



This is a repository copy of *Continuous-Time System Identification for Linear and Nonlinear Systems using Wavelet Decompositions*.

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/80503/>

Monograph:

Coca, D. and Billings, S.A. (1996) *Continuous-Time System Identification for Linear and Nonlinear Systems using Wavelet Decompositions*. Research Report. ACSE Research Report 611 . Department of Automatic Control and Systems Engineering

Reuse

Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher's website.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

629
.8
(S)

Continuous-Time System Identification for Linear and Nonlinear Systems using Wavelet Decompositions

D. Coca



S.A. Billings

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

Research Report No. 611

January 1996

200328359



Continuous-Time System Identification for Linear and Nonlinear Systems using Wavelet Decompositions

D. COCA and S.A. BILLINGS

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

Abstract

A new approach for estimating linear and nonlinear continuous-time models directly from noisy observations is introduced using wavelet decompositions. Results using both simulated and experimental data are included to demonstrate the performance of the new algorithm.

1 Introduction

In many practical situations it is required that the description of dynamical processes is given in terms of mathematical models so that proper analysis and design procedures can be carried out. Situations where such models are required can occur in engineering and a wide range of fields and cover a broad spectrum of applications including investigation of system behaviour under various conditions, forecasting, implementation of automatic control systems, simulation, failure detection and fault diagnosis.

Although most systems in the physical world are continuous and although the study of such systems has triggered important developments in many fields, continuous-time models are replaced by discrete approximations in many applications because these are better suited for numerical implementation. However, in many applications, it is desirable to use continuous-time models. The advantage of using such representations in self tuning control for example has been highlighted by Gawthrop (1982).

Discrete time models, regarded as discretisations of continuous processes can present difficulties because they are only valid for the selected sampling interval and the use of such models for control can generate problems related to an excessive sensitivity to errors

at high sampling rates. The sampling rate is critical especially in system identification, where the system of interest is actually unknown, so that the use of a given sampling rate gives rise to uncertainties with respect to the validity of the approximation.

The approaches followed in the past to obtain continuous-time models can be classified into two main categories. The first category includes the so called direct methods, which attempt to estimate the system parameters directly from the data, while the second category comprises all the indirect methods. Both approaches have been used in the past, with few exceptions, for the case of linear systems.

A major difficulty of direct identification of continuous-time models is that the derivatives of the system input-output signals are not measured directly and numerical differentiation may accentuate the effects of noise. This is why in the past, methods involving direct generation of time derivatives of the signals involved, either in a physically or computationally way have been regarded as reliable only in deterministic situations and for low noise levels.

Direct methods include the use of modulating functions (Shinbrot, 1957; Perdreauville and Goodson, 1966), digital filters (Sagara et al., 1991a,b) or the spectral representation of the signals which are expanded in terms of orthogonal functions such as piecewise continuous functions or orthogonal polynomials (Paraskevopoulos, 1985).

Indirect methods are usually performed in two steps. In the first instance a non-parametric model is identified from the discrete samples of the continuous lumped dynamical system. Most non-parametric representations of the system are based on the impulse response, step response or the frequency response function. At the second stage, using an appropriate inter-domain relation, the non-parametric model is transformed into an equivalent continuous-time model.

Another possibility is that instead of a non-parametric model, at the first stage, a discrete-time model of the system of interest is estimated initially. The discrete-time model is then used in the second stage to obtain the continuous-time version of this model. This is done by using for example the frequency response function of the discrete-time model to estimate equivalent linear or nonlinear continuous-time models by fitting parametric models to the frequency response data (Tsang and Billings, 1992; Swain and Billings, 1995).

A more in depth presentation of existing techniques for continuous-time system

identification can be found in the book by Unbehauen and Rao(1987) or in the survey papers by Young(1981), Unbehauen and Rao(1988).

In this paper, a new direct method of estimating linear or nonlinear continuous-time representations of a dynamical process is presented. The continuous-time model of the system under study is derived using the measured input-output signals and the corresponding derivatives are estimated directly from the noisy observations using a multiresolution wavelet decomposition of the signals. The method has the advantage that it can be applied in the more general context of nonlinear systems. The approach is tested on a simulated example and using real data collected from an electronic circuit.

2 Dynamical systems

The notion of a dynamical system covers any system evolving in time which can be described as a set of ordinary differential equations or by discrete mappings or difference equations. The actual notion of a dynamical system is actually larger and includes those systems which can be modelled by partial differential equations. In this context dynamical systems described by ordinary differential equations are sometimes called lumped dynamical systems.

Ordinary differential equations are used to model systems which evolve continuously in time. Such systems occur frequently in practice and can be represented in general by the following equations

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t)); \\ \mathbf{y}(t) &= \mathbf{h}(\mathbf{x}(t), \mathbf{u}(t));\end{aligned}\tag{1}$$

with $\mathbf{f} : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n$ representing the vector field and $\mathbf{h} : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^p$ the observation function. In equation (1) \mathbf{x} represents the state variable, a vector of dimension n , \mathbf{u} is the input or control, a vector of dimension m , and \mathbf{y} is the output, a vector of dimension p .

Equation (1) gives the so called state space representation of a continuous-time dynamical system.

The state space representation of a system can be converted, subject to some additional assumptions, into a set of nonlinear higher-order differential equations in the inputs and outputs.

$$P_i(y, \dot{y}, \dots, y^{(k)}) = Q_i(u, \dot{u}, \dots, u^{(k)}) \quad (2)$$

or alternatively by combining inputs and outputs into one vector of external variables the system can be described as

$$R_i(u, \dot{u}, \dots, u^{(k)}, y, \dot{y}, \dots, y^{(k)}) = 0 \quad (3)$$

where $u^{(j)}$, $y^{(j)}$ denote the j -th time derivative of the input function u respectively output function y and $i = 1, \dots, p$. Equation (2) can be obtained practically by eliminating in (1) the states x to obtain the direct relation between the inputs and outputs. Expression (2) is sometimes called the input-output description or the external differential representation of a dynamical system.

The inverse problem of obtaining a state space model from an input-output representation is known as the realisation problem. It should be noted here that the external differential representation given in equation (2) is one of the possible input-output representations of the dynamical system. Other input-output system descriptions include for example the Wiener-Volterra and Fliess series representation.

From equation (2) one may conclude that a continuous-time model of a system can be estimated from input-output data, collected experimentally from the system, providing the input and output derivatives involved can be accurately estimated and that an appropriate structure to approximate the input-output relation is used.

A method to estimate the required derivatives directly from the measured signals is proposed in the next section.

The choice of the approximation method includes the use of polynomial representations, widely used in system identification and control, neural networks or wavelet approximations. The choice as usual depends on the final application of the estimated model since in fact system identification is just one of the phases of activity integrated into a larger project and therefore cannot be considered as an isolated task. In this paper however, the model structure is derived using a wavelet approximation technique which

provides a powerful tool for the local approximation of functions.

3 Derivative estimation from discrete noisy data

The task of numerically differentiating discrete noisy data can prove to be difficult considering the fact that usually a numerical differentiation process tends to amplify the effects of noise so that higher-order derivatives can become strongly affected. The method proposed here to estimate the derivatives directly from data involves the use of a multiresolution approximation of the signal of interest based on B-Spline wavelets. The resulting representation of the signal takes the form of a series expansion in terms of smooth basis functions which can be easily differentiated to obtain the desired derivatives. Prior to this however the signal has to be smoothed so that the effects of noise are reduced and the signal and its derivatives can be determined with sufficient accuracy.

3.1 Wavelet approximations

A multiresolution decomposition of L^2 is a nested sequence of closed subspaces $\dots V_{-1} \subset V_0 \subset V_1 \subset V_2 \dots$ of L^2 having an empty intersection and a dense union and satisfying the translation and scaling properties

$$f(x) \in V_0 \Leftrightarrow f(2x) \in V_1 \tag{4}$$

$$f(x) \in V_0 \Leftrightarrow f(x - k) \in V_0$$

for all $j, k \in \mathbb{Z}$.

