A Direct Approach to Identification of Nonlinear Differential Models from Discrete Data

D. Coca

S.A. Billings

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

Research Report No.709
March 1998

A Direct Approach to Identification of Nonlinear Differential Models from Discrete Data

D. Coca and S.A. Billings
Department of Automatic Control and Systems Engineering,
University of Sheffield,
Sheffield S1 3JD, UK

Abstract

The paper introduces a direct approach to the identification of nonlinear differential equations from noisy input/output data. Both the parameter estimation and the structure determination problems are addressed. Central to the proposed methodology are two algorithms, a numerical differentiation algorithm involving fixed interval Kalman smoothing and an orthogonal regression routine used to perform model structure selection. The applicability of the identification procedure, which unlike most previous algorithms is not restricted to a special class of nonlinear systems, is demonstrated using simulated and experimental noise corrupted data.

1 Introduction

Although discrete-time models are best suited for numerical implementation, knowledge of the continuous-time differential representation of a dynamical system is essential for many applications in engineering and the physical sciences.

Compared with the problem of inferring the discrete-time model of a linear or nonlinear system from data, which is not a trivial problem, the identification of continuous-time models it is still regarded as a far more challenging task. In particular when the system of interest is nonlinear the existing approaches for identification and estimation of continuous-time systems (Unbehauen and Rao, 1987; Sinha and Rao, 1991; Young, 1981), many of which make use of special classes of modulating functions to eliminate the differential operators in the original equation, cannot be employed or their use is restricted to some special classes of nonlinear systems (Patra and Unbehauen, 1995).

Most of the existing methods also assume knowledge of the nonlinear differential representation of the dynamical system to within a set of unknown constant parameters. This limitation of the algorithms is very important since very often the only available information from a dynamical system is a set of discrete input/output observations (measurements). In these situations a possible solution is an indirect approach (Tsang and Billings, 1992) which involves the discrete-time identification of the system using polynomial NARMAX models, the computation of the generalised frequency response functions using the discrete model and finally the estimation of the nonlinear continuous-time representation from the frequency response data.

As an alternative solution, this paper introduces a direct identification approach in which a regression model is built using the measured signals and their higher order derivatives com-

puted directly from the noisy measurements.

The critical task of numerically differentiating the noisy measurements is performed using a fixed interval Kalman smoother. In order to implement the algorithm, a state-space model which exploits the regularity of the solutions of differential equations, is formulated.

When the model representation is known, the smoothed input and output signals and the corresponding higher-order derivatives can be used to derive a regression model and to estimate the unknown parameters. A similar direct approach to parameter estimation for continuous-time systems was introduced in Young (1993) using a Integrated Random Walk model as a basis for numerical differentiation. The state-space model used here to implement the Kalman Smoothing algorithm was initially proposed by Fioretti and Jetto (1989) and has been successfully used (Coca and Billings, 1998) for parameter estimation of Distributed Parameter Systems.

The main contribution of this work however is that it addresses an extremely important problem in nonlinear system identification, that of model structure determination. When the form of the nonlinear differential system is unknown a critical task is to determine the structure of the differential equation before estimating the unknown parameters. One possible solution to this important problem is to use non-parametric regression techniques. By considering a polynomial model structure, it is shown that the polynomial terms of the continuous-time model can be selected in a similar manner as for discrete-time NARMAX models (Chen and Billings, 1989) using an orthogonal forward regression algorithm (Chen et al., 1989). The main difference with respect to the traditional NARMAX identification algorithm is that in this case the form of the regression model precludes the implementation of noise models.

Using simulated and experimental noise contaminated data the applicability of this procedure to estimate the differential representation of a nonlinear system is demonstrated. Despite some structural differences between the simulated system and the estimated models, the number, and location of the equilibrium points as well as the eigenvalues associated with the equilibrium points of the original system are closely replicated by the models. This ensures that qualitatively the model and the original system will exhibit the same behaviour. Because of the need to use the fixed interval smoother the approach presented in this paper is mainly suited to off-line applications.

2 Differential Equations and Dynamical Systems

Most dynamical systems in the real world evolve continuously in time. The evolution of such systems is normally represented mathematically as the solution $\varphi(t, u, x_0)$ of a finite system of differential equations

$$\dot{x} = f(x, u) \tag{1}$$

where $x(t)$ is the vector of dimension n of state variables, $u(t)$ is a vector of dimension m of input variables, x_0 the a vector of initial conditions and $f : \mathbb{R}^{n+m} \rightarrow \mathbb{R}^n$ is a smooth, nonlinear mapping. Here smooth means C^l namely l times continuously differentiable.

If f and u are smooth functions the smoothness of the solution $\varphi(t)$ can also be analysed. Using a "bootstrap" technique it is easy to prove that if f and u are C^l and C^k functions

respectively, then the the solution $\varphi(t)$ is in $C^{\min(l,k)+1}$. This means that in general $\varphi(t)$ has one order of differentiability more than either f or u whichever has a lower regularity order. When f and u are analytic, the solution is a C^∞ function with respect to time. The regularity of the solution $\varphi(t)$ is an important aspect in the derivation and implementation of the numerical differentiation algorithm based on a fixed interval Kalman smoother.

The continuous-time system described in (1) is normally augmented by introducing an observation or measurement function

$$y = h(x, u) \tag{2}$$

where y is the output, a vector of dimension p and $h(\cdot)$ is a continuous function. Equations (1) and (2) represent the state-space description of a continuous-time system with outputs.

Subject to some additional assumptions the state-space description can be converted into a set of of nonlinear higher-order differential equations in the input and outputs

$$R(y, \dot{y}, ..., y^{(i)}, u, \dot{u}, ..., u^{(i-1)}) = 0 \tag{3}$$

where $u^{(i)}$, $y^{(i)}$ denote the i-th time derivative of the input and output variables and $R : \mathbb{R}^{2i+1} \rightarrow \mathbb{R}^p$. The higher-order differential equation (3) represents the external differential representation or the input/output equation of the dynamical system.

In practice the only information available from a system may be a set of input and output data which reflects how the input u affects the output y. In system identification it is of interest to model the input/output behaviour of the system as an input/output differential equation if for the given combination of inputs and outputs such a representation exists. In particular, in the present work a direct method of estimating both the structure and the unknown parameters in the following external differential representation

$$y^{(i)} = F(y, \dot{y}, ..., y^{(i-1)}, u, \dot{u}, ..., u^{(i-1)}), \tag{4}$$

which can often be derived from (3), is introduced. The approach involves estimation of the higher-order derivatives of the input/output signals in in equation (4) directly from the noisy measurements. An algorithm for generating the necessary derivatives is introduced in the following section.

