Relationship Between the Structure and Performance of Identified Nonlinear Polynomial Models

L A Aguirre and S A Billings

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

Research Report No 474

May 1993

Relationship Between the Structure and Performance of Identified Nonlinear Polynomial Models

LUIS A. AGUIRRE and S. A. BILLINGS

Department of Automatic Control and Systems Engineering
University of Sheffield
P.O. Box 600, Mappin Street — Sheffield S1 4DU - UK
e-mail:aguirre@acse.sheffield.ac.uk

Abstract

The effect of the structure of identified nonlinear polynomial models on the dynamical behaviour of such models is investigated. In particular, the effects of the sampling rate, the number of process terms and the order of the models is studied and input design is also considered. Bifurcation diagrams and Poincaré sections are used to assess global and local dynamical behaviour of the identified models which were estimated from data generated by the Duffing-Ueda equation. In several of the examples which are provided to illustrate the main points of the paper, the identified models are compared to the original system over a wide range of parameter values for which the Duffing-Ueda equation exhibits regular oscillations, period-doubling cascades, pitchfork bifurcations and chaos.

1 Introduction

Although most real systems are nonlinear, the classical approach to system identification has been to use linear models defined at specific operating points. Models obtained in this way are of limited utility because they fail to reproduce dynamical phenomena caused by nonlinearities such as limit cycles, bifurcations and chaos. Furthermore, in the case of nonlinear oscillations it is difficult to define an operating point around which the nonlinearities of the system may be *safely* neglected.

The necessity of identifying accurate nonlinear models has encouraged the development of techniques which are appropriate for this purpose. The NARMAX (nonlinear autoregressive moving average model with exogenous inputs) provides a basis for such a development (Billings and Leontaritis 1981, Leontaritis and Billings 1985). One of the main advantages of this technique is that if a polynomial expansion of the NARMAX model is chosen, such a model becomes linear-in-the parameters and various well known numerical techniques can be used to estimate the parameters.

The number of possible terms for a typical nonlinear polynomial of limited dynamical order and nonlinearity is enormous in most practical situations. Thus one of the major difficulties in the identification of nonlinear systems is how to choose the terms which are more relevant and which will capture the underlying dynamics adequately. Fortunately, some algorithms are currently available which select the terms that should be included in a nonlinear model (Billings *et al.* 1988, 1989, Haber and Unbehauen 1990).

Most term selecting procedures sort all the candidate terms in order of importance according to some particular criterion. This provides an answer to the question of *which* are the most important terms but does not indicate *how many* terms should be included. Kortmann and Unbehauen (1988) suggested an iterative procedure which uses information criteria to determine the number of statistically relevant terms in a nonlinear polynomial.

The availability of increasing computational resources and the appearance of model structures which are very flexible, introduces the danger of including an excessive number of terms in a model because simulations can be performed fast enough and, besides, 'large models seem to fit the data much better'.

It is generally accepted that overparametrization is to be avoided. In linear systems the consequences of including an excessive number of terms in a model may lead to nonminimal transfer functions and quasi-cancellation of poles and zeros. Although the dynamical effects of overparametrization in linear systems does not seem to be very drastic, in nonlinear systems, however, such effects are not well documented and may significantly alter the global and local dynamical behaviour of a model.

One of the purposes of this paper is to investigate the effects of overparametrization in NARMAX polynomials on the dynamical behaviour of such models. Besides the number of terms, other aspects are also considered such as the choice of the input, the sampling rate

and the dynamical order of the model. The relationship between these parameters and the effect of these on the dynamical behaviour of the identified models is investigated using the Duffing-Ueda equation. Over the parameter range considered, this equation exhibits a wide variety of dynamical regimes, namely period-one and period-three oscillations, period-doubling cascades, pitchfork bifurcations and chaos.

To assess the quality and dynamical properties of the estimated models, bifurcation diagrams and Poincaré sections of the chaotic attractors will be used. These tools are extremely useful in establishing the relationship among several of the parameters which characterize the structure of a nonlinear polynomial model.

The paper is organised as follows: the next section reviews some techniques for the identification of NARMAX models. In §3 input design is considered. Section 4 investigates the relationship between the sampling rate, the number of terms and the order of a model. Further insight into this relationship can be gained by comparing estimated models to discretised counterparts, this is done in §5. Section 6 deals with some aspects of overparametrization. Finally, the main conclusions are summarised in §7.

2 System Identification

In this section the main ideas concerning system identification are reviewed and references are provided for further reading. The techniques to be reviewed have been developed over the last ten years and currently compose a well established procedure for the identification of nonlinear systems (Billings and Chen 1989).

2.1 System representation

One of the most popular structures for linear models is the so-called autoregressive model with exogenous inputs (ARX)

$$\mathrm{y}(t) = -\sum_{i=1}^{n_y} a_i \, \mathrm{y}(t-i) \; + \; \sum_{i=1}^{n_u} b_i \, \mathrm{u}(t-i) \tag{1}$$

where $\{a_i\}_{i=1}^{n_y} \in \mathbb{R}$ and $\{b_i\}_{i=1}^{n_u} \in \mathbb{R}$ are the coefficients of the model, $\mathrm{y}(t) \in \mathbb{R}^{no}$ and $\mathrm{u}(t) \in \mathbb{R}^{ni}$ are the output and input vectors, respectively. In what follows only the monovariable ($no = ni = 1$) case is reviewed. The multivariable case is well documented

in the literature (Söderström and Stoica 1989) and details of the multivariable nonlinear case can be found in (Billings *et al.* 1989).

In the context of system identification and parameter estimation noise terms are usually required in order to produce unbiased estimates. This gives rise to the following difference equation

$$y(t) = -\sum_{i=1}^{n_y} a_i\, y(t-i) \;+\; \sum_{i=1}^{n_u} b_i\, u(t-i) \;+\; \sum_{i=1}^{n_e} c_i\, e(t-i) \;+\; e(t) \tag{2}$$

where $\{c_i\}_{i=1}^{n_e} \in \mathbb{R}$ and $e(t) \in \mathbb{R}$. This model is often referred to as an autoregressive moving average model with exogenous inputs (ARMAX). Many parameter estimation algorithms for linear systems have been developed based on this model (Södertröm and Stoica 1989). Similar representations can be derived if the system is nonlinear. Thus the nonlinear autoregressive model with exogenous inputs (NARX) is represented as

$$y(t) = F^{\ell}[y(t-1), \ldots, y(t-n_y), u(t-d), \ldots, u(t-d-n_u+1)] \tag{3}$$

where the delay $d \in \mathbb{Z}^+$ has been included and $F^{\ell}[\cdot]$ is some nonlinear function of $y(t)$ and $u(t)$ with degree of nonlinearity $\ell \in \mathbb{Z}^+$. Note that if $\ell = 1$ the resulting model is linear.

