Monograph:
Aguirre, L.A. and Billings, S.A. (1994) Retrieving Dynamical Invariants from Chaotic Data Using NARMAX Models. Research Report. ACSE Research Report 506 . Department of Automatic Control and Systems Engineering

Retrieving Dynamical Invariants from Chaotic Data Using NARMAX Models

L A Aguirre, and S A Billings

Department of Automatic Control and Systems Engineering
University of Sheffield
P.O. Box 600
Mappin Street
Sheffield S1 4DU
United Kingdom

Research Report No 506

February 1994

Retrieving Dynamical Invariants from Chaotic Data Using NARMAX Models

LUIS A. AGUIRRE[†] and S. A. BILLINGS

Department of Automatic Control and Systems Engineering
University of Sheffield
P.O. Box 600, Mappin Street — Sheffield S1 4DU - UK

Abstract

This paper is concerned with the estimation of dynamical invariants from relatively short and possibly noisy sets of chaotic data. In order to overcome the difficulties associated with the size and quality of the data records, a two-step procedure is investigated. Firstly NARMAX models are fitted to the data. Secondly, such models are used to generate longer and cleaner time sequences from which dynamical invariants such as Lyapunov exponents, correlation dimension, the geometry of the attractors, Poincaré maps and bifurcation diagrams can be estimated with relative ease. An additional advantage of this procedure is that because the models are global and have a simple structure, such models are amenable for analysis. It is shown that the location and stability of the fixed points of the original systems can be analytically recovered from the identified models. A number of examples are included which use the logistic and Hénon maps, Duffing and modified van der Pol oscillators, the Mackey-Glass delay system, Chua's circuit, the Lorenz and Rössler attractors. The identified models of these systems are provided including discrete multivariable models for Chua's double scroll, Lorenz and Rössler attractors which are used to reconstruct the trajectories in a three-dimensional state space.

[†]e-mail:aguirre@acse.sheffield.ac.uk

Taken's theorem gives sufficient conditions for equation (2) to hold, that is, in order to be able to infer dynamical invariants of the original system from the time series of a single variable, however no indication is given as to how to estimate the map f^T. A number of papers have been devoted to this goal and such methods can be separated into two major groups, namely *local* and *global* approximation techniques.

The local approaches usually begin by partitioning the embedding space into neighbourhoods $\{\mathcal{U}_i\}_{i=1}^{N_n}$ within which the dynamics can be appropriately described by a linear map $g^T : \mathbb{R}_e^d \rightarrow \mathbb{R}$ such that

$$y(k+T) \approx g_i^T(\mathrm{y}(k)) \quad \text{for } \mathrm{y}(k) \in \mathcal{U}_i,\ i = 1, \ldots, N_n \ . \tag{3}$$

Several choices for g^T have been suggested in the literature such as linear polynomials (Farmer and Sidorowich, 1987; Casdagli, 1991) which can be interpolated to obtain an approximation of the map f^T (Abarbanel et al., 1990). Simpler choices include *zeroth-order approximations*, also known as *local constant predictors* (Farmer and Sidorowich, 1987; Kennel and Isabelle, 1992) and a *weighted predictor* (Linsay, 1991).

A common difficulty of such approaches is that the data have to be separated into neighbourhoods. Thus given a point in the embedded space the closest neighbours to such a point must be found. It is well known that for many methods most of the CPU time is spent in searching for close neighbours in the embedding space within the data (Grassberger et al., 1991) and that the effort required to accomplish this grows exponentially with the embedding dimension.

One way of avoiding the need for constructing neighbourhoods is to fit global models to the data. In other words, all the data pertain to a single neighbouhood namely the entire embedding space. This considerable advantage has prompted some authors to investigate the estimation and use of global models (Cremers and Hüber, 1987; Crutchfield and McNamara, 1987; Kadtke et al., 1993; Aguirre and Billings, 1994a).

The use of global models, however, has two major difficulties namely i) the choice of a representation for the model which should be sufficiently complex to approximate the dynamics of f^T, and ii) the selection of the correct structure or basis within the chosen representation. In the case of global models the representation has to be nonlinear. Consequently the number and sizes of all possible model bases within a given representation become very large as

the parameters of polynomial models are currently widely available. One of the disadvantages of global polynomials, however, is that even for polynomial models of moderate order, the number of terms can become impractically large (Farmer and Sidorowich, 1988a; Casdagli, 1989). Using polynomials to forecast chaotic time series, Casdagli (1989) has reported that such predictors blow up in the iterative procedure and suggests that this is because polynomial predictors give bad approximants to the true dynamics except very close to the attractor. On the other hand, some of the problems related to global polynomials are believed to be connected to the structure of the models (Aguirre and Billings, 1993b) and promising results have been reported for some systems using nonlinear global polynomials with simplified structure (Kadtke et al., 1993; Aguirre and Billings, 1994b).

Rational models share with polynomials the advantage of being linear in the parameters. This feature makes it possible to use well known and numerically robust algorithms to estimate the parameters of such models. Moreover, rational models seem to extrapolate better than polynomials (Farmer and Sidorowich, 1988a).

The *radial basis function* (RBF) approach is a global interpolation technique with good localization properties and it is easy to implement as the algorithm is essentially independent of the dimension (Broomhead and Lowe, 1988; Casdagli, 1989; Whaba, 1992). However performance of radial basis functions depends critically upon the centres (Chen et al., 1990). For a few hundred data points the choice of the centres is a difficult task and the solution of the problem could become infeasible on standard workstations for more than 500 data points (Casdagli, 1989; Billings and Chen, 1992).

Local approximants are concerned with the mapping of a set of neighbouring points in a reconstructed state space into their future values. A major problem here is to select the neighbourhoods because such a choice is critical and there could be hundreds or even thousands of these (Farmer and Sidorowich, 1988b). The size of the neighbourhoods depends on the noise level and the complexity of the dynamics (Farmer and Sidorowich, 1991). In addition, local predictors are discontinuous and suffer from undesirable behaviour when long term intervals are computed.

A simple alternative to nonlinear modelling is the use of piecewise-linear representations (Billings and Voon, 1987). The discontinuities among the several linear models which compose a piecewise-linear model, can provide effects similar to those observed in nonlinear

models such as chaos (Mahfouz and Badrakhan, 1990a; Mahfouz and Badrakhan, 1990b). However, as other local representations, the final model is piecewise-linear and therefore discontinuous. Piecewise-linear models have been found to be unreliable indicators of the underlying dynamics in some cases (Billings and Voon, 1987), and a possible explanation for this is that such models violate the physically motivated hypothesis of smooth dynamical systems (Crutchfield and McNamara, 1987). Thus local predictors may not always be suitable for predicting invariant measures (Brown et al., 1991).

Smooth interpolation functions have been suggested as a way of alleviating the problem caused by discontinuities in piecewise-linear models (Johansen and Foss, 1993). Such functions have localised properties which confers to the final models composed in this way some similarities with radial basis functions. As would be expected, the quality of the approximation depends on the choice of several operating regimes where the system dynamics are approximately linear. This information has to be available *a priori* and is somewhat critical. The problem of selecting the operating points is similar to the choice of neighbourhoods and of centres in other approaches.

Other representations for modelling nonlinear systems include Legendre polynomials (Cremers and Hüber, 1987), neural networks (Elsner, 1992; Principe et al., 1992) and weighted maps (Stokbro and Umberger, 1992).

At present no particular representation can be regarded as the best for any application and "finding a good representation is largely a matter of trial and error" (Farmer and Sidorowich, 1988a). On the other hand, it seems that global polynomial models are in many respects simpler and therefore more convenient (Kadtke et al., 1993). Moreover, it has been pointed out that "global equations of motion are to be preferred over atlas[1] equations of motion when the former is available" (Crutchfield and McNamara, 1987).

This paper investigates the use of global polynomials with simplified structure to estimate dynamical invariants of strange attractors.

[1]Crutchfield and McNamara call *atlas* the dynamical equations obtained using data from *local* regions of the attractor.

2.2 Algorithms

Consider the nonlinear autoregressive moving average model with exogenous inputs (NARMAX) (Leontaritis and Billings, 1985a; Leontaritis and Billings, 1985b)

$$y(k) = F^{\ell}[\, y(k-1), \ldots, y(k-n_y), u(k-d), \quad \ldots \quad , u(k-d-n_u+1), e(k), \ldots \\ \ldots \quad , e(k-n_e)\,] \ , \tag{4}$$

where n_y, n_u and n_e are the maximum lags considered for the output, input and noise terms, respectively and d is the delay measured in sampling intervals, T_s. Moreover, $u(k)$ and $y(k)$ are respectively input and output time series obtained by sampling the continuous data $u(t)$ and $y(t)$ at T_s. Furthermore, $e(k)$ accounts for uncertainties, possible noise, unmodelled dynamics, etc. and $F^{\ell}[\cdot]$ is some nonlinear function of $y(k)$, $u(k)$ and $e(k)$ with nonlinearity degree $\ell \in \mathbb{Z}^+$. In this paper, the map $F^{\ell}[\cdot]$ is taken to be a polynomial of degree ℓ. In order to estimate the parameters of this map, equation (4) has to be expressed in prediction error form as

$$y(k) = \Psi^{\tau}(k-1)\hat{\Theta} \ + \ \xi(k) \ , \tag{5}$$

where

$$\begin{aligned} \Psi^{\tau}(k-1) &= \left[\Psi^{\tau}_{yu}(k-1) \ \ \Psi^{\tau}_{yu\xi}(k-1) \ \ \Psi^{\tau}_{\xi}(k-1)\right] \ , \\ \hat{\Theta} &= \left[\hat{\Theta}^{\tau}_{yu} \ \ \hat{\Theta}^{\tau}_{yu\xi} \ \ \hat{\Theta}^{\tau}_{\xi}\right]^{\tau} \ , \end{aligned} \tag{6}$$

and where $\Psi^{\tau}_{yu}(k-1)$ is a matrix which contains linear and nonlinear combinations of output and input terms up to and including time $t-1$. The matrices $\Psi^{\tau}_{yu\xi}(k-1)$ and $\Psi^{\tau}_{\xi}(k-1)$ are defined similarly. The parameters corresponding to each term in such matrices are the elements of the vectors $\hat{\Theta}_{yu}$, $\hat{\Theta}_{yu\xi}$ and $\hat{\Theta}_{\xi}$, respectively. Finally, $\xi(t)$ are the residuals which are defined as the difference between the measured data $y(t)$ and the one-step-ahead prediction $\Psi^{\tau}(k-1)\hat{\Theta}$. The parameter vector Θ can be estimated by minimizing the following cost function (Chen et al., 1989).

$$J_{\mathrm{LS}}(\hat{\Theta}) \doteq \| \, y(k) - \Psi^{\tau}(k-1)\hat{\Theta} \, \| \tag{7}$$

where $\|\cdot\|$ is the Euclidean norm. Moreover, least squares minimization is performed using orthogonal techniques in order to effectively overcome two major difficulties in nonlinear model identification, namely i) numerical ill-conditioning and ii) structure selection.

The ERR criterion used for selecting the most important terms in the model has been devised as a byproduct of the orthogonal parameter estimation procedure (Billings et al., 1989; Korenberg et al., 1988). In a recent paper it has been shown that the ERR criterion gives qualitatively similar results as higher order spectrum techniques in detecting nonlinear interactions within the underlying dynamics (Aguirre and Billings, 1994a). Moreover, a similar criterion has been used to generate radial basis functions with a small number of parameters (Mees, 1993).

In a multivariable model with r inputs and m outputs, the entries in equation (4) are vectors, that is $u(t) = [u_1(t) \ \ldots \ u_r(t)]^T$, $y(t) = [y_1(t) \ \ldots \ y_m(t)]^T$ and , $e(t) = [e_1(t) \ \ldots \ e_m(t)]^T$, and both structure detection and parameter estimation can be performed in a way which is analogous to the monovariable case (Billings et al., 1989)

Because the final models will be composed of a reduced number of terms, which is a small fraction of the total number of candidate terms, the models in this paper can be viewed as 'simplified' or 'concise' global polynomials. It is believed that these models overcome some of the practical difficulties usually reported for non-simplified polynomials.