3 The Identification Procedure

In this section the whole identification procedure is introduced beginning with the smoothing and numerical differentiation algorithm which is an essential part of the present approach.

3.1 Numerical Differentiation in the Presence of Noise

The problems of numerical differentiation of noisy signals are well known, especially that this operation can lead to the amplification of the noise if performed directly on the measured signals. Here an approach originally introduced in Fioretti and Jetto (1989), which is based

on a regularity assumption regarding the measured signals, is used. This is a relatively mild condition as far as the input and outputs of differential equations are concerned and hence is largely applicable for estimating the higher-order derivatives required in (4) directly from the (normally) noisy measurements.

Under the regularity assumption regarding the input/output signals consider the state vector $Y(t)$ composed of the output $y(t)$ and the derivatives $y^{(j)}$, $j = 1, ..., N$ (Fioretti and Jetto, 1989). Here y is assumed to be one-dimensional. When y is a vector, a similar model can be derived for each output variable.

By differentiating $Y(t)$ with respect to time, the following continuous state-space equations result

$$\dot{Y}(t) = DY(t) + G\frac{d^{N+1}y(t)}{dt^{N+1}} \tag{5}$$

where D is the $(N+1) \times (N+1)$ matrix

$$D = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 & \dots \\ 0 & 0 & 1 & \dots & 0 & \dots \\ \vdots & & & \ddots & \vdots & \\ 0 & 0 & 0 & \dots & 1 & \dots \\ \vdots & & & & \vdots & \ddots \end{bmatrix} \tag{6}$$

and $G = [0 \dots 0\, 1]^T$ is a $(N+1)$-dimensional vector.

Assuming Δt to be the sampling time, the discrete state space equations associated with (5) have the following form

$$Y(t + \Delta t) = AY(t) + W(t) \tag{7}$$

where A is the state transition matrix

$$A(\Delta t) = \begin{bmatrix} 1 & \Delta t & \Delta t^2/2! & \dots & \Delta t^N/N! & \dots \\ 0 & 1 & \Delta t & \dots & \Delta t^{(N-1)}/(N-1)! & \dots \\ \vdots & & & & \vdots & \\ 0 & 0 & 0 & \dots & 1 & \dots \\ \vdots & & & & \vdots & \ddots \end{bmatrix} \tag{8}$$

The state-space model (7) can be augmented with the following measurement equation

$$z(t) = HY(t) + e(t) = y(t) + e(t) \tag{9}$$

where $H = [1 \dots 0]$ is a $(N+1)$-dimensional vector and $e(t)$ represents measurement noise.

If $e(t)$ is a white noise sequence, equations (7) and (9) are in a form suitable for Kalman filter implementation. In particular, the fixed interval smoother described in Anderson and Moore (1979) provides an optimal solution to the state estimation problem. The state vector in this case consists of the measured output and the unobserved higher order derivatives. The remaining $(N+1)$th derivative of $y(t)$ in equation (5) will be treated as white noise so the term

$W(t)$ in (7) becomes a white noise sequence with the covariance matrix Q. It can be shown (Fioretti and Jetto, 1989) that the generic element of Q is given by

$$q_{ij} = \sigma_w^2 \frac{\Delta t^{2N+3-(i+j)}}{(N+1-i)!(N+1-j)!(2N+3-(i+j))} \tag{10}$$

where σ_w^2 was estimated as

$$\sigma_w^2 = \frac{q_N^2}{3\Delta t} \tag{11}$$

The quantity q_N, which formally represents the remainder of the Taylor series expansion of order N of the signal, can be approximated as

$$q_N = \frac{M}{\pi} \Delta t \frac{\omega_c^{N+2}}{4} \exp(\omega_c \Delta t) \tag{12}$$

In equation (12) M is the upper bound for the amplitude spectrum of the signal and $\omega_c = 2\pi f_c$ is such that f_c defines a cut-off frequency above which the power spectral density of the output signal is negligible. In practice by computing the FFT and the power spectral density P_ω of the measured signals both M and f_c can be determined directly from the data.

The fixed interval smoother was implemented here as a combination of two Kalman filters one running forward in time and one moving backwards. The state transition matrices in this case are A for the filter moving forward and A^{-1} for the backward moving filter.

If $Z = \{z(t_1), z(t_2), ..., z(t_n)\}$ is the set of available observations consider

$$Z^-(t_i) = \{z(t_1), z(t_2), ..., z(t_i)\} \tag{13}$$

$$Z^+(t_i) = \{z(t_i), z(t_{i+1}), ..., z(t_n)\} \tag{14}$$

as the sets of "past and present" and "present and future" observations at a given instant $t_i = i\Delta t$.

The smoothed state estimate at each time instant representing the minimum variance estimate given all the data, past and future, is given by the well known formulas

$$\hat{z}(t_i) = P(P_1^{-1}\hat{z}_1(t_i) + P_2^{-1}\hat{z}_2(t_i)) \tag{15}$$

$$P = \mathrm{cov}(\hat{z} - z) = (P_1^{-1} + P_2^{-1})^{-1} \tag{16}$$

where

$$\hat{z}_1(t) = \mathcal{E}\{z(t_i)|Z^-(t_i)\} \tag{17}$$

denotes the filter running forward in time and

$$\hat{z}_2(t) = \mathcal{E}\{z(t_i)|Z^+(t_i)\} \tag{18}$$

denotes the estimate produced by the filter running backwards in time and P_1 and P_2 are the covariance matrices corresponding to the forward and backward filters.

If the data sequence is short the two filters can be run forward and backwards more than once in order to achieve convergence. In this case at the end of each run the covariance matrices

associated with each filter are interchanged and so are the final state estimates. It follows that the final estimates of the forward filter become initial conditions for the filter running backwards and vice-versa.

The key assumption when smoothing is performed is that the forward-time and backward-time system descriptions are equivalent so that the filters can cooperate by combining in an optimal way the estimates obtained from two independent sets of measurements $Z^-(t_i)$ and $Z^+(t_i+1)$ (Lewis, 1986).

If however the signal or the associated derivatives have isolated discontinuities, eventually one of the models will fail to correctly describe the signal at that point. In this case it is possible to allow only one of the filters to produce the estimate, instead of combining them, by using an additional decision rule (Niedźwiecki and Sethares, 1995). This could provide a solution when the signals have isolated singularities.