Leontaritis and Billings (1985a and b) have rigorously proved that a nonlinear discrete system can always be represented as a NARX model around an equilibrium point if i) the response function of the system is finitely realizable and if ii) a linearized model of the system exists around the equilibrium point.

As for the linear case, in parameter estimation problems noise terms are usually required to avoid bias in the parameters, thus the nonlinear autoregressive moving average model with exogenous inputs (NARMAX) (Billings and Leontaritis 1981) is considered

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

where $e(t)$ accounts for uncertainties, possible noise, unmodelled dynamics, etc.

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 $e(t)$. For a theoretical justification for using polynomial NARMAX models see (Chen and Billings 1989).

2.2 Parameter estimation

Equation (4) can be expressed in polynomial form as

$$
\begin{aligned}
y(t) \;=\; & \Bigg[\sum_{i=1}^{n_y} \theta_i \; y(t-i) \;+\; \sum_{i=1}^{n_u} \theta_{n_y+i} \; u(t-i) \;+\; \sum_{i=1}^{n_y} \sum_{j=1}^{n_y} \theta_{i,j} \; y(t-i)y(t-j) \\
& + \; \sum_{i=1}^{n_y} \sum_{j=1}^{n_u} \theta_{i,n_y+j} \; y(t-i)u(t-j) \;+\; \sum_{i=1}^{n_u} \sum_{j=1}^{n_u} \theta_{n_y+i,n_y+j} \; u(t-i)u(t-j) \\
& + \; \textit{higher order terms up to degree } \ell \; \Bigg] \\
& + \; \Bigg[\sum_{i=1}^{n_y} \sum_{j=0}^{n_e} \overline{\theta}_{i,0,j} \; y(t-i)e(t-j) \;+\; \sum_{i=1}^{n_u} \sum_{j=0}^{n_e} \overline{\theta}_{0,i,n_e+j} \; u(t-i)e(t-j) \\
& + \; \sum_{i=1}^{n_y} \sum_{j=1}^{n_u} \sum_{k=0}^{n_e} \overline{\theta}_{i,j,k} \; y(t-i)u(t-j)e(t-k) \\
& + \; \textit{all possible combinations of } y(t),\; u(t) \textit{ and } e(t) \textit{ up to degree } \ell \; \Bigg] \\
& + \; \Bigg[\sum_{i=0}^{n_e} \overline{\overline{\theta}}_i \; e(t-i) \;+\; \sum_{i=0}^{n_e} \sum_{j=0}^{n_e} \overline{\overline{\theta}}_{i,j} \; e(t-i)e(t-j) \\
& + \; \textit{higher order terms up to degree } \ell \; \Bigg]
\end{aligned}
\tag{5}
$$

which can be rewritten as

$$
y(t) = \Psi_{yu}^T(t-1)\theta_{yu} \;+\; \Psi_{yue}^T(t)\theta_{yue} \;+\; \Psi_e^T(t)\theta_e \tag{6}
$$

where the superscript T denotes transposition and $\Psi_{yu}^T(t-1)$ includes all the output and input terms as well as all possible combinations up to degree ℓ and time t-1. The parameters of such terms are in the vector θ_{yu}. The other entries are defined likewise.

Equation (6) is unsuitable for estimating the parameter vector $\theta = [\theta_{yu} \; \theta_{yue} \; \theta_e]$ because the terms $e(t-i) \;\; i = 0, 1, \ldots, n_e$ are not known. To overcome this difficulty equation (6) can be written in the *predictor error* (PE) form

$$
\hat{y}(t) = \Psi_{yu}^T(t-1)\hat{\theta}_{yu} \;+\; \Psi_{yu\xi}^T(t-1)\hat{\theta}_{yu\xi} \;+\; \Psi_{\xi}^T(t-1)\hat{\theta}_{\xi} \;+\; \xi(t) \tag{7}
$$

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

$$
\xi(t) \doteq y(t) - \hat{y}(t) \tag{8}
$$

Finally, equation (7) can be expressed in the concise form

$$\hat{y}(t) = \left[\Psi_{yu}^T(t-1)\ \Psi_{yu\xi}^T(t-1)\ \Psi_{\xi}^T(t-1)\right] \begin{pmatrix} \hat{\theta}_{yu} \\ \hat{\theta}_{yu\xi} \\ \hat{\theta}_{\xi} \end{pmatrix} + \xi(t)$$

$$\hat{y}(t) = \Psi^T(t-1)\hat{\theta} + \xi(t) \tag{9}$$

Billings and Voon (1984) have investigated three alternative algorithms to estimate the parameter vector $\hat{\theta}$.

2.3 Orthogonal parameter estimation

One of the major difficulties in solving equation (9) is that such a set of equations is typically ill-conditioned, especially if the number of terms is large. To circumvent this problem orthogonal techniques may be used (Billings *et al.* 1988, Korenberg and Paarmann 1991). Consider a linear-in-the-parameters polynomial

$$y(t) = \sum_{i=1}^{n_\theta} \theta_i\ p_i(t) + \xi(t) \tag{10}$$

where the $\{p_i\}_{i=1}^{n_\theta}$ represent the different terms in the polynomial. It is noted that this equation is analogous to equation (9).

As a first step, equation (10) is transformed into the following equivalent auxiliary model

$$y(t) = \sum_{i=1}^{n_\theta} g_i\ w_i(t) + \xi(t) \tag{11}$$

where the $\{g_i\}_{i=1}^{n_\theta}$ are constant coefficients and the $\{w_i\}_{i=1}^{n_\theta}$ are constructed to be orthogonal over the data records such that

$$\overline{w_i(t)w_{j+1}(t)} = 0 \qquad i = 1, 2, \ldots, j \tag{12}$$

where the over-bar indicates time averaging.

The second step consists in estimating the coefficients $\{g_i\}_{i=1}^{n_\theta}$ and transform them back to the system parameters $\{\theta_i\}_{i=1}^{n_\theta}$. The parameters of the auxiliary model can be estimated by (Billings *et al.* 1988, 1989, Korenberg *et al.* 1988)

$$\hat{g}_i = \frac{\overline{y(t)w_i(t)}}{\overline{w_i^2(t)}} \qquad i = 1, 2, \ldots, n_\theta \tag{13}$$

provided that $\overline{w_i^2(t)} \neq 0$. The coefficients of the original equation can be obtained from $\{\hat{g}_i\}_{i=1}^{n_\theta}$ according to the formula

$$\hat{\theta}_m = \sum_{i=m}^{n_\theta} \hat{g}_i \, v_i \qquad\qquad m = 1, \ldots n_\theta \tag{14}$$

where

$$\begin{cases} v_m = 1 \\ v_i = -\sum_{r=m}^{i-1} \alpha_{ri} \, v_r \qquad i = m+1, \ldots, n_\theta \end{cases} \tag{15}$$

$$\alpha_{ij} = \frac{\overline{w_i(t) p_j(t)}}{\overline{w_i^2(t)}} \qquad i = 1, \ldots, j-1; \qquad j = 2, \ldots, n_\theta \tag{16}$$

