Effects of the Sampling Time on the Dynamics and Identification of Nonlinear Models

S A Billings and L A Aguirre

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

Research Report No 513

April 1994

Effects of the Sampling Time on the Dynamics and Identification of Nonlinear Models

S. A. BILLINGS and LUIS A. AGUIRRE

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

Abstract

This paper investigates how the choice of the sampling time affects the identification of nonlinear models. It is shown how the choice of the sampling interval can affect structure selection and parameter estimation in different ways so that a compromise is required to properly select an appropriate sampling time. The use of a nonlinear correlation function to aid in the choice of the sampling rate is investigated. Several numerical examples are included to illustrate the main points of the paper.

1 Introduction

Most papers on system identification are concerned with the problem of estimating a model from a set of given data and therefore do not usually investigate how properties of the data such as the frequency content of the input and the sampling time affect subsequent stages in the identification.

In the case of linear systems the choice of the input and of the sampling time seems to be well understood and perhaps this partly explains why little attention is usually devoted to these matters.

Unfortunately, in the case of nonlinear systems most of the procedures for selecting the input commonly used for linear systems do not apply. For example, the use of *pseudo random binary signal* (PRBS) inputs to excite nonlinear systems can cause loss of identifiability (Leontaritis and Billings, 1987) and in the design of input signals both the frequency content and the amplitude profile are of interest (Aguirre and Billings, 1994b).

Thus there seems to be a gap in the knowledge of how the quality of the input-output data affects other aspects in the identification of nonlinear systems. It is the aim of this paper to provide some results which hopefully will help bridge this gap. In particular, the effects of the sampling time on structure selection, parameter estimation and on the quality of the final model will be investigated.

The results suggest that in the identification of nonlinear systems a sampling interval which aids in model structure selection is not necessarily the best choice of sampling interval for parameter estimation and vice versa. A compromise in the selection of the sampling interval is therefore often required and a rule of thumb is introduced, based on higher order autocorrelation functions, which seems to provide the user with an appropriate range of sampling intervals for nonlinear system identification.

The outline of the paper is as follows. The next section provides some background related to the structure selection and parameter estimation of nonlinear models. Section 3 investigates the effects of the sampling time on the quality of the final models by comparing the dynamics of both estimated and analytically discretised models. Section 4 investigates how the sampling time influences the structure selection and parameter estimation algorithms. Section 5 discusses the use of a simple rule of thumb for choosing the sampling time. Finally, the main points of the paper are summarised in § 6.

2 Background

This section provides some background on the procedure used to select the structure and estimate the parameters of the identified models.

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

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

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 of length N obtained by sampling the continuous data $u(t)$ and $y(t)$ at T_s. The sequence $e(k)$ 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 but $F^{\ell}[\cdot]$ can be defined as a neural network, rational function expansions, etc. if required. In order to estimate the parameters of this map, equation (1) has to be expressed in prediction error form as

$$y(k) = \Psi^{\tau}(k-1)\hat{\Theta} \; + \; \xi(k) \;\; , \tag{2}$$

where

$$\begin{aligned} \Psi^{\tau}(k-1) &= \left[\Psi^{\tau}_{yu}(k-1) \;\; \Psi^{\tau}_{yu\xi}(k-1) \;\; \Psi^{\tau}_{\xi}(k-1)\right] \;\; , \\ \hat{\Theta} &= \left[\hat{\Theta}^{\tau}_{yu} \;\; \hat{\Theta}^{\tau}_{yu\xi} \;\; \hat{\Theta}^{\tau}_{\xi}\right]^{\tau} \;\; , \end{aligned} \tag{3}$$

and $\Psi_{yu}^{\tau}(k-1)$ is a matrix which contains linear and nonlinear combinations of output and input terms up to and including time $k-1$. The matrices $\Psi_{yu\xi}^{\tau}(k-1)$ and $\Psi_{\xi}^{\tau}(k-1)$ are defined similarly. The parameters corresponding to each term in these matrices are the elements of the vectors $\hat{\Theta}_{yu}$, $\hat{\Theta}_{yu\xi}$ and $\hat{\Theta}_{\xi}$, respectively. Finally, $\xi(k)$ are the residuals which are defined as the difference between the measured data $y(k)$ and the one-step-ahead prediction $\Psi^{\tau}(k-1)\hat{\Theta}$. The parameter vector Θ can be estimated using orthogonal techniques in order to effectively overcome numerical ill-conditioning and structure selection problems. Parameter estimation is performed for a linear-in-the-parameters model of the type (Billings et al., 1989; Korenberg et al., 1988)

$$y(k) = \sum_{i=1}^{n_\theta} g_i\, w_i(k) \;+\; \xi(k), \tag{4}$$

where $n_\theta = n_p + n_\xi$ is the number of (process plus noise) terms in the model, $\{g_i\}_{i=1}^{n_\theta}$ are constant parameters and the polynomials $\{w_i(k)\}_{i=1}^{n_\theta}$ are constructed to be orthogonal over the data records as follows

$$\begin{aligned} w_0(k) &= p_0(k) = 1 \\ w_m(k) &= p_m(k) - \sum_{i=0}^{m-1} \alpha_{im}\, w_i(k), \quad m = 1, 2, \ldots, n_\theta \ , \end{aligned} \tag{5}$$

where

$$\alpha_{im} = \frac{\sum_{k=1}^{N} p_m(k)\, w_i(k)}{\sum_{k=1}^{N} w_r(k)^2}, \quad 0 \le i \le m-1 \ . \tag{6}$$

The parameters of the auxiliary model in equation (4) can be estimated by

$$\hat{g}_m = \frac{\sum_{k=1}^{N} y(k)\, w_m(k)}{\sum_{k=1}^{N} w_m(k)^2} \ . \tag{7}$$

Finally, parameters of the model in equation (2) can be calculated from the $\{g_i\}_{i=1}^{n_\theta}$.

A criterion for selecting the most important terms in the model can be devised as a byproduct of the orthogonal parameter estimation procedure. The maximum *mean squared prediction error* (MSPE) is achieved when no terms are included in the model, that is, when $n_\theta = n_p + n_\xi = 0$. In this case the MSPE equals $\overline{y^2(k)}$ where the over-bar indicates time averaging. The reduction in the MSPE due to the inclusion of the ith term, $g_i\, w_i(k)$, in the auxiliary model of equation (4) is $g_i^2\, \overline{w_i^2(k)}$. Expressing this reduction as a fraction of the total MSPE yields the *error reduction ratio* (ERR) (Billings et al., 1989; Korenberg et al., 1988)

$$[\mathrm{ERR}]_i \doteq \frac{g_i^2 \sum_{k=1}^{N} w_i^2(k)}{\sum_{k=1}^{N} y^2(k)}, \quad i = 1, 2, \ldots, n_\theta \tag{8}$$

Hence those terms with large values of ERR are selected to form the model.

3 Effects of the Sampling Time on the Dynamics of Estimated NARMAX Polynomial Models

This section presents some results which illustrate how the sampling period influences the structure of estimated polynomials and therefore the final dynamics. The main objective is to verify how quantities such as the the number of process terms, n_p, and the maximum lag of the model $n_y = n_u$ vary as T_s is increased or decreased.