A factor which clearly determines the accuracy of the estimates is the order of the signal model used. In theory the estimation error tends to zero monotonically with N, the largest order of the derivatives considered if the data is sampled sufficiently fast relative to the frequency bandwidth of the signal. In practice, $N = 6, ..., 9$ can be chosen with good results. A value for N which is too large will generally slow down the computation and also increase the risk of numerical instability.

The fixed-interval smoother can be applied to obtain the optimal state estimate consisting of the inputs $u(t)$ and outputs $y(t)$ and the corresponding high-order derivatives of the inputs and outputs which appear in equation (4).

3.2 The Parameter Estimation Problem

Assume that the input/output equation

$$y^{(i)} = F(y, \dot{y}, ..., y^{(i-1)}, u, \dot{u}, ..., u^{(i-1)}, \theta) \tag{19}$$

is known with the exception of a set of constant parameters θ.

If the high-order derivatives of the input and output variables involved in (19) are computed from the sampled input/output data using the differentiation algorithm presented in the previous section, a regression model can be formulated in order to estimate the unknown parameters. For linear-in-the-parameters model equations the parameter vector can then be determined using least-squares. When expression (19) is nonlinear with respect to the parameter vector θ, alternative gradient-descent type algorithms can be used.

3.2.1 Example 1: Parameter Estimation for a Nonlinear Differential System

Consider the following nonlinear system

$$\ddot{y} = a_1\dot{y} + a_2y^2\dot{y} + a_3\dot{y}^2 + a_4y + b_1u \tag{20}$$

Based on the equivalent state-space model

$$\begin{aligned} \dot{x_1} &= x_2 \\ \dot{x_2} &= a_1x_2 + a_2x_1^2x_2 + a_3x_2^2 + a_4x_1 + b_1u \end{aligned} \tag{21}$$

the system with $a_1 = -3$, $a_2 = -1.5$, $a_3 = 2$ $a_4 = -2$ $b_1 = 3$ was numerically integrated using a fourth-order Runge-Kutta integration routine with the fixed integration step $\Delta t = 0.01$. The input u used in the simulation, which is illustrated in Fig 1a was a lowpass filtered white noise sequence. The cut-off frequency of the filter corresponded to 0.0025 of the original sampling frequency. After removing the transients the simulation data consisted of 10000 input/output data points sampled at $\Delta t = 0.01$. For parameter estimation only 1000 input/output data points sampled at $5\Delta t$ were actually used. In order to test the estimation scheme in a realistic way, the output signal was corrupted with white noise with standard deviation $\sigma = 0.0065$. The noisy output signal is illustrated in Fig.1b.

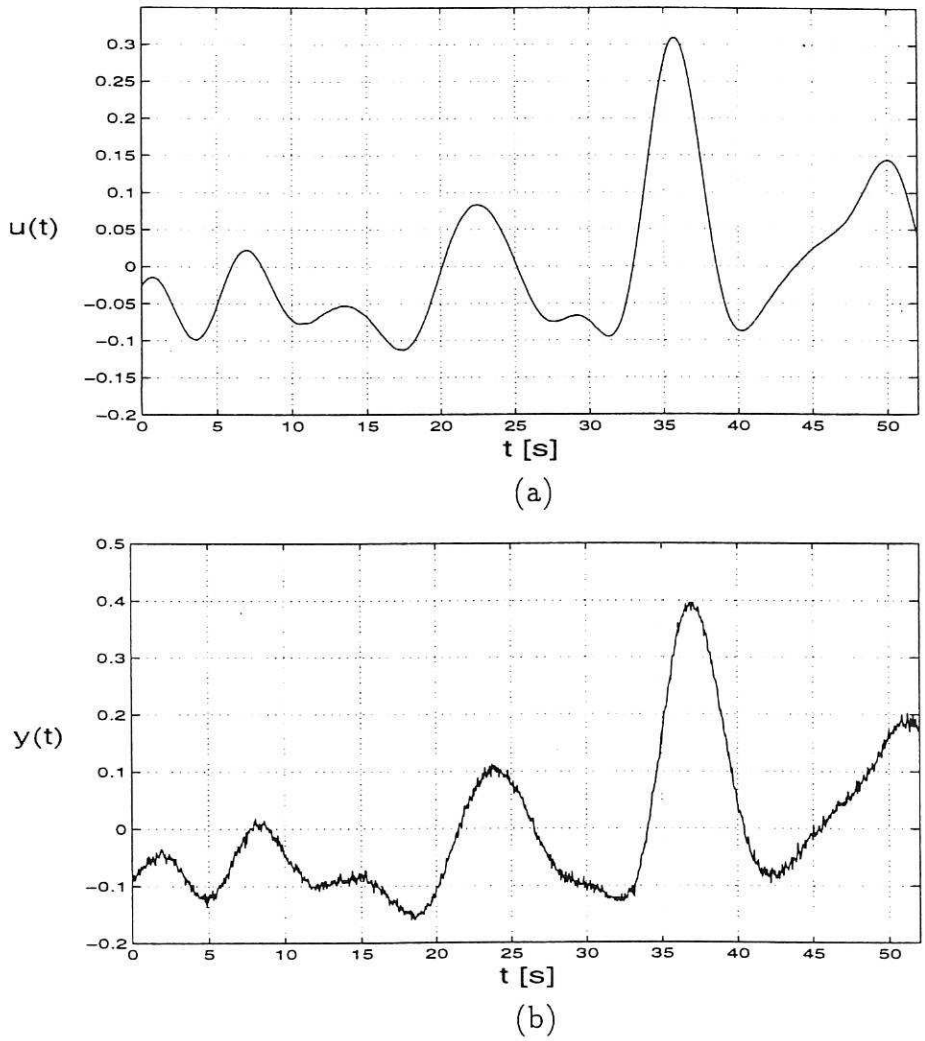


Figure 1: Identification Data: (a) Input Signal (b) Noisy Output Signal

The fixed interval Kalman smoother derived in Section (3.1) was used to smooth the output data and generate the higher order derivatives required. The parameters M and f_c required to compute the value of σ_w in equations (11), (12) were estimated by computing the power spectral density P_ω and the FFT of the output signal. In particular, f_c was selected such

that $P_\omega < \rho$ for $f > f_c$. Numerical experiments have shown that by choosing the threshold ρ around $(0.2\% - 0.5\%)P_\omega^{max}$ good estimates of the cut-off frequency f_c and indirectly of the parameter σ_w^2 can be obtained.