Any function f from L^2 can be approximated at resolution j by its orthogonal projection, denoted $P_j f$, on V_j .

The importance of the scaling property arises from the fact that a multiresolution approximation can be described by means of a single function ϕ and its translates and dilates. In this way at resolution j the projection P_j of a function f can be represented as a series which converges in the L^2 norm,

$$P_j f = \sum_k c_{j,k} \phi_{j,k} \quad (5)$$

where $\phi_{j,k} = 2^{j/2} \phi(2^j x - k)$.

The wavelet subspaces W_j can be introduced in the context of multiresolution analysis as the orthogonal complement of V_j with respect to the next resolution subspace V_{j+1}

$$V_{j+1} = V_j \oplus W_j \quad (6)$$

where \oplus denotes the orthogonal sum of subspaces. Each wavelet subspace W_j is generated by a single function $\psi(x)$ and its dilates and translates.

The projection $Q_j f$ of a function f on the wavelet subspace W_j also has a series representation in terms of the dilates and translates of the wavelet function $\psi(x)$ as follows

$$Q_j f = \sum_k d_{j,k} \psi_{j,k} \quad (7)$$

From equation (6) it follows that

$$P_{j+1} f = P_j f + Q_j f \quad (8)$$

This gives an alternative series representation of the projection of a function $f \in L^2$ as it was first introduced in (5) this time using both the scaling and wavelet functions

$$P_j f = \sum_k c_{j-l,k} \phi_{j-l,k} + \sum_k \sum_{i=j-l}^j d_{i,k} \psi_{i,k} \quad (9)$$

The projection $P_j f$ should be understood as the approximation of the function f at resolution j . The resolution controls in this case the degree of approximation. The higher the resolution the better the approximation.

Equation (9) is the result of a multiresolution decomposition of the initial series (5) expressed just in terms of scaling basis functions. This decomposition and the corresponding reconstruction can be performed in a fast way using the pyramidal decomposition and reconstruction algorithms (Mallat, 1989). These algorithms allow the computation of the coefficients involved in equation (9) from the coefficients in (5) and vice versa using just

simple algebraic operations as follows:

Decomposition algorithm

$$\begin{cases} c_{j-1,k} = \sum_l a_{l-2k} c_{j,l} \\ d_{j-1,k} = \sum_l b_{l-2k} c_{j,l} \end{cases} \quad (10)$$

Reconstruction algorithm

$$c_{j,l} = \sum_k [p_{l-2k} c_{j-1,k} + q_{l-2k} d_{j-1,k}] \quad (11)$$

Equations (10) describe a moving average process involving the scaling coefficients $c_{j,k}$ at resolution j and the decomposition sequences $\{a_k\}_{k \in \mathbb{Z}}$ and $\{b_k\}_{k \in \mathbb{Z}}$. The resulting data sequence has to be subsequently downsampled or decimated by a factor of two by taking every second point in the sequence in order to obtain the desired coefficients.

This decomposition can be continued by employing, at the next stage, the coefficients $c_{j-1,k}$ to perform the moving average procedure.

A similar moving average scheme is used in the reconstruction algorithm (11) this time with the weighting sequences $\{p_k\}_{k \in \mathbb{Z}}$, $\{q_k\}_{k \in \mathbb{Z}}$. In this case upsampling (inserting a zero between every two consecutive points in the input sequence $\{c_{j-1,k}\}_{j,k \in \mathbb{Z}}$ and $\{d_{j-1,k}\}_{j,k \in \mathbb{Z}}$ is required before the moving average scheme is performed.

The use of both algorithms can be understood from the next section concerning smoothing a discrete noisy data sequence.

3.2 Wavelet smoothing

It is well known that the problem of measurement is in most cases subject to errors. These errors are partly of a systematic and partly of an accidental nature. While the systematic errors may be due to some miscalibrations of the instrument employed for taking the measurements, accidental errors arise from interference effects, drift and several other sources and are subject to uncontrollable fluctuations which can normally be described by statistical laws. These accidental errors which occur during measurements are usually referred to as the noise on the measurements.

3.2.1 Using mutual information to smooth the data

A model of a sampled signal affected by measurement errors can be described by the equation

$$y_\epsilon(t_i) = y(t_i) + \epsilon(t_i) \quad (12)$$

where y is the function representing the signal of interest and $y_\epsilon(t_i)$ is the set of noisy observations obtained by uniformly sampling the analog signal. The random errors are represented by $\epsilon(t_i)$ and are assumed to satisfy

$$E[\epsilon(t_i)] = 0 \quad (13)$$

$$E[\epsilon(t_i)\epsilon(t_j)] = \delta_{i,j}$$

with E denoting the mathematical expectation.

The noise content of the signal can be described by the noise to signal ratio defined as

$$NSR\% = \frac{\sigma_\epsilon}{\sigma_y} \times 100 \quad (14)$$

where σ_y and σ_ϵ are the standard deviation of the signal and noise respectively.

The discrete noisy signal can be approximated by a wavelet series at resolution j leading to the following expression

$$y_\epsilon(t_i) = \sum_k c_{j-p,k} \phi_{j-p,k}(t_i) + \sum_k \sum_{l=j-p}^j d_{l,k} \psi_{l,k}(t_i) \quad (15)$$

Although the noise signal does not belong to the space of square integrable functions having infinite energy the right hand side series will converge in L^2 norm (Cambanis and Masry, 1994) since stationary random processes are square integrable over every finite interval.

Because of the presence of noise the coefficients of the wavelet series representation of the noisy signal will also be stochastic. The stochastic properties of the wavelet

coefficients can be characterised by calculating the expected value and variance of the coefficients (Coca and Billings, 1995).

The algorithm presented here makes use of the mutual information present in the signal to minimise the noise contribution to the wavelet coefficients (Coca and Billings, 1995). Such a procedure is equivalent to a smoothing operation performed over the noisy signal.

The noisy signal is in practice the result of a data acquisition procedure. At this stage the continuous signal produced by a physical process has to be sampled with a sampling frequency f_s which satisfies the Whittaker-Shannon condition $f_s > 2f_h$ where f_h is the highest frequency present in the continuous signal which is considered to be band limited. In most cases the performance of the data acquisition equipment can handle fairly high sampling rates and this allows the user to oversample the continuous signal $f_s > 2kf_h$.

At the first stage the initial data set is separated into k different subsets

$$\{y_\epsilon(t_{k(i-1)+p})\}_{i=1,n} = \{y_\epsilon^p(t_i)\}_{i=1,n} \quad (16)$$

with $k < f_s/2f_h$, $k \in \mathbb{Z}$, $p = 1, \dots, k$. This is done in fact by downsampling by k the original data sequence with k successive starting points $y_\epsilon(t_1), \dots, y_\epsilon(t_k)$. As a result $y_\epsilon^1(t_i), \dots, y_\epsilon^k(t_i)$, $i = 1, \dots, n$ represent k successive samples of the original data record stored in k different data sets. Because the sampling time $\delta t = 1/f_s$ is sufficiently small, k successive samples can be considered to have a linear variation since the Taylor series expansion of the signal around the central value within the k samples interval can be truncated to the first derivative. As a result the mean value of k successive samples of the noise-free signal is the central value of the interval that is the $(k+1)/2$ -th sample for k odd. This can be expressed as

$$\frac{1}{k} \sum_{p=0}^{k-1} y^p(t_i) \cong y^{(k+1)/2}(t_i) \quad (17)$$

for any $i = 1, \dots, n$. Each of the k signals can be represented independently as a wavelet series (15).