The basis of the approach is to estimate a family of models for several values of T_s. In order to assess the quality of each model, bifurcation diagrams are calculated. Such diagrams reveal dynamical features of the models operating over a wide range of parameters and are well suited for the purposes of model validation and comparison (Aguirre and Billings, 1994d). Subsequently, the models which best reproduce the bifurcation pattern of the original system are selected and the structure of such models is compared in order to verify if there is any relationship between the variables T_s, n_y and n_p. Two systems are considered in the simulations, namely the Duffing-Ueda oscillator (Ueda, 1985)

$$\ddot{y} + k\dot{y} + y^3 = A\cos(\omega t) \ , \tag{9}$$

where $k = 0.1$, and the modified van der Pol oscillator (Ueda and Akamatsu, 1981)

$$\ddot{y} + \mu(y^2 - 1)\dot{y} + y^3 = A\cos(\omega t) \ , \tag{10}$$

where $\mu = 0.2$. These relatively simple equations are capable of exhibiting a vast diversity of dynamical regimes and therefore have become benchmarks in the study of nonlinear dynamics (Moon, 1987).

It is assumed throughout that the structure of the noise model is adequate so that the process model is unbiased. This can be readily verified in practice by applying correlation validation tests (Billings and Voon, 1986). For the sake of simplicity it is further assumed that $n_u = n_y$.

For the Duffing-Ueda oscillator the three values of T_s were $\pi/30$, $\pi/60$ and $\pi/100$, where $3/\pi$ is the Nyquist rate. For the modified van der Pol oscillator the following values of the sampling time were considered $\pi/60$, $\pi/80$ and $\pi/100$, where $4/\pi$ is the Nyquist rate. It is helpful to realise that the number of samples per input period is $2\pi/\omega T_s$, thus for the Duffing-Ueda oscillator with $\omega = 1$ rad/s, $T_s = \pi/30$ results in 60 samples per input period. For the modified van der Pol oscillator, however, the input frequency was 4 rad/s, thus $T_s = \pi/80$ resulted in 40 samples per input period.

It is noted that the sampling time has been chosen as $T_s = \pi/K$, $K \in \mathbb{Z}^+$ in order to have well defined Poincaré sections and bifurcation diagrams.

In figure 1a–b the values of $n_y = n_u$ and n_p were plotted against T_s for the *best estimated models* of both oscillators. These figures suggest, for the systems and inputs considered, that

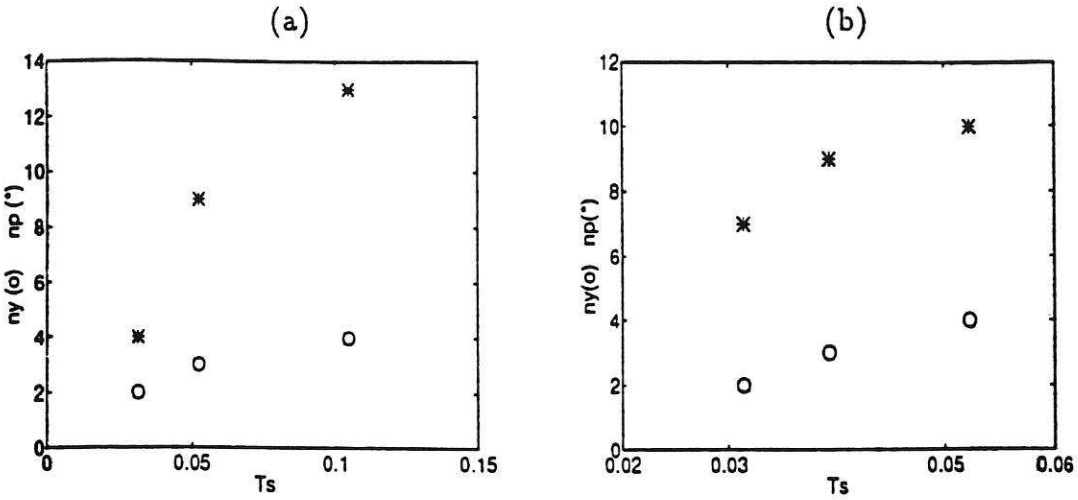

Figure 1: Values of $n_y = n_u$ and n_p as function of T_s for dynamically valid estimated models of (a) the Duffing-Ueda oscillator, and (b) the modified van der Pol oscillator

as the sampling interval is increased (sampling frequency is reduced) the best models tend to have more terms and require extra degrees of freedom to adequately capture the underlying dynamics. Similar tendencies have also been verified for inaccurate models with comparable bifurcation patterns.

It therefore seems appropriate to infer that the loss of accuracy due to slower sampling may, to some extent, be compensated by an increase in the number of terms in the model, n_p, and/or by an increase in the maximum lags considered, $n_y = n_u$, which is the number of degrees of freedom of the model.

It is worth pointing out that these results may also be interpreted from another point of view. Hence it also seems appropriate to conclude that if the data are deliberately over-sampled the complexity of the estimated models may be somewhat reduced. However, if the data are sampled too fast, successive measurements tend to be highly correlated and a number of practical problems arise such as ill-conditioning and lack of sufficient computational resources for recording and processing the data.

Analogous ideas have been described in the study of the relationship among the sampling time, the number of degrees of freedom and an *information redundancy function* of attracting sets (Fraser, 1989a). In that reference it was shown that a certain characteristic (the redundancy function) of a strange attractor may be increased (decreased) in two different and independent ways, namely i) by decreasing (increasing) the sampling period, or ii) by increasing (decreasing) the *embedding dimension* which is analogous to the maximum lag n_y. Such results seem to lend further support to the results summarised in figure 1a–b.

Similar results have been also observed for models of Chua's circuit (Aguirre and Billings, 1994a). In that case, however, the conjectured relationships are not as sharp as for the systems considered in this section because of the presence of the static nonlinearity in Chua's

circuit.

3.1 Estimated and discretised models

Discrete models estimated from input-output data obtained by sampling continuous signals are obviously approximate representations of the original system. Roughly, there are two main sources of errors involved in the identification. Firstly, a discrete model is being fitted to data which were generated by a system which, in principle, is continuous in time. Secondly, the estimated model is obtained from a finite amount of finite precision data and with no *a priori* knowledge of the parameters T_s, n_p, $n_y = n_u$, best input type, etc. which influence the results.

Further insight can be gained if the identified models are compared to discrete models obtained from the original equation by analytical discretisation. In this case the first source of errors is eliminated since discrete models are compared to discrete models with identical sampling periods. The *implicit Euler* or *backward difference* approximation defined as

$$\dot{y}(k) \approx \frac{y(k) - y(k - T_d)}{T_d} \tag{11}$$

can be used to obtain discretised models directly from equations (9) and (10).

Figure 2 shows the Poincaré sections of discretised and estimated models. The discretised models were obtained using equation (11), the original differential equation (9) and making the further approximation $y(k)^3 \approx y(k - T_d)^3$ for several values of the discretisation time, T_d. On the other hand, the identified models were estimated from data generated by equation (9) and sampled at $T_s = T_d$.