2.4 Structure selection

Polynomial models have a simple structure, provide a closed form for the model and can be used to represent a large class of nonlinear systems. Moreover, it has been said that polynomial models have good interpolation properties and that such models outperform other mathematical representations (Casdagli 1989). 'Polynomials are a good representation because the parameters can be linearly fit to minimize least-squares deviations, and because they arise naturally in Taylor expansions' (Farmer and Sidorowich 1988). A clear disadvantage however is the enormous number of terms a general nonlinear polynomial may have. In particular, the number of terms in the NARMAX model of equation (5) is given by

$$n_\theta = M + 1 \tag{17}$$

where

$$\begin{aligned} M &= \sum_{i=1}^{\ell} n_i \\ n_i &= \frac{n_{i-1}(n_y + n_u + n_e + i - 1)}{i} \qquad n_0 = 1 \end{aligned} \tag{18}$$

This illustrates how the number of terms in a polynomial grows very rapidly even for relatively low values of ℓ, n_y, n_u and n_e. Such an explosion of terms has been pointed out by several authors as the major disadvantage in using polynomials as mathematical

models for nonlinear dynamical systems (Farmer and Sidorowich 1988, Casdagli 1989, Grassberger *et al.* 1991, Giona *et al.* 1991).

Nevertheless, effective and elegant solutions to handle this problem are available, see (Billings *et al.* 1988) and the survey paper by Haber and Unbehauen (1990). One solution is the *error reduction ratio* (ERR) test (Billings *et al.* 1988, 1989, Korenberg *et al.* 1988). Consider the auxiliary model of equation (11). Multiplying this by itself and taking the time average gives

$$\overline{y^2(t)} = \sum_{i=1}^{n_\theta} g_i^2 \, \overline{w_i^2(t)} \; + \; \overline{\xi^2(t)} \tag{19}$$

It is assumed that $\xi(t)$ is a zero mean white sequence which is uncorrelated with the input and output data records and that the orthogonal property of equation (12) holds.

The maximum mean squared prediction error is achieved when no terms are included in the model, that is $n_\theta = 0$, and in this case equals $\overline{y^2(t)}$. Thus from equation (19) the reduction in the mean squared error due to the inclusion of the ith term, $g_i \; w_i(t)$, in the auxiliary model of equation (11) is $g_i^2 \; \overline{w_i^2(t)}$. Expressing this reduction as a percentage of the total mean squared error gives

$$[ERR]_i \doteq \frac{g_i^2 \; \overline{w_i^2(t)}}{\overline{y^2(t)}} \; \times \; 100 \qquad i = 1, 2, \ldots, n_\theta \tag{20}$$

The quantity $[ERR]_i$ 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 its 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, 1989, Korenberg *et al.* 1988).

Other structure selection methods such as the *stepwise regression* algorithm and the *log determinant ratio* test can be found in the literature (Billings and Voon 1986b, Leontaritis and Billings 1988).

2.5 Model validation

Once a model is estimated it should be submitted to a number of tests which should check if the model is adequate and, hopefully, will also provide a measure of goodness for the

model. Billings and Voon (1983; 1986a) have introduced high-order correlation functions to detect the presence of unmodelled terms in the residuals of discrete models. Thus if a model of a system is adequate the following conditions should hold

$$\begin{cases} \Phi_{\xi\xi}(\tau) \doteq E[\xi(t-\tau)\xi(t)] = \delta(0) & \\ \Phi_{u\xi}(\tau) \doteq E[u(t-\tau)\xi(t)] = 0, & \forall\tau \\ \Phi_{u^{2\prime}\xi}(\tau) \doteq E[(u^2(t-\tau) - \overline{u^2(t)})\xi(t)] = 0, & \forall\tau \\ \Phi_{u^{2\prime}\xi^2}(\tau) \doteq E[(u^2(t-\tau) - \overline{u^2(t)})\xi^2(t)] = 0, & \forall\tau \\ \Phi_{\xi(\xi u)}(\tau) \doteq E[\xi(t)\xi(t-1-\tau)u(t-1-\tau)] = 0, & \tau \geq 0 \end{cases} \tag{21}$$

where $\xi(t)$ are the residuals defined in equation (8), $u(t)$ is the input, the over-bar signifies mean value and $E[\cdot]$ denotes mathematical expectation.

In what follows the influence of the input and of the model structure on the quality of the final model will be of concern. In order to better assess this information, bifurcation diagrams and Poincaré sections will be used.

Bifurcation diagrams are well suited for validation purposes because they portray the bifurcation pattern of a system and clearly indicate the variety of dynamical regimes the model exhibits over the range of parameters being considered. Further, It has been shown that such diagrams are more sensitive to certain model variations than other tools developed for nonlinear dynamics such as Poincaré maps, Lyapunov exponents and fractal dimensions (Aguirre and Billings 1993). For details concerning bifurcation diagrams and Poincaré maps see (Hénon 1982, Guckenheimer and Holmes 1983, Parker and Chua 1989).

In order to facilitate the comparison of similar bifurcation diagrams, the following quality index will be used

$$J_b \doteq \sum_{i}^{N_b} (A_i - a_i)^2 w_i \tag{22}$$

where $A_i \in \mathbb{R}$ are the values of the control parameter A for which the system bifurcates. The a_i's are defined likewise for the model being validated and $\mathbf{w} = [w_1\ w_2\ \ldots\ w_{N_b}]$ is a vector of weights which can be chosen to reflect the particular needs of the problem. In this paper $\{w_i\}_{i=1}^{N_b} = 1$ and the summation is taken over all the (N_b) bifurcation points of interest.

2.6 The original nonlinear system

Consider the well known Duffing-Ueda equation (Ueda 1980)

$$\ddot{y} + k\dot{y} + y^3 = u(t) \tag{23}$$

This equation was originally proposed as a model for nonlinear oscillators and has become a bench test for the study of nonlinear dynamics. It has also been considered as a simple paradigm for chaotic dynamics in electrical science (Moon 1987). One of the main reasons for this is that in spite of being simple this model can produce a variety of dynamical regimes.

To obtain the bifurcation diagrams and Poincaré sections shown in this paper, the input was chosen to be of the form

$$u(t) = A\ cos(\omega t) \tag{24}$$

where the maximum input amplitude A was used as the control parameter and the Poincaré sections correspond to the input at phase zero.

The bifurcation diagram shown in figure 1a was obtained by taking $k = 0.1$, $\omega = 1$ rad/s and simulating equation (23) digitally using a fourth-order Runge-Kutta algorithm with an integration interval equal to $\pi/3000$ s. In the figure, the horizontal axis corresponds to the control parameter. As can be seen, varying this parameter in the range $4.5 \leq A \leq 12$ drives the system into a number of different dynamical regimes.