Using least-squares estimation the following parameters were obtained: $a_1 = -2.9015$, $a_2 - 1.3500$, $a_3 = 1.8771$, $a_4 = -1.9430$, $b_1 = 2.9003$.

The results obtained illustrate that the proposed approach can provide good estimates of the parameter vector of a continuous-time system in the presence of noise. The main advantage of this method is that it is not restricted to special classes of nonlinear systems (i.e. integrable or convolvable) which makes it suitable in situations than cannot be handled by other existing approaches (Patra and Unbehauen, 1995).

3.3 Model Structure Selection for Continuous-Time Models

In many practical applications no information regarding the structure of the observed dynamical system is available a priori. Identification therefore involves model structure selection prior to parameter estimation.

Similar to discrete-time NARMAX modelling the structure determination problem can be solved using non-parametric regression techniques. In practice this involves selecting a functional form of the multivariate nonlinear function $F(\cdot)$ and the specification of the order parameters representing the maximum derivative orders for the inputs and outputs which appear in (4). Because of the particular form of equation (4), if the order of the system is not known, a solution to determine the parameter i is to test a sequence of models of increasing order until a valid model which can reproduce the observed behaviour is found. Choosing the parameter which specifies the maximum derivative order for the input is not so critical since if this parameter is overestimated the structure selection algorithm should be able to select the significant terms involving the higher-order derivatives of u.

The order parameters indicate the number of variables of the function $F(\cdot)$ and are essential for defining a finite model set consisting of all linear and nonlinear terms that in theory can be used to implement the model.

Denote $X = \{y, \dot{y}, ..., y^{(i-1)}, u, \dot{u}, ..., u^{(i-1)}\}$ the vector of regression variables consisting of the inputs, outputs and higher-order derivatives of the inputs and outputs. If $F(\cdot)$ is initially assumed to be an r-th order multivariate polynomial $\Pi_r(X)$ the model set $\mathcal{M} = \{p_k(X)\}$ will consist of all possible terms of this polynomial. Of course a model which would include all these polynomial terms is unpractical because of the large number of parameters involved and also because in most cases a more complex model does not necessarily equate to a better model.

A practical solution to finding the significant terms to be included in the polynomial model structures and indeed for any linear-in-the-parameters functional $F(\cdot)$ is the orthogonal forward regression algorithm (Chen et al., 1989). The principle is to consider the inclusion or exclusion of each candidate term on the basis of its contribution to the reduction of the modelling error. This systematic approach uses the ERR (error reduction ratio) test (Billings and Chen, 1989) to evaluate the significance of each term and to determine its contribution to the approximation. The selection procedure is carried out until the approximation error becomes smaller than a given target value and/or the estimated model can reproduce sufficiently

well the observed dynamical behaviour.

The main difference between the discrete-time and continuous-time identification procedures is the absence of any noise models in the later case. Hence the importance of estimating accurately the derivatives of the input and output signals in order to ensure unbiased parameter estimates.

3.3.1 Example 2: Structure Selection and Parameter Estimation for a Nonlinear Differential System

Consider again the system (20) investigated in Section 3.2, Example 1. The same data set was used in this example in an attempt to both determine the model structure and estimate the parameters associated with the selected model.

The model set was implemented by selecting $F(\cdot)$ to be a cubic polynomial. A maximum derivative order $n_u = 1$ and $n_y = 1$ was chosen initially for u and y. According to these parameters a model set consisting of 20 terms was specified. Using the orthogonal forward estimator routine the following first order model (Model 1) consisting of 10 terms was identified from the data.

$$\begin{aligned} \dot{y} \;=\; & 1.08205\,u - 0.621515\,y + 1.91684\,yu - 12.1683\,y^2u \\ & -0.770119\,u^2 + 0.73712\times 10^{-2} - 0.975369\,y^2 - 4.16171\,u^3 \\ & +4.32898\,y^3 + 11.327\,yu^2 \end{aligned} \tag{22}$$

The model was integrated using a fourth-order Runge-Kutta routine with the same input which was used to generate the identification data, and with the input corresponding to the test data set. In Fig. (2) the results of the simulation using the first input signal are displayed. Note that although not perfect, there is a good match between the data simulated using the original and the estimated models. However, when the same model was simulated with a different, test input the solution diverged to infinity.

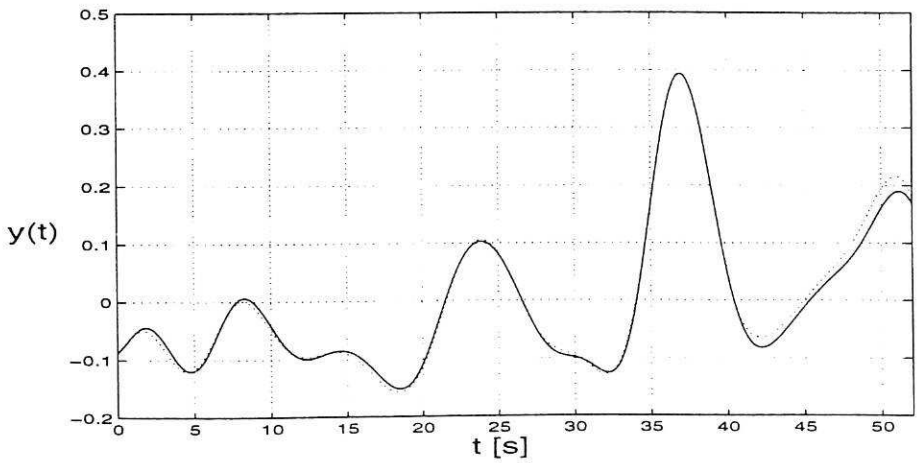


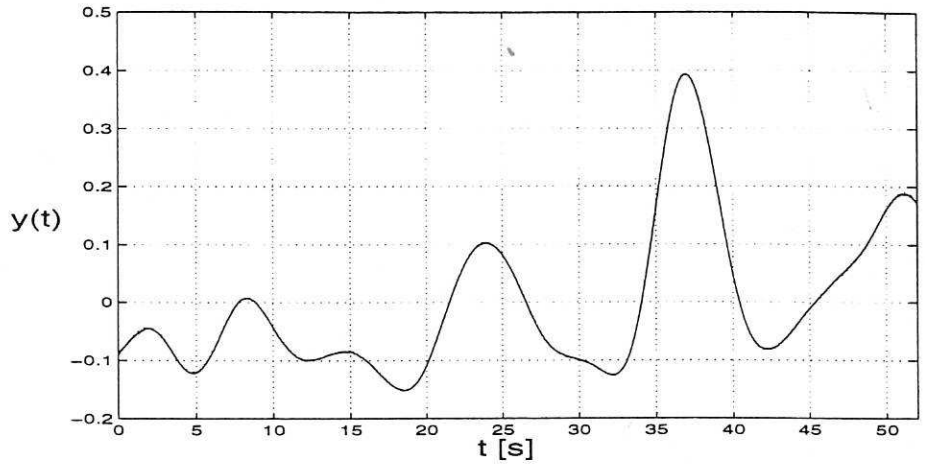
Figure 2: System Output (cont) and Model Predicted Output (dotted) for Model 1, eqn. (22)