Consider the noisy wavelet coefficients of the p -th signal at scale j to be denoted as $\{d_{j,k}^p\}$ and consider the properties of the mean

$$\bar{d}_{j,k} = \frac{1}{k} \sum_{p=1}^k \hat{d}_{j,k}^p = \frac{1}{k} \sum_{p=1}^k d_{j,k}^p + \frac{1}{k} \sum_{p=1}^k d\epsilon_{j,k}^p \quad (18)$$

where $\{d_{j,k}^p\}$ are the noise-free wavelet coefficients and $\{d\epsilon_{j,k}^p\}$ represents the stochastic part of the wavelet coefficients. Since the coefficients are the result of a linear transformation performed over the signal it follows that

$$\frac{1}{k} \sum_{p=1}^k d_{j,k}^p = d_{j,k}^{\frac{k+1}{2}} \quad (19)$$

for k odd. As a result the expected value of the mean is given by

$$\mathbf{E} \left[\bar{d}_{j,k} \right] = \mathbf{E} \left[d_{j,k}^{\frac{k+1}{2}} \right] + \frac{1}{k} \sum_{p=1}^k \mathbf{E} \left[d\epsilon_{j,k}^p \right] = d_{j,k}^{\frac{k+1}{2}} \quad (20)$$

and this provides an unbiased estimate of the wavelet coefficients corresponding to the signal having the index $p = (k + 1)/2$. The variance of the same variable is given by

$$\mathbf{E} \left[\left(\bar{d}_{j,k} - d_{j,k}^{\frac{k+1}{2}} \right)^2 \right] = \frac{1}{k^2} \sum_{p=1}^k \mathbf{E} \left[\left(d\epsilon_{j,k}^p \right)^2 \right] = \frac{\sigma_{d\epsilon_{j,k}}^2}{k} \quad (21)$$

since the noise components affecting each of the k signals can be considered uncorrelated. The results show that the effects of taking the mean, are that the variance deviation of the stochastic part of the wavelet coefficients is reduced k times. The overall effect is that the variance of the noise affecting the signal is also reduced by the same amount.

3.2.2 High frequency noise reduction

Another important property of the noisy wavelet coefficients is that the variance of the coefficients is theoretically scale invariant (Coca and Billings, 1995). In practice the variance decreases slightly with the scale due to the fact that in this case we are dealing with finite length signals. However for all practical purposes the variance of the wavelet coefficients can be considered invariant for a given range of scale.

The above behaviour of the wavelet coefficients yields a very effective method for separating the true signal $y(t)$ from the observed signal which from equation (12) is the true signal plus noise.

If the noisy signal is approximated in terms of a wavelet series, as in the previous section, it can be observed that above a certain scale j the wavelet coefficients do not have any tendency to become smaller but remain within a certain band of fairly constant amplitude. These coefficients will account for the high frequency components which are caused purely by noise. This is because the deterministic signal is by definition a bandlimited signal so the high frequency components of such a signal above the cut-off frequency make a negligible contribution to the spectrum of the signal. This in turn means that the wavelet coefficients of the noise-free signal corresponding to that frequency band should be very small, nearly zero.

By simply setting to zero and thus omitting the corresponding wavelet functions from the series representation, the signal will be sensibly smoothed. Because the genuine signal components may not be perfectly zero above the frequency we have chosen to define the noise band, it is anticipated that some errors will be introduced in estimating the deterministic signal. However these errors will generally be very small considering the fact that at higher scales the contribution of the signal is almost zero.

It is important to note that the above procedure has a global character since it involves all the coefficients representing the signal at a given scale. This in turn means that some nice properties of the wavelet functions are not yet fully exploited. The most important is the time frequency localisation offered by a multiresolution decomposition. This is a result of the time-frequency localisation of the wavelet function which acts as a window function in both the time and frequency domain and provides a local Fourier analysis which takes place at every single scale.

This feature is especially useful when dealing with deterministic signals with a time-varying frequency behaviour, with local oscillations or discontinuities. In such cases wavelets, due to the compact support (or at least spatial localisation) have the ability to locally adapt to the feature of the signal. The amplitude of the wavelet coefficients in this case reflects the contribution of each function to the approximation. A local high frequency burst of the signal can in this way be identified and isolated from the noise. In such cases instead of setting all the coefficients at the corresponding scale to zero, we can choose to preserve the high magnitude coefficients which represent a feature of the deterministic signal. However, if the multiresolution decomposition is not orthonormal, it is more correct to say that the corresponding wavelet functions are preserved since the

respective coefficients have to be re-estimated under the additional constraint that all other coefficients at the same scale are zero.

If high frequency smoothing is performed over the signal smoothed using the previous algorithm it will be more obvious which coefficients represent just noise since the reduction in the magnitude of the coefficients after the first algorithm has been performed will affect just the stochastic part of the coefficients.

The difference between the two algorithms described above is that while the first will reduce the noise at all frequency bands including the noise superimposed over the signal band the second algorithm removes the noise above the local characteristic frequency content of the signal.

3.3 Construction of B-spline multiresolution approximations

To smooth the discrete noisy data and estimate the derivatives a nonorthogonal multiresolution approximation was constructed using B-spline functions. A complete theory of B-spline multiresolution approximations can be found in (Chui, 1992). Only the basic properties will be presented below for completeness.

A B-spline multiresolution approximation is constructed by taking as a scaling function the m -th order cardinal B-spline function $\beta^m(x)$ with $m \in \mathbb{Z}$ which is defined recursively by the integral convolution

$$\beta^m(x) = \int_{-\infty}^{\infty} \beta^{m-1}(x-t)\beta^1(t)dt \quad (22)$$

where $\beta^1(x)$ is the characteristic function of the interval (called also the indicator function) $\chi(x)$.

$$\beta^1(t) = \chi(t) = \begin{cases} 1 & \text{if } x \in (0, 1) \\ 0 & \text{otherwise} \end{cases} \quad (23)$$

In practice the B-spline basis functions of higher order can be defined starting with the first order basis function (of degree zero) given explicitly in (23) by the following recursive algorithm

$$\beta^m(x) = \frac{x}{m-1}\beta^{m-1}(x) + \frac{m-x}{m-1}\beta^{m-1}(x-1) \quad (24)$$

Using this formula it is easy to see that the scaled and translated version of a B-spline function can be calculated using a similar recurrence relation

$$\beta_{j,k}^m(x) = \frac{2^j x - k}{m-1} \beta_{j,k}^{m-1}(x) + \frac{m+k-2^j x}{m-1} \beta_{j,k}^{m-1}(x-1) \quad (25)$$

where $\beta_{j,k}^m = 2^{j/2} \beta^m(2^j x - k)$. The B-spline function β^m (of order m) consists in fact of m nontrivial polynomial pieces of degree $m-1$ so that

$$\beta^m|_{[k-1,k]} = P_{m-1,k}, \quad k = 1, \dots, m \quad (26)$$

where $P_{m-1,k}$ is a polynomial of degree $m-1$. Denoting Π_n as the collection of all polynomials of degree at most n and with C^n the space of continuous functions having up to n continuous derivatives, allows us to define the subspace V_0 generated by $\phi(x) = \beta^m(x)$ as the subspace of all functions $f \in C^{m-2} \cap L^2(\mathbb{R})$ such that the restriction of f to any interval $[k-1, k]$ is a polynomial with degree at most $m-1$. Each polynomial piece can be computed analytically which gives an alternative way to compute the value of the B-spline functions at any point.

The counterpart of the scaling function, namely the wavelet function can be constructed (Chui, 1992) as a linear combination of the B-spline scaling functions. It can be shown that the compactly supported wavelets with minimum support that correspond to the m -th order cardinal B-spline are unique up to multiplication by a constant. The support of the m -th order B-spline wavelet is an interval of length $2m-1$ and all wavelets are symmetric for m even and anti symmetric for m odd.