The last step in any identification problem is the validation of the estimated models. Most 'conventional' approaches to model validation are not particularly attractive when the models are chaotic and therefore alternative invariants should be used to quantify the quality and adequacy of the estimated models (Aguirre and Billings, 1994b).

2.3 Relation to previous work

There are some similarities between the approach described in Sec. 2.2 and the more general setting stated in Sec. 1. In particular, the approach followed in this paper does not involve finding neighbourhoods, thus $N_n = 1$ and all the data belong to a unique 'neighbourhood', that is, $\mathbf{y}(k) \in \mathcal{U}_1 \ \ k = d_e, \ldots, N$. This reduces the number of data required to estimate the dynamics. Moreover, the delay time is taken to be equal to the sampling period, thus $\tau = T_s$.

There are, however, a number of important differences. Firstly, a NARMAX model

includes input terms. This enables fitting data from non-autonomous systems and therefore estimating input/output maps, see (Casdagli, 1992; Hunter, 1992) for related ideas on this subject. An immediate consequence of this is that for input/output systems, it is not required that the output be on any particular attractor. In fact a number of input waveforms can be used to enhance estimation accuracy (Aguirre and Billings, 1993b). Once an input/output model has been estimated, a particular input can be used to generate data on a specific attractor.

Another important difference is the presence of noise terms, that is, the moving average part of the model. It should be noted that equation (2) will only hold in the unlikely case when noise is absent. Any noise in the data or any imperfection in the estimate of the map f^T will result in an extra term in the right hand side of equation (2). Such a term would be responsible for modelling the mismatch introduced by the noise and unmodelled dynamics. It is a well known result in the theory of system identification that if such a term is omitted from the model structure, the estimate of the map f^T will become biased during parameter estimation (Söderström and Stoica, 1989) and nonlinear models are no exception to this rule

It seems that when the noise is white and enters the system as a purely additive component, the division of the data in neighbourhoods and subsequent estimation reduces the bias. This will not be the case however if the model is global or if the noise is correlated. Thus in order to avoid bias a model for the noise and uncertainties, $\Psi^T_{yu\xi}(k-1)\hat{\Theta}_{yu\xi} + \Psi^T_{\xi}(k-1)\hat{\Theta}_{\xi}$, is included in the model structure before proceeding to parameter estimation. Once parameters have been estimated, only the deterministic part of the model is used, namely $\Psi^T_{yu}(k-1)\hat{\Theta}_{yu}$. This procedure can handle moderate amounts of white and correlated noise. Although only white noise is considered in this paper, satisfactory results have been obtained when the noise was correlated (Aguirre and Billings, 1993b).

Summarising, equation (5) is a hybrid model since it is composed of a deterministic part and a stochastic component. The latter is only used during parameter estimation in order to avoid bias on the former. Therefore, in this paper the models used to generate the surrogate data are purely deterministic although the stochastic part of the models is also represented for clarity. Thus, the deterministic component of the identified models is an approximation to the dynamics, that is, $f^T \approx \Psi^T_{yu}(k-1)\hat{\Theta}_{yu}$ where $T = T_s$.

The NARMAX polynomials used in this paper should be viewed as simplified polynomials as opposed to other approaches which include every possible term in the model. Apart from becoming impractical due to the explosion in the number of terms, the dynamics of overparametrized polynomials are severely affected (Aguirre and Billings, 1993b). Consequently, alternative structure selection techniques have recently been investigated (Kadtke et al., 1993). It should be noted however, that whilst some of the approaches in that reference operate on a basis which originally contains a large number of terms and subsequently deletes those elements which are less important, the present work follows a 'constructive' approach since it is assumed that the model basis is initially an empty set and in each step the most important term among those remaining in the set of candidate terms is appended to the model. In this way a very large number of candidate terms can be considered without incurring numerical problems.

A helpful account of some of the main points on this and the previous section can be found in the literature (Casdagli et al., 1992).

3 Diagnosing Chaos

In general, the problem of diagnosing chaos can be reduced to estimating invariants which would suggest that the data are chaotic. For instance, positive Lyapunov exponents, non-integer dimensions, fractal structures in Poincaré sections and abrupt jump in the prediction error for long prediction times would suggest the presence of chaos. The main question is how to confidently estimate such properties from the data, especially when the available records are relatively short and possibly noisy. The techniques that have been suggested in the literature may be divided in two major groups.

Non-parametric methods. These include the use of tools which take the data and estimate dynamical invariants which, in turn, will give an indication of the presence of chaos. Such tools include power spectra, the largest Lyapunov exponent, the correlation dimension, reconstructed trajectories, Poincaré sections, relative rotation rates etc. Detailed description and application of these techniques can be found in the literature (Moon, 1987; Tufilaro et al., 1990; Denton and Diamond, 1991). For a recent comment of the practical difficulties in

using Lyapunov exponents and dimensions for diagnosing chaos see (Mitschke and Dämmig, 1993).

Two practical difficulties common to most of these approaches are the number of data points available and the noise present in the data. These aspects are briefly discussed in the following section.

Poincaré sections are very popular for detecting chaos because for a chaotic system the Poincaré section reveals the fractal structure of the attractor. However, in order to be able to distinguish between a fractal object and a fuzzy cloud of points a certain amount of data is necessary. Moon (1987) has suggested that a Poincaré section should consist of at least 4000 points before declaring a system chaotic. For non-autonomous systems this means 4×10^3 forcing periods which could amount to 4×10^5 data points.

Prediction-based techniques. Some methods try to diagnose chaos in a data set based upon prediction errors (Sugihara and May, 1990; Casdagli, 1991; Elsner, 1992; Kennel and Isabelle, 1992). Thus predictors are estimated from, say, the first half of the data records and used to predict over the last half. Chaos can, in principle, be diagnosed based on how the prediction errors behave as the prediction time is increased (Sugihara and May, 1990), or based on how the prediction errors related to the true data compare to the prediction errors obtained from 'faked' data which are random but have the same length and spectral magnitude as the original data (Kennel and Isabelle, 1992).

Regardless of which criterion is used to decide if the data are chaotic or not, predictions have to be made. Clearly, the viability of these approaches depends on how easily predictors can be estimated and on the convenience of making predictions. The two aforementioned techniques require dividing the reconstructed state space into neighbourhoods and subsequently finding the closest neighbours. This may require long data sets and is usually time-consuming. Moreover the choice of neighbours is known to be critical.

These problems seem to be somewhat alleviated in the NARMAX model approach. NARMAX models can be readily obtained from relatively short data sets with a moderate amount of noise. Once a predictor is estimated criteria and statistics such as the ones presented in (Sugihara and May, 1990; Kennel and Isabelle, 1992) can be used to diagnose chaos.

4 Data requirements

The length and quality of the data records are crucial in the problem of characterisation of strange attractors. At present, there seems to be no general rule which determines the amount of data required to learn the dynamics, to estimate Lyapunov exponents and the correlation dimension of attractors. However it is known that "in general the detailed diagnosis of chaotic dynamical systems requires long time series of high quality" (Ruelle, 1987).

Typical values of data length for learning the dynamics are 2×10^4 (Farmer and Sidorowich, 1987; Abarbanel et al., 1990) for systems of dimension 2 to 3, $1.2\times10^4 - 4\times10^4$ (Casdagli, 1991).

It has been argued that to estimate the Lyapunov exponents $10^3 - 10^4$ forcing periods should be used (Denton and Diamond, 1991). Another estimate is that the number of data points needed is around 30^D where D is the dimension of the system (Wolf et al., 1985) but in some cases at least twice this number of points was required (Abarbanel et al., 1990). Typical examples in the literature use $4\times10^4 - 6.4\times10^4$ (Eckmann et al., 1986) 1.6×10^4 (Wolf and Bessoir, 1991) and 2×10^4 data points (Ellner et al., 1991).

Fairly long time series are also required for estimating the correlation dimension. In fact, it has been pointed out that dimension calculations generally require larger data records (Wolf and Bessoir, 1991). For a strange attractor, if insufficient data is used the results would indicate the dimension of certain parts of the attractor rather than the dimension of the entire attractor (Denton and Diamond, 1991). However, results have been reported which suggest that consistent estimates of the correlation dimension can be obtained from data sequences with less than 1000 points (Abraham et al., 1986). On the other hand, there seems to be evidence that "spuriously small dimension estimates can be obtained from using too few, too finely sampled and too highly smoothed data" (Grassberger, 1986). Thus typical examples use $1.5\times10^4-2.5\times10^4$ (Grassberger and Procaccia, 1983) and $0.8\times10^4-30\times10^4$ data points (Atten et al., 1984). Thus there seems to be no agreed upon rule to determine the amount of data required to estimate dimensions with confidence but it appears that at least a few thousand points for low dimensional attractors are needed (Theiler, 1986; Havstad and Ehlers, 1989; Ruelle, 1990; Essex and Nerenberg, 1991).

It should be realised that the difficulties in obtaining long time series goes beyond prob-

lems such as storage and computation time. Indeed, it has been pointed out that for some real systems, stationarity cannot always be guaranteed even over relatively short periods of time. Examples of this include biological systems (May, 1987; Denton and Diamond, 1991), ecological and epidemological data (Schaffer, 1985; Sugihara and May, 1990).

These difficulties can be avoided by first estimating models and subsequently using such models for generating longer sequences of surrogate data. An advantage of using NARMAX models to do this is that typically $10^3 - 2 \times 10^3$ data points are sufficient to perform both structure selection and parameter estimation. Another advantage is that, because noise terms are allowed in the model during parameter estimation, the deterministic part of the model, $\Psi^{\tau}_{yu}(t-1)\hat{\Theta}_{yu}$, is rather robust to moderate amounts of white and correlated noise.

Therefore as long as a dynamically valid model is estimated, sufficiently long series of clean data can be generated in order to meet the requirements imposed by the estimation of Lyapunov exponents, correlation dimension and other invariants. This two-step procedure was originally suggested in connection with radial basis functions (Casdagli, 1989).

5 Estimation of Dynamical Invariants

In this section a number of examples are given to illustrate the performance of NARMAX polynomials in retrieving dynamical invariants from short and noisy data sets. It is worth stressing that although the stochastic component of some models is represented for greater clarity, only the deterministic part is actually used in estimating the invariants.