Beginning at $A \approx 4.86$ the system undergoes a period doubling (flip) bifurcation. This happens again at $A \approx 5.41$ and characterizes the well known period doubling route to chaos (Feigenbaum 1983). Another similar cascade begins at $A \approx 9.67$ preceding a different chaotic regime. Two chaotic windows can be distinguished at approximately $5.55 \leq A \leq 5.82$ and $9.94 \leq A \leq 11.64$. At $A \approx 6.61$ the system undergoes a supercritical pitchfork bifurcation and at $A \approx 9.67$ it undergoes a subcritical pitchfork bifurcation. The bifurcation diagram begins and ends with period-1 regimes and displays period-3 dynamics for $5.82 \leq A \leq 9.67$.

Figure 1b shows a Poincaré section of the attractor for $A = 11$. This figure clearly reveals the fractal structure of the chaotic (strange) attractor obtained for this input.

3 Input Design

The identification of non-autonomous systems includes choosing an appropriate input. Optimal input design is not a simple task for linear systems and for nonlinear systems the difficulties are considerably greater. Leontaritis and Billings (1987) have shown that certain inputs which perform very well for linear systems could lead to the loss of identifiability for nonlinear systems.

Although there seems to be no definite rule for designing inputs for nonlinear system identification, it is recognized that the input should fully excite the system nonlinear characteristics, or in other words, the input should be *persistently exciting*. In particular, independent gaussian and uniformly distributed sequences have been suggested (Leontaritis and Billings 1987). Haynes (1989) has suggested the use of composite inputs which can be obtained by superimposing a deterministic signal (usually a square wave) and a gaussian random variable of zero mean and variance σ^2.

In this section the influence of the input on estimated models will be investigated. Bifurcation diagrams will be used to assess the quality of the models estimated using the techniques reviewed in section 2.

3.1 Sinusoidal input

A characteristic feature of nonlinear systems is that energy may be transferred between different frequencies. Thus although the input may have frequencies over a limited range, the output could display a wide spectrum. This is particularly true for non-autonomous chaotic systems where a single sinusoidal input can drive the system into chaos which is characterized by a wide spectrum of frequencies. This means that an input with a relatively poor frequency content may excite all the modes of interest and consequently be considered persistently exciting in nonlinear applications.

For the Duffing-Ueda equation, the input $u_1(t) = 5.8cos(t)$ drives the system into chaos and so does the input $u_2(t) = 11cos(t)$. Figure 2 shows the input/output data for $u_2(t)$ sampled at $T_s = \pi/60$ s. These data were used to estimate a third-order NARMAX model with cubic nonlinearities and eight process terms which were chosen among 83 candidate terms using the ERR test described in section 2.4.

The bifurcation diagram and the Poincaré section of the attractor at $A=11$ are shown in figures 3a and 3c, respectively. Despite the bifurcation diagram of the estimated model being very accurate, $J_b=0.0124$, the attractor is rather distorted.

Adding a zero-mean gaussian distributed noise with variance $\sigma^2=10^{-5}$ to the output time series of figure 2, which has a variance of 3.4, a slightly different model was estimated with the same dynamical order and with the same number of terms as before, although some terms were different. The bifurcation diagram and Poincaré section for this model are shown in figures 3b and 3d, respectively. Although the bifurcation behaviour has slightly deteriorated, $J_b=0.0158$, the shape of the attractor at $A=11$ improved significantly. It should be noted that the bifurcation diagram conveys information about the *global* dynamical behaviour of the system over a range of values of the control parameter while the Poincaré section is a *local* portrait of the attractor for a fixed value of such a parameter.

This example illustrates that i) even the presence of *negligible* noise in the data can affect the performance of the structure selection algorithm, ii) the structure of the model influences the dynamical behaviour, and iii) bifurcation diagrams and Poincaré sections are useful in detecting this influence. It is noted that this information is not recognisable from standard criteria such as predictions and the variance of residuals. Indeed, the variance for the latter model was about 10^5 times the variance of the residuals produced by the model with the distorted attractor.

3.2 Square wave plus gaussian sequence inputs

This type of input has been suggested by Haynes (1989) who has also shown that such an input is more adequate than a pure gaussian sequence.

In the following example a square wave of amplitude ± 10 and frequency equal to 1 rad/s was superimposed to a gaussian-distributed random variable of zero mean and variance σ^2=9.1. Each value of the random variable was maintained during three sampling periods. This can be viewed as a rough presampling filtering of the input signal. The output data were generated by digitally simulating the differential equation (23) using a fourth-order Runge-Kutta algorithm with integration interval equal to $\pi/3000$ s.

Subsequently these data were sampled at $T_s = \pi/30$ s. The input/output data are shown in figure 4.

Taking $\ell = 3$, $n_y = n_u = 3$ and $n_e = 20$ several NARMAX polynomial models with increasing number of terms were estimated. Bifurcation diagrams were computed for each model. The bifurcation points are shown in figure 5. The horizontal axis indicates the number of terms in the model. The vertical axis corresponds to the input amplitude. The dashed lines represent the bifurcation points of the original system and have been labelled B1 to B9 for convenience. Thus B1→B2→B3 is the first period-doubling cascade, and so forth. In this diagram the observed bifurcation points of the identified models are indicated by symbols which also appear at the right side of the figure for reference. Hence bifurcation points corresponding to B1 are marked 'o', those corresponding to B2 are marked '+' and so on. If a model does not exhibit a particular bifurcation for the range of amplitudes considered, such a bifurcation point is omitted in the figure. On the other hand, if a model has spurious bifurcations which are not present in the original bifurcation diagram shown in figure 1, then an asterisk is placed at the bottom of the figure.

This figure shows that none of the estimated models exhibit all the nine bifurcations and, in addition, most models present spurious dynamical regimes. Moreover, the bifurcation points which are reproduced by the estimated models are severely misplaced.

3.3 Square wave with increasing amplitude plus gaussian sequence input

This type of input is similar to the previous one, the main difference is that the amplitude of the square wave in this case is gradually increased. In this example, the square wave amplitude was varied in the range ± 0.5 to ± 13 and the gaussian component had variance $\sigma^2 = 2.0$. The amplitude limits of the square wave were chosen to span over the range of values for which the control parameter was varied in the bifurcation diagrams.

The input/output data for this case are shown in figure 6 and the bifurcation structure of the estimated models are presented in figure 7.

As can be seen, the input/output data of figure 6 has made possible the identification of a family of models which are far better than those represented in figure 5.

3.4 Discussion

For nonlinear systems, as opposed to linear systems, the frequency contents of the output and input signals may be totally different. Thus, a single frequency input may produce an output with wide frequency content in some nonlinear systems. However, some difficulties may arise in a practical situation when a single frequency input is used to excite a nonlinear system. A certain input may not drive the system into chaos or any other regime for which the frequency spectrum of the output is wide enough. Consequently such inputs would not be persistently exciting. Moreover, even if the output is chaotic for a certain input, persistent excitation cannot be guaranteed. For instance, the sinusoidal input $u_1(t)$ drives the system into chaos but the output in this case has a rather poor frequency content. This can be explained by noticing that the largest *Lyapunov exponent* [1] of this attractor is smaller than the one for the attractor obtained with the sinusoidal input $u_2(t)$.