It is interesting to note that in fact scaling functions and wavelets can be considered as filter functions. The difference between them is that while the scaling function acts as a low pass filter, the corresponding wavelet function behaves like a bandpass filter. If a signal is represented using the multiresolution approximation approach, such as a wavelet series, the reconstructed signal is nothing other than the result of a linear filtering process. A usual requirement in such cases, to avoid distortions, is that the filter should have linear or at least generalised linear phase. It can be shown (Chui, 1992) that both the B-spline scaling and wavelet functions satisfy this requirement.

3.3.1 Estimation of the derivatives of a B-spline series

The derivatives of the signal can be determined, after smoothing has been performed, by simply differentiating the reconstructed B-spline series. Following de Boor (de Boor, 1978) The derivatives of a B-spline series at scale $j = 0$ can be written as

$$D \left[\sum_k c_k \beta_k^m \right] = \sum_k (c_k - c_{k-1}) \beta_k^{m-1} \quad (27)$$

This shows that the first derivative of a B-spline series can be found simply by differencing its B-spline coefficients, to obtain the coefficients of a B-spline series of one order lower which represents this derivative. Since formula (27) applies for infinite length signals, the derivatives of a finite B-spline series are obtained by making the series bi-infinite through adding B-spline functions with zero coefficients.

The formula to compute the derivatives of a B-spline series at any resolution j is as follows

$$D \left[\sum_k c_{j,k} \beta_{j,k}^m \right] = 2^j \sum_k (c_{j,k} - c_{j,k-1}) \beta_{j,k}^{m-1} \quad (28)$$

and can be derived easily from (27). By repeating the application of (28), the following formula of the i -th derivative of a spline series can be written:

$$D^i \left[\sum_k c_{j,k} \beta_{j,k}^m \right] = \sum_k c_{j,k}^{i+1} \beta_{j,k}^{m-1} \quad (29)$$

where

$$c_{j,k}^{i+1} = \begin{cases} c_{j,k} & \text{for } i = 0 \\ 2^{ij} (c_{j,k}^i - c_{j,k-1}^i) & \text{for } i > 0 \end{cases} \quad (30)$$

The whole procedure of obtaining the derivatives of a signal from discrete noisy observations can be summarised as follows:

- The noisy signal of interest is separated into k signals following the procedure described in (16) with k selected according to the oversampling factor: $k < f_s/2f_h$.
- Each signal is approximated by an m -th order B-spline series at resolution j . The

order of the B-spline series is chosen so that the signals resulting from subsequent differentiations are at least C^2 while the resolution j is selected in order that the discrete signal is approximated with very good accuracy.

- For each signal a wavelet decomposition algorithm is performed p times. Here p depends on the length of the data sequence and on the initial resolution level j .
- ✓ Perform a high frequency smoothing by setting to zero the wavelet coefficients at higher scales which reflect high frequency components above the characteristic frequency band of the noise-free signal.
- The signal is then reconstructed using the mean values of the wavelet coefficients corresponding to the k signals computed as in (18).
- Compute the derivatives of the reconstructed B-spline series using relation (30).

4 Numerical and experimental results

This section is devoted to numerical and experimental implementation of the proposed method. The simulated example uses the Duffing equation as a benchtest to assess the performance of the approach. Having estimated the derivatives from the simulated noisy input-output discrete observations the problem becomes a parameter estimation task if the structure of the input-output relation is considered known. The accuracy of the method can be then judged by comparing the estimated parameters with the ones used in the simulation.

The experimental example involves the use of a real data set recorded by sampling the signals from an electronic implementation of Chua's circuit.

4.1 Duffing-Ueda oscillator

Consider the Duffing-Ueda oscillator (Ueda, 1985) described by the following ordinary differential equation

$$\frac{d^2 y}{dt^2} + 0.1 \frac{dy}{dt} + y^3 = u(t) \quad (31)$$

which is an example of a periodically driven nonlinear oscillator with

$$u(t) = A \cos(\omega t). \quad (32)$$

Usually when such nonlinear oscillators are driven by periodic forces a variety of phenomena can occur depending both on the type of oscillator and the frequency and amplitude of the driving force. Such a periodic force applied to a passive oscillator can yield for example hysteresis, harmonic and subharmonic oscillations or chaotic motion.

The Duffing equation which can be used to model the hardening spring effect observed in many mechanical systems has emerged as one of the most important paradigms in the study of chaos.

Equation (31) driven by an input as described in (32) with $\omega = 1$ rad/sec and $A = 11$ was simulated using a fourth-order Runge-Kutta algorithm with an integration step equal to $\pi/300$ s. In this way 10000 data points of the output y were generated. White noise was then added to the simulated output signal so that the noise-to-signal ratio of the resulting signal y_ϵ , defined as in (14), is $NSR\% = 10\%$.

To estimate a continuous time model in the form of (31) the first step is to estimate smooth versions of y , \dot{y} and \ddot{y} directly from the noisy observations y_ϵ .

For the example chosen the smoothing procedure, summarised in the previous section, was performed with $j = 10$, $p = 8$, $m = 6$ and $k = 9$. The smoothness of the