From this figure it is clear that the increase in T_d tends to deteriorate the dynamics of the discretised models, as would be expected from equation (11). On the other hand, the estimated models are less sensitive to variations in T_s than the discretised counterparts to variations in T_d. This suggests that, within a limited range of sampling times, the identification procedure tends to compensate for the increase in T_s by including more terms in the model and reestimating the parameters.

Figure 3 summarises some features of figure 2. For each model in the latter figure a quality index was calculated taking into account both the bifurcation diagram structure and the shape of the strange attractor obtained for $A = 11$. Models discretised with $T_d = \pi/300$ and $T_d = \pi/15$ s have also been included in figure 3.

This figure illustrates that over a certain range of values of T_s the estimated models are more accurate than the discrete counterparts. Such an improvement is achieved due to the additional flexibility attained by incorporating some more terms and reestimating the parameters.

Nevertheless, there are lower and upper bounds on T_s and T_d beyond which the discretised

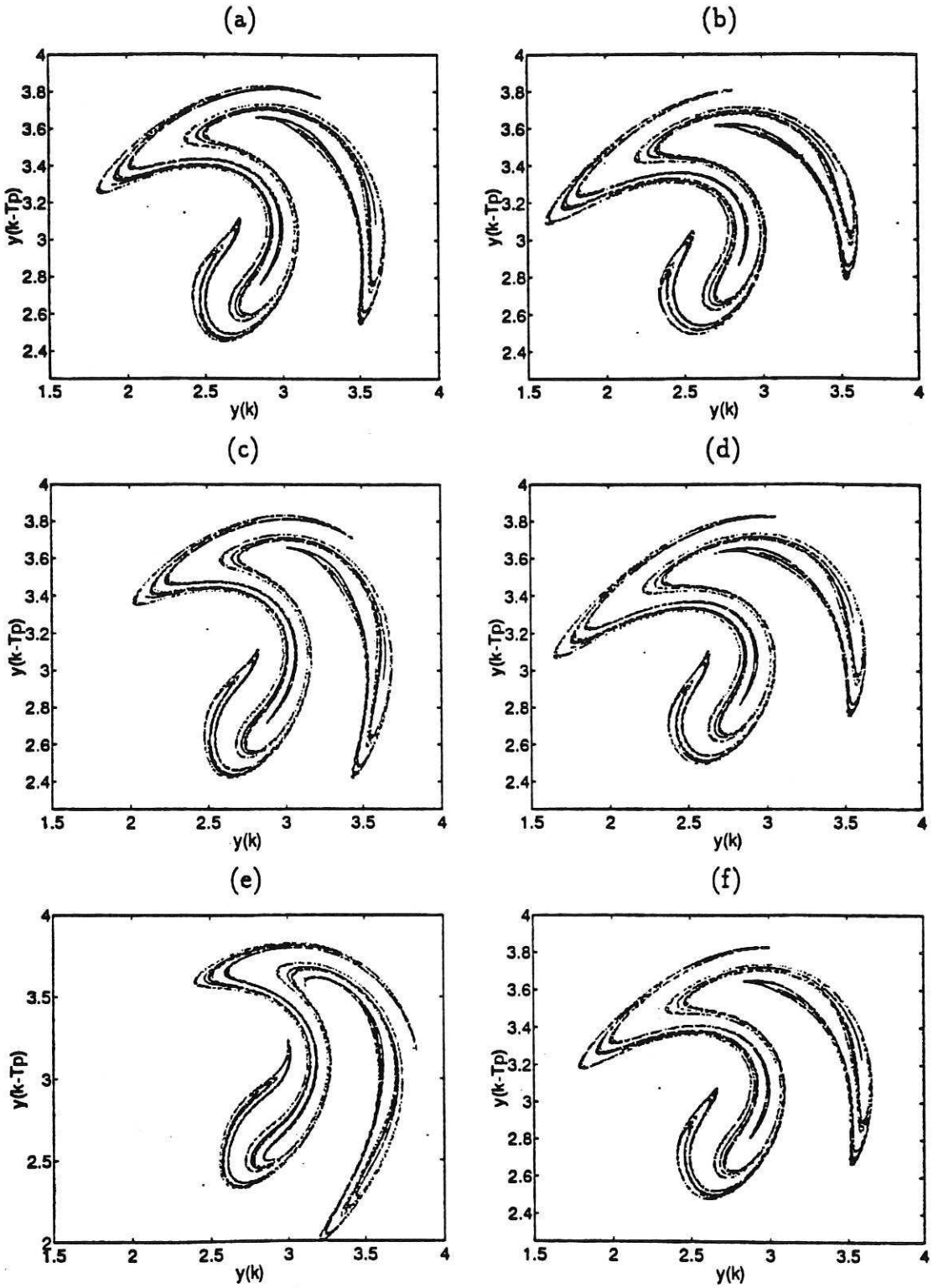

Figure 2: Poincaré sections for discretised and estimated models of the Duffing-Ueda attractor ($A=11$). (a) $T_d=\pi/100$, (b) $T_d=\pi/60$, (c) $T_d=\pi/30$, (d) $T_s=\pi/100$, (e) $T_s=\pi/60$ and (f) $T_d=\pi/30$.

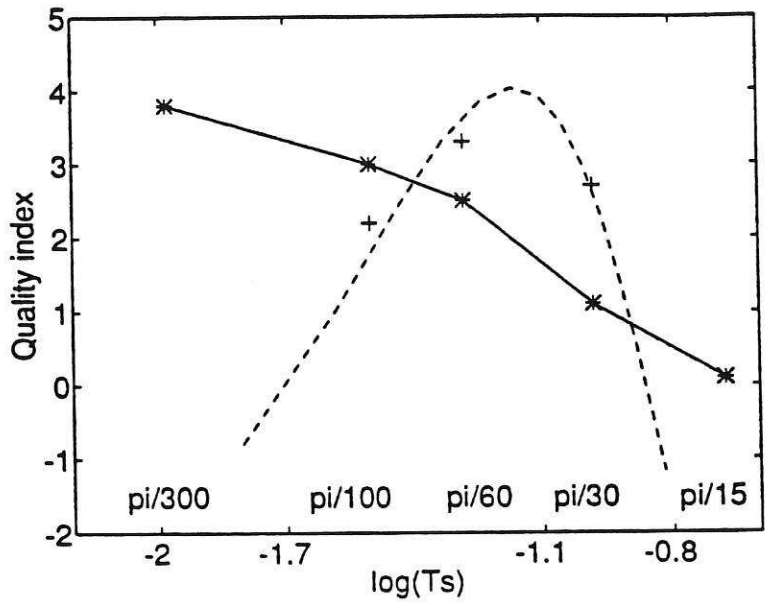

Figure 3: Quality index for (—) discretised and (- -) identified models of the Duffing-Ueda oscillator. The '$*$' indicate models discretised with $T_d = \pi/300, \pi/100, \pi/60, \pi/30, \pi/15$ and the '+' indicate the best models estimated from data sampled at $T_s = \pi/100, \pi/60, \pi/30$.

models are always better than the estimated ones for the Duffing-Ueda oscillator. Thus if T_d is decreased beyond the lower bound the discretised models will gradually tend to the continuous system since the approximation in equation (11) becomes more accurate as $T_d \rightarrow 0$. Besides, for oversampled data a number of numerical problems may arise in the parameter estimation.