5.1 Poincaré sections

Consider the Hénon map (Hénon, 1976)

$$\begin{cases} x(k) = 1 - 1.4\,x(k-1)^2 + y(k-1) \\ y(k) = 0.3\,x(k-1)\ . \end{cases} \tag{8}$$

Iterating this map and adding noise with variance $\sigma_e^2 = 9.2 \times 10^{-4}$ yields data with *signal to noise ratio* $\mathrm{SNR} = 20\log(\sigma_y^2/\sigma_e^2) = 55\,\mathrm{dB}$. The first return map for 1000 points taken from

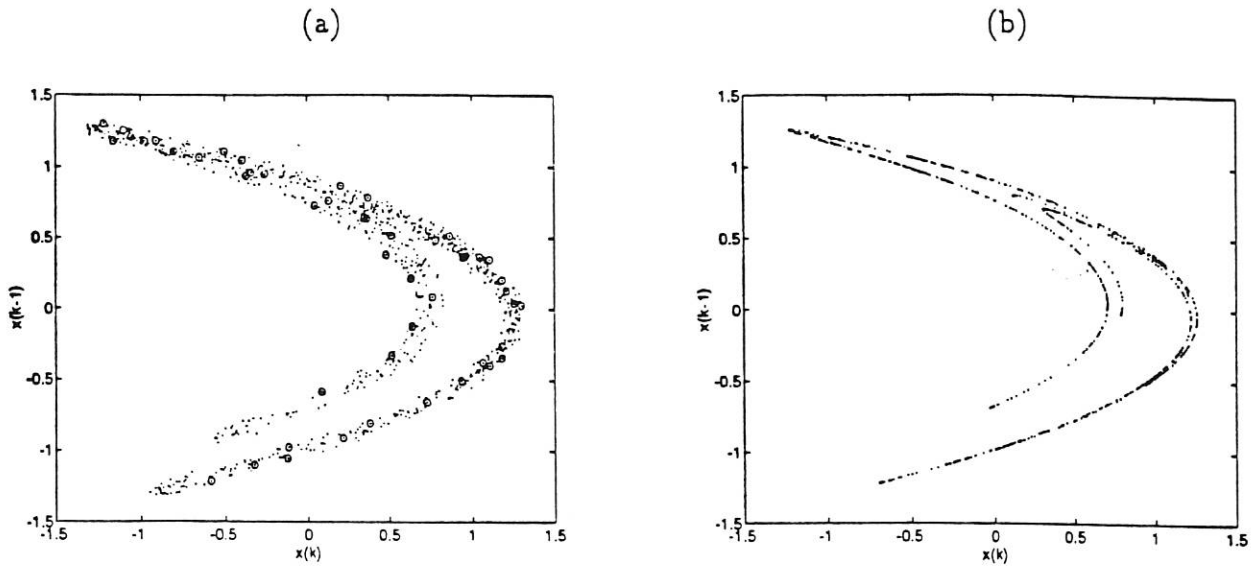

Figure 1: First return map for (a) the Hénon map contaminated with noise. Only the encircled data were used in the estimation, (b) the identified map of equation (9).

such data is shown in Fig 1a. The following model was estimated from the 50 points marked with circles

$$x(k) = 0.99433 - 1.3818\,x(k-1)^2 + 0.29302\,x(k-2)\ . \tag{9}$$

The first return map for this equation is shown in Fig. 1b and has a correlation dimension of $D_c = 1.11 \pm 0.216$ which shows good agreement with the original map for which $D_c = 1.21 \pm 0.01$. The correlation dimension estimated directly from 20000 data points with the same SNR as above was $D_c = 1.76 \pm 0.058$ revealing that the estimated value is quite sensitive to such levels of noise. Further improvement can be achieved by using more than 50 points, but the objective in this example was to show that the map can be estimated fairly accurately from a short and noisy time series.

In the case of driven oscillators, the practical reconstruction of Poincaré sections is restricted to controlled experiments because of the large amount of data required since only one point in such sections is obtained for each forcing period. Moreover, small amounts of noise often blur the delicate fractal structure of the attractor and the Poincaré sections tend to become fuzzy.

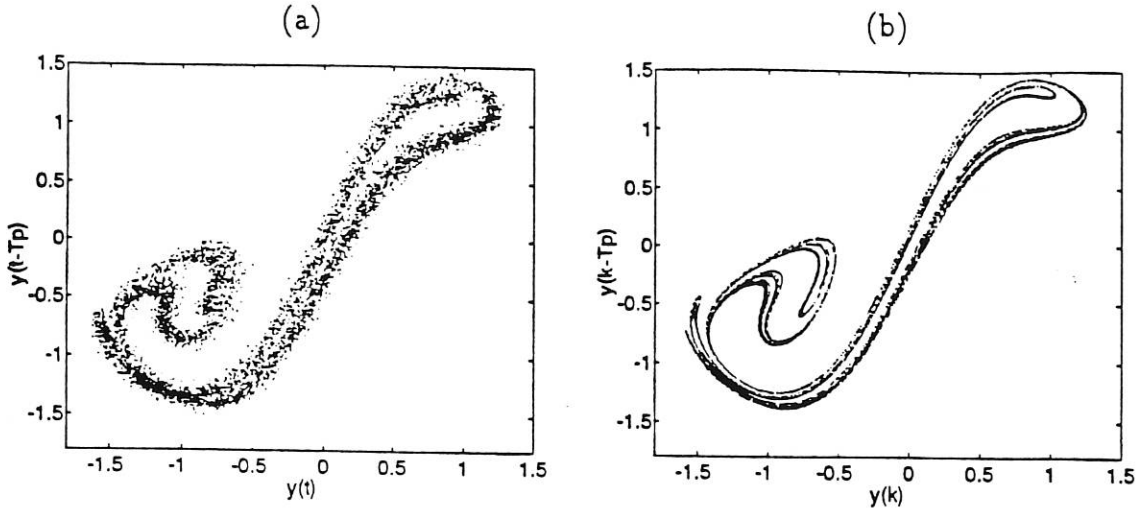

Figure 2: Poincaré sections (a) obtained from a noisy orbit of the Duffing-Holmes oscillator, (b) of the identified model of equation (11). $A=0.3$ and $\omega=1\,\mathrm{rad/s}$.

The well known Duffing-Holmes equation is commonly used to model mechanical oscillations arising in two-well potential problems (Moon, 1987). The equation which models this system is (Holmes, 1979)

$$\ddot{y} + \delta\,\dot{y} - \beta\,y + y^3 = A\cos(\omega\,t)\ . \tag{10}$$

For $\delta=0.15$, $\beta=1$, $A=0.3$ and $\omega=1\,\mathrm{rad/s}$, this system settles to a strange attractor. White noise with variance $\sigma_e^2=9\times10^{-4}$ was added to a sequence of $N=1500$ data points sampled at $T_s=\pi/15$. The Poincaré section obtained from a much longer time series but with the same SNR is shown in Fig. 2a. From such data the following model was obtained

$$\begin{aligned} y(k) \;=\; & 0.84725\,y(k-1) + 0.35713\,y(k-3) - 0.69431\times10^{-1}y(k-1)^3 \\ & +0.12780\times10^{-1}y(k-4) + 0.61319\times10^{-1}u(k-1) \\ & +0.40325\,y(k-2) - 0.24349\times10^{-3}y(k-1)y(k-2)y(k-5) \\ & -0.46215\,y(k-5) + 0.96339\times10^{-1}u(k-3) - 0.15316\,y(k-2)y(k-3)y(k-4) \\ & -0.73618\times10^{-2}y(k-1)^2y(k-5) + 0.71822\times10^{-1}y(k-1)y(k-3)y(k-4) \\ & +\Psi_\xi^T(t-1)\hat{\Theta}_\xi + \xi(t)\ . \end{aligned} \tag{11}$$

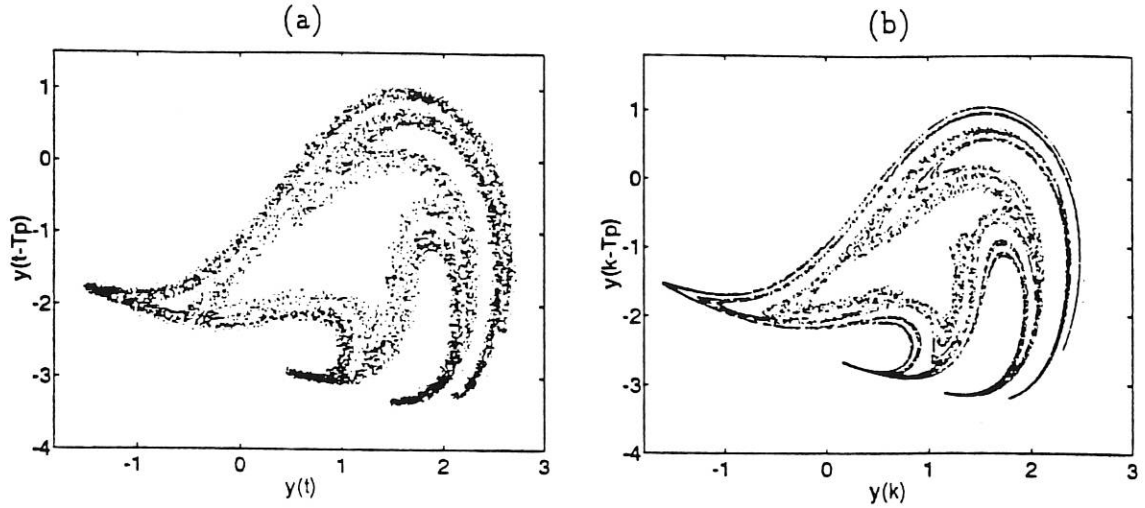

Figure 3: Poincaré sections (a) obtained from a noisy orbit of the modified van der Pol oscillator, (b) of the identified model of equation (13). $A=17$ and $\omega=4\,\mathrm{rad/s}$.

It should be noted that the data used in the identification of the model in equation (11) would yield only 50 points in Fig. 2a. An estimate of the original Poincaré section is shown in Fig. 2b which was obtained by iterating equation (11).

Similarly, consider the equation governing the dynamics of a forced oscillator with negative resistance (Ueda and Akamatsu, 1981)

$$\ddot{y} + \mu(y^2 - 1)\dot{y} + y^3 = A\cos(\omega t) \ . \tag{12}$$

This is known as the modified van der Pol oscillator (Moon, 1987) and has a self-sustained oscillation, that is, an oscillation for $A=0$ with $\omega=1.62\,\mathrm{rad/s}$. Taking $\mu=0.2$, $A=17$ and $\omega=4\,\mathrm{rad/s}$, this system settles to a strange attractor. Figure 3a shows the Poincaré section of this system when white noise with variance $\sigma_e^2=8.3\times10^{-4}$ (SNR=70.6 dB) was added to the measurements. Equation (12) was simulated and the data sampled at $T_s=\pi/80$. Noise with the same variance as above was added to 1500 data points which were used to identify the following NARMAX model

$$\begin{aligned} y(k) \ = \ & 0.83599\,y(k-1) + 0.87488\times10^{-1}y(k-4) + 0.68539\times10^{-1}u(k-2) \\ & +0.46776\times10^{-2}y(k-1)^3 - 0.47330\,y(k-6) + 0.12786\,y(k-2) \\ & +0.37341\,y(k-3) - 0.22840\times10^{-2}u(k-1) + 0.49504\times10^{-1}y(k-5) \end{aligned}$$

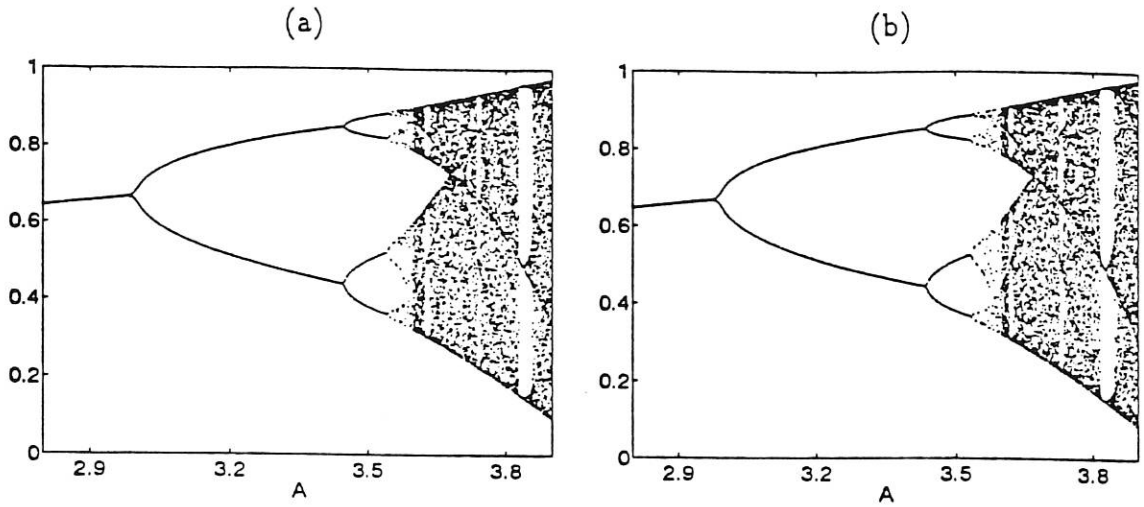

Figure 4: Bifurcation diagrama for (a) the logistic map, (b) the identified map of equation (15).

$$
\begin{aligned}
&-0.14841\times10^{-1}y(k-1)^2y(k-2)-0.81389\times10^{-1}u(k-3)+0.38305\times10^{-1}u(k-5)\\
&-0.13554\times10^{-1}u(k-4)+0.20404\times10^{-2}y(k-2)^2y(k-3)\\
&-0.34234\times10^{-2}y(k-1)y(k-6)^2+0.35999\times10^{-2}y(k-2)y(k-4)y(k-6)\\
&+\Psi_{\xi}^{T}(t-1)\hat{\Theta}_{\xi}+\xi(t)\ .
\end{aligned}
\tag{13}
$$

The Poincaré section obtained by iterating the deterministic part of equation (13) is shown in Fig. 3b. This model also has a self-sustained oscillation with $\omega = 1.56\,\mathrm{rad/s}$.