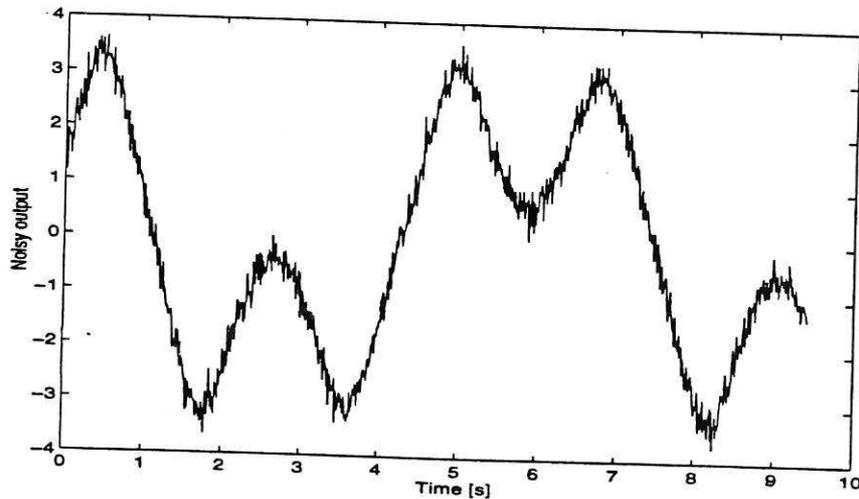


Figure 1: The simulated noisy output of the Duffing-Ueda oscillator

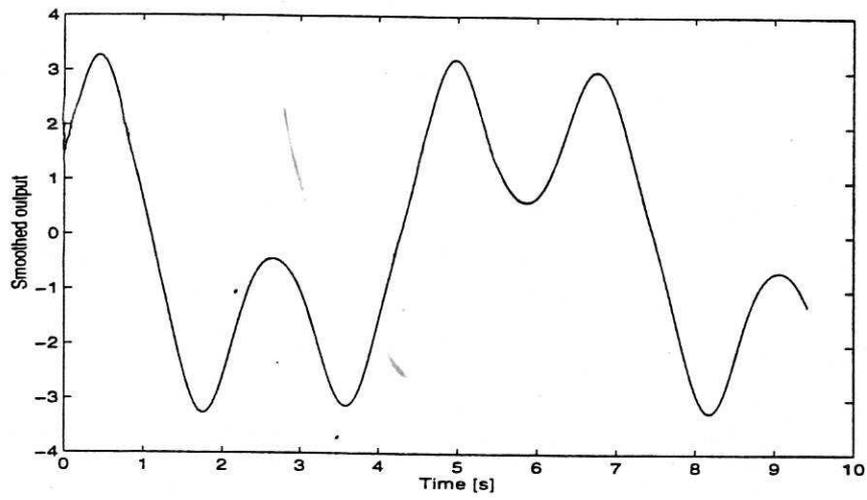


Figure 2: Smoothed output of the Duffing-Ueda oscillator

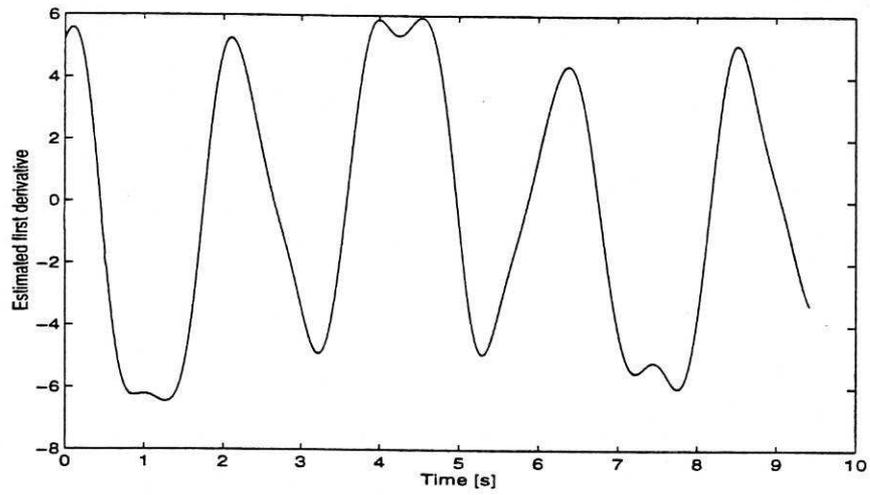


Figure 3: Smoothed first derivative

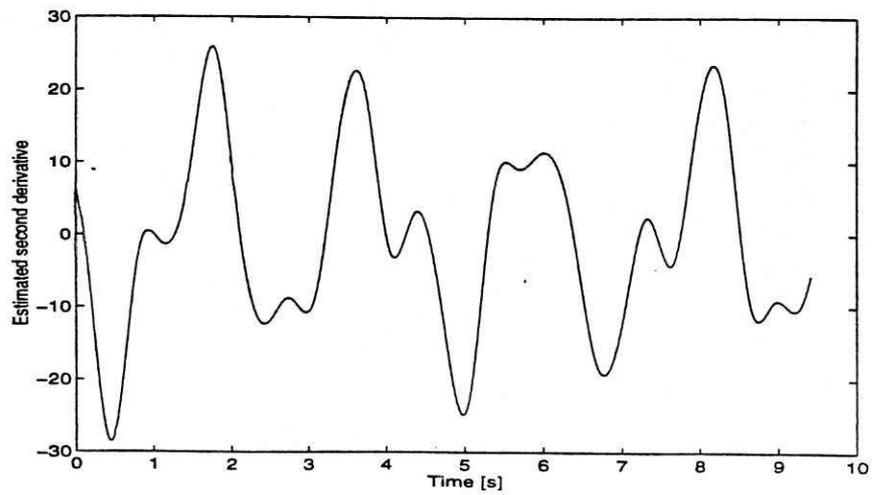


Figure 4: Smoothed second derivative

wavelet approximation is given by the order of the polynomial pieces (quintic for $m = 6$) which make up the B-spline scaling and wavelet functions.

One of the $k = 9$ signals resulting from downsampling the original signal is represented in Fig.(1).

Using the smoothed output and the corresponding first and second order derivatives represented in Fig.(2, 3, 4), the parameters of the continuous-time model in the form (31) can be estimated using a least-squares algorithm, leading to the following equation

$$0.9627 \frac{d^2 y}{dt} + 0.1017 \frac{dy}{dt} + 0.9687 y^3 = 11 \cos(t) \quad (33)$$

The coefficients in equation (33) show a very good agreement with the real coefficients used in simulation giving an absolute approximation error which is less than 4% of the real coefficient value.

4.2 Chua's circuit

In the last few years, a large number of papers have been devoted to the numerical, experimental and analytical aspects of the third order electrical circuit illustrated in Fig.(5a). This circuit, now widely known as the Double-Scroll or Chua's circuit, exhibits a great variety of nonlinear and chaotic dynamics which have been simulated, observed and mathematically proven (Chua, 1992; Matsumoto et al., 1993). This circuit consists of only one inductor (L), two capacitors (C_1, C_2), one linear resistor (G), and one nonlinear resistor (N_R) which has the piecewise linear characteristic $g(v_{C1})$ depicted in Fig.(6).

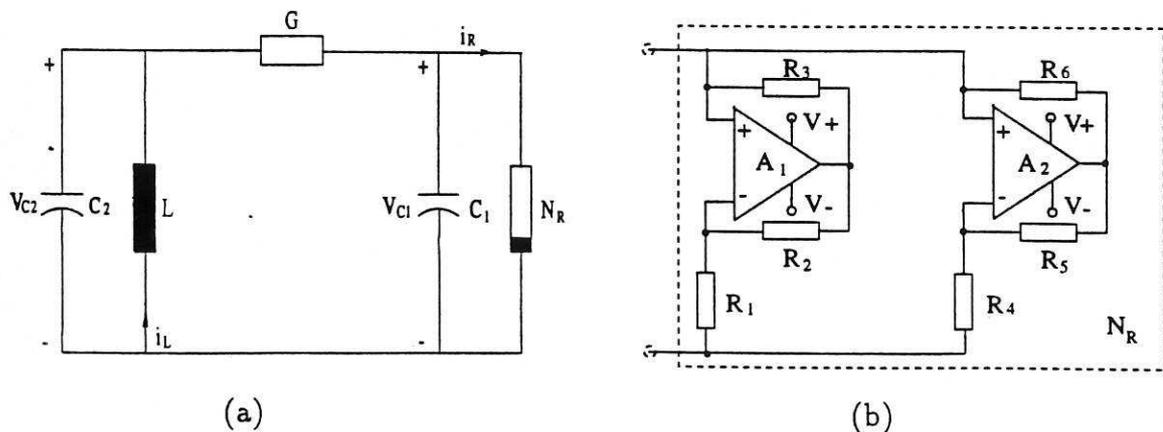


Figure 5: Chua's circuit(a). Nonlinear resistor implemented using op-amps(b)

The dynamical behaviour of the circuit is governed by the following system of three first order ordinary differential equations, derived using elementary laws from electric circuit theory

$$\begin{aligned} C_1 \frac{dv_{C_1}}{dt} &= G(v_{C_2} - v_{C_1}) - g(v_{C_1}) \\ C_2 \frac{dv_{C_2}}{dt} &= G(v_{C_1} - v_{C_2}) + i_L \\ L \frac{di_L}{dt} &= -v_{C_2} \end{aligned} \quad (34)$$

where

$$g(v_{C_1}) = \begin{cases} m_1 v_{C_1}, & |v_{C_1}| < B_p \\ m_0 v_{C_1} + B_p(m_1 - m_0), & v_{C_1} \geq B_p \\ m_0 v_{C_1} - B_p(m_1 - m_0), & v_{C_1} \leq -B_p \end{cases} \quad (35)$$

The circuit can be realized practically in a variety of ways using standard components. A single-chip integrated circuit (IC) is also available. The simple and robust implementation of Chua's circuit, described in Kennedy(1993b), and which makes use of commercially available components to realize the nonlinear resistor N_R using operational amplifiers (see Fig.(5b)) was used in the present study.