It is noted that the frequency spectra of the inputs shown in figures 5 and 7 are virtually *identical.* Nonetheless the latter input produces results which are clearly better than those obtained with the former input. It is interesting to note that the output in figure 6 clearly displays a wider variety of dynamical regimes than the output in figure 4. This illustrates that in designing an input for the identification of nonlinear systems amplitude considerations have also to be taken into account. This contrasts with the approach followed for linear systems where the frequency content of the input is of fundamental importance as opposed to considerations regarding the amplitude. This is unsurprising because linear systems do not transfer energy between two different frequencies and in real applications the input amplitude is usually required to be low in order not to excite the nonlinearities of the system undergoing experimentation.

For nonlinear systems, however, the requirements on the frequency content of the input may be somewhat relaxed due to the ability of such systems to transfer energy between different frequencies and actually *create* new frequencies at the output. On the other hand, greater attention must be paid to the *amplitude profile* of the input as illustrated in figures 5 and 7.

[1] Lyapunov exponents measure the average divergence of nearby trajectories along certain directions in state space (Wolf 1986). As the largest Lyapunov exponent is increased the more chaotic the attractor becomes and as a consequence the output signal will have a richer frequency content.

In what follows the noise-free input/output data shown in figure 6 will be employed using different sampling rates.

4 Influence of Model Structure

The influence of the model structure on the quality of the identified models is investigated in this section. In particular, three aspects of the model structure will be of concern, namely the sampling period, T_s, the total number of process terms in the model, n_p, and the maximum lag considered for the process model, n_y. It is assumed that the structure of the noise model is adequate and thus the process model is unbiased. This can be readily verified in practice by applying the correlation tests of equation (21). For the sake of simplicity it is further assumed that $n_u = n_y$.

In order to investigate how T_s, n_p and $n_y = n_u$ influence the dynamical properties of the estimated models, these parameters were varied within limited ranges and for each combination a model was identified using the techniques reviewed in section 2. Three values for T_s were considered, namely $\pi/30$, $\pi/60$ and $\pi/100$ s. The other two parameters were varied over the ranges $2 \leq n_y \leq 6$ and $4 \leq n_p \leq 14$. Subsequently, for each identified model a bifurcation diagram was obtained and the bifurcation quality index of equation (22) was also computed. The best models for $T_s = \pi/30$, $\pi/60$ and $\pi/100$ s respectively are

$$\begin{aligned} y(t) &= 2.0503y(t-1) - 1.1103y(t-2) + 0.056964y(t-3) \\ &+ 0.29799 \times 10^{-2} y(t-4) + 0.48803 \times 10^{-2} y(t-1)y(t-2)y(t-4) \\ &- 0.21922 \times 10^{-1} y(t-1)^3 + 0.73520 \times 10^{-2} u(t-2) \\ &+ 0.26938 \times 10^{-2} u(t-1) + 0.23293 \times 10^{-1} y(t-1)^2 y(t-2) \\ &- 0.031055y(t-1)y(t-2)y(t-3) + 0.022597y(t-1)y(t-3)^2 \\ &- 0.79855 \times 10^{-2} y(t-1)y(t-3)y(t-4) + 0.12640 \times 10^{-3} u(t-3) \end{aligned} \tag{25}$$

$$\begin{aligned} y(t) &= 2.1579y(t-1) - 1.3203y(t-2) + 0.16239y(t-3) \\ &+ 0.22480 \times 10^{-3} y(t-3)^3 - 0.48196 \times 10^{-2} y(t-1)^3 \end{aligned}$$

$$
\begin{aligned}
&+ \ 0.19463 \times 10^{-2} u(t-2) + 0.34160 \times 10^{-3} u(t-1) \\
&+ \ 0.35230 \times 10^{-2} y(t-1)^2 y(t-2) \\
&- \ 0.12162 \times 10^{-2} y(t-1) y(t-2) y(t-3)
\end{aligned} \tag{26}
$$

$$
\begin{aligned}
y(t) \ &= \ 1.9969 y(t-1) - 0.99697 y(t-2) - 0.98014 \times 10^{-3} y(t-1)^3 \\
&+ \ 0.97769 \times 10^{-3} u(t-2)
\end{aligned} \tag{27}
$$

As can be seen, for the model of equation (25), $n_p = 13$, $n_y = n_u = 4$, for the model of equation (26), $n_p = 9$, $n_y = n_u = 3$ and for the model of equation (27), $n_p = 4$, $n_y = n_u = 2$. The bifurcation quality index for these models were respectively 0.023, 0.012 and 0.180.

Nearly seventy models were estimated corresponding to several combinations of the maximum lag n_y, the number of process model terms n_p and the sampling period T_s. It should be noted that the model of equation (25) was the only model estimated from data sampled at $T_s = \pi/30$ s which had a complete bifurcation diagram in the sense that the model exhibited all of the bifurcation points of the original model and had no spurious bifurcations although a chaotic window is beginning to develop for $8.65 \leq A \leq 8.95$. This spurious regime will be fully established as the number of terms in the model increases. This phenomenon will be discussed in section 5.

The model of equation (27) was the only model identified from data sampled at $T_s = \pi/100$ s. After including the fourth term, the structure selection algorithm reached the limit $\sum ERR_i = 1$ and therefore stopped the term-selection procedure.

Twenty nine models were estimated from the data sampled at $T_s = \pi/60$ s. Fifteen of these models had complete and non-spurious bifurcation diagrams. Thus these results favour the use of sampling periods around $T_s = \pi/60$ s for identification purposes.

In figure 8 $n_y = n_u$ and n_p are plotted against T_s for the models of equations (25)-(27). This figure reveals that, for the systems and inputs considered in the present study, as the sampling time is increased 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. To illustrate this point, consider the bifurcation diagrams of the models obtained with

$T_s = \pi/30$ s, $n_y = n_u = 3$, $n_p = 11$ and $T_s = \pi/60$ s, $n_y = n_u = 3$, $n_p = 7$ which are shown in figures 9a and 9b, respectively. Clearly the overall bifurcation behaviour is very similar but the model estimated from the data which has been sampled faster has fewer terms.

Since the models of equations (25)-(27) have comparable bifurcation behaviours and from the analysis of figure 8, it 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 noted that in practice the increase in $n_y = n_u$ usually implies a larger number of terms in models of comparable performance.

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 oversampled 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 resorces 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 1989). In this reference it has been reported 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* [2] which is analogous to the maximum lag n_y.

