
Monograph:
Aguirre, L.A., Mendes, E.M. and Billings, S.A. (1994) Smoothing data with Local Instabilities for the Identification of Chaotic Systems. Research Report. ACSE Research Report 502 . Department of Automatic Control and Systems Engineering

---

Smoothing Data with Local Instabilities for the Identification of Chaotic Systems

L A Aguirre, E M Mendes 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 502

January 1994

Smoothing Data with Local Instabilities for the Identification of Chaotic Systems

LUIS A. AGUIRRE[†], EDUARDO M. MENDES[‡], S. A. BILLINGS

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

Abstract

Chaotic systems are characterised by local divergence of nearby orbits in state space. This provokes sensitive dependence on initial conditions and in turn drastically limits the accuracy of long term predictions. This has important implications in the filtering of data generated by such systems. In the present paper the use of global smoothers for chaotic data is investigated. The ultimate objective is to be able to identify dynamically valid models from smoothed data when the identification from the original noisy data has completely failed. The objective is to produce identified models which faithfully reproduce the dynamical invariants of the original system such as the geometry of the attractors in state space, the largest Lyapunov exponent, fractal dimensions and Poincaré sections. Numerical examples are included which illustrate the main points of the paper.

1 Introduction

Chaotic systems have attracted a great deal of attention in the last three decades. Systems and models which undergo chaotic regimes for a rather wide range of operating conditions have been found in virtually every branch of science and engineering.

[†]e-mail:aguirre@acse.shef.ac.uk

[‡]e-mail:mendes@acse.shef.ac.uk

In the evolution of the study of chaotic systems, several distinct but sometimes co-existent phases can be distinguished. In the first phase, chaos was recognised as a deterministic dynamical regime which could be responsible for fluctuations that hitherto had been regarded as noise and therefore modelled as stochastic processes (Lorenz, 1963). In a subsequent phase, it was necessary both to develop criteria to detect chaotic dynamics and to establish dynamical invariants to quantify chaos (Guckenheimer, 1982; Eckmann and Ruelle, 1985; Denton and Diamond, 1991). Having succeeded in diagnosing chaos, the next step was to build models which would learn the dynamics from data on the strange attractor. In this respect a number of model structures have been investigated such as local linear mappings (Farmer and Sidorowich, 1987; Crutchfield and McNamara, 1987), radial basis functions (Broomhead and Lowe, 1988; Casdagli, 1989), neural networks (Elsner, 1992) and global nonlinear polynomials (Aguirre and Billings, 1993a; Kadtke et al., 1993).

A difficulty which appears to be common to all these approaches is that realistically noise will be present in the data. In particular, it has been conjectured that the local divergence of nearby orbits in a chaotic system imposes a natural limit on the accuracy of prediction-based identification algorithms when the data are noisy (Aguirre and Billings, 1993a). Consequently, there has been great motivation to develop filtering techniques for chaotic systems. This comprises one of the current phases in the investigation of chaos and is evident in the recent literature (Mitschke, 1990; Chennaoui et al., 1990; Schreiber and Grassberger, 1991; Broomhead et al., 1992; Davies, 1992; Grassberger et al., 1993; Holzfuss and Kadtke, 1993).

Some authors assume that some kind of *a priori* knowledge concerning the original system is available such as a piece of noise-free data (Marteau and Abarbanel, 1991), the structure of the maps describing the underlying dynamics (Davies, 1992), or even the complete maps, that is, structure and parameters are known (Hammel, 1990; Ozaki, 1993). However, a clear limitation in any real noise reduction problem is that the underlying dynamics are not usually known *a priori* and the map has to be estimated (learned) from the noisy data as an integral part of the noise reduction process. Consequently, the noise will pose limitations on the amount of noise which can effectively be eliminated. In the field of nonlinear dynamics, the main objective of filtering a chaotic time series is to enable the reconstruction and estimation of dynamical invariants such as Poincaré sections, Lyapunov exponents and fractal

dimensions. In the present paper, however, the ultimate objective is to enable the identification of dynamically valid models which would reproduce the aforementioned invariants from sequences of filtered data.

In a first attempt to solve this identification problem, global nonlinear predictors were used to filter the data (Aguirre and Billings, 1993a). In such a procedure the noise was separated from the signal by means of (very) short-term predictions. Because a chaotic predictor actually amplifies uncertainties in some 'direction' in state space, the aforementioned approach cannot be used to filter the noise by successive passes through the data and this therefore limits the achievable noise reduction.

This paper investigates the use of global smoothers. Unlike prediction-based techniques, smoothers are well suited for filtering chaotic data via successive noise-reduction iterations. Both global linear and nonlinear smoothers are considered. Although linear smoothers tend to perform well in general, results suggest that in some cases the nonlinear smoothers outperform the linear counterparts mainly in two situations, namely i) for a small number of passes through the data, and ii) for higher noise levels.

Furthermore, in order to investigate a somewhat more realistic scenario, in the present paper no *a priori* knowledge is assumed. The starting point is a set of noisy data from which no valid model could be identified. The final objective is to be able to identify a dynamically valid model from a set of data obtained from the original records via smoothing.

The paper is organised as follows. Section 2 briefly presents some basic concepts and algorithms. In § existing filtering techniques are reviewed. In § 4 the use of global smoothers is presented, an iterative procedure is suggested and some requirements on the smoother are established. Section 5 presents the numerical examples. The main points of the paper are summarised in § 6.

2 Preliminaries

This section briefly reviews some basic concepts which will be used in the estimation and use of smoothers.

2.1 Sensitive dependence on initial conditions

The sensitive dependence on initial conditions is one of the most peculiar features of a chaotic system. Such a phenomenon arises due to the local divergence of nearby orbits in the state space. Therefore, a chaotic system is, in a sense, locally unstable but globally stable. The rate of contraction and expansion of the system orbits in state space is measured by the Lyapunov exponents. For a chaotic system at least one such exponent is positive, thus indicating local instabilities and therefore sensitive dependence on initial conditions.

For dissipative systems the overall rate of contraction in all directions in state space is negative. In other words, the summation of all the spectrum of Lyapunov exponents of a dissipative system is less than zero. If a system is chaotic, the orbits in state space will diverge locally in the directions defined by the eigenvectors of the Jacobian of the system, along a given trajectory on the attractor, corresponding to the positive Lyapunov exponents. Conversely, orbits will contract in directions associated with negative Lyapunov exponents. The 'directions' in which nearby orbits diverge and contract can be thought of as stable and unstable 'directions' (or manifolds), respectively.

For a chaotic system the stable and unstable manifolds intersect an infinite number of times. Consequently the dynamics are globally stable although nearby orbits diverge everywhere on the attractor along the unstable manifold. This explains why prediction-based techniques do not perform well in the case of chaotic data, because although noise would be reduced in some directions, it would actually be magnified in others.

2.2 Identification of chaotic systems

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

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

where $e(t)$ accounts for uncertainties, possible noise, unmodelled dynamics, etc. and the delay $d \in \mathbb{Z}^+$ has been included and $F^{\ell}[\cdot]$ is some nonlinear function of $y(t)$, $u(t)$ and $e(t)$

with degree of nonlinearity $\ell \in \mathbb{Z}^+$. Note that if $\ell = 1$ the resulting model is linear.

Clearly, many options are possible for $F^\ell[\cdot]$. In this work $F^\ell[\cdot]$ is assumed to be a polynomial function of $y(t)$, $u(t)$ and $e(t)$. A theoretical justification for using polynomial NARMAX models has been given in (Chen and Billings, 1989).

Equation (1) can be expressed as

$$y(t) = \Psi_{yu}^{\mathrm{T}}(t-1)\Theta_{yu} \; + \; \Psi_{yue}^{\mathrm{T}}(t)\Theta_{yue} \; + \; \Psi_{e}^{\mathrm{T}}(t)\Theta_{e} \tag{2}$$

where the superscript T denotes transposition and $\Psi_{yu}^{\mathrm{T}}(t-1)$ includes all the output and input terms as well as all possible combinations up to degree ℓ and up to time $t-1$. The coefficients of such terms are in the parameter vector Θ_{yu}. The other entries are defined likewise. Equation (2) can be written in the *predictor error* (PE) form

$$\begin{aligned} y(t) &= \left[\Psi_{yu}^{\mathrm{T}}(t-1) \; \Psi_{yu\xi}^{\mathrm{T}}(t-1) \; \Psi_{\xi}^{\mathrm{T}}(t-1)\right] \begin{pmatrix} \hat{\Theta}_{yu} \\ \hat{\Theta}_{yu\xi} \\ \hat{\Theta}_{\xi} \end{pmatrix} \; + \; \xi(t) \\ y(t) &= \Psi^{\mathrm{T}}(t-1)\hat{\Theta} \; + \; \xi(t) \end{aligned} \tag{3}$$

where $\xi(t)$ is the residual at time t and is defined as

$$\xi(t) \doteq y(t) - \hat{y}(t) \tag{4}$$

The following cost function can be defined based on equation 3

$$J_{\mathrm{LS}}(\hat{\Theta}) \doteq \| \, y(t) - \Psi^{\mathrm{T}}(t-1)\hat{\Theta} \, \| \tag{5}$$

where $\| \cdot \|$ is the Euclidean norm. A typical *least squares* (LS) problem is to find $\hat{\Theta}$ such that $J_{LS}(\hat{\Theta})$ is minimised (Chen et al., 1989).