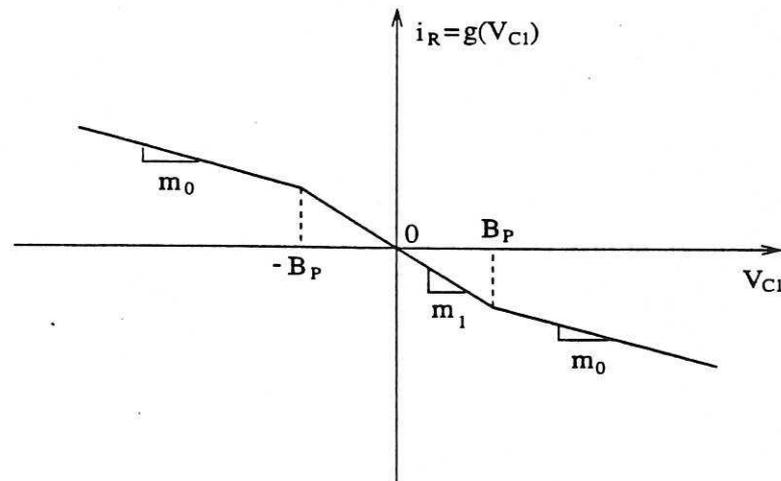


Figure 6: Nonlinear resistor characteristics.

Equations (34) can be transformed by rescaling (Matsumoto et al., 1993)

$$\begin{aligned} x &= v_{C_1}/B_p, & y &= v_{C_2}/B_p, & z &= i_L/(B_p G), & \tau &= tG/C_2, \\ a &= m_1/G, & b &= m_0/G, & \alpha &= C_2/C_1, & \beta &= C_2/(LG^2), \end{aligned} \quad (36)$$

to yield the following form

$$\begin{aligned} \frac{dx}{dt} &= \alpha(y - h(x)) \\ \frac{dy}{dt} &= x - y + z \\ \frac{dz}{dt} &= -\beta y \end{aligned} \quad (37)$$

with

$$h(x) = \begin{cases} ax, & |x| < 1 \\ bx + a - b, & x \geq 1 \\ bx - a + b, & x \leq -1 \end{cases} \quad (38)$$

A data set consisting of 50,000 samples, obtained by sampling and recording the three state variables from the circuit with a sampling time $dt = 2\mu s$, (Dedieu and Ogorzałek, 1994) was used for the estimation of nonlinear continuous-time models.

Two continuous-time models were estimated, one involving all three state variables which requires only the estimation of the first derivatives of each signal and a second using just a single measured state variable namely the current i_L through the inductor, in which the derivatives of the signal up to third order have to be estimated.

The derivatives of all the signals were estimated using a multiresolution decomposition of the signals as in the previous example. 10,000 data points of the measured time series were used to perform wavelet smoothing and to estimate the first order derivatives of each signal plus the second and third order derivatives of the signal current i_L .

A state space continuous-time model taking the form of a third order system of differential equations

$$\begin{aligned}
\frac{dv_{C_1}}{dt} &= F_1(v_{C_1}, v_{C_2}, i_L); \\
\frac{dv_{C_2}}{dt} &= F_2(v_{C_1}, v_{C_2}, i_L); \\
\frac{di_L}{dt} &= F_3(v_{C_1}, v_{C_2}, i_L);
\end{aligned} \tag{39}$$

can be estimated by approximating the three dimensional vector field $F(v_{C_1}, v_{C_2}, i_L)$ in equation (39) using the measured signals and the corresponding first order derivatives. Usually, when the order of the system is not known a priori, this parameter has to be selected iteratively. However when the system is proven to be chaotic the initial model order has to be three, a necessary condition for the existence of chaos in a system governed by ordinary differential equations.

In this example the vector field was approximated using a wavelet approximation approach. Each component of the vector field was approximated as a linear combination of three dimensional B-spline scaling and wavelet functions formed as a tensor product of the one dimensional basis functions. The number of parameters to be estimated, which is equal with the number of basis functions, is chosen iteratively. Starting with a resolution level j which gives the initial number of basis functions needed (scaling basis functions), more basis functions (wavelet basis functions) can be added subsequently to improve the degree of approximation. The order of the basis functions, $m = 4$ for this example, can also be selected iteratively starting with a low order which is later increased if the approximation error does not decrease significantly with the addition of new basis functions.

More explicitly, the continuous-time model takes the form

$$\begin{aligned}
\frac{dv_{C_1}}{dt} &= \sum_{k_1, k_2, k_3} c_{j_1, \dots, j_3, k_1, \dots, k_3}^1 \Phi_{j_1, \dots, j_3, k_1, \dots, k_3}(v_{C_1}, v_{C_2}, i_L) \\
&+ \sum_{m=1}^{2^3-1} \sum_{l_1, l_2, l_3} d_{i_1, \dots, i_3, k_1, \dots, k_3}^{l_1, m} \Psi_{i_1, \dots, i_3, k_1, \dots, k_3}^m(v_{C_1}, v_{C_2}, i_L)
\end{aligned}$$

$$\begin{aligned}
\frac{dv_{C_2}}{dt} &= \sum_{k_1, k_2, k_3} c_{j_1, \dots, j_3, k_1, \dots, k_3}^2 \Phi_{j_1, \dots, j_3, k_1, \dots, k_3}(v_{C_1}, v_{C_2}, i_L) \\
&+ \sum_{m=1}^{2^3-1} \sum_{i_1, i_2, i_3=j_1, j_2, j_3}^{l_1, l_2, l_3} \sum_{k_1, k_2, k_3} d_{i_1, \dots, i_3, k_1, \dots, k_3}^{2, m} \Psi_{i_1, \dots, i_3, k_1, \dots, k_3}^m(v_{C_1}, v_{C_2}, i_L) \\
\frac{di_L}{dt} &= \sum_{k_1, k_2, k_3} c_{j_1, \dots, j_3, k_1, \dots, k_3}^3 \Phi_{j_1, \dots, j_3, k_1, \dots, k_3}(v_{C_1}, v_{C_2}, i_L) \\
&+ \sum_{m=1}^{2^3-1} \sum_{i_1, i_2, i_3=j_1, j_2, j_3}^{l_1, l_2, l_3} \sum_{k_1, k_2, k_3} d_{i_1, \dots, i_3, k_1, \dots, k_3}^{3, m} \Psi_{i_1, \dots, i_3, k_1, \dots, k_3}^m(v_{C_1}, v_{C_2}, i_L)
\end{aligned} \tag{40}$$

with

$$\begin{aligned}
\Phi_{j_1, \dots, j_3, k_1, \dots, k_3}(x^1, x^2, x^3) &= \phi_{j_1, k_1}(x^1) \phi_{j_2, k_2}(x^2) \phi_{j_3, k_3}(x^3) \\
\Psi_{i_1, \dots, i_3, k_1, \dots, k_3}^1(x^1, x^2, x^3) &= \phi_{i_1, k_1}(x^1) \phi_{i_2, k_2}(x^2) \psi_{i_3, k_3}(x^3) \\
\Psi_{i_1, \dots, i_3, k_1, \dots, k_3}^2(x^1, x^2, x^3) &= \phi_{i_1, k_1}(x^1) \psi_{i_2, k_2}(x^2) \phi_{i_3, k_3}(x^3) \\
\Psi_{i_1, \dots, i_3, k_1, \dots, k_3}^3(x^1, x^2, x^3) &= \phi_{i_1, k_1}(x^1) \psi_{i_2, k_2}(x^2) \psi_{i_3, k_3}(x^3) \\
\Psi_{i_1, \dots, i_3, k_1, \dots, k_3}^4(x^1, x^2, x^3) &= \psi_{i_1, k_1}(x^1) \phi_{i_2, k_2}(x^2) \phi_{i_3, k_3}(x^3) \\
\Psi_{i_1, \dots, i_3, k_1, \dots, k_3}^5(x^1, x^2, x^3) &= \psi_{i_1, k_1}(x^1) \phi_{i_2, k_2}(x^2) \psi_{i_3, k_3}(x^3) \\
\Psi_{i_1, \dots, i_3, k_1, \dots, k_3}^6(x^1, x^2, x^3) &= \psi_{i_1, k_1}(x^1) \psi_{i_2, k_2}(x^2) \phi_{i_3, k_3}(x^3) \\
\Psi_{i_1, \dots, i_3, k_1, \dots, k_3}^7(x^1, x^2, x^3) &= \psi_{i_1, k_1}(x^1) \psi_{i_2, k_2}(x^2) \psi_{i_3, k_3}(x^3)
\end{aligned} \tag{41}$$

The coefficients in expression (40) were estimated by means of a least-squares algorithm and the resulting model simulated using a fourth order Runge-Kutta integration routine with an integration step $dt_{int} = 0.001$ which due to time rescaling is equivalent to

the initial sampling time $dt = 2\mu s$.