[2] Consider a time series $x(t)$, $t \in \mathbb{Z}^+$. The vector $\mathbf{x}_m(t) = [x(t)\ x(t-1)\ \ldots\ x(t-m-1)]^T$ can be represented as a point in an $\mathbb{R}^m$ pseudo-state-space where m is the *embedding dimension*. The sequence $\mathbf{x}_m(t)\ \mathbf{x}_m(t-1)\ \ldots$ is a *reconstruction* of $x(t)$ in such a space. It has been proved that if $m \geq 2n+1$ qualitative and quantitative dynamical properties of the n*th*-order system which generated $x(t)$ can be recovered from the reconstructed sequence (Packard *et al.* 1980, Takens 1981)

5 Estimated and discretised models

Discrete models estimated from sampled input/output data of continuous systems 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 was 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

$$\dot{y}(k) \doteq \lim_{T_s \to 0} \frac{y(k) - y(k-1)}{T_s} \tag{28}$$

can be used to obtain discretised models directly from equation (23).

Figures 10a, 10c and 10e are the bifurcation diagrams of the discretised models with $T_s = \pi/100$, $T_s = \pi/60$ and $T_s = \pi/30$ s. Figures 10b, 10d and 10f correspond to the models estimated from data sampled also at such rates, see equations (27), (26) and (25). Figures 10g, 10i, 10k and 10h, 10j and 10l are the respective Poincaré sections.

From these figures it is clear that the increase in T_s tends to deteriorate the dynamics of the discretised models, as would be expected from equation (28) and that such degradation is best revealed by the Poincaré sections. Increasing T_s also has the effect of shifting the bifurcation points to the left. Thus the discretised models with larger T_s will tend to bifurcate at lower values of the control parameter A.

On the other hand, the identification procedure tends to compensate for the increase in T_s by including more terms in the model and reestimating the parameters. Consequently the estimated models are less sensitive to variations in T_s than the discretised counterparts (this can be better appreciated by comparing the Poincaré sections). This indicates that different criteria should be used to chose suitable values for T_s in identification and

discretization applications. Furthermore, classical tools such as predictions, correlation functions, variance of residuals, power spectra, etc. are rather insensitive to variations of T_s (Aguirre and Billings 1992).

Figure 11 summarizes many features of figures 10a–l. For each model in the latter figures a quality index was calculated taking into account both the bifurcation diagram structure and the shape of the strange attractor for $A = 11$. Models discretised with $T_s = \pi/300$ and $T_s = \pi/15$ s have also been included in figure 11.

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 beyond which the discretised models are always better than the estimated ones. Thus if T_s is decreased beyond the lower bound the discretised models will gradually tend to the continuous system since the approximation in equation (28) becomes more accurate as $T_s \to 0$. Besides, for oversampled data a number of numerical problems may arise in the parameter estimation.

An example of this is the model estimated from data sampled at $T_s = \pi/100$ s. Due to numerical problems only four terms were included. Deterioration in the bifurcation diagram can be verified by noticing that the chaotic window at $A \approx 5.8$ is much narrower than the original one, compare figures 10a–b.

On the other hand, 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 is lost and therefore cannot be retrieved by the estimation algorithm. In this case the estimated models will also tend to be worse than the discretised counterparts. It is noted that none of the models estimated from data sampled at $T_s = \pi/15$ s had complete bifurcation diagrams.

6 Model Overparametrization

A major difficulty in the identification of nonlinear systems is the huge number of terms and parameters required to model relatively simple nonlinearities. This is the chief dis-

advantage of classical methods. For instance, the identification via Wiener's method of a simple system containing a second-order nonlinearity would require the evaluation of, typically, 10^{10} coefficients (Billings 1980). Even for polynomials the number of terms may become impractically large even for moderate values of ℓ, n_y, n_u and n_e, see equation (18).

The structure selection algorithm described briefly in §2.4 and many other algorithms available in the literature aim at enabling the experimentalist to choose systematically the terms which are more important and therefore estimate parsimonious models.

Unfortunately, there is a tendency to allow more terms than necessary in a model. Possible reasons for this are i) inability to adequately choose the truly important terms, ii) some structure selection algorithms may be too time-consuming, iii) the variance of the residuals usually decreases monotonically as the number of terms in a model is gradually increased, iv) large models are more flexible and consequently tend to fit the estimation data better. It should be noted that flexible models which fit the data better do not necessarily capture the underlying dynamics appropriately. Consequently items iii and iv are misleading and may well be dangerous.

Recently, alternative model structures have been investigated such as neural networks and radial basis functions. Such models are typically very flexible and this usually enables a very good fit to the data. However, these models can become too complex very easily and therefore the reasons listed above are likely to be verified in such cases.

This section investigates the consequences of overparametrization in NARMAX polynomial models. It is believed that the conclusions also apply to different model structures.

6.1 Minimum and maximum number of terms

The minimum number of terms which should be included in a model can be roughly thought of as the minimum number of terms required such that the estimated model passes some validity test. In other words the minimum number of terms is directly associated with the size of the simplest *valid* model.

In practice, this number can be found by gradually allowing extra terms in a model and verifying if the final model is valid or not using, for instance, the correlation tests of equation (21). Such terms could be selected based on a criterion such as ERR described

in §2.4 When the estimated model is nonlinear, it is usually difficult to determine the best size for a model and more terms than strictly necessary are often included.

If the data were generated by a discrete map, the number of terms and the order of such a system would determine the 'correct' structure of the model to be identified. However in practice the data usually proceed from continuous systems whose mathematical representations are unknown. Thus it seems that no maximum number of terms can be confidently assumed *a priori*. What is clear however is that such a number depends on the sampling period (note that this only applies for data generated by continuous systems).

In order to further investigate this point, consider the family of models estimated from data similar to those of figure 6. The only difference is that in this case the data were sampled at $T_s = \pi/60$ s. Fixing the maximum lags allowed as $n_y = n_u$=3, a number of models with different numbers of terms were estimated from the data. The bifurcation points of these models are shown in figure 12.

Comparing figures 7 and 12 reveals that $T_s = \pi/60$ s is a better choice for the sampling period in this example. A clear difference between these figures is that all of the models represented in figure 12 exhibit all the bifurcation points.

It is worth pointing out that the models with six, seven, twelve, thirteen and fourteen terms have spurious bifurcation points, see asterisks at the bottom line of figure 12. The evident displacement of the bifurcation points of the models with six and seven terms would certainly prevent such models from being considered adequate. By contrast, the models with twelve, thirteen and fourteen terms exhibit very accurate bifuraction behaviour but such models also present spurious or ghost bifurcations which indicate dynamical regimes which are qualitatively different from the original system.

Based on figure 12 it is clear that for the chosen values of T_s and the maximum lag $n_y = n_u$, the minimum and maximum number of terms are eight and eleven, respectively. This figure illustrates that a model which is unnecessarily complex may induce spurious dynamical regimes and this has important implications for model structure selection.