To circumvent numerical problems in equation (3) due to ill-conditioning, orthogonal techniques may be used (Billings et al., 1988).

Effective solutions to handle the problem of determining the structure of nonlinear models are available in the literature (Billings et al., 1988; Haber and Unbehauen, 1990). One solution is based on the *error reduction ratio* (ERR) test which provides an indication of which terms to include in the model. Two advantages of this approach are i) it does not

require the estimation of a complete model to determine the significance of a candidate term and the respective statistical contribution to the output, and ii) the ERR test is derived as a by-product of the orthogonal estimation algorithm. For details see (Billings et al., 1988; Korenberg et al., 1988).

Inspection of equation (5) reveals that the parameter vector is estimated based on the predictive characteristics of the *one-step-ahead* (OSA) predictor $\Psi^{\tau}(t-1)\hat{\Theta}$. Even when the data come from a chaotic system, the aforementioned techniques can still be used because the period of time (one sample) over which predictions are made is very short and therefore the effects of the positive Lyapunov exponents is negligible. On the other hand, when the uncertainty (noise) in the data is larger than a certain critical value, the effect of the extreme sensitivity to initial conditions is such that the accuracy of the OSA predictions deteriorates to the extent that good models cannot be obtained (Aguirre and Billings, 1993a).

2.3 Validation of dynamical models

It has been shown that if a nonlinear model adequately fits a set of data the following conditions should hold (Billings and Voon, 1986)

$$\begin{cases} \Phi_{\xi\xi}(\tau) \doteq \mathrm{E}\{\xi(t-\tau)\xi(t)\} = \delta(0) & \\ \Phi_{u\xi}(\tau) \doteq \mathrm{E}\{u(t-\tau)\xi(t)\} = 0, & \forall\tau \\ \Phi_{u^{2'}\xi}(\tau) \doteq \mathrm{E}\{(u^2(t-\tau)-\overline{u^2(t)})\xi(t)\} = 0, & \forall\tau \\ \Phi_{u^{2'}\xi^2}(\tau) \doteq \mathrm{E}\{(u^2(t-\tau)-\overline{u^2(t)})\xi^2(t)\} = 0, & \forall\tau \\ \Phi_{\xi(\xi u)}(\tau) \doteq \mathrm{E}\{\xi(t)\xi(t-1-\tau)u(t-1-\tau)\} = 0, & \tau \geq 0 \end{cases} \tag{6}$$

where $\delta(0)$ is the Kronecker delta, $\xi(t)$ are the residuals defined in equation (4), $u(t)$ is the input, the over-bar signifies mean value and $\mathrm{E}\{\cdot\}$ denotes mathematical expectation.

If the data are a single time series the model will not have input terms and the adequacy of the system should be assessed using the following correlation functions (Billings and Tao, 1991)

$$\begin{cases} \Phi_{\xi\xi}(\tau) \doteq \mathrm{E}\{(\xi(t)-\overline{\xi(t)})(\xi(t-\tau)-\overline{\xi(t)})\} = \delta(0) \\ \Phi_{\xi\xi^2}(\tau) \doteq \mathrm{E}\{(\xi(t)-\overline{\xi(t)})(\xi(t-\tau)-\overline{\xi(t)})^2\} = 0 \\ \Phi_{\xi^{2'}\xi^{2'}}(\tau) \doteq \mathrm{E}\{(\xi^2(t)-\overline{\xi^2(t)})(\xi^2(t-\tau)-\overline{\xi^2(t)})\} = \delta(0) \end{cases} \tag{7}$$

For sufficiently long data sets the 95 % confidence bands are approximately $\pm 1.96\, N^{-0.5}$ and any significant correlation will be indicated by one or more points of the function lying outside the respective confidence band. Thus if the correlation functions of a model fall within the respective confidence bands, such a model is said to be unbiased in the sense that no dynamics (linear or nonlinear) have been left in the residuals.

Correlation tests can be readily computed from the data and the residuals but such tests only provide *statistical* information. Thus a model which is statistically valid may turn out to be *dynamically* inappropriate, or in other words, the correlation tests do not necessarily indicate if a particular model has captured and indeed is able to reproduce dynamical invariants of the original system. In order to validate the estimated models dynamically, it is necessary to compare dynamical invariants developed for the analysis of nonlinear dynamics such as Poincaré maps, Lyapunov exponents and fractal dimensions (Aguirre and Billings, 1994; Principe et al., 1992). In this paper the largest Lyapunov exponent, λ_1 (Wolf, 1986), and the correlation dimension, D_c (Grassberger and Procaccia, 1983), were used as quantitative invariants to validate the models.

3 Filtering and noise reduction — existing techniques

Throughout this paper it is assumed that the noise is purely additive, or in other words the noise is entirely observational. This has become a standard procedure in the literature (Casdagli et al., 1991; Grassberger et al., 1991) because "while there are situations where, say, parametric or nonlinear fluctuation coupling are appropriate, experience has shown that the additive form is adequate for most modeling purposes" (Crutchfield and McNamara, 1987).

Given a chaotic time series $x(t)$, it is desired to filter the measured data $y(t)=x(t)+e(t)$, where $e(t)$ is the additive noise, in order to recover $x(t)$. This is useful in 'cleaning' Poincaré sections and embedded attractors which have been blurred by noise.

Another aspect of this problem is to find a 'noise-reduced' orbit $\bar{y}(t)$ from which invariants such as λ_1, D_c and the attractor geometry can be more accurately estimated than if the noisy data $y(t)$ were used. This is sometimes referred to as *statistical noise reduction* as opposed to recovering $x(t)$ from $y(t)$ which has been called *detailed noise reduction* (Farmer and

Sidorowich, 1991). In this paper, the objective is to be able to identify dynamically valid models from $\bar{y}(t)$.

Filtering based on model predicted outputs, whilst reducing the noise content in the data, will not guarantee that $\bar{y}(t)$ remains close to $y(t)$ (and ultimately close to $x(t)$) if the latter is chaotic. Thus, to ensure that $\bar{y}(t)$ remains close to $y(t)$, the following cost function can be used

$$J_{\mathrm{NR}} = \sum_{k=1}^{N} \{J_1 [\bar{y}(k) - g_k(\bar{y}(k-1))] \; + \; J_2 [\bar{y}(k) - y(k)]\} \tag{8}$$

where N is the number of points in the data, $J_1[\cdot]$ and $J_2[\cdot]$ indicate functions which are usually metric norms and $g_k(\cdot)$ are linear maps which describe the dynamics in a neighbourhood of a point on the true orbit. Clearly $J_1[\cdot]$ penalizes deviations from the true deterministic dynamics described by $g_k(\cdot)$ while $J_2[\cdot]$ guarantees that the cleaned orbit remains close to the measured orbit.

In particular, the following cost functions have been used (Kostelich and Yorke, 1988; Kostelich and Yorke, 1990)

$$J_1[\cdot] = \| \bar{y}(k) - g_k(\bar{y}(k-1)) \|^2 + \| \bar{y}(k+1) - g_k(\bar{y}(k)) \|^2 \tag{9}$$

where $\| \cdot \|$ is the Euclidean norm, and

$$J_2[\cdot] = \| \bar{y}(k) - y(k) \|^2 \tag{10}$$

Another option is to choose $J_2[\cdot]$ as above and

$$J_1[\cdot] = 2 \| g_k(\bar{y}(k)) - \bar{y}(k+1) \|^{\mathrm{T}} \; \mu_k \tag{11}$$

where μ_k are Lagrange multipliers (Farmer and Sidorowich, 1991). The minimization of J_{NR} is a typically ill-conditioned problem, especially for chaotic time series (Farmer and Sidorowich, 1991). Some improvement in the numerical conditioning however can be attained at the expense of performance (Kostelich and Yorke, 1988).

