
&- 267.4688\psi_{0,0}^{[4]}(y(t-2)) - 249.4105\phi_{0,-3}^{[4]}(y(t-1))\phi_{0,0}^{[4]}(y(t-2)) \\[4pt]
&- 0.1921\phi_{0,-1}^{[2]}(u(t-1))
\end{aligned} \tag{28}
$$

To test the performance of the identified model (28), an input signal $u(t) = u_i(t) + w_i(t)$, where $u_i(t) = A_i \sin(2\pi t / 50)$ for $A_1 = 0.8$ and $A_2 = 0.5$, and $w_i(t)$ was a random sequence that was uniformly distributed in $[-c_i, c_i]$ for $c_1 = 0.2$ and $c_2 = 0.5$, was used to drive the model (28). The model output from (28) was recovered to its original amplitude and this was compared with the output from the original system (26), see Figures 4 and 5, which clearly indicate that the identified model (28) provide an excellent representation for the original TARSO system (26).

Figure 4   Performance of the identified model (28) for the TARSO system (26) by setting the input signal $u(t) = u_1(t) + w_1(t)$, where $u_1(t) = 0.8\sin(2\pi t / 50)$ and $w_1(t)$ was a random sequence uniformly distributed in [-0.2, 0.2]. (a) The input. (b) A comparison between the recovered signal from the model (28) and the output from the original system (26). (The solid line in (b) indicates the real output, and the dashed line indicates the simulated output from the identified model (28)).
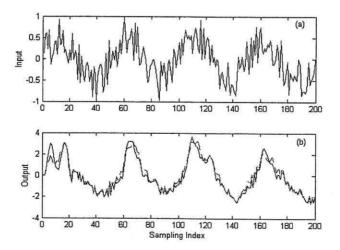

Figure 5   Performance of the identified model (28) for the TARSO system (26) by setting the input signal $u(t) = u_2(t) + w_2(t)$, where $u_2(t) = 0.5\sin(2\pi t / 50)$ and $w_2(t)$ was a random sequence uniformly distributed in [-0.5, 0.5]. (a) The input. (b) A comparison between the recovered signal from the model (28) and the output from the original system (26). (The solid line in (b) indicates the real output, and the dashed line indicates the simulated output from the identified model (28)).

Example 3 A piecewise rational system (Pearson, 1999) described by the model

$$y(t) = \frac{y(t-1)y(t-2)}{a^2 + y^2(t-1) + y^2(t-2)} + \frac{\text{sgn}(u(t-1))u(t-1)}{b^2 + c^2 u^2(t-1)} + \varepsilon(t) \tag{29}$$

where $a=b=1$, $c=0.1$, $\text{sgn}(u)=1$ for $u \geq 0$ and $\text{sgn}(u)=0$ for $u<0$, $\varepsilon(t)$ is a Gaussian white noise with a standard variation $\sigma_\varepsilon^2 = 0.01$. This system was simulated by setting the input signal $u(t)$ as a random sequence uniformly distributed in [-10, 10] and 500 input-output data points were used for model estimation. Using the information that $\underline{u} \leq u(t) \leq \overline{u}$ and $\underline{y} \leq y(t) \leq \overline{y}$, where $\overline{u} = -\underline{u} = 10$, $\underline{y} = -6$ and $\overline{y} = 6$, the observational data were normalized. Three most significant variables, $x_1(t) = y(t-1)$, $x_2(t) = y(t-2)$ and $x_3(t) = u(t-1)$ were chosen using a variable selection algorithm (Wei et al., 2004) for this system and the wavelet-NARX model with respect to the normalized variables for the rational system (29) was initially set to be the same as (27) for Example 2. An identified wavelet-NARAX model was found to be

$$y(t) = f(y(t-1), y(t-2), u(t-1))$$

$$= -0.0033\psi_{1,0}^{[1]}(y(t-2)) - 0.0040\phi_{2,1}^{[1]}(y(t-1))\phi_{2,2}^{[1]}(y(t-2)) - 0.0063\phi_{2,2}^{[1]}(y(t-1))\phi_{2,1}^{[1]}(y(t-2))$$

$$+ 0.0014\phi_{2,2}^{[1]}(y(t-1))\phi_{2,3}^{[1]}(y(t-2)) - 0.0032\phi_{2,3}^{[1]}(y(t-1))\phi_{2,1}^{[1]}(y(t-2))$$

$$+ 0.5983\phi_{0,-1}^{[2]}(u(t-1)) + 1.3329\phi_{0,0}^{[2]}(u(t-1)) + 0.0180\psi_{2,-2}^{[2]}(u(t-1)) + 0.0041\psi_{2,2}^{[2]}(u(t-1))$$