By increasing the system order to $n_y = 2$ a different model set consisting of 35 terms was defined. The model selected in this case (Model 2) by the orthogonal forward regression

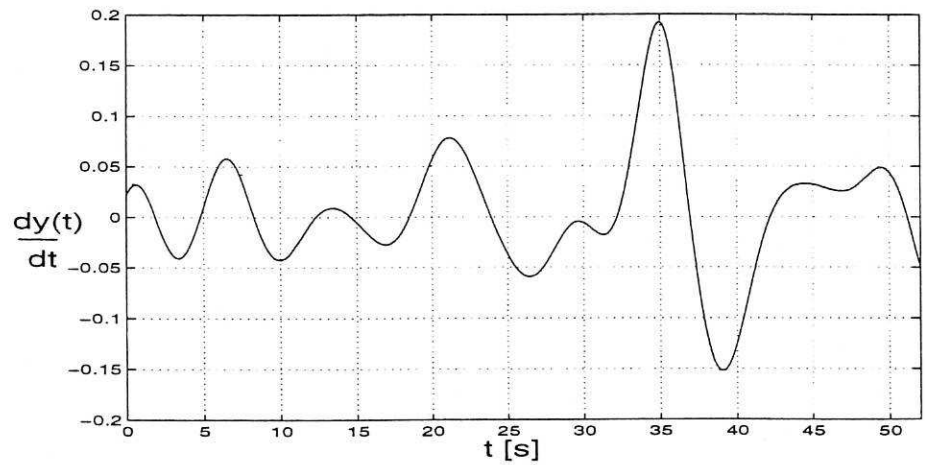
procedure consisted of 8 terms

$$\begin{aligned}\ddot{y} &= -1.9368\,y + 1.9533\,\dot{y}^2 + 0.10913\,yu - 0.41612\,y^2u \\ &\quad -0.92173\times 10^{-5} - 1.4661\,y\dot{y}u - 2.9267\,\dot{y} + 2.9144\,u\end{aligned} \tag{23}$$

The model simulated output is an excellent match over the identification (Figs. 3a,b) and test data sets (Fig. 4ab).



(a)



(b)

Figure 3: Simulation results using the identification input signal: (a) y and (b) $\dot{y}$ for Model 2, eqn. (23) (dotted) and the original system (cont)

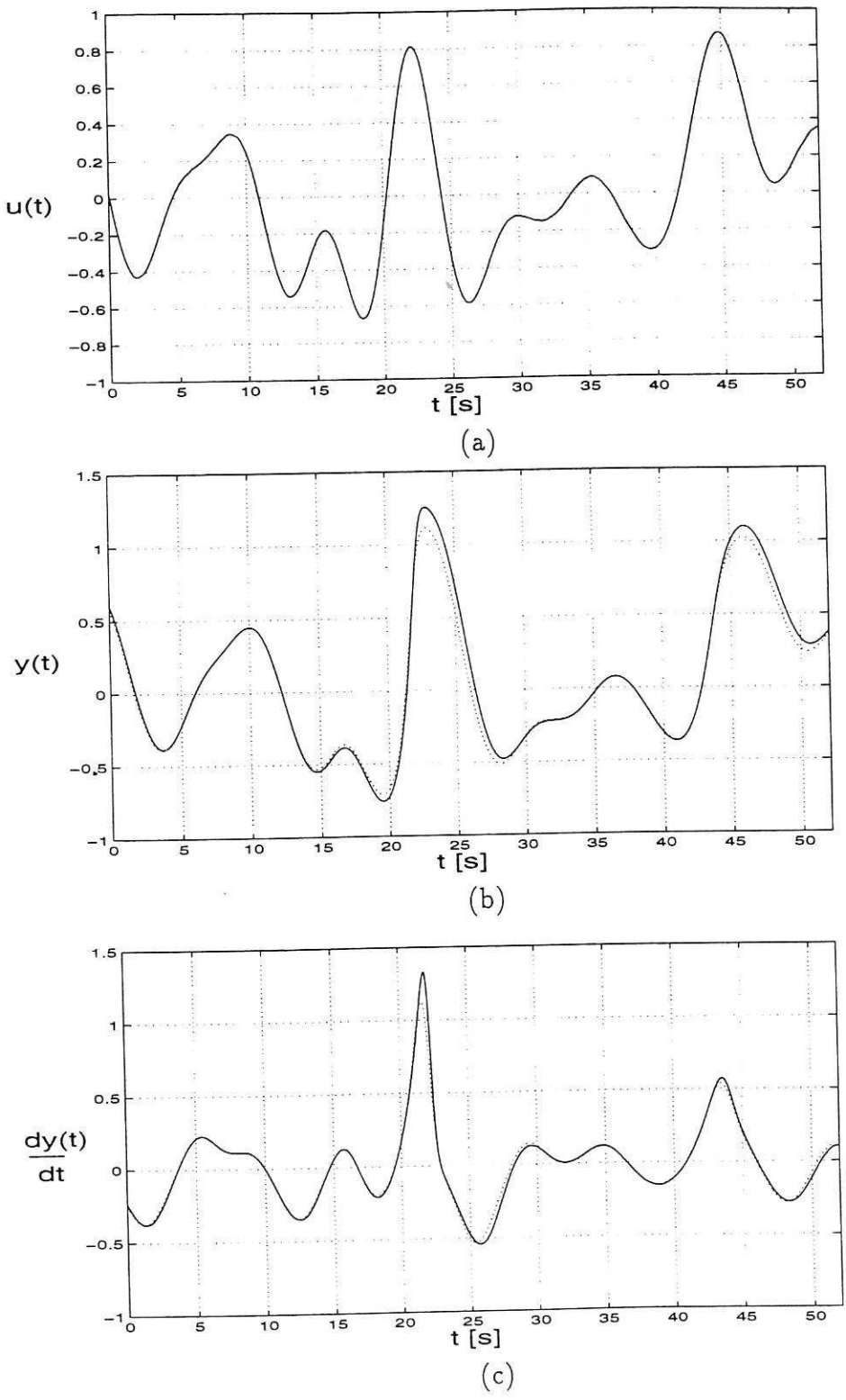


Figure 4: Simulation results using (a) the test input signal: (b) y and (c) $\dot{y}$ for Model 2 (dotted) and the original system (continuous)

However, the model structure picked up by the structure selection algorithm differs from that of the original system. An important question in this case is how to assess the validity of this model apart from just comparing the model predicted output.