In the field of system identification, improving the *signal to noise ratio* (SNR) is also of interest because this facilitates both the unbiased estimation of the parameter vector and the correct determination of the model structure. The chief idea is to estimate the noise-free

data and then use this estimate to perform the parameter estimation. A way of doing this is to use the following predictor which can be derived from equation (3)

$$\hat{y}(t) = \Psi^{\mathrm{T}}_{\hat{y}u}(t-1)\hat{\Theta}_{yu} \tag{12}$$

It should be realised that in the last equation the parameter vector $\hat{\Theta}_{yu}$ was estimated from the original noisy data as is indicated by the absence of the hat on the subscript y. On the other hand, the matrix $\Psi^{\mathrm{T}}_{\hat{y}u}(t-1)$ was formed using predicted values of the data, that is $\hat{y}(t)$ up to and including time $t-1$. Because $\hat{y}(t)$ is an estimate of $x(t)$, equation (12) can be used in suboptimal parameter estimation schemes (Billings and Voon, 1984).

However, if the data were chaotic after a few iterations $\hat{y}(t)$ would not be an accurate estimate of $x(t)$ because of the sensitive dependence on initial conditions. Therefore the use of $\hat{y}(t)$ in suboptimal schemes seems somewhat restricted for chaotic systems.

In order to measure the effectiveness of filtering schemes the following quantities have been used in the literature (Kostelich and Yorke, 1988)

$$E_{\mathrm{dyn}} = \sqrt{\frac{1}{N}\sum_{k=n_y}^{N}\xi^2(k)} \tag{13}$$

and

$$E_{\mathrm{true}} = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(x(k)-y(k))^2} \tag{14}$$

E_{dyn} measures the distance from determinacy, that is, how far is the noise-reduced data from being purely deterministic. The second index, E_{true}, measures the distance from the original noise-free orbit, $x(t)$. Clearly, this index is of theoretical interest only since in practice $x(t)$ will not be available. The performance of the filtering scheme can then be expressed as a ratio of the error before and after noise reduction. The indices in equations (13) and (14) will be used in the examples of § 5.

3.1 The resetting filter

The following predictor has been suggested to overcome some of the difficulties associated with the filtering of chaotic data (Aguirre and Billings, 1993a)

$$\hat{y}(t) = \Psi^{\mathrm{T}}_{yu}(t-1)\,\hat{\Theta}_{yu} + \Psi^{\mathrm{T}}_{\xi}(t-1)\,\hat{\Theta}_{\xi} \tag{15}$$

It should be noted that in this case $\hat{y}(t)$ is predicted based on previous values of the measured data $y(s)$, $s \leq t-1$, and not based on previously predicted values such as in equation (12). Moreover, since this predictor is used to predict only one step into the future, the predicted value $\hat{y}(t)$ is, in most cases, guaranteed to remain close to the data $y(t)$. This can be interpreted as being a consequence of the *resetting* effect achieved by using measured data to initialise the predictor at each step. The predictor in equation (15) will be referred to as the *resetting filter* (RF) and it is adequate for filtering chaotic signals.

The qualitative effect attained by the resetting filter is, in some respects, analogous to other methods. This can be verified by considering equation (8). It is worth noticing that the resetting effect of the RF guarantees that $J_2[\cdot]$ is kept small. Moreover, the parameter vector of the RF is obtained by minimising J_{LS} in equation (5), which is clearly analogous to $J_1[\cdot]$ in equation (8). The main difference is that whilst $g_k(\cdot)$ usually represents local linear maps, $\Psi^{\text{T}}(t-1)\,\hat{\Theta}$ is a global nonlinear map which may include inputs and residual in addition to output terms. The use of global models for signal processing and modelling of chaotic systems has recently attracted some attention (Aguirre and Billings, 1993a; Holzfuss and Kadtke, 1993; Kadtke et al., 1993).

Predictor-based filtering for chaotic systems will not work in general because of the inability of making long-term accurate predictions along the unstable manifold. Therefore in such directions, the filter would actually amplify the noise (Schreiber and Grassberger, 1991). The same is valid for the RF, but to a much lesser extent because of the resetting effect which will guarantee that any noise amplification along the unstable manifold is kept to a minimum. However, if several passes through the data are required to attain the desired level of noise reduction, it is inevitable that the effect of positive Lyapunov exponents be manifest. Consequently, the filtered data may not resemble the original sequence and, in fact, might have a greater noise content than the raw data.

In the next section the use of a global nonlinear smoothers is investigated. The objective is to develop an algorithm suitable for reducing the noise even if a large number of passes through the data is required.

4 Data Smoothing

This section describes the structure of the global nonlinear smoother and an iterative procedure which can be followed to achieve noise reduction. The last part of this section discusses how to verify that the suggested noise reduction procedure is statistically sound.

4.1 Global nonlinear smoothers

The difficulty with the resetting filter in equation (15) is that it only uses past information to predict the future. However, the dynamics can only be predicted with any certainty as $t \rightarrow \infty$ along the stable manifold. Conversely, the dynamics can only be 'predicted' along the unstable manifold in reverse time, that is as $t \rightarrow -\infty$ (Schreiber and Grassberger, 1991). In other words, in order to estimate $y(t)$, future information as well as past information is required (Schreiber and Grassberger, 1991).

This motivates the search for NARMAX smoothers of the form

$$\begin{aligned} y(t) \;=\; F_s^{\ell}[\, & y(t-n_y),\ldots,y(t-1),y(t+1),\ldots,y(t+n_y), \\ & u(t-d-n_u+1),\ldots,u(t-d),u(t+d),\ldots,u(t+d+n_u-1), \\ & \xi(t-1),\ldots,\xi(t-n_e)\,] + \xi(t) \end{aligned} \tag{16}$$

It will be helpful to express the last equation as

$$y(t) = F_s^{\ell}\left[y(t_-^+),u(t_-^+),\xi(t-1)\right] + \xi(t) \tag{17}$$

where $y(t_-^+)$ and $u(t_-^+)$ indicate terms which relate to the past *and* future, with respect to time t, $\xi(t-1)$ represents the residuals up to time $t-1$ and $\xi(t)$ is the residual at time t.

It should be noted that equations analogous to (3)-(5) can be derived for the smoother in equation (16). Moreover, the ERR criterion, used to select the most important terms to compose a NARMAX model, can also be used to select the structure of the the smoother and the same least squares algorithm can be used to estimate the parameters.

From a dynamical point of view, the smoother in equation (16) will also succeed in predicting the dynamics along the unstable directions because it contains terms which relate to the future. Such terms will enable 'predicting' in reverse time since orbits converge along the unstable manifold as $T \rightarrow -\infty$.

4.2 The iterative procedure

As mentioned before, neither the structure nor the parameters of F_s^ℓ are assumed known. Therefore, it is necessary to select the structure and estimate the parameters of the smoother before reducing noise. Thus, each noise reduction iteration comprises two steps. In the first step a smoother is fitted to the data and in the second step the resulting map is operated on the data in order to reduce noise. Because noise will not be amplified along the unstable manifold, this two-step procedure can be repeated as many times as necessary, beginning with the set of data smoothed in the preceding step.

This procedure can be represented as follows. Let $y^{(0)}(t) = y(t)$ be the original set of data. The index in parenthesis indicates the iteration in which the data were generated. Thus, in the first iteration a map is fitted to the raw data

$$y^{(0)}(t) = F_s^{(0)}\left[y^{(0)}(t_-^+), u(t_-^+), \xi^{(0)}(t-1)\right] + \xi^{(0)}(t) \tag{18}$$

It should be noted that $\xi^{(0)}(t)$ are the residuals obtained during the parameter estimation performed for $F_s^{(0)}$. The data are then smoothed using $F_s^{(0)}$ to obtain the first sequence of noise-reduced data

$$y^{(1)}(t) = F_s^{(0)}\left[y^{(0)}(t_-^+), u(t_-^+), \xi^{(0)}(t-1)\right] \tag{19}$$

This completes the first iteration. The second iteration begins by fitting a new smoother to the data $y^{(1)}(t)$, thus

$$y^{(1)}(t) = F_s^{(1)}\left[y^{(1)}(t_-^+), u(t_-^+), \xi^{(1)}(t-1)\right] + \xi^{(1)}(t) \tag{20}$$

Conversely, the new reduced-noise data are

$$y^{(2)}(t) = F_s^{(1)}\left[y^{(1)}(t_-^+), u(t_-^+), \xi^{(1)}(t-1)\right] \tag{21}$$

and so on.