$$- 0.00470\psi_{2,3}^{[2]}(u(t-1)) - 0.0023\psi_{3,6}^{[2]}(u(t-1)) + 0.0095\phi_{1,1}^{[2]}(y(t-1))\phi_{1,-1}^{[2]}(y(t-2))$$

$$+ 0.0354\phi_{1,0}^{[2]}(y(t-1))\phi_{1,0}^{[2]}(y(t-2)) + 0.0195\phi_{1,-1}^{[2]}(y(t-1))\phi_{1,1}^{[2]}(y(t-2))$$

$$- 8.2873\psi_{0,-5}^{[4]}(u(t-1)) + 12.7305\psi_{0,-1}^{[4]}(u(t-1)) - 7.6666\phi_{0,-3}^{[4]}(y(t-1))\phi_{0,-3}^{[4]}(y(t-2))$$

$$- 1.1709\phi_{0,-2}^{[4]}(y(t-1))\phi_{0,-1}^{[4]}(y(t-2)) - 1.1270\phi_{0,-1}^{[4]}(y(t-1))\phi_{0,-2}^{[4]}(y(t-2))$$

$$- 7.9817\phi_{0,0}^{[4]}(y(t-1))\phi_{0,0}^{[4]}(y(t-2)) \tag{30}$$

where $\phi_{j,k}^{[m]}(x_p(t)) = 2^{j/2}\phi^{[m]}(2^j x_p(t) - k)$ and $\psi_{j,k}^{[m]}(x_p(t)) = 2^{j/2}\psi^{[m]}(2^j x_p(t) - k)$ are the mth order B-spline and wavelet functions.

To inspect the performance of the identified model (30), an input signal $u(t) = u_1(t) + u_2(t)$, where $u_1(t) = 7.5\sin(2\pi t/50)$ and $u_2(t)$ was a random sequence that was uniformly distributed in $[-2.5, 2.5]$, was used to drive the model (30). The model output from (30) was recovered to its original amplitude and this was compared with the output from the original system (29), see Figures 6. The result clearly indicates that the identified model (30) provide an excellent representation for the original rational system (29).
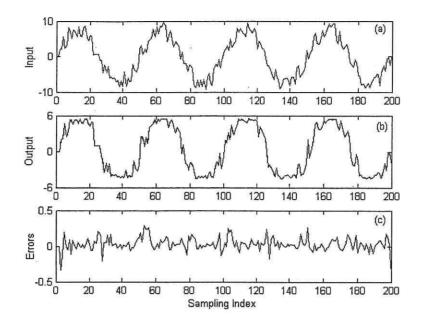
Figure 6   Performance of the identified model (30) for the TARSO system (29) by setting the input signal $u(t) = u_1(t) + u_2(t)$, where $u_1(t) = 7.5\sin(2\pi t / 50)$ and $u_2(t)$ was a random sequence uniformly distributed in[-2.5, 2.5]. (a) The input. (b) A comparison between the recovered signal from the model (30) and the output from the original system (29). (The solid line in (b) indicates the real output, and the dashed line indicates the simulated output from the identified model (30)). (c) Model predicted errors.

Example 4   A multi-mode hinging hyperplane ARX (HHARX) system (Roll, 2003) described by the piecewise ARX model

$$y(t) = 1.1515 - 0.5557\,y(t-1) - 0.3370\,y(t-2) + 1.3795u(t-1) + \sum_{i=1}^{4} \lambda_i \max\{\xi(t),0\} \qquad (31)$$

where $\lambda_1 = \lambda_2 = 1$, $\lambda_3 = \lambda_4 = -1$ and

$$\xi_1(t) = -0.3918 + 0.0538\,y(t-1) + 1.5396\,y(t-2) - 0.2742u(t-1) \qquad (31a)$$

$$\xi_2(t) = -0.0293 + 0.3517\,y(t-1) + 0.1249\,y(t-2) - 1.1005u(t-1) \qquad (31b)$$

$$\xi_3(t) = -0.3097 - 1.3018\,y(t-1) + 0.3074\,y(t-2) + 1.3506u(t-1) \qquad (31c)$$

$$\xi_4(t) = 0.9393 - 0.0503\,y(t-1) + 0.7998\,y(t-2) - 1.1129u(t-1) \qquad (31d)$$