3.4 Model Validation

Ideally in nonlinear system identification the model obtained as a result of the identification procedure should be globally valid that is the model and the system should behave similarly when excited by identical input signals. However, the use of non-parametric regression techniques to determine the form of the nonlinear function $F(\cdot)$ means that in many cases this function is only a local approximation over the interval (hypercube) containing the regression data. This is evident especially when the class of functions used to derive $F(\cdot)$ have local support (the functions tend to zero rapidly outside a closed interval known as the support of the function) such as Radial Basis Functions or wavelet functions for example. In these cases the nonlinear mapping is also locally supported that is $F(\cdot)$ vanishes rapidly outside the original interval of approximation.

A class of model structures which can provide global models are polynomials. Using this class of functions it is possible to obtain a global model for a polynomial system if the structure selection and parameter estimation routine is able to include in the model the same polynomial terms which describe the original system and to provide reasonably accurate parameter estimates. However when the underlying mathematical relationship which governs the original system is not polynomial, in general, a polynomial model once again provides only a local approximation despite the fact that the approximating polynomial $F(\cdot)$ is not a locally supported function. A simple intuitive argument is that although a sine function for example can be locally approximated by a polynomial it is obvious that such an approximation is no longer valid outside the original domain of approximation.

In Example 2, although the original system is polynomial, the model structure selected using the orthogonal forward regression routine, differs from that of the original system. Despite this the model can still predict quite well on a different set of data. An obvious problem is how to compare the two dynamical systems in order to decide model adequacy.

In the qualitative theory of dynamical systems the number, location and stability of the equilibrium points are essential for classifying and comparing dynamical systems. In system identification this is a possible approach to assess the validity of the estimated models.

In this case from the equivalent state-space form of the Model 2 (23)

$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= -1.9368\,x_1 + 1.9533\,x_2^2 + 0.10913\,x_1 u - 0.41612\,x_1^2 u \\ &\quad -0.92173\times 10^{-5} - 1.4661\,x_1 x_2 u - 2.9267\,x_2 + 2.9144\,u \end{aligned} \tag{24}$$

it is clear that the estimated model has a single equilibrium point $(x_1, x_2) = (-4.7591e-06, 0)$ which is very close to that of the original system $(x_1, x_2) = (0, 0)$. The eigenvalues associated with the the linearisation of equation (24) around the origin are $(\lambda_1, \lambda_2) = (-1.9158 - 1.0109)$ are also close to those corresponding to the original system $(\lambda_1, \lambda_2) = (-2 - 1)$. The invariance of the number and location of the equilibrium points coupled with the fact that the eigenvalues of

the linearisation around the equilibrium point are nearly identical, means that qualitatively the behaviour of the two systems will be essentially the same. The numerical simulations computed for Example 2 have also demonstrated that, from a quantitative point of view, the model was able to predict well with respect to both the estimation and the test data. In particular it is important to note that the amplitude of the input signal used to generate the test data was significantly larger that the amplitude of the input used to generate the estimation data. As a result the spatial domain spanned by the test data was larger than the initial domain of approximation. In conclusion, although not perfect, the identified model is a very good approximation of the original system.

3.4.1 A Model Sequencing Strategy for Improved Structure Selection

Ideally of course, the structure selection algorithm should have selected a correct model structure in Example 2. The errors associated with the numerical differentiation process, although small, seem to be the main reason for the structural inaccuracy in this case.

A possible way to increase the robustness of the selection algorithm for polynomial model structures could be the use of a slightly different model sequencing strategy. The idea is to exclude from the initial model set all the terms representing cross products of the regression variables. A subsequent model set can be appended with cross-product terms corresponding to the input variables, cross-product terms corresponding to the output variables and finally with cross-product terms including both input and output variables.

This sequence of nested model structures $\mathcal{M}_1 \subset \mathcal{M}_2 \subset \mathcal{M}_3 \subset \ldots$ naturally leads to a sequence of models $\mathbf{m_1}, \mathbf{m_2}, \mathbf{m_3}, \ldots$ of increasing complexity which can be selected and tested for each model. This approach while reducing the risk of estimating a more complex model than necessary can also lead to more accurate model structures.

As a justification of this approach note that the cross-product terms are correlated with each regression variable involved in the cross-product and in turn with the left-hand-side of equation (4). Hence when a full model set is unnecessarily considered, there are chances that the selection procedure which is based on the goodness of fit criterion (ERR) can wrongly select such cross-product terms instead of true terms containing those regression variables alone.

This approach was successfully tested on the simulated system considered in Example 2. The following model was selected from a model set which excluded any terms that were a combination of variables from $\{u, \dot{u}\}$ and from $\{y, \dot{y}\}$:

$$\begin{aligned} \ddot{y} \;=\; & -1.96982\,y + 1.85868\,\dot{y}^2 + 0.22054\times 10^{-3} \\ & -0.208908\,y\dot{y}^2 - 2.98658\,\dot{y} + 2.94791\,u \end{aligned} \tag{25}$$

Note that in this case the selected structure is more accurate than the initial selection.