If T_s is increased beyond the upper bound, because of the sampling theorem, the original frequency content and consequently the information in the data about the original dynamics will be lost and cannot therefore be retrieved by the estimation algorithm. In this case the estimated models will, for this system, also tend to be worse than the discretised counterparts.

In the case of the modified van der Pol oscillator, the discretised models were unstable for approximately $T_d > \pi/2000$. For this oscillator, the compensation associated with the identification procedure was verified from the fact that by including a few more terms and restimating parameters the identified models would in many instances become stable.

4 Effects of the Sampling Time on Structure Selection and Parameter Estimation

The choice of the sampling period is usually linked to the frequency content of the data. The data should be sampled sufficiently fast in order to guarantee that all frequency components are well represented in the final data set. Although this is an important observation, it does not account for a number of practical aspects such as the fact that the sampling time affects the performance of the structure selection and parameter estimation algorithms and that there could be significant nonlinear interactions in the data which would probably require a faster sampling rate. The former aspect is discussed in this section and the latter will be addressed in section 5.

4.1 Structure selection

The objective of this section is to report a few results which suggest that within practical bounds some structure selection schemes perform better at lower sampling rates. For the sake of clarity such results will be presented in the form of examples.

In system identification the effects of oversampling are usually associated with numerical problems during parameter estimation which arise as a consequence of the conditioning of $\Psi\Psi^{\mathrm{T}}$, see equation (2). In practice, however, data oversampling will also pose problems for selecting the model structure. The following examples illustrate this point.

Example 4.1

In order to illustrate that for very small sampling times the structure of the model becomes increasingly difficult to discern, consider the simple model

$$a_1\dot{y} + a_2 y = u \ , \tag{12}$$

and the discretisation formula

$$\dot{y}(k) = \frac{y(k) - y(k - T_{\mathrm{d}})}{T_{\mathrm{d}}} + e_{\mathrm{d}}(k) \ , \tag{13}$$

where $e_{\mathrm{d}}(k)$ denotes the discretisation error due to the approximation in equation (11). Substituting equation (13) into (12) yields

$$y(k) = \frac{a_1}{a_1 + T_{\mathrm{d}}a_2}\, y(k - T_{\mathrm{d}}) + \frac{T_{\mathrm{d}}}{a_1 + T_{\mathrm{d}}a_2}\, u(k) + \frac{T_{\mathrm{d}}a_1}{a_1 + T_{\mathrm{d}}a_2}\, e_{\mathrm{d}}(k) \ . \tag{14}$$

Clearly, in the limit as $T_{\mathrm{d}} \to 0$ equation (14) becomes

$$y(k) = \frac{a_1}{a_1} y(k) \ , \tag{15}$$

which is an obvious result and illustrates that for $T_d \rightarrow 0$ the structure becomes 'unidentifiable'. □

In practice the sampling time will not reach the limit $T_d \rightarrow 0$. However, for small values of T_s accurate structure selection is precluded mainly for two reasons. Firstly, for sufficiently small values of T_s all of the terms of the same general form become practically indistinguishable, for instance $y(k-T_s)y(k-2T_s)u(k-3T_s)$ would for all practical purposes be equivalent to, say, $y(k-2T_s)^2u(k-T_s)$. It is also possible to exploit this property to improve *coarse* structure selection (Aguirre and Billings, 1994c). Secondly, numerical problems arise when T_s is too small and such difficulties are reflected in poor performance of the structure selection algorithm as shown in the next example.

Example 4.2

In order to gain some insight into the effects of oversampling on the error reduction ratio defined in equation (8), this ratio will be analytically calculated for the term $p_1(k)=y(k-T_s)$ which is usually selected to compose the structure of most models estimated from data produced by a continuous process. Thus taking $m=1$ in equation (5) gives $w_1(k)=p_1(k)$.

The coefficient of $w_1(k)$ in the auxiliary model is given by equation (7) as

$$\hat{g}_1 = \frac{\sum_{k=1}^{N} y(k)\, w_1(k)}{\sum_{k=1}^{N} w_1(k)^2} \quad , \tag{16}$$

and the error reduction ratio is given by equation (8) as

$$[\mathrm{ERR}]_1 = \frac{\sum_{k=1}^{N} g_1\, y(k-T_s)^2}{\sum_{k=1}^{N} y(k)^2} \quad . \tag{17}$$

Substituting the estimated value of g_1 given by equation (16) into equation (17) yields

$$[\mathrm{ERR}]_1 = \frac{\sum_{k=1}^{N} \frac{\sum_{k=1}^{N} y(k)\, w_1(k)}{\sum_{k=1}^{N} w_1(k)^2}\, y(k-T_s)^2}{\sum_{k=1}^{N} y(k)^2} \quad . \tag{18}$$

Bearing in mind that $w_1(k)=y(k-T_s)$ and taking the limit as $T_s \rightarrow 0$ in the last equation, yields

$$\lim_{T_s \rightarrow 0} [\mathrm{ERR}]_1 = 1 \quad . \tag{19}$$

The interpretation of the last equation is simple. In the limit $T_s \rightarrow 0$, only one term is necessary to explain the measured data, namely $y(k-T_s)$. This result is analogous to the one in example 4.1. □

In practice T_s will not be zero and therefore $[\mathrm{ERR}]_1 \neq 1$. However, as shown in the example above, if the sampling time is too short $[\mathrm{ERR}]_1 \approx 1$. This characterises a numerically ill-conditioned problem because

$$\sum_{i=1}^{n_\theta} [\mathrm{ERR}]_i = 1 \ , \tag{20}$$

and if $[\mathrm{ERR}]_1 \approx 1$ the values of ERR corresponding to the all the other candidate terms are very small. Consequently, it becomes difficult to select the structure in such a situation. When the data are sampled slower, the ERR values are better distributed thus facilitating structure selection. The following example illustrates this point.

Example 4.3

This example uses the modified van der Pol model in equation (10). Discretised models for this oscillator were obtained by using equations (10) and (11), and by making the further approximations $y(k)^3 \approx y(k-T_d)^3$ and $y(k)^2 y(k-T_d) \approx y(k-T_d)^2 y(k-2T_d)$. The discretised models have the five following terms: $y(k-T_d)$, $y(k-2T_d)$, $y(k-T_d)^3$, $y(k-T_d)^2 y(k-2T_d)$ and $u(k-T_d)$. This set of terms will be referred to as the *discretised model structure* (DMS) and will be used as a benchmark to assess the adequacy of the structure of estimated models for this system.

A set of data was obtained by integrating equation (10) and deliberately oversampled at $T_s = \pi/200$. Such data were subsequently decimated to obtain records of the same length ($N = 1000$) with $T_s = \pi/100, \pi/40$ and $\pi/20$. The maximum lag considered was $n_y = 3$. Table 1 shows the first ten terms selected using the *error reduction ratio* (ERR) and table 2 was obtained in the same way but for noisy data with a *signal to noise ratio* (SNR) of $20\log(\sigma_y^2/\sigma_e^2) \approx 70$ dB, where σ_e^2 is the noise variance.