Note that by setting the input signal $u(t) = (-1)^t$, the output of the system is *subharmonic*, that is, the output oscillation is at a fraction of the input oscillation frequency (Tong, 1990), see Figure 7.
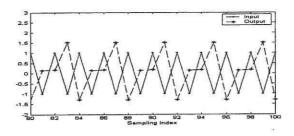
Figure 7 The HHARX system (31) is subharmonic driven by the input $u(t) = (-1)^t$

This system was simulated by setting the input signal $u(t) = \sin(2\pi t / 50)$ and 500 input-output data points were collected, the first 300 points with additive noise $\varepsilon(t) \sim N(0, 0.02^2)$ were used for model estimation and the others for model validation. Using the information $\underline{u} \leq u(t) \leq \overline{u}$ and $\underline{y} \leq y(t) \leq \overline{y}$, where $\overline{u} = -\underline{u} = 1$, $\underline{y} = -1.5$ and $\overline{y} = 2.5$, the observational data were normalized. The three most significant variables, $x_1(t) = y(t-1)$, $x_2(t) = y(t-2)$ and $x_3(t) = u(t-1)$ were chosen using a variable selection algorithm (Wei et al., 2004) for this system and the wavelet-NARX model with respect to the normalized variables for the HHARX system (31) was initially set to be the same as (27) for Example 2. An identified wavelet-NARAX model to represent the input-output relation of the given observational data set for the HHARX system was found to be

$$y(t) = f(y(t-1), y(t-2), u(t-1))$$

$$= f^{[1]}(y(t-1), y(t-2), u)t-1)) + f^{[2]}(y(t-1), y(t-2), u)t-1))$$

$$+ f^{[4]}(y(t-1), y(t-2), u)t-1))$$

$$= -0.0013\psi_{5,15}^{[1]}(y(t-1)) + 0.0005\psi_{5,19}^{[1]}(y(t-1)) - 0.0291\psi_{0,0}^{[1]}(y(t-2))$$

$$- 0.0096\psi_{3,3}^{[1]}(y(t-2)) - 0.0007\psi_{5,28}^{[1]}(u(t-1)) + 0.0025\psi_{4,3}^{[1]}(u(t-1))$$

$$- 0.0023\psi_{2,2}^{[1]}(y(t-1))$$

$$+ 1.2958\phi_{0,0}^{[2]}(y(t-1)) - 0.0006\psi_{4,11}^{[2]}(y(t-1)) - 0.0038\psi_{2,0}^{[2]}(y(t-2))$$

$$+ 0.0026\psi_{2,1}^{[2]}(u(t-1)) + 1.0816\phi_{1,-1}^{[2]}(y(t-1))\phi_{1,1}^{[2]}(u(t-1))$$

$$+ 0.2270\phi_{1,-1}^{[2]}(y(t-1))\phi_{1,0}^{[2]}(u(t-1)) + 0.4408\phi_{1,0}^{[2]}(y(t-1))\phi_{1,1}^{[2]}(y(t-2))$$

$$- 0.3311\phi_{1,1}^{[2]}(y(t-1))\phi_{1,0}^{[2]}(y(t-2)) - 0.1221\phi_{1,-1}^{[2]}(y(t-2))\phi_{1,0}^{[2]}(u(t-1))$$

$$+ 5.6145\phi_{0,-1}^{[4]}(y(t-1))\phi_{0,-3}^{[4]}(y(t-2)) - 337.9477\phi_{0,0}^{[4]}(y(t-1))\phi_{0,-3}^{[4]}(u(t-1))$$

$$- 17.3846\phi_{0,0}^{[4]}(y(t-2))\phi_{0,-2}^{[4]}(u(t-1)) \tag{32}$$

The model predicted output over the test data set was compared with the original measurements and this is shown in Figure 8, which clearly indicates that the identified model (32) provides an excellent representation for the observational data set. To inspect how the output signal is recovered from the wavelet decompositions with respect to the first, second, and fourth order B-spline and associated wavelet functions, Figure 9 shows the recovered signals from these decompositions in (32) with respect to the three types B-spline wavelets.
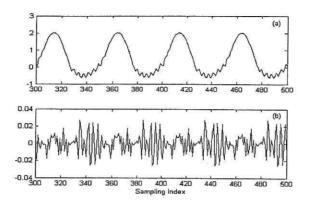
16

Figure 8 Performance of the identified model (32) over the test data set for the HHARX system (31). (a) Overlap of the output based on the model (32) and the measurements. (b) Discrepancy between the model predicted output and the measurements.
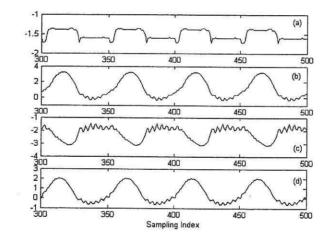

Figure 9 Signals recovered from $f^{[1]}$, $f^{[2]}$, $f^{[4]}$ and $f^{[1]} + f^{[2]} + f^{[4]}$ based on the model (32). (a) Recovered from $f^{[1]}$. (b) Recovered from $f^{[2]}$. (c) Recovered from $f^{[4]}$. (d) Superimposition $f^{[1]} + f^{[2]} + f^{[4]}$.