3.5 Practical Identification Of an Electronic Circuit from Noisy Measurements

This section concerns the identification of the differential representation of a dynamical system from experimental measurements. The system considered in this case is a nonlinear electronic

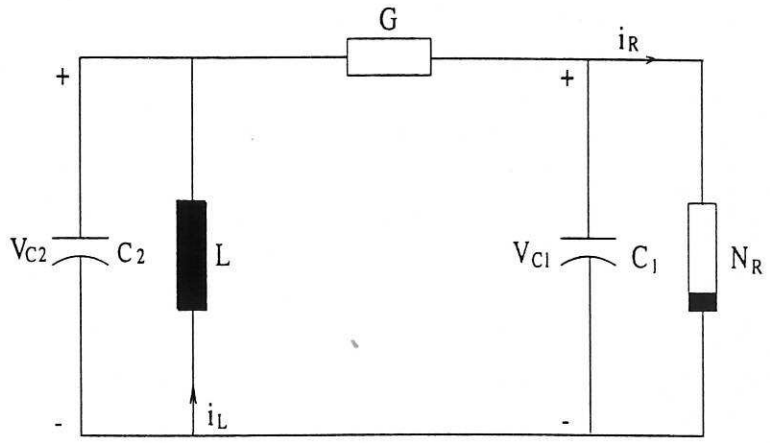


Figure 5: Chua's circuit

circuit widely known as the Double Scroll or Chua's circuit (Chua, 1992; Matsumoto et al., 1993). The circuit which is a source of rich and complex dynamical behaviour including chaos is also renowned for the number of practical applications that exploit its properties.

The circuit Fig (5) which consists of one inductor (L), two capacitors (C_1, C_2) one linear resistor (G) and one nonlinear resistor (N_R) with a piecewise linear characteristic $g(v_{C1})$ can be described mathematically by a system of three differential equations

$$\begin{aligned} C_1\frac{dv_{C_1}}{dt} &= G(v_{C_2} - v_{C_1}) - g(v_{C_1}) \\ C_2\frac{dv_{C_2}}{dt} &= G(v_{C_1} - v_{C_2}) + i_L \\ L\frac{di_L}{dt} &= -v_{C_2} \end{aligned} \tag{26}$$

where

$$g(v_{C_1}) = \begin{cases} m_1 v_{C_1}, & |v_{C_1}| < B_p \\ m_0 v_{C_1} + B_p(m_1 - m_0), & v_{C_1} \geq B_p \\ m_0 v_{C_1} - B_p(m_1 - m_0), & v_{C_1} \leq -B_p \end{cases} \tag{27}$$

The circuit can be implemented in many ways using standard components. The implementation described in Kennedy (1993) will be considered here. The data set used for identification was measured by Dedieu and Ogorzalek (1994) who investigated a control method for this circuit.

The data consisted of three time series of 10000 data points each, representing the current i_L through the inductor L, the voltage V_{C1} across C_1 and the voltage v_{C2} across C_2. The measurements were simultaneously taken using a sampling frequency of 500KHz.

For identification the original data was decimated by 5 resulting in three time series of 2000 data points each. The numerical differentiation algorithm introduced in Section (3.1) was used on the raw data to compute the higher order derivatives of the experimental signals. The values of M and f_c which are essential to determine the value of σ_w in (10) were determined by computing the power spectral density and the FFT of the signals to give $M^{i_L} = 30$, $M^{v_{C1}} = 28$,

$M^{v_{C2}} = 10$, $f_c^{i_L} = 2.2$KHz, $f_c^{v_{C1}} = 3$KHz, $f_c^{v_{C2}} = 2.2$KHz. The resulting values for σ_w were $1.7620 \cdot 10^{43}$, $5.2428 \cdot 10^{44}$, $5.8734 \cdot 10^{42}$ for i_L, v_{C1} and v_{C2} respectively.

Using the smoothed signals and the first order derivatives the following type of model was sought

$$\begin{aligned} \dot{v}_{C1} &= f_1(v_{C1}, v_{C2}, i_L) \\ \dot{v}_{C2} &= f_2(v_{C1}, v_{C2}, i_L) \\ \dot{i}_L &= f_3(v_{C1}, v_{C2}, i_L) \end{aligned} \tag{28}$$

To determine f_1, f_2, f_3 in equation (28) a cubic polynomial model structure was defined. The selection of the polynomial terms that made up the final model and the parameter estimation was carried out using the orthogonal forward regression routine.

The final model

$$\begin{aligned} \dot{v}_{C1} &= +0.42234283\times10^{+4}\, i_L^3 + 0.97625632\times10^{+5}\, i_L v_{C2} - 0.22819645\times10^{+7}\, v_{C2}^2 \\ &\quad -0.9162641\times10^{+3}\, i_L^2 + 0.29095643\times10^{+4}\, v_{C1} i_L - 0.48678716\times10^{+6}\, i_L^2 v_{C2} \\ &\quad +0.29273897\times10^{+4} + 0.28158381\times10^{+6}\, v_{C2} - 0.20227579\times10^{+5}\, i_L \\ &\quad -0.42994379\times10^{+4}\, v_{C1} - 0.24869384\times10^{+9}\, v_{C2}^3 + 0.28640418\times10^{+6}\, v_{C2} v_{C1}^2 \\ &\quad +0.1889272\times10^{+8}\, i_L v_{C2}^2 + 0.15655216\times10^{+4}\, v_{C1}^2 \\ \\ \dot{v}_{C2} &= +0.20091875\times10^{+4}\, v_{C1}^3 + 0.12570149\times10^{+4}\, i_L v_{C1} + 0.14734927\times10^{+3} \\ &\quad -0.11643835\times10^{+5}\, v_{C2} v_{C1} + 0.94986386\times10^{+6}\, i_L v_{C2}^2 + 0.11739572\times10^{+4}\, v_{C1} \\ &\quad +0.11968095\times10^{+5}\, v_{C2} - 0.70067849\times10^{+5}\, v_{C2}^2 - 0.34656118\times10^{+8}\, v_{C2}^3 \\ &\quad -0.18521636\times10^{+7}\, v_{C2}^2 v_{C1} - 0.40810209\times10^{+3}\, i_L + 0.48761216\times10^{+4}\, i_L v_{C1}^2 \\ &\quad +0.10189376\times10^{+6}\, i_L v_{C2} v_{C1} + 0.10745656\times10^{+6}\, v_{C2} v_{C1}^2 \\ \\ \dot{i}_L &= -0.30361124\times10^{+5}\, v_{C1}^3 + 0.23893292\times10^{+4} + 0.25051101\times10^{+5}\, v_{C1} \\ &\quad -0.19136995\times10^{+4}\, i_L + 0.22300821\times10^{+9}\, v_{C2}^3 - 0.38465654\times10^{+5}\, v_{C2} \\ &\quad -0.612364\times10^{+7}\, i_L v_{C2}^2 - 0.77528023\times10^{+7}\, v_{C2}^2 v_{C1} - 0.63514572\times10^{+4}\, v_{C1}^2 \end{aligned} \tag{29}$$

was simulated using a fourth-order Runge-Kutta integration routine. The model predicted output was used to plot the strange attractor in Fig (6b) which is an excellent match for the original attractor plotted in Fig (6a).