Figure 4 shows the *Akaike information criterion* (AIC) (Akaike, 1974) corresponding to table 2. Although a clear global minimum is hardly recognisable in this figure, it seems that the optimal number of terms in the models suggested by AIC(4) are 2 for $T_s = \pi/100$, 2 or 7 for $T_s = \pi/40$ and 5 for $T_s = \pi/20$. The same values were obtained for the *final prediction error* (FPE) (Akaike, 1974), Khundrin's *law of iterated logarithm criterion* (LILC) (Hannan and Quinn, 1979) and the *Bayesian information criterion* (BIC) (Kashyap, 1977). It is worth pointing out that the first five terms of the last column in both tables correspond to the five terms which compose the DMS thus suggesting that in this example the data set corresponding to a shorter sampling interval enabled somewhat improved structure selection. This illustrates that in some cases decimating the data helps discern the important terms in the model. In practice, however, the DMS will not usually be known and the inclusion of a few more terms and a slight increase of the maximum lag considered often helps handle noisy data sets better. Nonetheless, even in practical situations decimating the data slightly

Table 1: Model structures for models estimated using different sampling times for the modified van der Pol oscillator. The noise-free case.

$T_s = \pi/100$	$T_s = \pi/40$	$T_s = \pi/20$
$y(k-1)$	$y(k-1)$	$y(k-1)$
$y(k-2)$	$y(k-2)$	$y(k-2)$
$y(k-3)$	$y(k-3)$	$u(k-1)$
$y(k-1)y(k-3)u(k-3)$	$y(k-1)y(k-2)u(k-3)$	$y(k-1)^3$
$y(k-2)^2u(k-2)$	$y(k-1)^2u(k-1)$	$y(k-1)^2y(k-2)$
$y(k-1)y(k-3)^2$	$y(k-3)u(k-3)^2$	$y(k-1)y(k-2)y(k-3)$
$y(k-2)^2y(k-3)$	$y(k-1)^2y(k-3)$	$y(k-1)y(k-3)^2$
$y(k-1)u(k-1)^2$	$u(k-3)$	$y(k-1)y(k-2)^2$
$y(k-2)u(k-3)^2$	$y(k-1)^3$	$y(k-2)^3$
$u(k-3)$	$u(k-1)$	$y(k-1)^2y(k-3)$

Table 2: Model structures for models estimated using different sampling times for the modified van der Pol oscillator. The noise-corrupted case, SNR $\approx$ 70 dB.

$T_s = \pi/100$	$T_s = \pi/40$	$T_s = \pi/20$
$y(k-1)$	$y(k-1)$	$y(k-1)$
$y(k-3)$	$y(k-2)$	$y(k-2)$
$u(k-3)$	$u(k-2)$	$u(k-1)$
$y(k-2)$	$y(k-1)^3$	$y(k-1)^3$
$y(k-1)^3$	$y(k-3)$	$y(k-1)^2y(k-2)$
$u(k-2)$	$y(k-1)^2u(k-3)$	$y(k-3)$
$y(k-2)^3$	$u(k-3)$	$u(k-3)$
$y(k-1)^2u(k-2)$	$y(k-1)^2y(k-2)$	$y(k-1)y(k-2)^2$
$y(k-3)^2u(k-1)$	$y(k-1)y(k-2)u(k-3)$	$y(k-1)^2u(k-2)$
$y(k-3)^3$	$y(k-2)^3$	$y(k-2)y(k-3)^2$

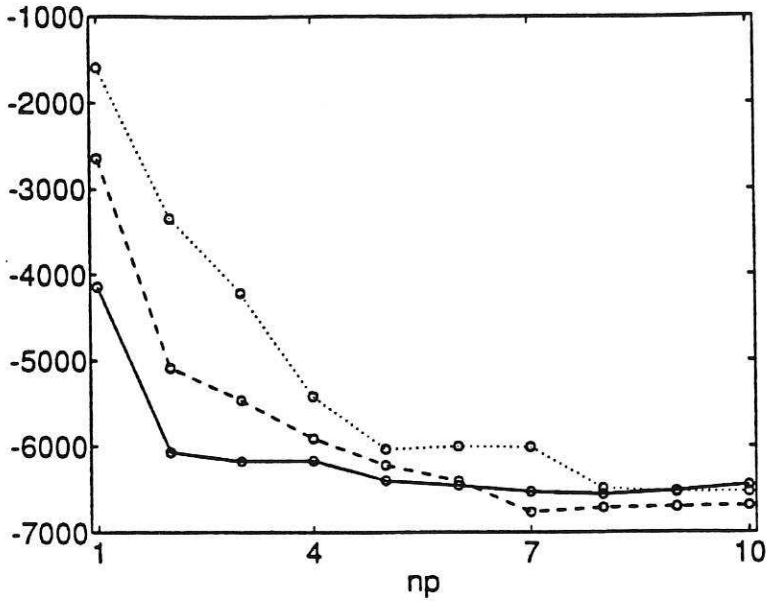

Figure 4: AIC(4) for models of the modified van der Pol oscillator estimated from data sampled at (—) $T_s = \pi/100$, (- -) $T_s = \pi/40$ and (···) $T_s = \pi/20$.

seems to enhance structure selection. Similar results have been observed for the Duffing-Ueda oscillator. □

4.2 Parameter estimation

The objective in this section is to investigate how parameter estimates are affected by the sampling time. As before, the parameters of the respective discretised model are used as standards to which the estimated parameters are compared.

Example 4.4

This example illustrates that improved parameter estimation may be attained by using data sampled slightly faster. To this end, equation (14) is rewritten as follows

$$y(k) = \Psi^T \Theta + \frac{T_d a_1}{a_1 + T_d a_2} e_d(k) \ . \tag{21}$$

The least squares estimate of Θ is given by

$$\begin{aligned} \hat{\Theta} &= [\Psi \Psi^T]^{-1} \Psi \, y(k) \ , \\ &= \mathbf{A} \, y(k) \ . \end{aligned} \tag{22}$$

The bias is defined as

$$
\begin{aligned}
B &= \mathrm{E}\{\hat{\Theta}\} - \Theta \ , \\
&= \mathrm{E}\{\mathrm{A}\, y(k)\} - \Theta \ , \\
&= \mathrm{E}\{\mathrm{A}\,(\Psi^{\mathrm{T}}\Theta + \frac{T_{\mathrm{d}} a_1}{a_1 + T_{\mathrm{d}} a_2} e_{\mathrm{d}}(k))\} - \Theta \ , \\
&= \frac{T_{\mathrm{d}} a_1}{a_1 + T_{\mathrm{d}} a_2} \mathrm{E}\{\mathrm{A}\, e_{\mathrm{d}}(k)\} \ , \qquad (23)
\end{aligned}
$$

where $\mathrm{E}\{\cdot\}$ denotes mathematical expectation.