Comparing the Poincaré sections in Figs. 1b, 2b and 3b, reconstructed from the respective NARMAX models, to the original maps (not shown) shows that good agreement in such cases is possible. Moreover, less than 2000 noisy data points were used to learn the dynamics successfully.

5.2 Bifurcation diagrams

Consider the logistic equation (May, 1976)

$$
y(k) = A\,[1-y(k-1)]\,y(k-1)\ . \tag{14}
$$

This equation displays a variety of dynamical regimes as the bifurcation parameter is varied in the interval $2.8 \geq A \geq 3.9$. The bifurcation diagram for this system is shown in

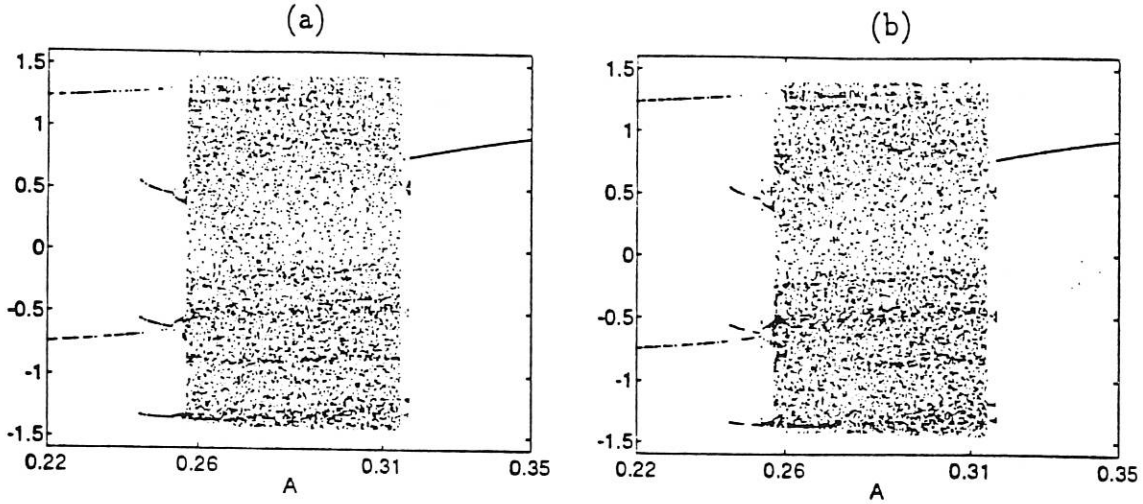

Figure 5: Bifurcation diagram of (a) the Duffing-Holmes oscillator, (b) the identified model of equation (11). $\omega = 1\,\mathrm{rad/s}$.

Fig. 4a. Equation (14) with $A = 3.8$ was used to generate 1750 data points to which white noise was added to give in a SNR=37.9 dB. Having fixed the structure, parameter estimation yielded

$$y(k) = 3.8492\,y(k-1) - 3.8380\,y(k-1)^2 \ . \tag{15}$$

In order to generate a bifurcation diagram from this model, equation (15) should be written in the same form of equation (14). This yields

$$y(k) = A\,(1.0029 - y(k-1))\,y(k-1) \tag{16}$$

for which the bifurcation diagram if Fig. 4b can be readily obtained. Comparison of Figs. 4a and 4b shows good agreement.

Figures 5a and 5b show the bifurcation diagrams of the Duffing-Holmes system in equation (10) and of the identified model shown in equation (11). Similarly, Figs. 6a and 6b display the bifurcation sequence of the oscillators described by equations (12) and (13), respectively. These figures show that the identified models do reproduce the major bifurcation patterns of the original systems. This is usually more demanding than, for instance, requiring that a model should reproduce invariants associated with particular attractors (Aguirre and Billings, 1994b).

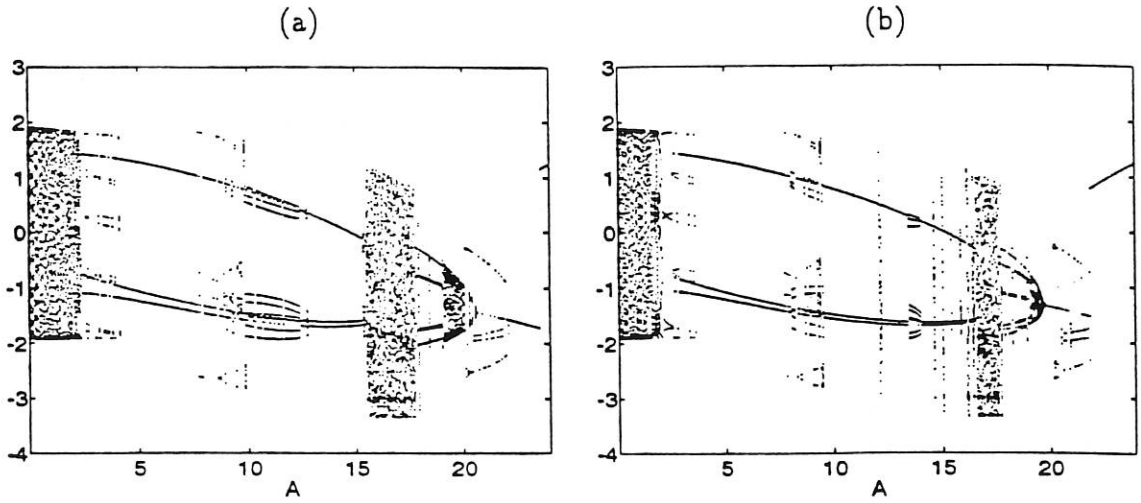

Figure 6: Bifurcation diagram of (a) the modified van der Pol oscillator, (b) the identified model of equation (13). $\omega = 4\,\mathrm{rad/s}$.

The identified models used to generate the above bifurcation diagrams were estimated from data on one particular attractor. Nevertheless, such models reproduce bifurcation sequences which span over a fairly wide parameter range. It has been shown, however, that the bifurcation diagrams of identified models of non-autonomous systems can be improved if the input amplitude is chosen such that the original system is better excited (Aguirre and Billings, 1993b).

5.3 Original and embedded trajectories

This section reports some results concerning the use of NARMAX polynomials in reproducing embedded and original trajectories of strange attractors. To investigate this three well known attractors are considered, namely Chua's double scroll, Lorenz and Rössler attractors. If only one variable of the original system is measured then a monovariable NARMAX polynomial is used and the iterated discrete-time output can be used to reconstruct the geometry of the original attractor in the embedding space. On the other hand, if all variables are measured, multivariable NARMAX models can be fitted to the data and the iterated discrete-time outputs can be used to reconstruct the original attractor geometry in state-space.

Consider the set of normalised equations governing the dynamics of Chua's circuit (Chua, 1992; Madan, 1993; Chua and Hasler, 1993)

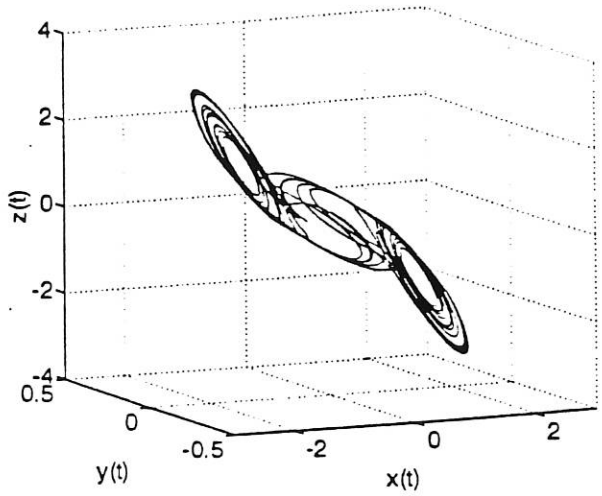

Figure 7: The double scroll Chua's attractor.

$$\begin{cases} \dot{x} = \alpha(y - h(x)) \\ \dot{y} = x - y + z \\ \dot{z} = -\beta y \end{cases} \quad , \quad h(x) = \begin{cases} m_1 x + (m_0 - m_1) & x \geq 1 \\ m_0 x & |x| \leq 1 \\ m_1 x - (m_0 - m_1) & x \leq -1 \end{cases} , \tag{17}$$

where $m_0 = -1/7$ and $m_1 = 2/7$. Varying the parameters α and β the circuit displays several regular and chaotic regimes. This system is a particular case of the more general *unfolded Chua's circuit* (Chua, 1993). The well known double scroll attractor, for instance, is obtained for $\alpha = 9$ and $\beta = 100/7$, see Fig. 7. Monovariable models for this attractor can be identified from each component individually (Aguirre and Billings, 1994a). In particular, the following model was estimated from 1750 measurements of the z component sampled at $T_s = 0.15$ and with SNR=42.7 dB

$$\begin{aligned} z(k) = {} & 0.17799 \times 10\, z(k-1) + 0.46703 z(k-3) z(k-4)^2 - 0.87039 z(k-3) \\ & + 0.91053 \times 10^{-1} z(k-1)^2 z(k-4) + 0.63361\, z(k-5) \\ & + 0.13195 \times 10^{-2} z(k-4)^2 z(k-5) + 0.57917 z(k-1) z(k-4) z(k-5) \\ & - 0.34940\, z(k-4) - 0.91067 \times 10^{-1} z(k-1)^3 + 0.81258 \times 10^{-1} z(k-1)^2 z(k-2) \\ & - 0.10603\, z(k-5)^3 - 0.16898\, z(k-1)^2 z(k-5) + 0.73569 \times 10^{-1} z(k-2) z(k-5)^2 \\ & - 0.12979 \times 10^{-1} z(k-1) z(k-2)^2 - 0.39490 z(k-1) z(k-5)^2 \\ & - 0.31911 \times 10^{-1} z(k-2) z(k-4)^2 + 0.22167 \times 10^{-1} z(k-1)^2 z(k-3) \\ & + 0.56608 z(k-3) z(k-5)^2 - 0.78674 z(k-3) z(k-4) z(k-5) \end{aligned}$$

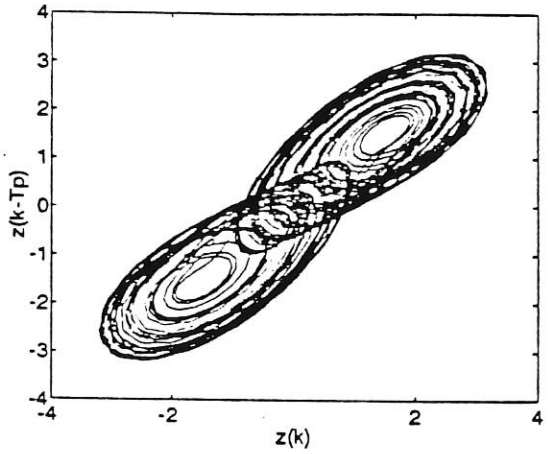

Figure 8: Embedded trajectory of the double scroll attractor obtained from equation $T_p = 2$. This figure compares very well with an original trajectory embedded in the same way.

$$-0.40725z(k-1)z(k-4)^2 + 0.19246z(k-1)z(k-3)z(k-5)$$
$$-0.15206z(k-3)^2z(k-5) + \Psi_{\xi}^{T}(t-1)\hat{\Theta}_{\xi} + \xi(t) \ . \qquad (18)$$

The deterministic part of this model can be iterated to produce the embedded attractor shown in Fig. 8 which is very similar to a trajectory on the original attractor embedded in the same way.