4.3 Requirements on the smoother

Equations (19) and (21) reveal that the smoothed data are obtained by dropping the last terms in equations (18) and (20), respectively. In other words, the above procedure for data smoothing achieves noise reduction by leaving out whatever cannot be explained in the data. This, of course, assumes that the raw data is composed by a 'predictable' part (the clean orbit) and by an 'unpredictable' component (the noise). It should be noted that in this context the terms 'predictable' and 'unpredictable' should be understood in a statistical sense and should not be confused with the unpredictability arising in purely deterministic systems due to sensitive dependence on initial conditions. The latter type of unpredictability is being handled by the smoother terms which convey future information whereas the former is precisely the noise which is to be attenuated.

Assume that it is desired to reduce the noise by passing through the data only once, as is usually the case with the resetting filter. Thus, the final data are $y^{(1)}(t)$ and should be free from any kind of bias. From equation (19) it becomes clear that this is equivalent to impose that the smoother $F_s^{(0)}$ be unbiased. It has been shown that this can be verified if the corresponding residual sequence, $\xi^{(0)}(t)$, satisfies the set of correlation tests presented in § 2.3 (Billings and Voon, 1986; Billings and Tao, 1991).

In this paper it is assumed that noise can be reduced by successive passes through the data. Hence, suppose the data $y^{(N_p)}(t)$ was obtained after N_p passes as described in § 4.3. Then,

$$\begin{aligned}
y^{(0)}(t) &= F_s^{(0)}\left[Y^{(0)}\right] + \xi^{(0)}(t) \\
y^{(1)}(t) &= F_s^{(1)}\left[F_s^{(0)}\left[Y^{(0)}\right], \bullet\right] + \xi^{(1)}(t) \\
&\vdots \quad \vdots \qquad \vdots \quad \vdots \quad \vdots \\
y^{(N_p-1)}(t) &= F_s^{(N_p-1)}\left[\cdots F_s^{(1)}\left[F_s^{(0)}\left[Y^{(0)}\right], \bullet\right], \cdots \bullet\right] + \xi^{(N_p-1)}(t)
\end{aligned} \tag{22}$$

where $[Y^{(0)}] = [y^{(0)}(t_-^+), u(t_-^+), \xi^{(0)}(t-1)]$, and the $\bullet$ represents $[u(t_-^+), \xi^{(i)}(t-1)]$. It should be noted that if the sequence $\xi^{(N_p-1)}(t)$ passes the correlation tests, the only thing that can be inferred is that in the last iteration of the noise reduction procedure no dynamics have been left unmodelled. However, this test would not reveal if any dynamics were left unmodelled in previous iterations. To ensure that all the iterations are taken into account, the entire

procedure has to be referred to the original data, $y^{(0)}(t)$. This can be done by expressing the data $y^{(N_p)}(t)$ in terms of $y^{(0)}(t)$. Note that

$$\begin{aligned} y^{(N_p-1)}(t) &= y^{(N_p-2)}(t) - \xi^{(N_p-2)}(t) \\ y^{(N_p-1)}(t) &= y^{(N_p-3)}(t) - \xi^{(N_p-3)}(t) - \xi^{(N_p-2)}(t) \\ &\vdots \quad \vdots \qquad \vdots \quad \vdots \quad \vdots \\ y^{(N_p-1)}(t) &= y^{(0)}(t) - \sum_{i=0}^{N_p-2} \xi^{(i)}(t) \end{aligned} \tag{23}$$

thus substituting the last equation in (23) in the last expression in (22) yields

$$y^{(0)}(t) = F_s^{(N_p-1)}\left[\cdots F_s^{(1)}\left[F_s^{(0)}\left[Y^{(0)}\right]\bullet,\right]\cdots\bullet,\right] + \sum_{i=0}^{N_p-1} \xi^{(i)}(t) \tag{24}$$

The last equation has been obtained by considering the iterative procedure. However, such an equation can be interpreted as representing a one-iteration noise reduction procedure using the smoother $F_s^{(N_p-1)}\left[\cdots F_s^{(1)}\left[F_s^{(0)}[\cdot],\bullet\right],\cdots\bullet\right]$ on the data $\left[\cdot Y^{(0)}\right]$. Therefore, such a 'composite smoother' should be unbiased in order to guarantee that only the inexplainable part of the data has been attenuated.

It should be noted that equation (24) is in the same general format as the equations for which the correlation tests in § 2.3 were developed (Billings and Voon, 1986; Billings and Tao, 1991). Therefore, the entire smoothing procedure will be statistically sound (unbiased) if

$$\xi_s(t) = \sum_{i=0}^{N_p-1} \xi^{(i)}(t) \tag{25}$$

passes the correlation tests. It is stressed that $\xi_s(t)$ can be computed very easily. To see this, note that the final smoothed data sequence is

$$y^{(N_p)}(t) = F_s^{(N_p-1)}\left[\cdots F_s^{(1)}\left[F_s^{(0)}\left[Y^{(0)}\right],\bullet\right],\cdots\bullet\right] \tag{26}$$

Substituting the last equation and (25) in (24) yields

$$y^{(0)}(t) = y^{(N_p)}(t) + \xi_s(t) \tag{27}$$

Thus $\xi_s(t)$ can be obtained by subtracting the final smoothed data from the original sequence. Equations (25) and (27) also reveal that the raw data can be viewed as the composition of the cleaned data and the residuals eliminated in each iteration.

The structure of a nonlinear model may heavily influence the dynamics (Aguirre and Billings, 1993b) but the structure of the resetting filter is comparatively unimportant as long as the data remain unbiased (Aguirre and Billings, 1993a). This is also valid for the global smoothers. Therefore, as long as $\xi_s(t)$ in equation (25) passes the respective correlation tests in § 2.3, there are no severe restrictions on the structure of the smoothers. It is stressed that this will *not* be the case for the identified models.

There is, however, one important difference concerning the structures of a resetting filter and that of a smoother. Because a resetting filter cannot be used to reduce the noise in the data via a large number of iterations, the noise terms of the resetting filter play an important role. Thus if the respective noise model is deficient, the filtered data will be biased and that would be indicated by the correlation tests. On the other hand, the entire noise model in the smoothers can be dispensed with because although some noise will certainly not be removed from the data in the first iterations, after many passes however the remaining noise will eventually be reduced to acceptable levels. But in any case the final $\xi_s(t)$ must pass the correlation tests.

5 Numerical results

The procedure outlined in the preceding section is illustrated by means of three examples. The first example considers the model of the well-known Chua's circuit operating in the double-scroll regime. In the second example, the data were generated using the Mackey-Glass delay equation, which was originally suggested as a model for the growth of blood cells in patients with leukemia. Finally, the third example uses the Duffing-Holmes equation which serves as a model of mechanical vibrations in two-well-potential problems.

5.1 The double-scroll attractor

The normalised equations of Chua's circuit can be written as (Chua et al., 1986)

$$\begin{cases} \dot{x} = \alpha(y - h(x)) \\ \dot{y} = x - y + z \\ \dot{z} = -\beta y \end{cases} \qquad h(x) = \begin{cases} m_1 x + (m_0 - m_1) & x \geq 1 \\ m_0 x & |\, x \,| \leq 1 \\ m_1 x - (m_0 - m_1) & x \leq -1 \end{cases} \tag{28}$$

In what follows $m_0 = -1/7$ and $m_1 = 2/7$. Varying the parameters α and β the circuit displays several regular and chaotic regimes. The well known double scroll attractor, for instance, is obtained for $\alpha = 9$ and $\beta = 100/7$ and has the largest Lyapunov exponent and the Lyapunov dimension equal to $\lambda_1 = 0.23$ and $D_L = 2.13$, respectively (Matsumoto et al., 1985).

White noise with variance $\sigma_e^2 = 0.021$ was added to a data sequence with 1750 points corresponding to the z component sampled at $T_s = 0.15$. The resulting data had a *signal to noise ratio* (SNR) equal to $20\log(2.62/0.021) \approx 42$ dB and is shown in figure 1.

The techniques mentioned in § 2.2 were then used to identify NARMA polynomial models from the data. The parameters used were $\ell = 4$, $n_y = 5$, $n_u = 0$ and $n_e = 20$. The total number of candidate process terms was 125 and the most significant were chosen based on the ERR test. Several different models were estimated by varying the number of terms allowed in the model but no model was able to learn and therefore reproduce the dynamical invariants of the circuit.