It should be noted that $e_{\mathrm{d}}(k)$ is not white noise but the discretisation error and that this error will usually have a nonzero mean and could easily become correlated with A. Consequently the bias in equation (23) will not be zero and will be proportional to $T_{\mathrm{d}} a_1/(a_1 + T_{\mathrm{d}} a_2)$, thus diminishing as T_{d} is decreased. Unlike in examples 4.1 and 4.2, here the limit $T_{\mathrm{d}} \to 0$ cannot be considered because in such a limit the matrix $\Psi\Psi^{\mathrm{T}}$ is singular and equation (22) would not hold. □

Example 4.5

This example uses the modified van der Pol oscillator. In order to verify how the sampling time influences the accuracy of the parameter estimates, the original equation (10) was discretised for various values of T_{d}. Subsequently, equation (10) was integrated to produce data records which were sampled at corresponding values of T_{s} to be used in the identification. Before estimating parameters however, the structure of the identified models was set to equal the DMS plus a set of noise terms included to guarantee unbiased estimates. Thus all the models had the same structure and the corresponding parameters could be compared.

Figure 5a shows the parameter values of both the discretised and estimated models for various values of T_{d} and T_{s} which were varied in the range $[\pi/200,\ \pi/20]$. The same simulation was carried out for noisy data with SNR$\approx$70 dB, the results are shown in figure 5b.

With the exception of the coefficient of the term $u(k - T_{\mathrm{s}})$, all other estimated coefficients converge to the discretised counterparts as $T_{\mathrm{s}} = T_{\mathrm{d}} \to 0$ and diverge as$T_{\mathrm{s}} = T_{\mathrm{d}} \to \infty$. This tendency is somewhat blurred by the noise but can still be distinguished in figure 5b. Analogous results have been verified for the Duffing-Ueda oscillator. □

4.3 Discussion

The main objectives of the two previous sections were to investigate the influence of the sampling time not only on the dynamics of the final estimated model but also on the performance

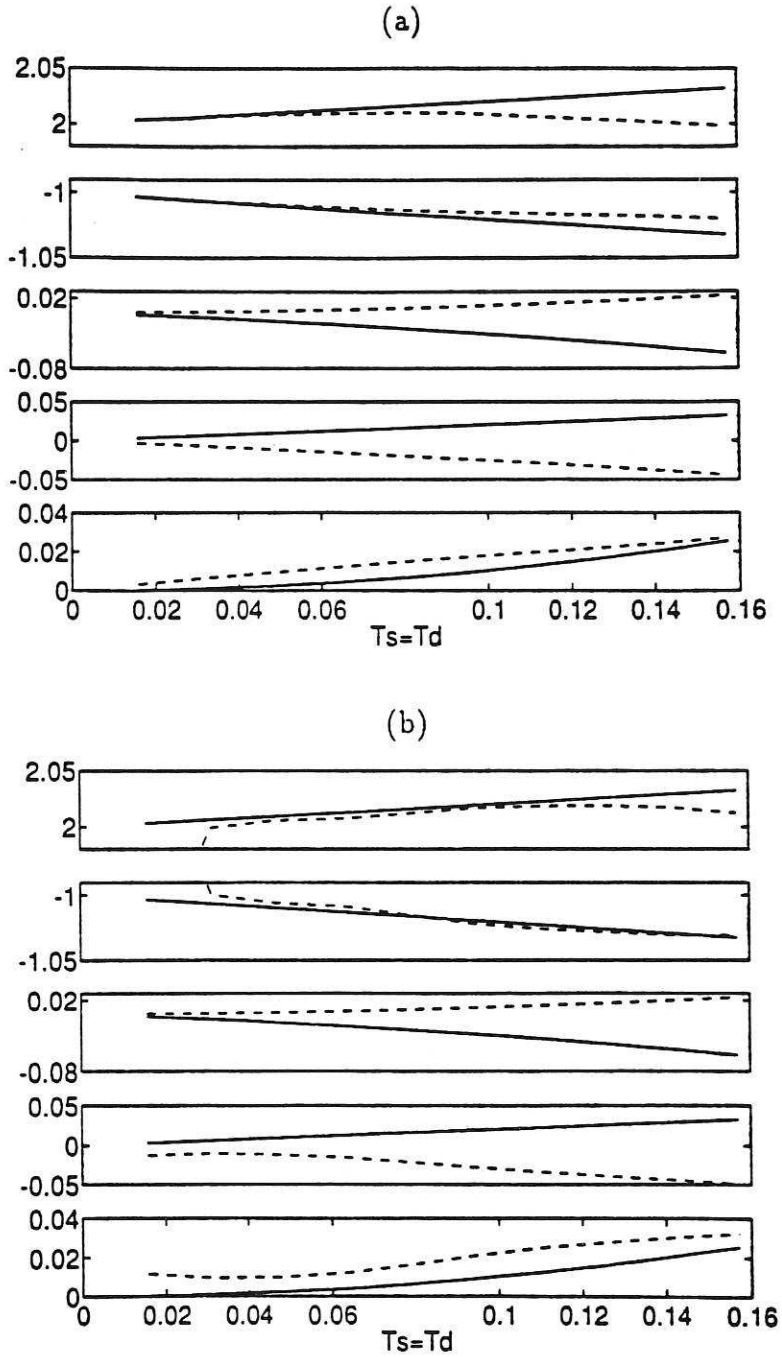

Figure 5: Parameters of discretised (—) and estimated (- -) models of the modified van der Pol oscillator for various values of T_d and T_s. The plots correspond to the parameters of the following terms (from top to bottom) $y(k-T_s)$, $y(k-2T_s)$, $y(k-T_s)^3$, $y(k-T_s)^2y(k-2T_s)$ and $u(k-T_s)$. (a) noise-free data, (b) noise-corrupted data, SNR$\approx$70 dB.

of the algorithms used to perform structure selection and parameter estimation.

In view of the results reported above it seems appropriate to infer that parameter estimation and structure selection are influenced by the choice of the sampling time in antagonistic ways. Thus, within a practical range of values, short sampling times seem to favour accurate parameter estimation for a given structure whilst precluding the correct selection of such a structure. Conversely, longer sampling times appear to enhance structure selection but in this case parameter estimation accuracy may deteriorate.

It becomes clear that if the system was adequately excited, the information needed for performing both structure selection and parameter estimation is contained in the data. However, some results seem to support the conjecture that the information related to the structure of the model and that related to the parameters are present in the data but at different time scales. Therefore, the sampling time can be viewed as a parameter which can be judiciously chosen by the practitioner in order to place a greater weight either on structure selection or on parameter estimation.

The conjecture that the structure of the dynamics is associated with the *macro scales* of the data seems to find further support in the field of nonlinear dynamics. It is well known that the fractal dimension of an attractor, which is a typical measure of the structure of the attractor, can be estimated from an embedded time series (Grassberger and Procaccia, 1983). Whilst there are restrictions on the dimension of the embedding space, in theory there are no restrictions on the sampling time (Takens, 1980). In practice because of the finite length and the noise on the data, the sampling time cannot be excessively large, however it can still be chosen many times longer than what would be required for accurate parameter estimation (Liebert and Schuster, 1989).