A multivariable model can be obtained for the double scroll Chua's attractor if the three variables are used. Thus, the data were sampled at $T_s = 0.07$ and white noise was added to the x, y and z components resulting in signal-to-noise ratios equal to 72.9, 39.9 and 75.5 dB respectively, see Fig. 9a. Such data were used to identify the following model

$$\begin{aligned}
x(k) &= 0.11282\times 10x(k-1) + 0.55867\,y(k-1) - 0.47190\times 10^{-1}x(k-1)^3 \\
&\quad +0.39895\times 10^{-1}y(k-1)z(k-1)^2 - 0.31229\times 10^{-2}z(k-1)^3 \\
&\quad +0.18363\times 10^{-1}z(k-1) + \Psi_{\xi_x\xi_y\xi_z}^{T}(t-1)\hat{\Theta}_{\xi_x\xi_y\xi_z} + \xi_x(t) \\
y(k) &= 0.91948\,y(k-1) - 0.10392\times 10^{-3}z(k-1)^3 + 0.70843\times 10^{-1}x(k-1) \\
&\quad +0.67800\times 10^{-1}z(k-1) - 0.13424\times 10^{-2}x(k-1)^3 \\
&\quad +0.44206\times 10^{-3}x(k-1)^2y(k-1) + \Psi_{\xi_x\xi_y\xi_z}^{T}(t-1)\hat{\Theta}_{\xi_x\xi_y\xi_z} + \xi_y(t) \\
z(k) &= 0.96628\,z(k-1) - 0.95854\,y(k-1) - 0.36719\times 10^{-1}x(k-1)
\end{aligned}$$

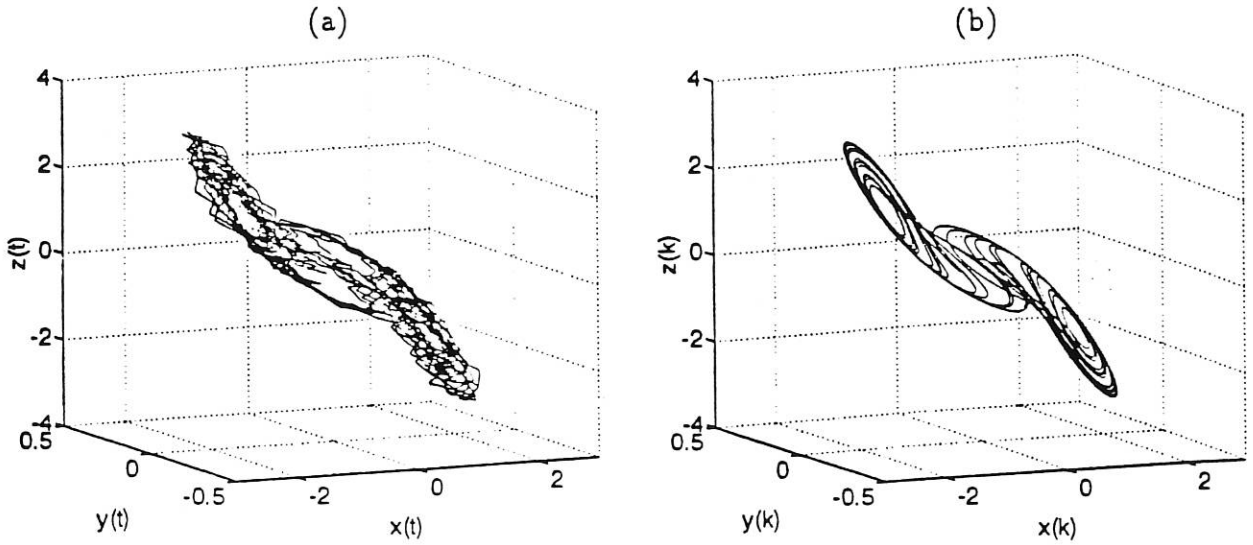

Figure 9: (a) Noisy trajectory used for identification, SNR=72.9, 39.9 and 75.5 dB for x, y and z components, respectively. (b) Double scroll Chua's attractor reconstructed from the identified model in equation (19).

$$-0.55765\times10^{-1}y(k-1)^3 + 0.10333\times10^{-2}x(k-1)^3$$
$$+0.20536\times10^{-2}x(k-1)y(k-1)z(k-1) + \Psi^T_{\xi_x\xi_y\xi_z}(t-1)\hat{\Theta}_{\xi_x\xi_y\xi_z}+\xi_z(t)\ . \qquad (19)$$

This estimated model settles to a strange attractor which resembles the original double scroll Chua's attractor, see Fig. 9b.

The Lorenz equations are (Lorenz, 1963)

$$\begin{cases} \dot{x} = \sigma(y - x) \\ \dot{y} = \rho x - y - xz \\ \dot{z} = xy - \beta z\ . \end{cases} \qquad (20)$$

Choosing $\sigma = 10$, $\beta = 8/3$ and $\rho = 28$, the solutions of equations 20 settles to the well known Lorenz attractor shown in Fig. 10a. Such data were sampled at $T_s = 0.025$ and white noise was added to 1000 data points. The resulting data had a signal to noise ratio of 37 dB and were used to identify the model

$$x(k) = 0.85919\,x(k-1) + 0.22489\,y(k-1) - 0.2833\times10^{-2}x(k-1)z(k-1)$$

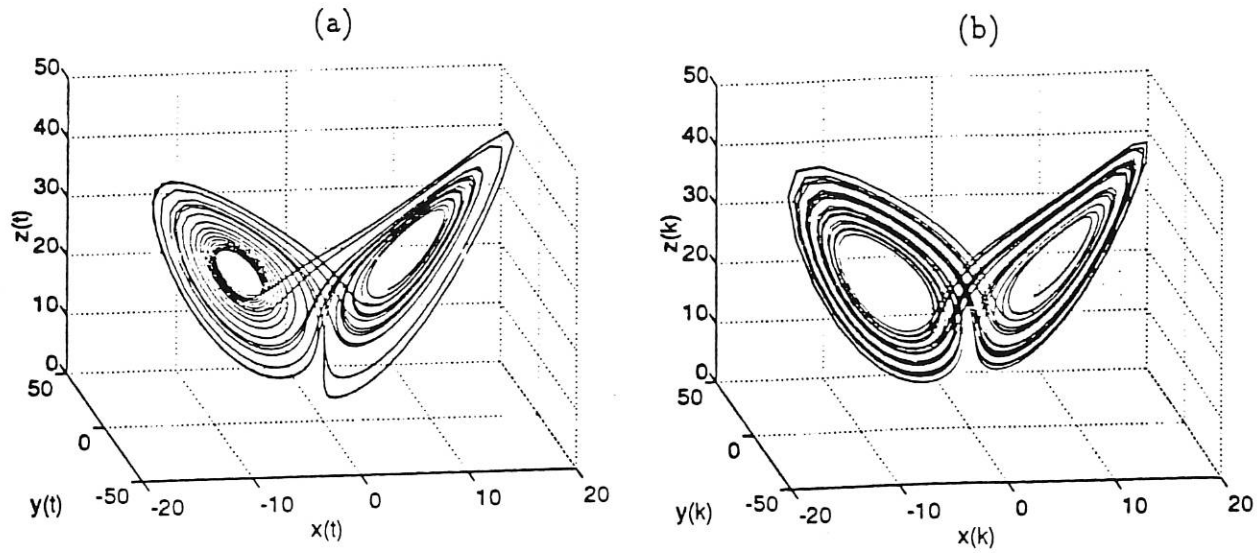

Figure 10: The Lorenz attractor (a) original, (b) reconstructed from the identified model in equation (21).

$$
\begin{aligned}
& -0.34598\times10^{-3}y(k-1)z(k-1) - 0.12927\times10^{-1} - 0.69159\times10^{-3}x(k-1)y(k-1) \\
& +0.73257\times10^{-3}x(k-1)^2 + \Psi^T_{\xi_x\xi_y\xi_z}(t-1)\hat{\Theta}_{\xi_x\xi_y\xi_z} + \xi_x(t) \\
y(k) \;=\; & 1.109\,y(k-1) - 0.6713\times10^{-2}y(k-1)z(k-1) - 0.18688\times10^{-1}x(k-1)z(k-1) \\
& +0.54947\,x(k-1) - 0.33705\times10^{-4}z(k-1)^2 + 0.67054\times10^{-4}x(k-1)y(k-1) \\
& -0.36965\times10^{-3}x(k-1)^2 + \Psi^T_{\xi_x\xi_y\xi_z}(t-1)\hat{\Theta}_{\xi_x\xi_y\xi_z} + \xi_y(t) \\
z(k) \;=\; & 1.0077\,z(k-1) + 0.93353\times10^{-2}x(k-1)y(k-1) - 0.21708\times10^{-2}z(k-1)^2 \\
& +0.72469\times10^{-2}x(k-1)^2 + 0.76919\times10^{-2}y(k-1)^2 - 0.47834 \\
& +0.48760\times10^{-4}y(k-1)z(k-1) + \Psi^T_{\xi_x\xi_y\xi_z}(t-1)\hat{\Theta}_{\xi_x\xi_y\xi_z} + \xi_z(t)\ . \qquad (21)
\end{aligned}
$$

This model can be iterated to obtain time series for the three components x, y and z. Plotting such data in state space reproduces the original attractor fairly well as can be seen in Fig. 10b.

Consider the set of equations originally suggested by Rössler (Rössler, 1976)

$$
\begin{cases}
\dot{x} = -(y + z) \\
\dot{y} = x + \alpha y \\
\dot{z} = \alpha + z(x - \mu)\ .
\end{cases} \qquad (22)
$$

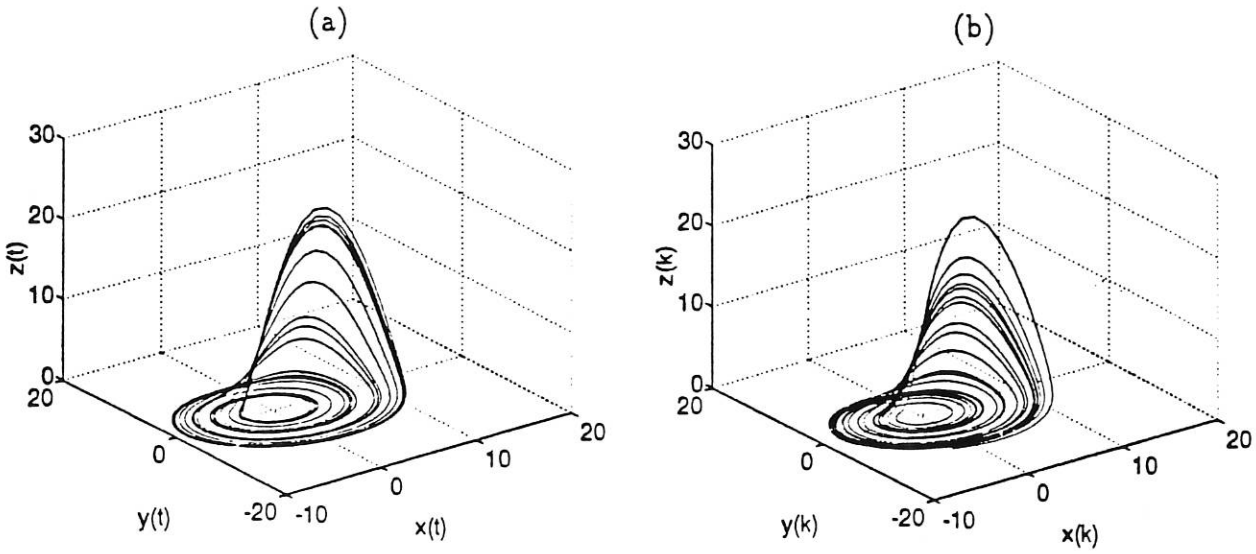

Figure 11: The Rössler attractor (a) original, (b) reconstructed from the identified model in equation (23).