5.   Conclusions

A new identification approach for hybrid systems has been introduced based on a class of unified wavelet-NARX models, where a hybrid system is described using some mixed multiresolution wavelet decompositions with respect to different types of wavelet and associated scaling functions. Three types of wavelet and scaling functions, the first, second and fourth order B-spline and associated wavelet functions, were considered in the present study. First order B-spline (the Haar) scaling and wavelet functions, which are piecewise constant, in the mixed wavelet multiresolution decompositions, are particularly appropriate to detect discontinuous dynamics in a system. Second order B-spline and wavelet functions, which are piecewise ramp type functions, are capable of capturing linear trends in the dynamics including sharp varying trends. Fourth order B-spline and wavelet functions, which are smooth, can describe the system behaviour that varies smoothly. The combination of these

17

different type wavelet and scaling functions makes the mixed wavelet decomposition capable of effectively describing severely nonlinear systems, including systems with jumps or even discontinuities as shown in the examples provided in Section 4.

At first sight, such a unified mixed wavelet-NARX model may involve a great number of wavelet terms for a high-dimensional system. In most cases, however, many of the model terms are redundant and only a small number of significant terms are necessary to describe a given nonlinear system with a good accuracy. More fortunately, an efficient model structure and term detection algorithm can be employed to decide which model terms should be included in the model and finally a parsimonious model can often be obtained.

Note that for modelling a hybrid system using the mixed wavelet-NARX model, no a priori knowledge was required for the system dynamics, but only an assumption that the system input and output were bounded in finite intervals. The new unified wavelet-NARX models are therefore flexible and effective while modelling hybrid and other complex systems.

Acknowledgment

References

Batruni, R. (1991). A multilayer neural networks with piecewise-linear structure and back-propagation learning. *IEEE Transactions on Neural Networks*, **2(3)**, 395-403.

Bemporad, A., Ferrari-Trecate, G., and Morari, M. (2000). Observability and Controllability of piecewise affine and hybrid systems. *IEEE Transactions on Automatic Control*, **45(10)**, 1864-1876.

Bemporad, A., Garulli, A., Paoletti, S. and Vicino, A. (2003). A greedy approach to identification of piecewise affine models. In O. Maler and A. Pnuli (Eds.), *Hybrid Systems*: *Computation Control*. Lecture Notes in Computer Science, Vol. 2623, Berlin: Springer, pp. 97-112.

Bemporad, A. and Morari, M. (1999). Control of systems integrating of logic, dynamics, and constraints. *Automatica*, **35(3)**, 407-427.

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

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

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

Billings, S.A., Wei, H.L., Mei, S.S. and Guo, L.Z.(2003). Identification of spatio-temporal systems using multiresolution wavelet models (submitted for publication).

Branicky, M.S., Borkar, V.S. and Mitter, S.K. (1998). A unified framework for hybrid control: model and optimal control theory. *IEEE Transactions on Automatic Control*, **43(1)**, 31-45.

Breiman, L (1993). Hinging hyperplanes for regression, classification, and function approximation. *IEEE Transactions on Information Theory*, **39(3)**, 999-1013.

Brockett, R.W. (1993). Hybrid models for motion control systems. In H.L. Trentelman and J.C. Willems (eds.),

Essays in Control. Boston: Birkhauser. pp. 29-53.

Camlibel, M.K., Heemels, W.P.M.H., van der Schaft, A.J. and Schumacher, J.M. (2003). Switched networks and complementarity. *IEEE Transactions on Circuits and Systems I-Fundamental Theory and Applications*, **50(8)**, 1036-1046.

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

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

Chen, Z.H. (1993). Fitting multivariate regression functions by interaction spline models. *Journal of the Royal Statistical Society, Series B (Methodological)*, **55(2)**, 473-491.

Chua, L.O. and Deng, A.C. (1988). Canonical piecewise-linear representation. *IEEE Transactions on Circuits and Systems* , **35(1)**, 101-111.