The results of this section must be interpreted bearing in mind that they are valid within practical bounds on the sampling time. For instance, if T_s is too small the matrix $\Psi\Psi^T$ becomes ill-conditioned and the accuracy of parameter estimates will deteriorate instead of improve. It should also be pointed out that the results reported are concerned with model identification. This remark is important because in some applications such as the estimation of correlation dimensions, optimum results are obtained when the data window spanned by the model, that is $T_s(n_y - 1)$, is constant (Martinerie et al., 1992).

The results discussed above suggest that, given a a set of 'well-sampled data', structure selection could be performed on decimated data and after choosing a particular structure, the parameters could be estimated from the original records. An alternative approach would be to choose the sampling time as a trade-off between adequate structure selection and accurate parameter estimation. A rule of thumb for this is suggested in the following section.

5 A Rule of Thumb for Selecting the Sampling Time

The practical difficulty in selecting the sampling time can be appreciated by considering two extreme situations. Firstly, if T_s is too short the data become highly correlated (singular), that is $y(k) \approx y(k - T_s)$, and a number of numerical problems due to ill-conditioning arise. This phenomenon is referred to as *redundance*. On the other hand, if T_s is too long adjacent data points tend to become uncorrelated, especially if the data are chaotic. This phenomenon is known as *irrelevance* (Casdagli et al., 1991; Fraser, 1989b).

Ways of determining the sampling time for nonlinear dynamical reconstructions include the use of information theory (Fraser and Swinney, 1986), the correlation time (Albano et al., 1988; Abarbanel et al., 1990) and the reconstruction expansion (Rosenstein et al., 1994). These and other approaches have been briefly reviewed and compared in (Rosenstein et al., 1994). The rationale behind these methods is that although adjacent points in the data should not be uncorrelated, this would indicate an undesirable loss of information, the sampling time should be long enough as to avoid overcorrelation which would indicate that most of the information contained in a subsequent measurement is redundant. These approaches are not without difficulties. Against the former technique it has been pointed out that the algorithm is rather cumbersome (Liebert and Schuster, 1989) and against the latter it has been said that using the autocorrelation function for determining the correlation time would only take into account linear correlations in the data (Casdagli et al., 1991; Fraser, 1989b).

This section considers a nonlinear correlation function which has been developed to detect nonlinear correlations in the residuals of dynamical autonomous models (Billings and Tao, 1991). Consider the correlation functions

$$\begin{aligned} \Phi_{yy}(\tau_c) &= \mathrm{E}\{(y(k) - \overline{y(k)})(y(k-\tau_c) - \overline{y(k)})\} \ , \quad \tau_c = 0, 1, \dots \ , \\ \Phi_{y^{2'}y^{2'}}(\tau_c) &= \mathrm{E}\{(y^2(k) - \overline{y^2(k)})(y^2(k-\tau_c) - \overline{y^2(k)})\} \ , \quad \tau_c = 0, 1, \dots \ , \end{aligned} \tag{24}$$

where $\mathrm{E}\{\cdot\}$ is the mathematical expectation and the overbars indicate averaging with respect to time. While Φ_{yy} detects linear correlations, $\Phi_{y^{2'}y^{2'}}$ was designed to detect nonlinear correlations in the data (Billings and Tao, 1991).

The following procedure for determining the sampling time has produced improved results in a number of identification problems. First, compute the aforementioned correlation functions using the time series to be used. In the case of input output models, the time series corresponding to the system output should be used. Then define

$$\tau_m = \min\{\tau_y, \tau_{y^{2'}}\} \ , \tag{25}$$

where τ_y is the time of the first minimum of Φ_{yy} and $\tau_{y^{2'}}$ is defined analogously.

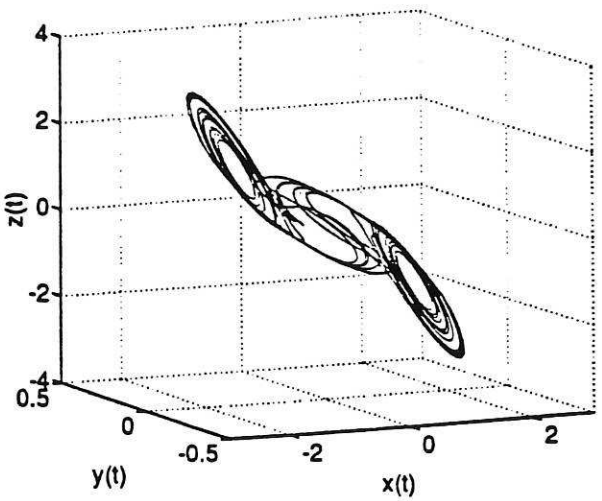

Figure 6: The double scroll Chua's attractor.

Finally, the sampling time can be chosen as follows

$$\frac{\tau_m}{20} \leq T_s \leq \frac{\tau_m}{10} \ . \tag{26}$$

It is noted that in some cases the upper bound on T_s can be somewhat relaxed, say, $\tau_m/5$. The following example illustrates some features of this procedure.

Example 5.1

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

$$\begin{cases} \dot{x} = \alpha(y - h(x)) \\ \dot{y} = x - y + z \\ \dot{z} = -\beta y \end{cases} \quad , \quad h(x) = \begin{cases} m_1 x + (m_0 - m_1) & x \geq 1 \\ m_0 x & |\, x \,| \leq 1 \\ m_1 x - (m_0 - m_1) & x \leq -1 \ , \end{cases} \tag{27}$$

where $m_0 = -1/7$ 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 $\beta = 100/7$, see figure 6. The identification of monovariable models for this attractor estimated from each component individually has been considered in (Aguirre and Billings, 1994a). This example illustrates the selection of the sampling time based on the rule of thumb in equations (25)–(26).

The correlation functions in equation (24) were calculated for the three time series generated by equation (27), that is, for the data measured from the x, y and z components sampled at $T_s = 0.01$. The results are shown in figure 7.

As can be seen from this figure, the information conveyed by the linear autocorrelation function $\Phi_{yy}(\tau_c)$ is very similar for the three components. However, $\Phi_{y^2'y^2'}(\tau_c)$ seems to suggest that the data measured from the y component should be sampled faster than the