The strange attractor displayed by these equations for $\alpha = 0.2$ and $\mu = 5.7$ is shown in Fig. 11a. These data were sampled at $T_s = 0.025$ and white noise was added to each component as for the previous examples. The resulting data with SNR=39 dB and $N = 1000$ were used to estimate the model

$$
\begin{aligned}
x(k) &= 0.99982\,x(k-1) - 0.24884\times10^{-1}y(k-1) - 0.10431\times10^{-4}z(k-1)^2 \\
&\quad -0.51675\times10^{-2}y(k-1)z(k-1) - 0.32205\times10^{-2}x(k-1)z(k-1) \\
&\quad +\Psi^{\mathrm{T}}_{\xi_x\xi_y\xi_z}(t-1)\hat{\Theta}_{\xi_x\xi_y\xi_z} + \xi_x(t) \\
y(k) &= 1.0048\,y(k-1) + 0.25786\times10^{-1}x(k-1) - 0.64615\times10^{-3}y(k-1)z(k-1) \\
&\quad -0.81375\times10^{-3}z(k-1) - 0.98027\times10^{-4}x(k-1)x(k-1) \\
&\quad +\Psi^{\mathrm{T}}_{\xi_x\xi_y\xi_z}(t-1)\hat{\Theta}_{\xi_x\xi_y\xi_z} + \xi_y(t) \\
z(k) &= 0.86209\,z(k-1) + 0.24939\times10^{-1}x(k-1)z(k-1) + 0.81343\times10^{-3}y(k-1)z(k-1) \\
&\quad +0.17808\times10^{-3}x(k-1)^2 - 0.61844\times10^{-3}z(k-1)^2 \\
&\quad +\Psi^{\mathrm{T}}_{\xi_x\xi_y\xi_z}(t-1)\hat{\Theta}_{\xi_x\xi_y\xi_z} + \xi_z(t)\ .
\end{aligned}
\tag{23}
$$

The reconstructed attractor obtained by iterating this model is shown in Fig. 11b. The above results illustrate that the estimated NARMAX models successfully reproduce the

geometry of the original attractors in state space. Quantitative invariants of such attractors can also be estimated using the identified models as reported in the next section.

5.4 Lyapunov exponents and correlation dimension

A major difficulty in estimating Lyapunov exponents and measures of fractal dimensions is the amount and quality of the data required, see Sec. 4. This section reports some results which suggest that a viable way of estimating such invariants from short and possibly noisy sequences of data is via system identification.

The results concern NARMAX polynomial models although any other representation could, at least in principle, be used to generate the surrogate data as long as the models used are dynamically valid. However, the use of some 'simple' model structures seem more adequate for some applications. Examples of this are discussed below.

One way of calculating Lyapunov exponents is by evaluating the Jacobian of the system along a trajectory on the attractor. In many approaches the Jacobian (and consequently the tangent space) has to be estimated from data a great number of times. This is usually done with local linear maps. Although the calculations involved are straightforward, the whole procedure is time consuming. Moreover, the estimation of the tangent space using linear maps is rather sensitive to noise and it has been pointed out that global models are better suited because they offer significant noise-averaging characteristics since the entire data set is fitted simultaneously (Kadtke et al., 1993).

A further advantage of using simple global model structures is that the Jacobian can be easily determined analytically from the estimated models. In particular, a polynomial NAR[2] model (see equation (4)) of order n_y can be represented as

$$\mathbf{y}(k) = \mathbf{f}(\mathbf{y}(k-1)) \tag{24}$$

[2] It should be noted that only the autoregressive part of the model is needed here.

where $\mathbf{y} \in \mathbb{R}^{n_y}$. Thus

$$\begin{bmatrix} y(k-n_y+1) \\ y(k-n_y+2) \\ \vdots \\ y(k) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ \frac{\mathrm{f}}{y(k-n_y)} & \frac{\mathrm{f}}{y(k-n_y+1)} & \frac{\mathrm{f}}{y(k-n_y+2)} & \cdots & \frac{\mathrm{f}}{y(k-1)} \end{bmatrix} \begin{bmatrix} y(k-n_y) \\ y(k-n_y+1) \\ \vdots \\ y(k-1) \end{bmatrix} , \quad (25)$$

where $\mathrm{f}/y(k-i)=0$ if $\mathrm{f}(\cdot)$ does not include $y(k-i)$. Thus the Jacobian of f is

$$\mathrm{Df} = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ \frac{\partial \mathrm{f}}{\partial y(k-n_y)} & \frac{\partial \mathrm{f}}{\partial y(k-n_y+1)} & \frac{\partial \mathrm{f}}{\partial y(k-n_y+2)} & \cdots & \frac{\partial \mathrm{f}}{\partial y(k-1)} \end{bmatrix} . \quad (26)$$

If a global model is used, the structure and parameters of such a model are estimated only once using the entire data set. Further, the Jacobian in equation (26) can be easily obtained analytically. This can also be done only once. The Jacobian can then be promptly evaluated at specific points taken along an orbit on the attractor, $\mathrm{Df}(\mathbf{y})$, as many times as necessary.

The accuracy of Lyapunov exponents can be improved by using more terms in the Taylor expansion of f around y (Brown et al., 1991). It should be noted that usually only the Jacobian of f is used which corresponds to the first term in the series. For the same reasons discussed above, the use of higher-order derivatives of f, that is, $\mathrm{D}^{(2)}\mathrm{f}(\mathbf{y})$, $\mathrm{D}^{(3)}\mathrm{f}(\mathbf{y}), \dots$ is greatly facilitated if global models with simple structure are obtained first.

Consider the delay system (Mackey and Glass, 1977)

$$\dot{y}(t) = \frac{a\, y(t-\tau)}{1+y^c(t-\tau)} - b\, y(t) \ , \quad (27)$$

where τ is the delay and the following values have been used $a=0.2$, $b=0.1$, $c=10$ and $\tau=30$. Equation (27) was originally suggested to model the creation of blood cells in patients with leukemia. Because of the delay τ this system is of high dimension (Farmer, 1982). Using a sequence of 234 data points sampled at $T_s=3$ the following model was obtained

$$
\begin{aligned}
y(k) \;=\; & 0.24662\times 10\, y(k-1) - 0.16423\times 10\, y(k-2) + 0.60992\, y(k-3) \\
& +0.73012\times 10^{-1} y(k-5)^2 y(k-10)^2 + 0.38566\, y(k-3)y(k-10) \\
& +0.66999\, y(k-1)y(k-10)^2 + 0.88364\, y(k-1)^3 - 0.67300\, y(k-4)y(k-10)^2 \\
& -0.11929\times 10\, y(k-1)^2 - 0.50451\times 10^{-1} y(k-4)y(k-5) - 0.24765\, y(k-1)^4 \\
& +0.42081\, y(k-4)y(k-9)y(k-10)^2 - 0.70406\, y(k-1)y(k-10)^3 \\
& -0.14089\, y(k-5)y(k-8)^2 + 0.14807\, y(k-1)y(k-7)y(k-10) \\
& +\Psi_{\xi}^{T}(t-1)\hat{\Theta}_{\xi} + \xi(t) \;. \qquad (28)
\end{aligned}
$$

This model has $\lambda_1 = 0.006$, $\lambda_2 = 0.002$ and $D_c = 3.02 \pm 0.228$ which compare fairly well to the original values.

Table 1 lists the values of the largest Lyapunov exponent and of the correlation dimension estimated from noise-free data obtained from both the original and the identified models. It is noted that the data length and noise levels in the table correspond to the records used in the identification and not to the data from which the invariants were actually estimated.

5.5 Location and stability of fixed points

To determine the location and stability of fixed points is often the first step in the analysis of nonlinear systems (Guckenheimer and Holmes, 1983). In the case of non-autonomous oscillators, a first step in the analysis is the determination of the fixed points of the respective autonomous system. This section investigates the feasibility of using estimated NARMAX models to determine the location and stability of the fixed points of the original systems. Two examples are considered, a monovariable model of the Duffing-Holmes oscillator and a multivariable model for the double scroll Chua's attractor.

Table 1. Estimated values of λ_1 and D_c from identified models

System	Original			Estimated			N/FP^c	SNR^a (dB)
	Eq.	$\lambda_1{}^b$	D_c	Eq.	$\lambda_1{}^b$	D_c		
Duffing-Ued.	*d*	0.11	2.19 ±0.020	*d*	0.12	2.34 ±0.050	1900/15.8	∞
Duffing-Hol.	(10)	0.20	2.40 ±0.019	(11)	0.19	2.37 ±0.031	1500/50	58.5
van der Pol	(12)	0.33	2.18 ±0.028	(13)	0.31	2.20 ±0.044	1500/9.4	70.6
Chua's spiral	(17)	0.11	1.88 ±0.010	*e*	0.11	1.87 ±0.021	1750	∞
Chua's DS-1D	(17)	0.23	1.99 ±0.023	(18)	0.23	2.03 ±0.019	1750	42.7
Chua's DS-3D	(17)	0.23	1.99 ±0.023	(19)	0.24	2.08 ±0.009	1000	39.9[f]
Lorenz	(20)	0.90[g]	2.01 ±0.017	(21)	0.90	2.06±0.021	1000	37
Rössler	(22)	0.074	1.91±0.002[h]	(23)	0.066	1.89±0.010[i]	1000	39
Mackey-Glass	(27)	0.007	3.0 ±0.2	(28)	0.006	3.02 ±0.228	234	∞
		0.004[j]			0.002[j]			

[a] $20\log_{10}(\sigma_y^2/\sigma_e^2)$ [b] Computed using $\log_e$ [c] Number of forcing periods

[d] (Aguirre and Billings, 1994b) [e] (Aguirre and Billings, 1994a) [f] For the *y* component

[g] (Peitgen et al., 1992) [h] Lyapunov dimension $D_L = 2.01$ [i] $D_L = 2.02$

[j] Second Lyapunov exponent, λ_2

5.5.1 The monovariable case

The Duffing-Holmes system in equation (10) can be expressed in state space as

$$\begin{cases} \dot{y} = x \\ \dot{x} = -\delta\, x + \beta\, y - y^3 + A\cos(\omega\, t) \ . \end{cases}$$

The fixed points of this system with $A = 0$ are $(\bar{x}, \bar{y}) = (0,0); (0, +\sqrt{\beta})$ and $(0, -\sqrt{\beta})$. The characteristic equation of the Jacobian of this system is $\lambda_c^2 + \delta\,\lambda_c + 3\,y^2 - \beta = 0$. Thus for the values $\delta = 0.15$ and $\beta = 1$, it becomes clear that the aforementioned fixed points are a saddle and two sinks, respectively. The location of such fixed points as well as the fixed points of estimated models are listed in Table 2.

Table 2. Fixed points — the autonomous case

Fixed Points	Original (y-coordinate)	Estimated	
		Eq. (29)	Eq. (11)
Saddle	0.0	0.0	0.0
Sinks	± 1	± 0.99928	± 0.99964

From a sequence of 1500 data points taken on the strange attractor attained with $u(t) = 0.3\cos(t)$, the following model was estimated for $n_y = 2$

$$\begin{aligned} y(k) &= [a_1 + a_3\, y(k-1)^2 \quad a_2 + a_4\, y(k-2)^2]\,[y(k-1)\; y(k-2)]^{\mathrm{T}} + b_1 u(k-1) + b_2 u(k-2) \\ &\quad + \Psi_{\xi}^{\mathrm{T}}(t-1)\hat{\Theta}_{\xi} + \xi(t)\ , \\ &= 0.20122\times 10\, y(k-1) - 0.96973\, y(k-2) - 0.42851\times 10^{-1} y(k-1)^3 \\ &\quad + 0.44832\times 10^{-1} u(k-1) - 0.20161\times 10^{-2} u(k-2) + 0.31973\times 10^{-3} y(k-2)^3 \\ &\quad + \Psi_{\xi}^{\mathrm{T}}(t-1)\hat{\Theta}_{\xi} + \xi(t)\ . \end{aligned} \tag{29}$$

The fixed points of the estimated model in equation (29) for $u(k-1) = u(k-2) = 0$ are $\bar{y} = 0, +0.99928$ and -0.99928. It should be noted that because the model in equation (29) is monovariable and has been estimated from data of the y-component, the fixed points obtained from this model correspond to the fixed points of the y-component in the original system.