In order to overcome this difficulty, the resetting filter (see § 3.1) was successfully used on an oversampled ($T_s = 0.015$) realisation of the raw data. The raw data were oversampled in order to enhance the performance of the resetting filter. This was necessary because no good model was identified (from the filtered data) when the raw data was not oversampled. In many situations, however, it will not be possible to perform the filtering on oversampled data sets.

Finally, the nonlinear smoothing procedure suggested in § 4 was used to reduce the noise in the original raw data, that is, the data which were *not* oversampled. The data obtained after 20 iterations ($N_p = 20$) are shown in figure 2 and were subsequently used in the identification. Using the same parameters as before, that is, $\ell = 4$, $n_y = 5$, $n_u = 0$ and $n_e = 20$ the following model was identified

$$
\begin{aligned}
z(k) &= 0.42909\times 10z(k-1) - 0.75174\times 10z(k-2) + 0.67768\times 10z(k-3) \\
&- 0.31722\times 10z(k-4) + 0.63239z(k-5) - 0.11485\times 10^{-1}z(k-5)^3 \\
&+ 0.40830\times 10^{-1}z(k-1)z(k-5)^2 + 0.13464\times 10z(k-1)^2z(k-2) - 0.49701z(k-3)^2z(k-5) \\
&- 0.40878z(k-1)^3 - 0.16186z(k-3)z(k-5)^2 - 0.40198z(k-1)z(k-4)^2 \\
&+ 0.61566z(k-1)z(k-3)z(k-4) + 0.82185z(k-3)z(k-4)z(k-5)
\end{aligned}
$$

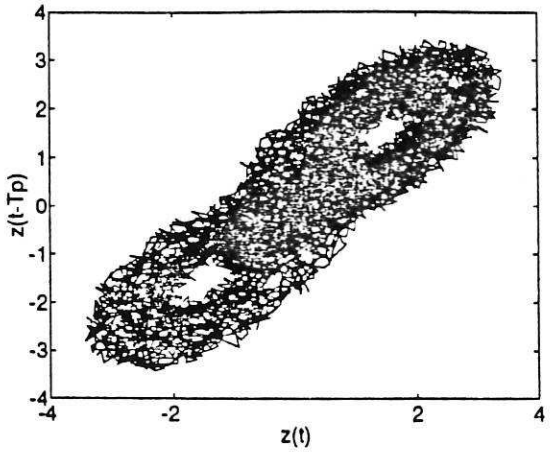

Figure 1: Embedded trajectory for the double scroll. Noisy data

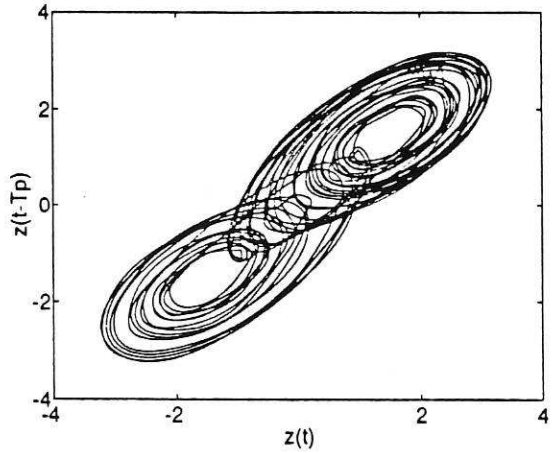

Figure 2: Embedded trajectory for the double scroll. Smoothed data.

$$
\begin{aligned}
- \quad & 0.66712\times 10^{-1} z(k-2)z(k-4)z(k-5) - 0.21150 z(k-4)^3 - 0.92219 z(k-1)z(k-2)^2 \\
+ \quad & 0.13481 z(k-2)^2 z(k-4) - 0.44309 z(k-1)^2 z(k-3) + 0.16095 z(k-2)^3 \\
+ \quad & \Psi_\xi^T(k-1)\hat{\Theta}_\xi + \xi(k)
\end{aligned} \tag{29}
$$

with $\sigma_\xi^2 = 0.411\times 10^{-5}$. For this model $\lambda_1 = 0.21$ and $D_c = 2.04 \pm 0.013$. It is noted that the Lyapunov dimension D_L serves as an upper bound for the correlation dimension D_c, thus this model satisfactorily reproduces the invariants of the double scroll attractor.

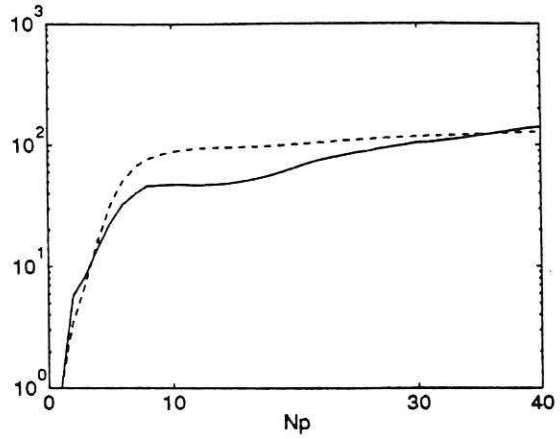

Figure 3: $E_{\rm dyn}(1)/E_{\rm dyn}(N_{\rm p})$ as a function of $N_{\rm p}$. (—) Linear smoother, (- -) nonlinear smoother.

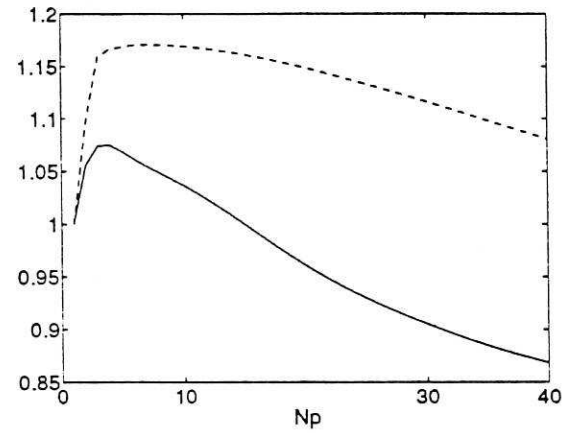

Figure 4: $E_{\rm true}(1)/E_{\rm true}(N_{\rm p})$ as a function of $N_{\rm p}$. (—) Linear smoother, (- -) nonlinear smoother.

5.2 The Mackey-Glass model

Consider the delay system (Mackey and Glass, 1977)

$$\dot{y}(t) = \frac{a\, y(t-\tau)}{1+y^c(t-\tau)} - b\, y(t) \tag{30}$$

where τ is the delay. For the following parameter values $a=0.2$, $b=0.1$, $c=10$ and $\tau=30$ equation (30) settles to a chaotic attractor with the Lyapunov spectrum $\lambda_1=0.0074\pm0.0007$, $\lambda_2=0.0038\pm0.0007$, $\lambda_3=0.0015\pm0.0008$ and $\lambda_4=-0.017\pm0.003$ (Sano and Sawada, 1985) and Lyapunov dimension (Kaplan and Yorke, 1979) $D_{\rm L}=3.64$.

Equation (30) was simulated to generate 700 data points sampled at $T_{\rm s}=1$. Zero-mean Gaussian noise with variance $\sigma_e^2=9\times10^{-4}$ was added to the data. The resulting records with