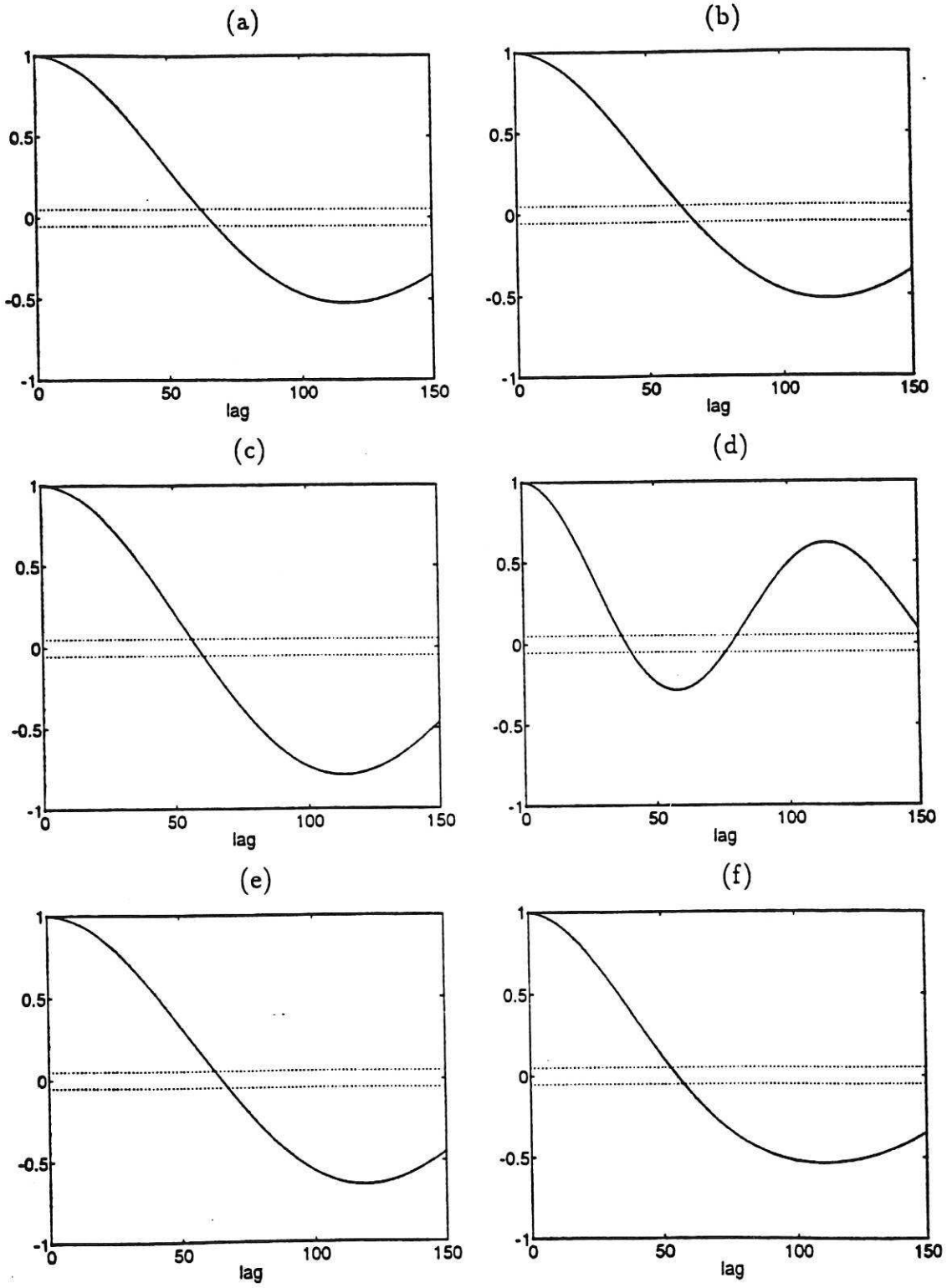

Figure 7: $\Phi_{yy}(\tau_c)$ for the double scroll calculated from the components (a) x, (c) y and (e) z. $\Phi_{y^2'y^2'}(\tau_c)$ calculated from the components (b) x, (d) y and (f) z. The dotted lines indicate the 95% confidence bands.

Table 3: Estimated values of τ_y and τ_{y^2} for the double scroll

Component	τ_y	τ_{y^2}
x	1.29	1.18
y	1.14	0.59
z	1.20	1.11

data measured from x and z. This can also be verified from table 3. The values in this table were obtained by multiplying $T_s = 0.01$ (the sampling time of the original data) by the number of lags corresponding to the first minima of the correlation functions plotted in figure 7. Thus it appears that because of certain nonlinear interactions, which were not detected by $\Phi_{yy}(\tau_c)$, in the data from the y component, such data should be sampled faster. In fact, whilst it was possible to identify models from the x and z components sampled at $T_s = 0.15$, that is approximately $\tau_m/7.3$, which reproduced the dynamical invariants of the double scroll the only valid models estimated from the y component were obtained from data sampled at $T_s = 0.07$ (Aguirre and Billings, 1994a) which is approximately $\tau_m/8.4$. Two things should be noted i) this is an example of a rather common situation for which the upper bound on T_s can be relaxed as mentioned after equation (26), and ii) shorter sampling intervals, but still in the range of equation (26), also yield good results.

□

6 Conclusions

The effects of the sampling time in the identification of nonlinear models have been investigated. A number of results seem to support the following conclusions: i) shorter sampling times usually favour identifying good models with less terms, ii) if the sampling time is too short structure selection is hampered in spite of an increase in the accuracy of the parameter estimates, iii) longer sampling times which degrade the accuracy of the parameter estimates usually aid structure selection, and iv) the last two items suggest that structure selection and parameter estimation should be traded off in practice and one way of achieving such a compromise would be by a judicious choice of the sampling time.

A simple rule of thumb for choosing the sampling time has been investigated. Such a rule uses correlation functions which can be readily computed from the data. Unlike many other rules which select the sampling time based on the correlation time of the linear autocorrelation function, this paper considers both linear and nonlinear correlations in the data. As a consequence it has been verified that for most nonlinear systems the sampling times suggested by nonlinear correlation functions are shorter than those recommended by

rules based solely on linear correlations.

The results of this paper have been verified using a number of benchmark systems of which the modified van der Pol and Duffing-Ueda oscillators and Chua's circuit have been reported. In order to gain further insight, analytically discretised models have been used for comparison. Although such models will not be available in most practical situations it is believed that the results and conjectures which have been discussed are rather general and would therefore apply to a much wider range of problems concerning the identification of nonlinear systems.

ACKNOWLEDGMENTS

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.

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). Dynamical effects of overparametrization in nonlinear models. (Submitted for publication).

Aguirre, L. A. and Billings, S. A. (1994c). Improved structure selection for nonlinear models based on term clustering. (submitted for publication).

Aguirre, L. A. and Billings, S. A. (1994d). Validating identified nonlinear models with chaotic dynamics. *Int. J. Bifurcation and Chaos*, 4(1):109–125.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Automat. Contr.*, 19(6):716–723.

Albano, A. M., Muench, J., Schwartz, C., Mees, A. I., and Rapp, P. E. (1988). Singular-value decomposition and the Grassberger-Procaccia algorithm. *Phys. Rev. A*, 38(6):3017–3026.

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.

Liebert, W. and Schuster, H. G. (1989). Proper choice of the time delay for the analysis of chaotic series. *Phys. Lett.*, 142A(2,3):107–111.

Martinerie, J. M., Albano, A. M., Mees, A. I., and Rapp, P. E. (1992). Mutual information, strange attractors, and the optimal estimation of dimension. *Phys. Rev. A*, 45(10):7058–7064.

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

Rosenstein, M. T., Collins, J. J., and De Luca, C. J. (1994). Reconstruction expansion as a geometry-based framework for choosing proper delay times. *Physica D*, (to appear).

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.

Ueda, Y. (1985). Random phenomena resulting from nonlinearity in the system described by Duffing's equation. *Int. J. Non-Linear Mech.*, 20(5/6):481–491.

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