To assess the stability of each fixed point a state space representation should be used since the model is of second-order. Thus taking the state vector $\mathbf{y}(k-1) = [y(k-2)\; y(k-1)]^{\mathrm{T}}$, the eigenvalues of the Jacobian of this model are given by the roots of the following quadratic equation

$$\lambda_{\mathrm{d}}^2 - (a_1 + 3\, a_3\, y(k-1)^2)\,\lambda_{\mathrm{d}} - (a_2 + 3\, a_4\, y(k-2)^2) = 0\ . \tag{30}$$

Solving for λ_{d} in the last equation also reveals that the fixed points are a saddle and two sinks. This agrees with the analytical results obtained from the original system.

The model in equation (29) was obtained from noise-free data and that enabled estimating such a model with a relatively small maximum lag, namely $n_y = 2$. In practice, however,

because of the presence of noise larger values of n_y will usually be required if the model is monovariable. The final state space will be in the $\mathbb{R}^{n_y}$ and therefore the fixed points of the resulting models will have more eigenvalues than the original system if $n_y > 2$. A similar problem is known in the estimation of the spectrum of Lyapunov exponents. In fact, some techniques have been put forward to distinguish between genuine and spurious Lyapunov exponents (Stoop and Parisi, 1991; Parlitz, 1992). However, it seems that the estimated models would still be useful for assessing the location and stability of the original fixed points even when n_y is greater than the order of the original system. It is believed, although further investigation is needed in this direction, that similar techniques used to detect spurious Lyapunov exponents could also be used to discern which eigenvalues are not genuine.

Consider the model in equation (11) which was identified from data generated by the Duffing-Holmes system and subsequently corrupted with white noise. Following the same procedure as above, the fixed points of this model for $u(k-1) = u(k-3) = 0$ are $\bar{y} = 0, +0.99964$ and -0.99964. The eigenvalues of the Jacobian of this model computed at such fixed points also reveal that the first one is a saddle-type and the other two are sinks. The eigenvalues are shown in Table 3. It is noted that the classification of the fixed points would not be altered even if it was not known *a priori* which were the genuine eigenvalues. It is remarkable that the genuine eigenvalues of the model in equation (11) are very accurate.

It should be observed that in order to compare the rates of contraction and expansion of nearby orbits the eigenvalues of the Jacobian related to the continuous system were mapped in the discrete domain using $\lambda_d = e^{\lambda_c T_s}$.

The results in Table 3 clearly reveal that it is viable to estimate not only the fixed points and related stability but also the rates of contraction and expansion of orbits in the neighbourhood of such fixed points. This can be done by straightforward calculations due to the simple structure of the models identified.

Table 3. Eigenvalues of the Jacobian evaluated at the fixed points for the Duffing-Holmes original and identified models.

Fixed Points	Original		Estimated		
	Continuous Domain (λ_c)	Discrete Domain ($e^{\lambda_c T_s}$)	Eq. (29)	Eq. (11)	
				genuine	spurious
Saddle	-1.0778	0.7979	0.7999	0.7977	-0.220 $\pm j$ 0.780
	0.9271	1.2143	1.2123	1.2161	-0.7262
Sinks	-0.075 $\pm j$ 1.412	0.942 $\pm j$ 0.287	0.942 $\pm j$ 0.286	0.947 $\pm j$ 0.287	-0.235 $\pm j$ 0.789
					-0.7158

5.5.2 The multivariable case

To investigate how the location and stability of fixed points can be recovered by means of a multivariable NAR (see footnote 2) polynomial model, consider the Chua system, equation (17), and the identified model (19).

The fixed points of the original system are $(x, y, z) = (0, 0, 0)$, $(1.5, 0, -1.5)$ and (-1.5,0,1.5). These fixed points are saddles with indices one two and two respectively. This information is conveyed by the eigenstructure of the Jacobian evaluated at such fixed points.

If the estimated model is multivariable it is possible, in principle, to recover the coordinates of each fixed point along the x, y and z components. In the monovariable case, only the coordinate related to the component used to identify the model can be recovered, see Table 2. Another difference is that whereas in the monovariable case the fixed point can be determined analytically, for multivariable models a set of nonlinear algebraic equations must be solved. Although this is still a simple problem, it cannot always be solved without resorting to numerical methods. Thus the fixed points of model in equation (19) are $(x, y, z) = (0, 0, 0)$, (1.5767,-6.927×10^{-4},-1.5767) and (-1.5767,6.927×10^{-4},1.5767) which are also saddles with indices one, two and two respectively. These results are summarised in Table 4.

It is worth pointing out that apparently the unstable eigenvalues seem to be estimated more accurately than the stable ones. This is more evident for the nontrivial fixed points. It seems appropriate to suggest that such inaccuracy may be attributed to the fact that, in

steady state, no trajectories evolve in the directions of the stable manifolds (the inlets of the nontrivial fixed points) and therefore there is less information concerning the rates of contraction is such directions in the records. On the other hand, because the trajectories 'jump' from the unstable manifolds of one non-trivial fixed point to the other, there seems to be motion in the direction of the stable manifold or inlet of the trivial fixed point and consequently the eigenvalue associated with this direction is better estimated.

Table 4. Eigenvalues of the Jacobian evaluated at the fixed points for the Chua's double scroll original and identified models.

Original			Estimated	
Fixed Points	Continuous Domain (λ_c)	Discrete Domain ($e^{\lambda_c T_s}$)	Fixed Points	Eq. (19)
(0,0,0)	2.2174	1.1679	(0,0,0)	1.1741
	-0.9658$\pm j$2.7112	0.9179$\pm j$0.1763		0.9199$\pm j$0.1888
(1.5,0,-1.5)[a]	-3.9421	0.7589	(1.5767,0.0007,	0.6825
	0.1854$\pm j$3.0470	0.9901$\pm j$2.1444	-1.5767)	0.9903$\pm j$2.1486

[a] The stability of the fixed point (-1.5,0,1.5) is identical and therefore has been omitted.

A similar problem can be observed for the Lorenz system which has fixed points at $(x, y, z) = (0, 0, 0)$, $(8.4853, 8.4853, 27)$ and (-8.4853,-8.4853,27). The fixed points of the identified model in equation (21) are (-0.0001,0.0012,1.7736), (8.4080,8.3644,25.7312) and (-8.5872,-8.5065,26.1559). Clearly the identified model does not have the trivial fixed point and it has also become unsymmetrical. It is conjectured that because there are no trajectories which evolve 'close' to the origin, the dynamics in the neighbourhood of the trivial fixed point of the original system are not well represented in the data. Consequently such a fixed point becomes 'unobservable' to the identification algorithm[3]. It is interesting to note that this is not the case for the double scroll since the dynamics evolve in the vicinity of the three fixed points although not along all the directions as discussed above.

[3]We are grateful to Carlos Fonseca for this observation.

6 Limitations

There are two main limitations to the use of NARMAX polynomials in modelling nonlinear dynamical systems, namely high levels of noise and 'fast dynamics'.

It is believed that noise in the data will pose greater difficulties to prediction-based estimation algorithms if such data are chaotic and that there is a maximum level of noise which can be handled adequately by such algorithms (Aguirre and Billings, 1993a). If good models cannot be obtained because of high noise levels, filtering techniques can be used to alleviate such a problem. A number of noise-reduction techniques for chaotic data have been suggested in the literature which could be used to pre-filter the data before identification. In particular, filtering techniques which are also based on NARMAX polynomials and the algorithms in Sec. 2.2 have recently been investigated in the context of system identification (Aguirre and Billings, 1993a; Aguirre et al., 1994).

The second aforementioned difficulty is related to the underlying dynamics. Some of the problems related to polynomial model bases can be avoided by the judicious choice of the terms which compose such bases. However, polynomial models may fail to give good results if the dynamics present abrupt 'jumps'. In such cases different representations should be tested. It should be noted that bases composed by rational functions, radial basis functions and neural nets can also be represented as NARMAX models (Billings and Chen, 1989; Chen et al., 1990).

7 Conclusions

This paper has investigated the use of NARMAX models as a means of estimating dynamical invariants from short and noisy time series which would be inadequate for estimating such invariants by using most current techniques. Thus a two step procedure has been investigated in which global models are first estimated from data records which may have less than two thousand points and be noise-corrupted. The deterministic part of the identified models can then be used to produce an, in principle, unlimited amount of clean data from which dynamical invariants can be estimated more comfortably.

It has been shown that, because the estimated models are global, such models are apt

for analysis. In particular, the Jacobian of the model can be analytically obtained and subsequently evaluated along trajectories in order to estimate the Lyapunov exponents and the stability of the fixed points which can also be obtained analytically from the identified model.

Multivariable models have been estimated from state space data and subsequently used to reconstruct the three-dimensional Chua's double scroll, Lorenz and Rössler attractors. Whereas such models require that more variables be measured, greater robustness against noise has been observed in the multivariable case.

Although polynomial models are not adequate to model all types of dynamical systems, it is hoped that the results presented in this paper will draw attention to the fact that, as long as the polynomials are 'simplified' or 'concise' *and* possess the right structure, such models can describe the dynamics of a number of chaotic systems and therefore be used not only in modelling chaos by also in diagnosing and quantifying chaotic dynamics.

ACKNOWLEDGMENTS

The assistance of Eduardo Mendes is greatly appreciated. LAA gratefully acknowledges financial support from CNPq (Brazil) under grant 200597/90-6. SAB gratefully acknowledges that part of this work was funded by SERC under contract GR/H35286.

References

Abarbanel, H. D. I., Brown, R., and Kadtke, J. B. (1990). Prediction in chaotic nonlinear systems: Methods for time series with broadband Fourier spectra. *Phys. Rev. A*, 41(4):1782–1807.

Abraham, N. B., Albano, A. M., Das, B., De Guzman, G., Yong, S., Gioggia, R. S., Puccioni, G. P., and Tredicce, J. R. (1986). Calculating the dimension of attractors from small data sets. *Phys. Lett.*, 114A(5):217–221.

Aguirre, L. A. and Billings, S. A. (1993a). Identification of models for chaotic systems from noisy data: implication for performance and nonlinear filtering. (Submitted for publication).

Aguirre, L. A. and Billings, S. A. (1993b). Relationship between the structure and performance of identified nonlinear polynomial models. *(Submitted for publication).*

Aguirre, L. A. and Billings, S. A. (1994a). Discrete reconstruction of stange attractors in Chua's circuit. *Int. J. Bifurcation and Chaos*, (in press).

Aguirre, L. A. and Billings, S. A. (1994b). Validating identified nonlinear models with chaotic dynamics. *Int. J. Bifurcation and Chaos*, (in press).

Aguirre, L. A., Mendes, E. M., and Billings, S. A. (1994). Smoothing data with local instabilities for the identification of chaotic systems. (submitted for publication).

Atten, P., Caputo, J. G., Malraison, B., and Gagne, Y. (1984). Determination of attractor dimension of various flows. *Journal de Mécanique Théorique et Appliquée*, (in French), Numéro spécial:133–156.

Billings, S. A. (1980). Identification of nonlinear systems — a survey. *IEE Proceedigs Pt. D*, 127(6):272–285.

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.

Billings, S. A. and Chen, S. (1992). Neural networks and system identification. In Warwick, K., Irwing, G. W., and Hunt, K. J., editors, *Neural Netwoks for Systems and Control*, chapter 9, pages 181–205. Peter Peregrinus, London.