Figure 13 shows the bifurcation quality index, J_b, of the models identified from the same data as the models of figure 12 for different values of maximum lag $n_y = n_u$ and total number of process model terms n_p. It is noted that only models with complete and non-spurious bifurcation diagrams were considered in this figure.

This reveals that i) fixing the order of the system, $n_y = n_u$, there is an optimum number of terms beyond which the bifurcation diagram begins to deteriorate, and ii) there seems to be an optimum value for $n_y = n_u$ beyond which the bifurcation diagram also deteriorates.

Figures 12 and 13 indicate that there is a minimum number of terms to be included in the model in order to be able to capture the system dynamics, that there is a maximum number of terms beyond which spurious dynamical regimes are induced in the model and that in between these limits lies an optimum [3] number of terms.

Figures 14a–c and 15a–c show the bifurcation diagrams and Poincaré sections of the models estimated with $T_s = \pi/30$ s, $n_y = n_u = 4$, $n_p = 14$ and $T_s = \pi/60$ s, $n_y = n_u = 3$, $n_p = 14$, respectively. In the bifurcation diagram of figure 14a, spurious supercritical and subcritical flip bifurcations occur at $A \approx 7.2$ and $A \approx 7.6$, respectively. Furthermore, a spurious chaotic window is observed at approximately $8.65 \leq A \leq 8.95$. Figure 14c shows the Poincaré section of the attractor at A=8.85. This clearly reveals that a well developed spurious chaotic regime has been induced as a consequence of overparametrization. It should be noted that this chaotic regime has been observed for all of the estimated models with $T_s = \pi/30$ s, $n_y = n_u = 4$ and $14 \leq n_p \leq 17$. The spurious flip bifurcations have been observed for models with $n_p = 12$, 14 and 15.

A similar situation has also been verified for models estimated from data sampled at $T_s = \pi/60$ s. The respective diagrams are presented in figures 15a–c.

Figures 14 and 15 also illustrate that the Poincaré section of the attractor at $A = 11$ is rather insensitive to overparametrization. Thus although the extra terms included in these models have induced spurious bifurcations many of the original dynamical regimes remain very much the same. In other words, the behaviour of the identified model can be quite different to the behaviour of the underlying system. This highlights the need for a thorough and systematic validation procedure. In spite of this, validation is sometimes based solely on how well a model predicts over a different set of data, see for example Pottmann *et al.* (1993), Masri *et al.* (1993) and most papers which use neural networks

[3] This is not a mathematical assertion but rather an observation inferred from figures 12 and 13 where the optimum is in terms of the quality index, J_b, defined in equation (22). This optimum is related to ERR, thus the use of different criteria to select the terms may lead to different optima.

to model systems. But predictive model performance does not give any indication about how adequate the structure of the identified model is. An additional validation step would be to check for correlation in the residuals via the set of equations (21). These tests indicate if the minimum number of terms has been included in order to guarantee unbiased estimates but do not reveal if a model is overparameterised or not.

The best way to use such tests therefore is to increase the model complexity untill the tests are satisfied. But even then some mechanism for deciding the order in which terms are added to the model is required. Structure detection, or the selection and inclusion of only significant model terms can therefore be critically important when estimating any form of model for nonlinear systems. Bifurcation diagrams and Poincaré sections, which give detailed information concerning local and global dynamical behaviour, are well suited for validation (Aguirre and Billigns 1993). These two methods are very valuable when used as tools to measure the quality and sensitivity of model estimation procedures with respect to all the factors which influence the estimated model.

6.2 True identification versus data fitting

The parameters in a NARMAX model are estimated by minimizing, see equation (9),

$$J_1 \doteq \| y(t) - \Phi^T(t-1)\hat{\theta} \| \tag{29}$$

where the $\| \cdot \|$ is the Euclidean norm.

Intuitively, the more terms included in a model, the more 'flexible' such a model will be and a better fit to the estimation data will also be possible. This is particularly true in nonlinear applications because the nonlinearities in the data will require a greater flexibility from the models to adjust to the records. This is performed by minimizing J_1 over the estimation data set. Implicit in this process is the hope that a model which fits the data well will also capture the underlying dynamics of the true system in some appropriate manner.

Increasing the number of terms in a polynomial model enhances the ability of the model to fit the estimation data. Figure 16 shows that the variance of the residuals decreases monotonically as the number of terms in the model increases. This is normally all the information that is available to the algorithm during term selection and

parameter estimation and it is therefore understandable why there is a tendency towards overparametrization. This is also one of the reasons why the validation should always be performed using a different data set, a testing set.

To illustrate this point further, consider the model of equation (26). This was the best estimated model. Using this model to predict over the estimation data yielded the time series shown in figure 17a. This is a typical scenario for a chaotic system, namely the short term predictions are very accurate but gradually the model output drifts away from the measurements as a consequence of the sensitive dependence on initial conditions caused by local divergence of trajectories in the state space along the directions associated with the positive Lyapunov exponents.

Allowing four extra terms and reestimating the model parameters gives a thirteen-term model which predicts the estimation data extremely well, see figure 17b. However this model, as noted in figure 12, includes spurious bifurcations. Indeed, the thirteen-term model exhibits spurious chaotic attractors at $8.61 \leq A \leq 8.85$ and spurious period-one motions at $8.85 \leq A \leq 9.10$ and at $9.22 \leq A \leq 9.31$. This clearly illustrates that a model which fits the estimation data better is not necessarily the best model. This is a very important point. In a lot of the literature on estimation and especially neural networks, the claim is often made that one model produces superior predictions compared with another and must therefore be better. The prediction over a chaotic data set, see figure 17a, provides a good illustration of the folly of such a conjecture. The true model of the system is expected to exhibit poor long-term predictions. This is an inherent and fundamental property of systems which display extreme sensitivity on initial conditions. Any model therefore which produces apparently excellent predictions, as in figure 17b, cannot possibly be representative of the underlying system, it is just a curve fit to one piece of data.

Recently it has been stated that neural 'networks with an excess of adjustable parameters may be expected to tune to experimental noise, and therefore although the training set is captured essentially perfectly, the predictions on new data may be poor' (Adomaitis *et al.* 1990). These authors have also reported that 'pruned' networks often outperform the unpruned networks. Similar results concerning the size of neural networks have been reported by Masri *et al.* (1993).