The well known Double-Scroll chaotic attractor represented in Fig.(7a) using the noisy measured time series compares very well (if the noise affecting the measured signals is ignored) with the attractor produced by simulating the estimated continuous time model plotted in Fig.(7b).

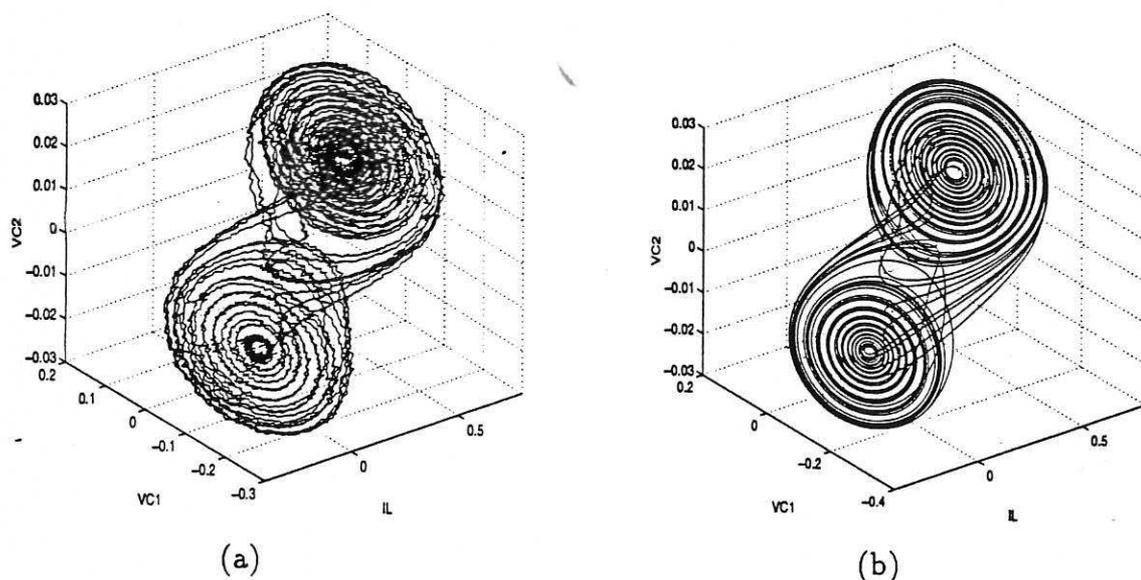


Figure 7: Double-Scroll attractor plotted using the measured signals (a) and the signals resulting from the simulation of the continuous-time model (b).

A second model was estimated using just a single time series. In many practical situations just a single output can be recorded from the observed system in which case the external differential representation has to be estimated.

To estimate a continuous time model in this form, the current signal i_L was used to generate the higher order derivatives.

Usually the input-output relation describing the dynamical system is more complex and therefore more difficult to approximate. In this particular case it is easy to see that by rewriting the system's equations (37) in the input-output form

$$\frac{d^3 z}{dt^3} = \frac{d^2 z}{dt^2} + (\alpha + \beta) \frac{dz}{dt} + \alpha \beta h \left(-\frac{1}{\beta} \frac{d^2 z}{dt^2} - \frac{1}{\beta} \frac{dz}{dt} - z \right); \quad (42)$$

the piecewise linear characteristic $g(v_{C1})$ of the nonlinear resistor N_R , that is $h(x)$ as in (38) after rescaling, becomes a function which depends on three variables with the break

points confined in two planes of equations

$$\begin{aligned}\Pi_1 : \quad & \frac{1}{\beta} \frac{d^2 z}{dt^2} + \frac{1}{\beta} \frac{dz}{dt} + z = 1; \\ \Pi_2 : \quad & \frac{1}{\beta} \frac{d^2 z}{dt^2} + \frac{1}{\beta} \frac{dz}{dt} + z = -1;\end{aligned}\tag{43}$$

Using a wavelet approximation a continuous time model was estimated in the form

$$\begin{aligned}\frac{d^3 i_L}{dt^3} = & \sum_{k_1, k_2, k_3} c_{j_1, \dots, j_3, k_1, \dots, k_3} \Phi_{j_1, \dots, j_3, k_1, \dots, k_3} \left(i_L, \frac{di_L}{dt}, \frac{d^2 i_L}{dt^2} \right) \\ & + \sum_{m=1}^{2^3-1} \sum_{l_1, l_2, l_3} \sum_{k_1, k_2, k_3} d_{i_1, \dots, i_3, k_1, \dots, k_3}^m \Psi_{i_1, \dots, i_3, k_1, \dots, k_3}^m \left(i_L, \frac{di_L}{dt}, \frac{d^2 i_L}{dt^2} \right)\end{aligned}\tag{44}$$

where the basis functions are defined in (41).

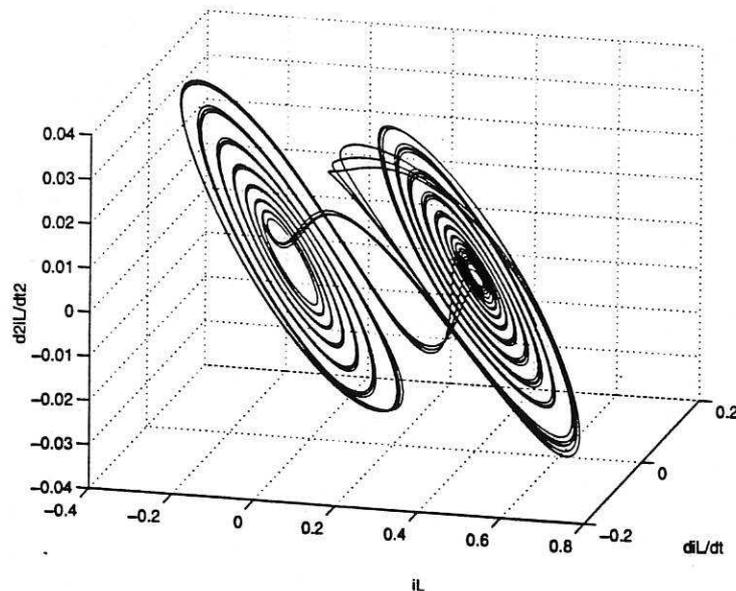


Figure 8: The current signal and the corresponding first and second order derivatives resulting from simulation of the estimated external differential representation of Chua's circuit.

The resulting model was then simulated using a fourth order Runge-Kutta algorithm with the integration step $dt_{int} = 0.001$. In Fig.(8) the resulting trajectory was plotted using the output and the corresponding first and second order derivatives of the output.