Billings, S. A., Chen, S., and Korenberg, M. J. (1989). Identification of MIMO nonlinear systems using a forward-regression orthogonal estimator. *Int. J. Control*, 49(6):2157–2189.

Billings, S. A. and Voon, W. S. F. (1984). Least squares parameter estimation algorithms for nonlinear systems. *Int. J. Systems Sci.*, 15(6):601–615.

Billings, S. A. and Voon, W. S. F. (1987). Piecewise linear identification of non-linear systems. *Int. J. Control*, 46(1):215–235.

Broomhead, D. S. and Lowe, D. (1988). Multivariable functional interpolation and adaptive networks. *Complex Systems*, 2:321–355.

Brown, R., Bryant, P., and Abarbanel, H. D. I. (1991). Computing the Lyapunov spectrum of a dynamical system from an observed time series. *Phys. Rev. A*, 43(6):2787–2806.

Casdagli, M. (1989). Nonlinear prediction of chaotic time series. *Physica D*, 35:335–356.

Casdagli, M. (1991). Chaos and deterministic *versus* stochastic non-linear modelling. *J. R. Stat. Soc. B*, 54(2):303–328.

Casdagli, M. (1992). A dynamical systems approach to modeling input-output systems. In Casdagli, M. and Eubank, S., editors, *Nonlinear Modeling and Forecasting*, pages 265–281. Addison Wesley, New York.

Casdagli, M., Jardins, D. D., Eubank, S., Farmer, J. D., Gibson, J., Theiler, J., and Hunter, N. (1992). Nonlinear modeling of chaotic time series: theory and applications. In Kim, J. H. and Stringer, J., editors, *Applied Chaos*, pages 335–380. John Wiley & Sons., New York.

Chen, S., Billings, S. A., Cowan, C. F. N., and Grant, P. M. (1990). Practical identification of NARMAX models using radial basis functions. *Int. J. Control*, 52(6):1327–1350.

Chen, S., Billings, S. A., and Luo, W. (1989). Orthogonal least squares methods and their application to nonlinear system identification. *Int. J. Control*, 50(5):1873–1896.

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

Chua, L. O. (1993). Global unfolding of Chua's circuit. *IEICE Trans. on Fundamentals of Electronics, Communications and Computer Sciences*, 76-A(5):704–734.

Chua, L. O. and Hasler, M. (1993). (Guest Editors). Speciall issue on Chaos in nonlinear electronic circuits. *IEEE Trans. Circuits Syst.*, 40(10–11).

Cremers, J. and Hüber, A. (1987). Construction of differential equations from experimental data. *Z. Naturforsch*, 42a:797–802.

Crutchfield, J. P. and McNamara, B. S. (1987). Equations of motion from a data series. *Complex Systems*, 1:417–452.

Denton, T. A. and Diamond, G. A. (1991). Can the analytic techniques of nonlinear dynamics distinguish periodic, random and chaotic signals? *Compt. Biol. Med.*, 21(4):243–264.

Eckmann, J. P., Kamphorst, S. O., Ruelle, D., and Ciliberto, S. (1986). Liapunov exponents from time series. *Phys. Rev. A*, 34(6):4971–4979.

Ellner, S., Gallant, A. R., McCaffrey, D., and Nychka, D. (1991). Convergence rates and data requirements for Jacobian-based estimates of Lyapunov exponents from data. *Phys. Lett.*, 153A(6,7):357–363.

Elsner, J. B. (1992). Predicting time series using a neural network as a method of distinguishing chaos from noise. *J. Phys. A: Math. Gen.*, 25:843–850.

Essex, C. and Nerenberg, M. A. H. (1991). Comments on 'Deterministic chaos: the science and the fiction' by D. Ruelle. *Proc. R. Soc. Lond. A*, 435:287–292.

Farmer, J. and Sidorowich, J. (1987). Predicting chaotic time series. *Phys. Rev. Lett.*, 59(8):845–848.

Farmer, J. and Sidorowich, J. (1988a). Exploiting chaos to predict the future and reduce noise. In Lee, Y., editor, *Evolution, Learning and Cognition*. World Scientific, Singapore.

Farmer, J. and Sidorowich, J. (1988b). Predicting chaotic dynamics. In Kelso, J. A. S., Mandell, A. J., and Shlesinger, M. F., editors, *Dynamic patterns in complex systems*, pages 265–292. World Scientific, Singapore.

Farmer, J. D. (1982). Chaotic attractors of an infinite-dimensional dynamical system. *Physica D*, 4:366–393.

Farmer, J. D. and Sidorowich, J. J. (1991). Optimal shadowing and noise reduction. *Physica D*, 47:373–392.

Giona, M., Lentini, F., and Cimagalli, V. (1991). Functional reconstruction and local prediction of chaotic time series. *Phys. Rev. A*, 44(6):3496–3502.

Grassberger, P. (1986). Do climatic attractors exist? *Nature*, 323:609–612.

Grassberger, P. and Procaccia, I. (1983). Measuring the strangeness of strange attractors. *Physica D*, 9:189–208.

Grassberger, P., Schreiber, J., and Schaffrath, C. (1991). Nonlinear time sequence analysis. *Int. J. Bifurcation and Chaos*, 1(3):521–547.

Guckenheimer, J. and Holmes, P. (1983). *Nonlinear oscillations, dynamical systems, and bifurcation of vector fields.* Springer-Verlag, New York.

Havstad, J. W. and Ehlers, C. L. (1989). Attractor dimensions of nonstationary dynamical systems from small data sets. *Phys. Rev. A*, 39(2):845–853.

Hénon, M. (1976). A two-dimensional map with a strange attractor. *Commun. Math. Phys.*, 50:69–77.

Holmes, P. J. (1979). A nonlinear oscillator with a strange attractor. *Philos. Trans. Royal Soc. London A*, 292:419–448.

Hunter, N. F. (1992). Application of nonlinear time-series models to driven systems. In Casdagli, M. and Eubank, S., editors, *Nonlinear Modeling and Forecasting*, pages 467–491. Addison Wesley, New York.

Johansen, T. A. and Foss, B. A. (1993). Constructing NARMAX models using ARMAX models. *Int. J. Control*, 58(5):1125–1153.

Kadtke, J. B., Brush, J., and Holzfuss, J. (1993). Global dynamical equations and Lyapunov exponents from noisy chaotic time series. *Int. J. Bifurcation and Chaos*, 3(3):607–616.

Kennel, M. B. and Isabelle, S. (1992). Method to distinguish possible chaos from coloured noise and to determine embedding parameters. *Phys. Rev. A*, 46(6):3111–3118.

Korenberg, M. J., Billings, S. A., Liu, Y. P., and McIlroy, P. J. (1988). Orthogonal parameter estimation algorithm for nonlinear stochastic systems. *Int. J. Control*, 48(1):193–210.

Leontaritis, I. J. and Billings, S. A. (1985a). Input-output parametric models for nonlinear systems part I: deterministic nonlinear systems. *Int. J. Control*, 41(2):303–328.

Leontaritis, I. J. and Billings, S. A. (1985b). Input-output parametric models for nonlinear systems part II: stochastic nonlinear systems. *Int. J. Control*, 41(2):329–344.

Linsay, P. S. (1991). An efficient method of forecasting chaotic time series using linear regression. *Phys. Lett.*, 153A(6,7):353–356.

Lorenz, E. (1963). Deterministic nonperiodic flow. *J. Atmos. Sci.*, 20:282–293.

Mackey, M. C. and Glass, L. (1977). Oscillation and chaos in physiological control systems. *Science*, 197:287–289.

Madan, R. A. (1993). *Chua's circuit: a paradigm for chaos.* World Scientific Publishers, Singapore.

Mahfouz, I. A. and Badrakhan, F. (1990a). Chaotic behaviour of some piecewise-linear systems part I: Systems with set up spring or with unsymetric elasticity. *J. Sound Vibr.*, 143(2):255–288.

Mahfouz, I. A. and Badrakhan, F. (1990b). Chaotic behaviour of some piecewise-linear systems part II: Systems with clearance. *J. Sound Vibr.*, 143(2):289–328.

May, R. M. (1976). Simple mathematical models with very complicated dynamics. *Nature*, 261:459–467.

May, R. M. (1987). Chaos and the dynamics of biological populations. *Proc. R. Soc. Lond. A*, 413:27–44.

Mees, A. (1993). Parsimonious dynamical reconstruction. *Int. J. Bifurcation and Chaos*, 3(3):669–675.

Mitschke, F. and Dämmig, M. (1993). Chaos versus noise in experimental data. *Int. J. Bifurcation and Chaos*, 3(3):693–702.

Moon, F. C. (1987). *Chaotic Vibrations - an introduction for applied scientists and engineers.* John Willey and Sons, New York.

Packard, N. H., Crutchfiled, J. P., Farmer, J. D., and Shaw, R. S. (1980). Geometry from a time series. *Phys. Rev. Lett.*, 45(9):712–716.

Parlitz, U. (1992). Identification of true and spurious Lyapunov exponents from time series. *Int. J. Bifurcation and Chaos*, 2(1):155–165.

Peitgen, H. O., Jürgens, H., and Saupe, D. (1992). *Chaos and Fractals - New frontiers of science*. Springer-Verlag, Berlin.

Principe, J. C., Rathie, A., and Kuo, J. M. (1992). Prediction of chaotic time series with neural networks and the issue of dynamic modeling. *Int. J. Bifurcation and Chaos*, 2(4):989–996.

Rössler, O. E. (1976). An equation for continuous chaos. *Phys. Lett.*, 57A(5):397–398.

Ruelle, D. (1987). Diagnosis of dynamical systems with fluctuating parameters. *Proc. R. Soc. Lond. A*, 413:5–8.

Ruelle, D. (1990). Deterministic chaos: the science and the fiction. *Proc. R. Soc. Lond. A*, 427:241–248.

Sauer, T., Yorke, J. A., and Casdagli, M. (1991). Embedology. *J. of Statistical Physics*, 65(3/4):579–616.

Schaffer, W. M. (1985). Order and chaos in ecological systems. *Ecology*, 66(1):93–106.

Söderström, T. and Stoica, P. (1989). *System Identification*. Prentice Hall, London.

Stokbro, L. and Umberger, D. K. (1992). Forecasting with weighted maps. In Casdagli, M. and Eubank, S., editors, *Nonlinear Modeling and Forecasting*, pages 73–93. Addison Wesley, New York.

Stoop, R. and Parisi, J. (1991). Calculation of Lyapunov exponents avoiding spurious elements. *Physica D*, 50:89–94.

Sugihara, G. and May, R. M. (1990). Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature*, 344:734–741.

Takens, F. (1980). Detecting strange attractors in turbulence. In Rand, D. A. and Young, L. S., editors, *Dynamical systems and turbulence, Lecture Notes in Mathematics, vol. 898*, pages 366–381. Springer Verlag, Berlin.

Theiler, J. (1986). Spurious dimension from correlation algorithms applied to limited time-series data. *Phys. Rev. A*, 34(3):2427–2432.

Tufilaro, N. B., Solari, H. G., and Gilmore, R. (1990). Relative rotation rates: fingerprints for strange attractors. *Phys. Rev. A*, 41(10):5117–5720.

Ueda, Y. and Akamatsu, N. (1981). Chaotically transitional phenomena in the forced negative-resistance oscillator. *IEEE Trans. Circuits Syst.*, 28(3):217–224.

Whaba, G. (1992). Multivariabe function and operator estimation, based on smoothing splines and reproducing kernels. In Casdagli, M. and Eubank, S., editors, *Nonlinear Modeling and Forecasting*, pages 95–112. Addison Wesley, New York.

Wolf, A. and Bessoir, T. (1991). Diagnosing chaos in the space circle. *Physica D*, 50:239–258.

Wolf, A., Swift, J. B., Swinney, H. L., and Vastano, J. A. (1985). Determining Lyapunov exponents from a time series. *Physica D*, 16:285–317.