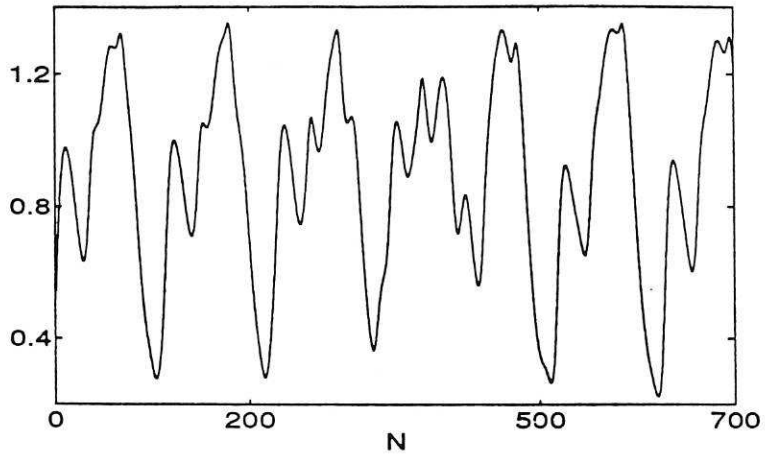

Figure 5: Noise-free original time series for the Mackey-Glass model

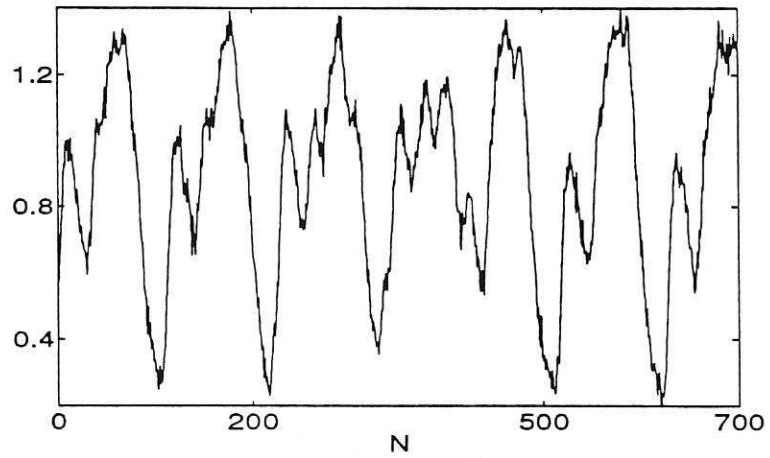

Figure 6: Noisy (raw) time series for the Mackey-Glass model

SNR=40.2 dB were smoothed following the procedure outlined in § 4. The performance of the filtering procedure is shown in figures 3–4 and suggests $N_p = 10$.

Figures 5–7 respectively show the noise-free, noisy and smoothed data for the Mackey-Glass model and figure 8 shows the resulting data filtered following the RF approach. The performance of predictor-based and smoother-based approaches is compared in figure 9–10, where it becomes clear that, for predictors, the 'distance' between the raw data and the resulting sequence deteriorates as the number of passes increases.

In order to enable proper structure detection, the smoothed data in figure 7 were decimated in such a way that for the final time series $N = 234$ and $T_s = 3$ s. From such data the following model was estimated

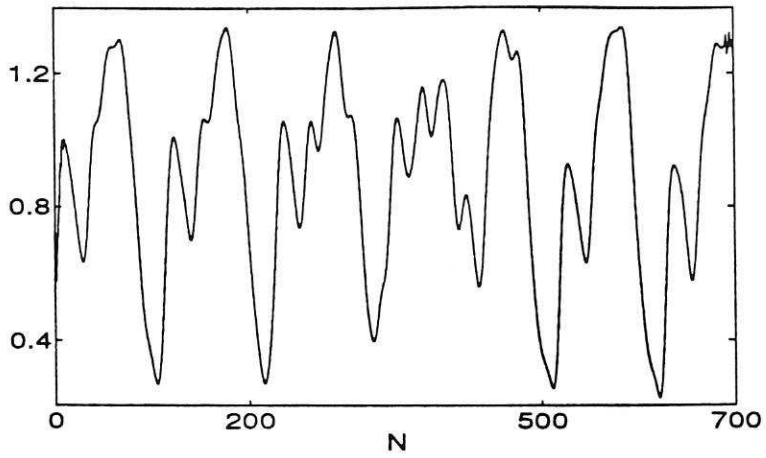

Figure 7: Smoothed time series for the Mackey-Glass model. Terms with both negative and positive lags were used to compose the smoother, $N_p = 10$.

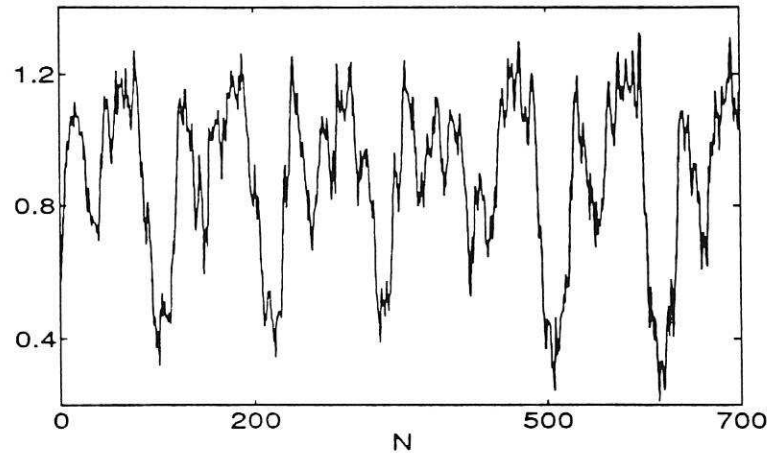

Figure 8: Filtered time series for the Mackey-Glass model. Only terms with negative lags were used to compose the resetting filter, $N_p = 10$.

$$
\begin{aligned}
y(k) &= 0.27014\times 10\, y(k-1) - 0.34077\times 10\, y(k-2) + 0.25289\times 10 y(k-3) - 1.0504\, y(k-4) \\
&+ 0.16879\, y(k-5) - 0.44936\times 10^{-1} y(k-8)^4 - 0.65114\times 10^{-1} y(k-8) \\
&+ 0.33689\, y(k-8) y(k-9)^2 - 0.42007\times 10^{-1} y(k-9) - 0.12001\times 10\, y(k-9)^3 \\
&+ 0.34626\, y(k-9)^4 + 0.64559\, y(k-9)^2 + \Psi_\xi^T(k-1)\hat{\Theta}_\xi + \xi(k) \qquad (31)
\end{aligned}
$$

with $\ell = 4$, $n_y = 9$, $n_u = 0$, $n_e = 20$ and $\sigma_\xi^2 = 0.120 \times 10^{-3}$. For this model $\lambda_1 = 0.0051$, $\lambda_2 = 0.0005$, $\lambda_3 = -0.0011$ and $\lambda_4 = -0.0102$ and $D_L = 3.43$. This model presents dynamics

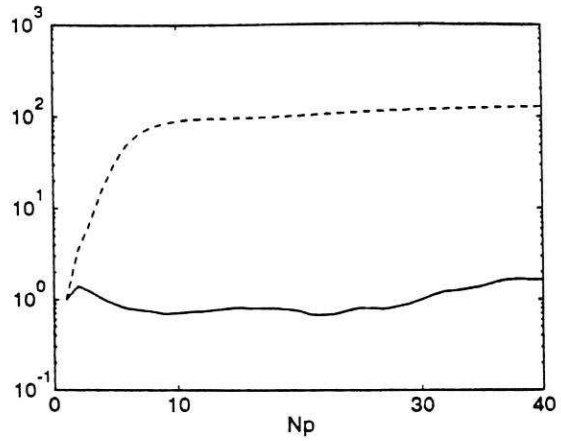

Figure 9: $E_{\rm dyn}(1)/E_{\rm dyn}(N_{\rm p})$ as a function of $N_{\rm p}$. (—) Prediction-based noise reduction, (- -) smoother-based noise reduction.

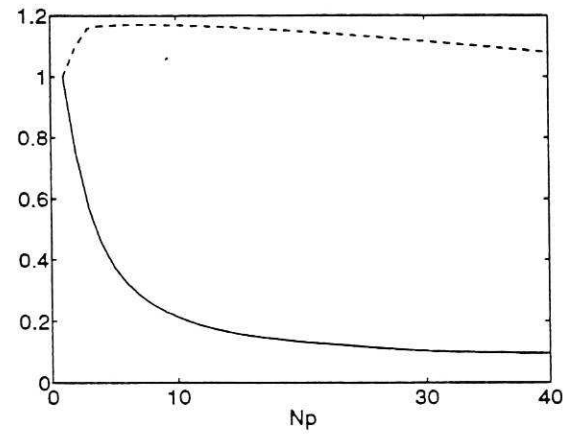

Figure 10: $E_{\rm true}(1)/E_{\rm true}(N_{\rm p})$ as a function of $N_{\rm p}$. (—) Prediction-based noise reduction, (- -) smoother-based noise reduction.

which are similar to those of the original system. No such approximation has been achieved using the raw data to carry out the identification.

Figures 11a–c shows the correlation tests for the smoothed data. As can be seen, the residual sequence $\xi_{\rm s}(t)$ is white with 95% confidence. This gives an indication that no detectable dynamics have been eliminated in the noise-reduction procedure.