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

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

Coca, D. and Billings, S.A.(2001). Non-linear system identification using wavelet multiresolution models. *International Journal of Control*, **74(18)**,1718-1736.

Daubechies, I. (1992). *Ten Lectures on Wavelets*. Philaelphia, Pennsylvania : Society for Industrial and Applied Mathematics.

David, R. and Alla, H. (1994). Petri nets for modelling of dynamics—a survey. *Automatica*, **30(2)**, 175-202.

Ferrari-Trecate, G., Muselli, M., Liberati, D. and Morari, M. (2001). Identification of piecewise affine and hybrid systems. In Proceedings of the American Control Conference. Arlington, VA, June 25-27, pp3521-3526.

Ferrari-Trecate, G., Muselli, M., Liberati, D. and Morari, M. (2003). A clustering technique for the identification of piecewise affine systems. *Automatica*, **39(2)**, 205-217.

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

Heemels, W.P.M.H., Schumacher, J.M. and Weiland, S. (2000). Linear compleemntarity systems. *SIAM Journal of Applied Mathematics*, **60(4)**, 1234-1269.

Heemels, W.P.M.H., De Schutter, B. and Bemporad, A. (2001). Equivalence of hybrid dynamical models. *Automatica*, **37(7)**, 1085-1091.

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

Johansen, T.A. and Foss, B.A.(1993). Constructing NARMAX models using ARMAX modes. *International Journal of Control*, **58(5)**, 1125-1153.

Johansen, T.A. and Foss, B.A.(1995). Identification of nonlinear system structure and parameters using regime decomposition. *Automatica*, **31(2)**, 321-326.

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). *International Journal of Control*, **41(2)**,303-344.

Mallat,S.G. (1989). A theory for multiresolution signal decomposition: the wavelet representation, *IEEE*

Transactions on Pattern Analysis and Machine Intelligence, **11**(7), 674-693.

Meyer, Y. (1993). *Wavelets: Algorithms and Applications*. Philaelphia, Pennsylvania: SIAM.

Morse, A.S., Pantelides, C.C., Sastry, S.S. and Schumacher, J.M. (1999). Introduction to the special issue on hybrid systems. *Automatica*, **35** (3), 347-348.

Pearson, R.K. (1999). *Descrete-Time Dynamic Models*. Oxford: Oxford University Press.

Roll, J. (2003). *Local and Piecewise Affine Approaches to System Identification*. PhD Thesis, Department of Electrical Engineering, Linköping University, Sweden.

Roll, J., Bemporad, A. and Ljung, L. (2004). Identification of piecewise affine systems via mixed-integer programming. *Automatica*, **40**(1), 37-50.

Skeppstedt, A., Ljung, L. and Millnert, M. (1992). Construction of composite models from observed data. *International Journal of Control*, **55**(1), 141-152.

Tavernini, L. (1987). Differential automata and their discrete simulators. *Nonlinear Analysis, Theory, Methods and Applications*, **11**(6), 665-683.

Tittus, M. and Egardt, B. (1998). Control design for integrator hybrid systems. *IEEE Transactions on Automatic Control*, **43**(4), 491-500.

Tong, H.(1983). *Threshold Models in Non-Linear Time Series Analysis*. Lecture Note in Statistics, Vol. **21**, Berlin Heideiberg: Springer.

Tong, H.(1990). *Non-Linear Time Series*. Oxford: Oxford University Press.

Van der Schaft, A.J. and Schumacher, J.M. (1998). Complementarity modelling of hybrid systems. *IEEE Transactions on Automatic Control*, **43**(4), 483-490.

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

Wei, H.L. and Billings, S.A.(2003). Identification and reconstruction of chaotic systems using multiresolution wavelet decompositions (submitted for publication).

Wei, H.L. and Billings, S.A. (2004). A unified wavelet-based modelling framework for nonlinear system identification: the WANARX model structure. *Accepted for publication by International Journal of Control*.

Wei, H.L., Billings, S.A. and Balikhin, M.A.(2002). Wavelet-based nonparametric additive models for nonlinear system identification. In *Proceedings of Annual Conference of Chinese Automation and Computer Society in the United Kingdom*, Beijing, P.R. China, Sep 20-22, 2002, pp20-25.

Wei, H.L., Billings, S.A. and Liu J. (2004). Term and variable selection for nonlinear system identification. *International Journal of Control*, **77**(1), 86-110.

Zhang, Q. and Benveniste, A. (1992). Wavelet networks. *IEEE Transactions on Neural Networks*, **3**(6), 889-898.