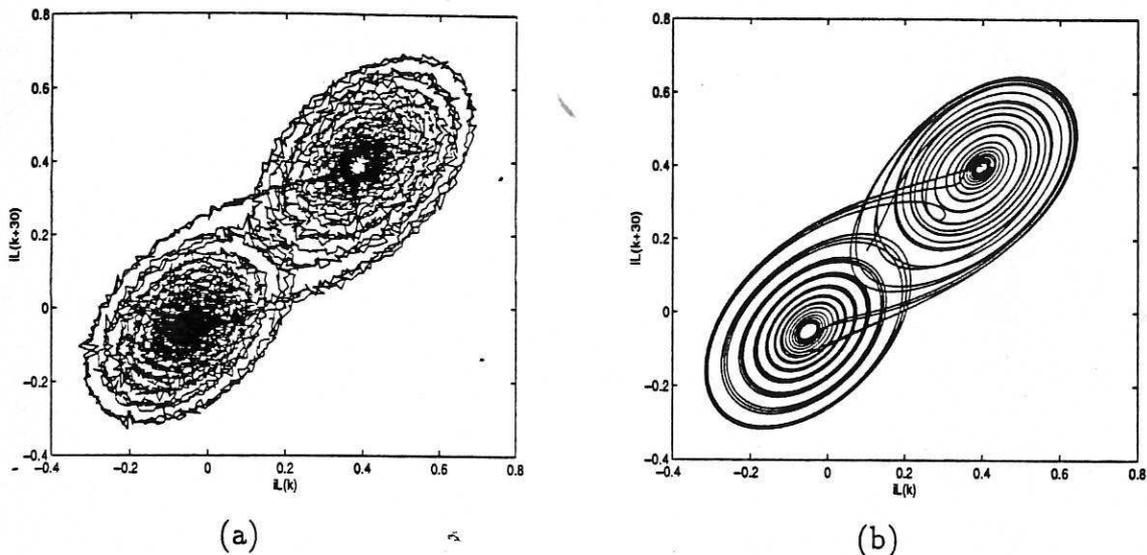


Figure 9: Double-Scroll attractor obtained by embedding the measured current signal i_L (a) and the associated simulated signal using the estimated continuous-time model eqn.(44) (b).

Using an embedding method both the measured current signal and the simulated signal were used to plot the Double-Scroll shown in Figs.(9a) and (9b) which illustrate excellent agreement between the embedded trajectories..

5 Conclusion

A new approach to estimate continuous-time nonlinear models directly from discrete noisy observations has been introduced based on an application of wavelet theory.

Using a nonorthogonal multiresolution wavelet decomposition based on B-spline wavelets the observed noisy signals are smoothed and then differentiated to provide the information needed to approximate the continuous-time description of the system of interest as a set of ordinary differential equations. The method was tested on both simulated

and experimental data produced by nonlinear systems and was shown to perform very well even in the presence of noise

6 Acknowledgments

SAB gratefully acknowledges that part of this work was supported by ESPRC. DC would like to express his thanks to the University of Sheffield for the award of a scholarship which made this study possible. Both authors express their gratitude to H. Dedieu who provided the experimental data used to illustrate their approach.

References

- Akansu, A. N. (1992). *Multiresolution signal decomposition. Transforms, subbands, and wavelets*. Academic Press, London.
- Cambanis, S. and Masry, E. (1994). Wavelet approximations of deterministic and random signals: convergence properties and rates. *IEEE Transactions On Information Theory*, 40(4):1013-1029.
- Chua, L. O. (1992). The genesis of Chua's circuit. *Archiv für Elektronik und Übertragungstechnik*, 46(4):250-257.
- Chui, C. K. (1992). *An introduction to wavelets*. Academic Press, New York.
- Coca, D. and Billings, S. A. (1995). Estimating derivatives from noisy data: A wavelet multiresolution decomposition approach. *Submitted to Int. J. of Control*.
- de Boor, C. (1978). *A practical guide to spline*. Springer Verlag.
- Dedieu, H. and Ogorzałek, M. (1994). Controlling chaos in Chua's circuit via sampled inputs. *Int. J. of Bifurcation and Chaos*, 4(2):447-455.
- Gawthrop, P. J. (1982). A continuous-time approach to discrete-time self-tuning control. *Optimal Control Applications & Methods*, 3:399-414.

- Kennedy, M. P. (1993a). Three steps to chaos-Part I: Evolution. *IEEE Transactions on Circuits and Systems-1: Fundamental theory and Applications*, 40(10):640-656.
- Kennedy, M. P. (1993b). Three steps to chaos - Part II: A Chua's circuit primer. *IEEE Transactions on Circuits and Systems-1: Fundamental theory and Applications*, 40(10):657-674.
- Ljung, L. (1987). *System identification-Theory for the user*. Prentice Hall Inc., Englewood Cliffs, New Jersey.
- Mallat, S. G. (1989). A theory of multiresolution signal decomposition; the wavelet representation. *IEEE Pattern Anal. and Machine Intelligence*, 11:674-693.
- Matsumoto, T., Komuro, M., Kokubu, H., and Tokunaga, R. (1993). *Bifurcations. Sights sounds and mathematics*. Springer-Verlag Tokyo.
- Meyer, Y. (1993). *Wavelets and operators*. Cambridge studies in advanced mathematics. Cambridge Univ. Press.
- Paraskevopoulos, P. N. (1985). System analysis and synthesis via orthogonal expansions in systems and control. *IMACS Trans.; Special issue on orthogonal expansions in systems and control*, 27:453-469.
- Perdreauville, P. J. and Goodson, R. E. (1966). Identification of systems described by partial differential equations. *Journal of Basic Eng., Trans. ASME, Series D*, 88:463-468.
- Ronald, A., DeVore, Jawerth, B., and Popov, V. (1992). Compression of wavelet decompositions. *American Journal of Mathematics*, (114):737-785.
- Sagara, S., Zi-Jiang, Y., and Wada, K. (1991a). Identification of continuous systems from noisy sampled input-output data. *Proceedings of IFAC Symposium on Identification and System Parameter estimation, Budapest, Hungary.*, pages 603-608.
- Sagara, S., Zi-Jiang, Y., and Wada, K. (1991b). Filtering approaches to identification of continuous systems. *Proceedings of IFAC Symposium on Identification and System Parameter Estimation, Budapest, Hungary.*, pages 305-310.

- Shinbrot, M. (1957). On the analysis of linear and nonlinear systems. *Trans. ASME*, 79:547-552.
- Swain, A. K. and Billings, S. A. (1995). Weighted complex orthogonal estimator for identifying linear and nonlinear continuous time models from generalised frequency response functions. *Submitted to Mech. Systems and Signal Processing*.
- Tsang, K. M. and Billings, S. A. (1992). Reconstruction of linear and nonlinear continuous-time models from discrete-time sampled data systems. *Mech. Systems and Signal Processing*, 6(1):69-84.
- Ueda, Y. (1985). Random phenomena resulting from nonlinearity in the system described by Duffing equation. *Int. J. Non-Linear Mech.*, 20(5/6):481-491.
- Unbehauen, H. and Rao, G. P. (1987). *Identification of continuous systems*, volume 10 of *North-Holland Systems and Control Series*. North-Holland, Amsterdam.
- Unbehauen, H. and Rao, G. P. (1988). Continuous-time approaches to system identification-A survey. *Automatica*, 26(1):23-35.
- Unser, M. and Aldroubi, A. (1992). Polynomial splines and wavelets; a signal processing perspective. In Chui, C. K., editor, *Wavelets. A tutorial in theory and applications*, pages 91-122. Academic Press, New York.
- Wahlberg, B., Ljung, L., and Söderström, T. (1991). On sampling and reconstruction of continuous-time stochastic systems. *Proceedings of IFAC Symposium on Identification and System Parameter Estimation, Budapest, Hungary.*, pages 281-285.
- Young, P. C. (1981). Parameter estimation for continuous-time models - a survey. *Automatica*, 17(1):23-39.
- Young, P. C. (1984). *Recursive estimation and time series analysis - an introduction*. Springer Verlag, Berlin.
- Young, P. C., Chotai, A., and Tych, W. (1991). Adaptive true digital control (TDC) of continuous-time systems using delta operator models. *Proceedings of IFAC Symposium on Identification and System Parameter Estimation, Budapest, Hungary.*, 1:273-280.