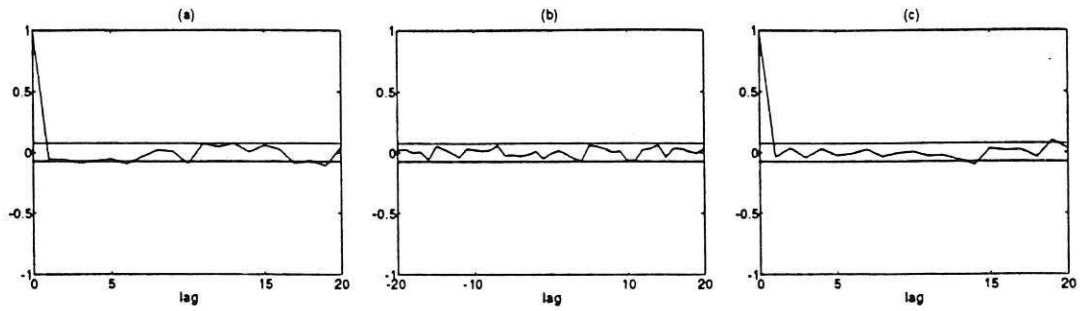

Figure 11: Correlation tests of the residuals ξ_s (a) $\Phi_{\xi_s\xi_s}(\tau)$, (b) $\Phi_{\xi_s\xi_s^2}(\tau)$ and (c) $\Phi_{\xi_s^2\xi_s^2}(\tau)$.

5.3 The Duffing-Holmes oscillator

In the field of nonlinear dynamics Duffing's equation has become a standard benchmark. This example considers the Duffing-Holmes equation (Holmes, 1979)

$$\ddot{y} + \delta\dot{y} - \beta y + y^3 = A\cos(\omega t) \tag{32}$$

A Poincaré section of this system for $\delta = 0.15$, $\beta = 1$, $A = 0.30$ and $\omega = 1\,\text{rad/s}$ is shown in figure 12. The following invariants for this attractor were estimated $\lambda_1 = 0.20$ and $D_c = 2.40 \pm 0.019$.

Equation (32) was used to generate data which were sampled at $T_s = \pi/15$. White Gaussian noise with variance $\sigma_e^2 = 0.01$ was added to 2000 data points. A Poincaré section taken from data with such a level of noise is shown in figure 13. The resulting data sequence, with SNR=37.6 dB, was subsequently used to identify NARMAX models with 20 linear noise terms and $\ell = 3$ and $n_y = n_u = 3$. None of the identified models succeeded in reproducing the dynamical features of the original attractor.

The procedure outlined in § 4.2 was used to smooth the data. The structure of the smoothers was defined by $\ell = 3$, $n_y = n_u = 10$ and $n_e = 0$. Moreover, the smoothers had 20 terms, ten with negative lags and ten with positive lags. Further, twenty passes through the data were performed, that is, $N_p = 20$.

Using $\ell = 3$ and $n_y = n_u = 3$ as before, the following model was identified from the smoothed data sequence $y^{(20)}(t)$

$$\begin{aligned} y(k) &= 0.12764\times 10 y(k-1) + 0.48698 y(k-2) - 0.14994\times 10^{-1} y(k-1)^3 \\ &+ 0.73064\times 10^{-1} u(k-1) - 0.69537 y(k-3) \end{aligned}$$

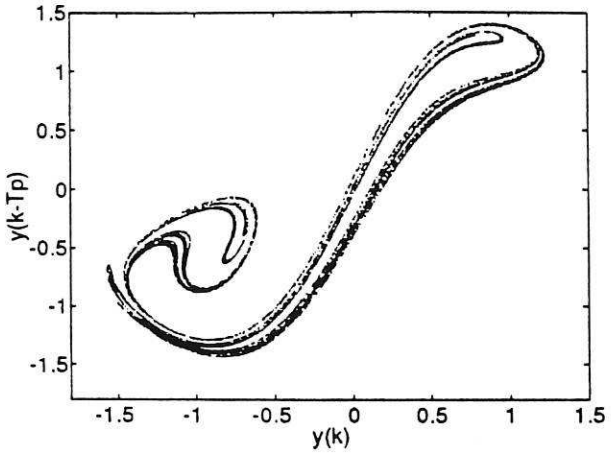

Figure 12: Poincaré section of the Duffing-Holmes attractor.

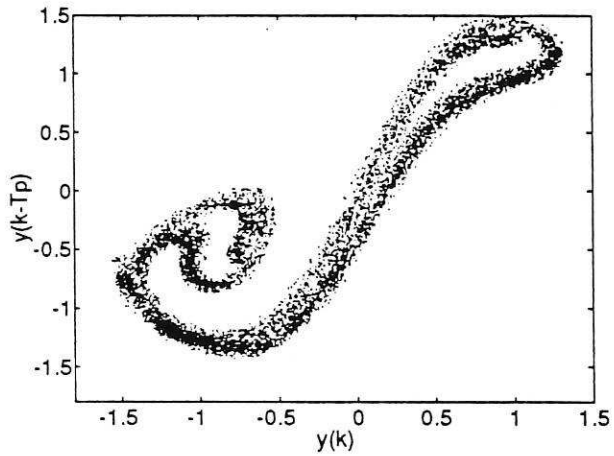

Figure 13: Poincaré section of the Duffing-Holmes attractor corrupted with noise.

$$- \quad 0.56888\times 10^{-1}y(k-1)^2y(k-3)+\Psi_\xi^T(k-1)\hat{\Theta}_\xi+\xi(k) \tag{33}$$

with $\sigma_\xi^2 = 0.149\times 10^{-3}$. For this model $\lambda_1 = 0.18$ and $D_c = 2.45 \pm 0.108$. A Poincaré section of this model is shown in figure 14. Clearly, the model in equation (33) adequately reproduces the dynamics of the strange attractor both qualitatively and quantitatively.

Figures 15–16 show the performance of the noise-reduction procedure. Such diagrams also include the performance of linear smoothers which were estimated and used in the same way as the nonlinear counterparts. The only difference was that in the former only linear terms of the form $y(k-i)$, $i=-10,\ldots,10$ were allowed to compose the smoother. Such figures reveal that in this example the nonlinear smoothers outperform the linear counterparts. In

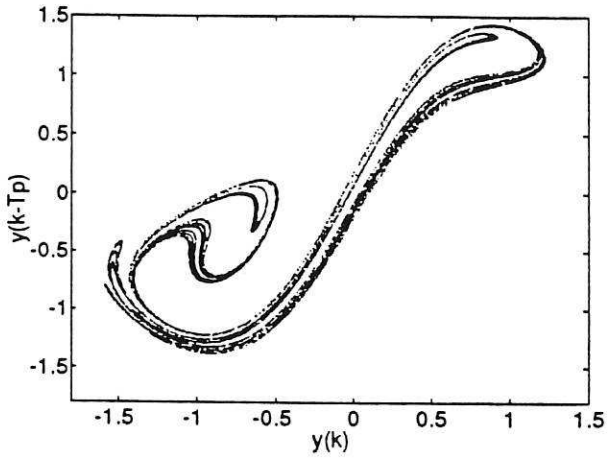

Figure 14: Poincaré section for the identified model of the Duffing-Holmes oscillator.

particular it can be seen that for a large number of noise-reduction iterations ($N_p > 35$), the linear smoothers seem to reduce more noise than the nonlinear smoothers. However, such a gain is offset by the fact that the 'distance' between the noise-reduced data obtained with the linear smoothers and the clean orbit tends to deteriorate slightly for the linear smoothers as N_p is increased.

Finally, equation (32) was used to generate non-chaotic data by taking $A = 0.22$ and the remaining parameters as before. Input-output data with $N = 2000$, $T_s = \pi/15$ and SNR=9.8 dB were used as raw data. Figures 17–18 compare the performance of nonlinear smoothers and resetting filters on such data. Clearly, the resetting filter outperforms the smoother when the data are regular. Moreover, figure 18 reveals that the filtered data obtained with the resetting filter (see § 3.1) does not diverge from the raw data as the number of passes increases (compare with figure 10). This is because such data are non-chaotic and therefore do not have inherent local instabilities which would provoke sensitive dependence on initial conditions and thereby would preclude the use of prediction-based filtering techniques.