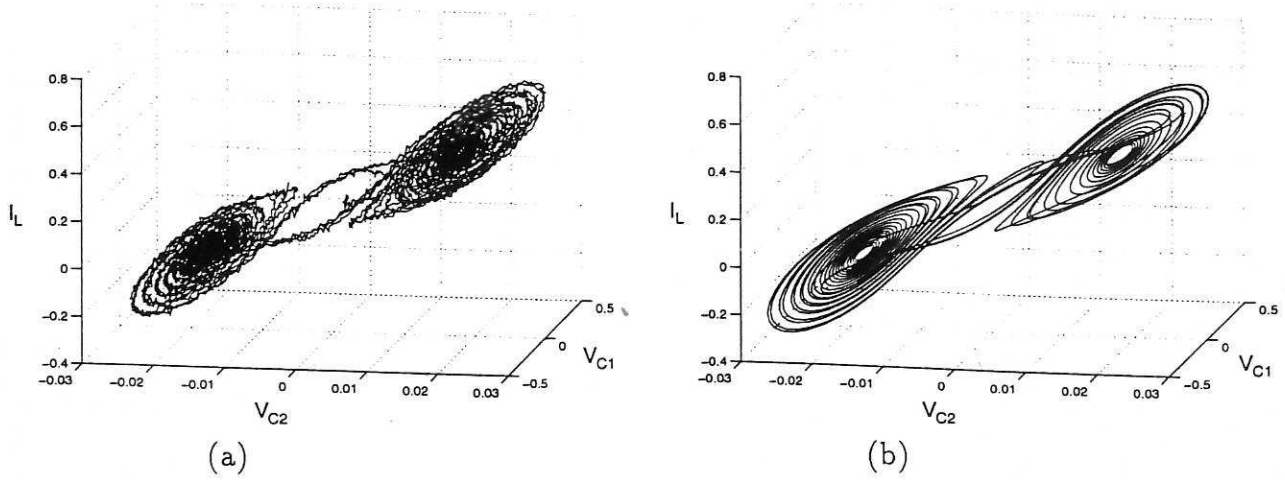


Figure 6: Double-Scroll Attractor reconstructed using (a) the experimental measurements and (b) the model predicted output

4 Conclusions

The problem of identifying the differential representation of general nonlinear systems has been studied. The new algorithm has the advantage that it does not assume a special form for the dynamical system. A state-space model which exploits a regularity assumption regarding measured input and output data forms the basis of a fixed interval Kalman smoother which is used to generate the required time derivatives.

When the form of the nonlinear differential equation is known the derivatives of the inputs and outputs can be used to build a regression model and the unknown parameters can be estimated by means of a least-squares algorithm or other optimisation technique.

The paper has also addressed two fundamental problems in nonlinear system identification that of structure selection and of model validity. Numerical and experimental results have illustrated that using an orthogonal forward estimator the structure of the nonlinear differential equations describing the dynamics of the observed system can be determined with sufficient accuracy.

In the simulated example, despite the fact that the model structure did not exactly match that of the original system, the model could reproduce the behaviour of the original system. This emphasises the fact that the structure alone is not the best way of validating identified systems. Dynamical invariants such as the number, location and stability of the equilibrium points can provide in this case a far better criteria for judging the model adequacy.

The identification procedure was also tested using experimental data collected from a nonlinear electronic circuit (Chua's circuit) which exhibits chaotic dynamics. The model derived from the noisy measurements using the approach advocated in this paper was able to reproduce very well the complicated dynamics displayed by this circuit.

5 Acknowledgements

The authors express their gratitude to H. Dedieu who provided the experimental data set used in this study. SAB gratefully acknowledges the support of EPSRC for this work.

References

Anderson, B. D. O. and Moore, J. B. (1979). *Optimal filtering.* Prentice-Hall, Englewood Cliffs, N. J.

Billings, S. A. and Chen, S. (1989). Extended model set, global data and threshold model identification of severely nonlinear systems. *Int. J. Control*, 50(5):1897–1923.

Chen, S. and Billings, S. A. (1989). Representations of non-linear systems: the NARMAX model. *Int. J. Control*, 49(3):1013–1032.

Chen, S., Billings, S. A., and Luo, W. (1989). Orthogonal least-squares methods and their applications to nonlinear system identification. *Int. J. Control*, 50:1873–1896.

Chua, L. O. (1992). The genesis of Chua's circuit. *Archiv für Elektronik und Übertragungstechnik*, 46(4):250–257.

Coca, D. and Billings, S. A. (1997). Continuous-time system identification for linear and nonlinear systems using wavelet decompositions. *Int. J. of Bifurcation and Chaos*, 7(1):87–96.

Coca, D. and Billings, S. A. (1998). Direct parameter identification of distributed parameter systems. *Submitted for publication.*

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.

Fioretti, S. and Jetto, L. (1989). Accurate derivative estimation from noisy data: A state-space approach. *Int. J. Systems Sci.*, 20(1):33–53.

Kennedy, M. P. (1993). 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.

Lewis, F. L. (1986). *Optimal estimation.* Wiley, New York.

Matsumoto, T., Komuro, M., Kokubu, H., and Tokunaga, R. (1993). *Bifurcations. Sights sounds and mathematics.* Springer-Verlag Tokyo.

Meditch, J. S. (1973). A survey of data smoothing for linear and nonlinear dynamic systems. *Automatica*, 9:151–162.

Niedźwiecki, M. and Sethares, A. (1995). Smoothing of discontinuous signals: The competitive approach. *IEEE Trans. on Signal Processing*, 43(1):1–13.

Patra, A. and Unbehauen, H. (1995). Identification of a class of nonlinear continuous-time systems using Hartley modulating functions. *Int. J. of Control*, 62(6):1431–1451.

Sinha, N. K. and Rao, G. P. (1991). *Identification of continuous-time systems.* Kulwer Academic, Dordrecht.

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.

Unbehauen, H. and Rao, G. P. (1987). *Identification of continuous systems*, volume 10 of *North-Holland Systems and Control Series.* North-Holland, Amsterdam.

Young, P. C. (1981). Parameter estimation for continuous-time models - A survey. *Automatica*, 17(1):23–39.

Young, P. C., Foster, M., and Lees, M. (1993). A direct approach to the identification and estimation of continuous-time systems from discrete-time data based on fixed interval smoothing. In *Proceedings of the 12-th Triennial World Congress of the International Federation of Automatic Control*, pages 647–650. Pergamon Oxford, UK.