6 Conclusions

A procedure for reducing the noise in chaotic data has been presented. Such data are characterised by local instabilities and arise in many real systems over a wide range of operating conditions. It has been pointed out that prediction-based filtering techniques are not ap-

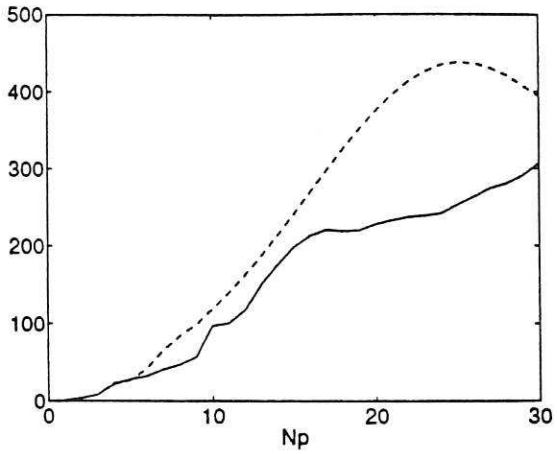

Figure 15: $E_{\mathrm{dyn}}(1)/E_{\mathrm{dyn}}(N_{\mathrm{p}})$ as a function of N_{p}. (—) Linear smoother, (- -) nonlinear smoother.

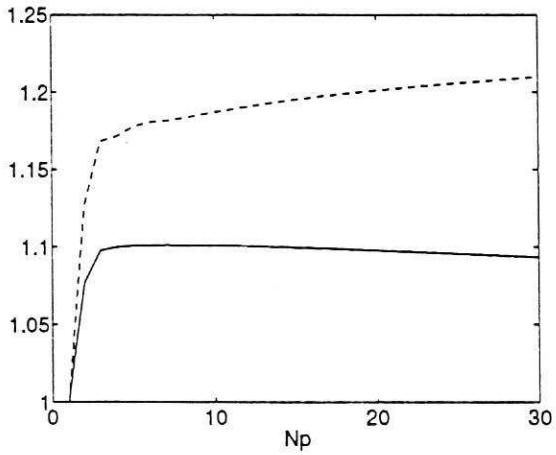

Figure 16: $E_{\mathrm{true}}(1)/E_{\mathrm{true}}(N_{\mathrm{p}})$ as a function of N_{p}. (—) Linear smoother, (- -) nonlinear smoother.

propriate for filtering this kind of data following iterative procedures which require several passes through the data in order to achieve the desired noise reduction. This is because although such schemes would reduce noise in some directions in state space, noise would actually be amplified in the 'directions' associated with the positive Lyapunov exponents of the system.

The use of global linear and nonlinear smoothers has been investigated. Because the smoothers contain terms related to the future, good 'prediction' can be attained along the 'directions' related to positive Lyapunov exponents. Consequently, greater noise reduction can be achieved by successive passes through the data.

The smoothers are estimated directly from the data and no initial knowledge is assumed.

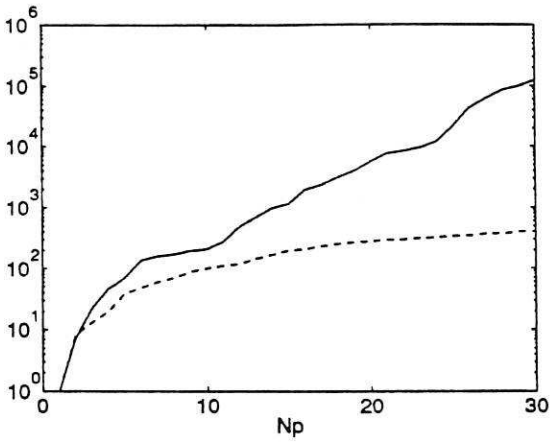

Figure 17: $E_{\mathrm{dyn}}(1)/E_{\mathrm{dyn}}(N_{\mathrm{p}})$ as a function of N_{p}. (—) Prediction-based noise reduction, (- -) smoother-based noise reduction.

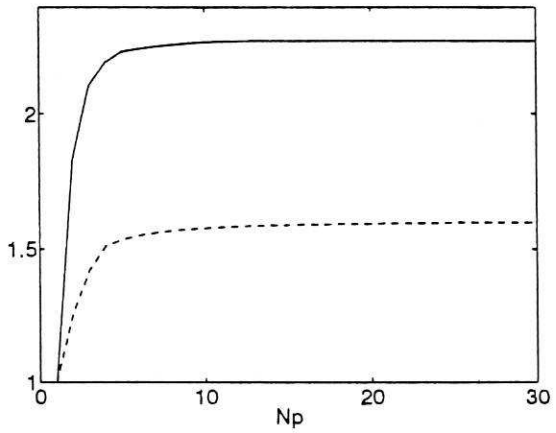

Figure 18: $E_{\mathrm{true}}(1)/E_{\mathrm{true}}(N_{\mathrm{p}})$ as a function of N_{p}. (—) Prediction-based noise reduction, (- -) smoother-based noise reduction.

Moreover, structure selection and parameter estimation is achieved by trivial modifications on well established algorithms.

Although the linear smoothers performed well in many situations, nonlinear smoothers seem to yield better results in some cases for higher noise levels and for a rather limited number of passes through the data. However, there are a number of factors which seem to affect the performance such as the maximum lag and the number of noise terms used in the smoothers. The investigation of how such parameters influence the noise reduction procedure has not been exhausted in this paper and remains a matter for further research.

The suggested procedure has been tested on models of well known nonlinear systems

which display chaotic dynamics for a wide range of parameters. The usefulness of the noise reduction procedure has been assessed in the context of model identification. Therefore, the ultimate objective has been to enable the identification of nonlinear models from smoothed data in situations where the use of the raw data had been ineffective.

ACKNOWLEDGMENTS

LAA and EMM gratefully acknowledge financial support from CNPq (Brazil) under grants 200597/90-6 and 201557/91-6, respectively. SAB gratefully acknowledges that part of this work was funded by SERC under contract GR/H35286.

References

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. (1994). Validating identified nonlinear models with chaotic dynamics. *Int. J. Bifurcation and Chaos*, (in press).

Billings, S. A., Korenberg, M. J., and Chen, S. (1988). Identification of nonlinear output affine systems using an orthogonal least squares algorithm. *Int. J. Systems Sci.*, 19(8):1559–1568.

Billings, S. A. and Tao, Q. H. (1991). Model validation tests for nonlinear signal processing applications. *Int. J. Control*, 54:157–194.

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. (1986). Correlation based model validity tests for nonlinear models. *Int. J. Control*, 44(1):235–244.

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

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

Grassberger, P., Hegger, R., Kantz, H., and Schaffrath, C. (1993). On noise reduction methods for chaotic data. *Chaos*, 3(2):127–141.

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. (1982). Noise in chaotic systems. *Nature*, 298:358–361.

Haber, H. and Unbehauen, H. (1990). Structure identification of nonlinear dynamic systems — A survey on input/output approaches. *Automatica*, 26(4):651–677.

Hammel, S. H. (1990). A noise reduction method for chaotic systems. *Phys. Lett.*, 148 A(8,9):421–428.

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

Holzfuss, J. and Kadtke, J. (1993). Global nonlinear noise reduction using radial basis functions. *Int. J. Bifurcation and Chaos*, 3(3):589–596.

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.

Kaplan, J. L. and Yorke, J. A. (1979). Chaotic behavior of multidimensional difference equations. In Peitgen, H. O. and Walther, H. O., editors, *Functional differential equations and approximation of fixed points, Lecture Notes in Mathematics, vol. 730*, pages 204–227. Springer Verlag, Berlin.

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.

Kostelich, E. J. and Yorke, J. A. (1988). Noise reduction in dynamical systems. *Phys. Rev. A*, 38(3):1649–1652.

Kostelich, E. J. and Yorke, J. A. (1990). Noise reduction: Finding the simplest dynamical system consistent with the data. *Physica D*, 41:183–196.

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.

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.

Marteau, P. F. and Abarbanel, H. D. I. (1991). Noise reduction in chaotic times series using scaled probabilistic methods. *J. Nonlinear Sci.*, 1:313–343.

Matsumoto, T., Chua, L. O., and Komuro, M. (1985). The double scroll. *IEEE Trans. Circuits Syst.*, 32(8):798–818.

Mitschke, F. (1990). Acausal filters for chaotic signals. *Phys. Rev. A*, 41(2):1169–1171.

Ozaki, T. (1993). A local linearization approach to nonlinear filtering. *Int. J. Control*, 57(1):75–96.

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.

Sano, M. and Sawada, Y. (1985). Measurement of the Lyapunov spectrum from a chaotic time serires. *Phys. Rev. Lett.*, 55(10):1082–1085.

Schreiber, T. and Grassberger, P. (1991). A simple noise-reduction method for real data. *Phys. Lett.*, 160 A(5):411–418.

Wolf, A. (1986). Quantifying chaos with Lyapunov exponents. In Holden, A., editor, *Chaos*, pages 273–290. Manchester University Press, Manchester.