6.3 Information criteria

It is clear that the number of terms in a model should be trade-off with the goodness of fit attained with such a model. Four information criteria are listed below which are composed of two terms. The first term measures how well the model fits the estimation data and the second term penalises models with a large number of terms. The criteria considered were

(a) The *final prediction error* (FPE) (Akaike 1974)

$$FPE = Nln\left[\sigma^2(n_p)\right] + Nln\frac{N+n_p}{N-n_p} \tag{30}$$

(b) *Akaike's information criterion* (AIC) (Akaike 1974)

$$AIC(\alpha) = Nln\left[\sigma^2(n_p)\right] + \alpha n_p \qquad \alpha > 0 \tag{31}$$

(c) Khundrin's *law of iterated logarithm criterion* (LILC) (Hannan and Quinn 1979)

$$AIC(\alpha) = Nln\left[\sigma^2(n_p)\right] + 2n_p ln\ lnN \tag{32}$$

(d) The *Bayesian information criterion* (BIC) (Kashyap 1977)

$$AIC(\alpha) = Nln\left[\sigma^2(n_p)\right] + n_p lnN \tag{33}$$

where N is the number of data points and $\sigma^2(n_p)$ is the variance of the residuals associated to the n_p-term model. These and other criteria have been reviewed in (Gooijer *et al.* 1985) and have been suggested as a means of selecting the best number of terms in nonlinear polynomial models by Kortmann and Unbehauen (1988).

Tables 1 and 2 show the values of the four information criteria when applied to the families of models obtained for $T_s = \pi/30$ s, $n_y = 4$, $7 \leq n_p \leq 18$ and $T_s = \pi/60$ s, $n_y = 3$, $6 \leq n_p \leq 14$. For the first family $N = 900$ and for the second $N = 1800$. It should be noted that, based on bifurcation information, the best individuals of these families were found to be the models of equations (25) and (26) respectively.

Table 1. Information criteria for $T_s = \pi/30$ s, $n_y = 4$

n_p	FPE	AIC(4)	LILC	BIC
7	-8523	-8509	-8510	-8489
8	-10021	-10005	-10007	-9983
9	-10361	-10343	-10345	-10318
10	-10538	-10518	-10520	-10490
11	-10645	-10623	-10625	-10592
12	-10895	-10871	-10873	-10837
13	-10896	-10870	-10872	-10833
14	-10914	-10886	-10889	-10847
15	-10922	-10892	-10895	-10850
16	-10929	-10897	-10899	-10852
17	-10933	-10899	-10902	-10851
18	-10967	-10931	-10935	-10881

In the first case, AIC(4), LILC and BIC indicated that the correct structure was that of the model with twelve terms. Though the bifurcation diagram of this model is similar to the diagram of the best model, it presents two spurious bifurcations, namely a supercritical flip at $A \approx 7.28$ and a subcritical flip at $A \approx 7.52$.

The information criteria were originally proposed and applied in the context of linear systems. The correct structure would then be indicated by the value of n_p for which the criteria reached a minimum value (Gooijer *et al.* 1985). As can be seen from Table 1, AIC(4), LILC and BIC have a *local minimum* at $n_p = 12$ which is clearly not the global minimum. Besides, The FPE criterion does not have a minimum in the considered range although its first derivative is nearly null for $n_p = 12$ and the BIC has another local minimum at $n_p = 16$.

Table 2. Information criteria for $T_s = \pi/60$ s, $n_y = 3$

n_p	FPE	AIC(4)	LILC	BIC
6	-24897	-24885	-24885	-24864
7	-26196	-26182	-26182	-26158
8	-27019	-27003	-27003	-26975
9	-27109	-27091	-27091	-27059
10	-27108	-27088	-27088	-27053
11	-27647	-27625	-27624	-27586
12	-27993	-27969	-27968	-27927
13	-28014	-27988	-27988	-27943
14	-28228	-28200	-28199	-28151

A similar situation can be observed in Table 2 where the four criteria have a *local minimum* at $n_p = 9$ which was exactly the best model obtained from bifurcation analysis. Again, no global minimum was observed within the considered range.

6.4 The model structure space (MSS)

Consider an $n_y = n_u$th-order NARX model with n_p process terms which has been estimated from data sampled at T_s s. The structure of this model can be characterised, to some extent, by the numbers T_s, n_p and $n_y = n_u$ and, hence, can be represented as a point in an $\mathbb{R} \times \mathbb{N}^2$ space which will be called the *model structure space* (MSS). Clearly, different model sets may result in the same selected structure and can therefore be represented as a single point in the MSS.

From the preceding sections it is clear that there are lower and upper bounds on the parameters which determine the model structure. Thus $T_{s.min} \leq T_s \leq T_{s.max}$, $n_{p.min} \leq n_p \leq n_{p.max}$ and $n_{y.min} \leq n_y \leq n_{y.max}$. Such bounds exist due to different reasons such as i) theoretical, $T_{s.max}$ is limited by the sampling theorem, ii) numerical, if $T_s \ll T_{s.min}$ the problem may become numerically ill-conditioned, iii) dynamical, if $n_p < n_{p.min}$ the model may not capture the dynamics of the original system and if $n_p > n_{p.max}$ spurious dynamical regimes may be induced, and so on.

For the examples presented earlier, the following practical values can be attributed to the limits: $T_{s.min} \approx \pi/100$ s, $T_{s.max} \approx \pi/30$ s, $n_{y.min} = 2$, $n_{y.max} = 5$, $n_{p.min} = 4$ and $n_{p.max} = 13$.

These limits define a polyhedron in the MSS, see figure 18, where the best models estimated for $T_s = \pi/30$, $\pi/60$ and $\pi/100$ s were represented by black dots. The circles or white dots represent the inaccurate models corresponding to figures 9, 14 and 15. The oval dots are the projections of the model structures on the planes $T_s \times n_y$, $T_s \times n_p$ and $n_p \times n_y$.

One of the principal results reported so far is that the limit values of n_p and $n_y = n_u$ depend on T_s, as illustrated in figure 8. Specifically, $n_{p.max}$ and $n_{y.max}$ decrease as T_s is made smaller. Taking this into account a subregion $\mathcal{Q}$ of the MSS can be expected to contain the best estimated models (equations (25)-(27)). This has can be verified from figure 18. Moreover, the best models presented in figure 13 are also in $\mathcal{Q}$ and the projections fall on the shaded regions on the projection planes.

Figure 18 might give the impression that the subspace $\mathcal{Q}$ has a rather constant transversal section for $n_y \geq 4$. This is because $\mathcal{Q}$ is only defined for integer values of n_y and n_p and, in the scales used, it would be difficult to represent the narrowing of $\mathcal{Q}$ along the n_y direction. This can be better appreciated in figure 13 by noticing that the bifurcation characteristics of the models deteriorate for $n_y \geq 4$.

Because the best models are confined to $\mathcal{Q}$, it would be helpful to have rough estimates for the limit values of T_s, n_y and n_p. This is discussed in the following.

$T_{s.max}$: This is limited by the sampling theorem as $T_{s.max} \leq 1/2f_m$ s, where f_m is the highest frequency of interest in the data.

$T_{s.min}$: If T_s is taken too small, consecutive samples in the data will be highly correlated. The choice of $T_{s.min}$ can be based on the first zero-crossing of the autocorrelation function of the data or on information theory (Fraser and Swinney 1986). $T_{s.min}$ can, of course, be also determined based on f_m, say, as $T_{s.min} \geq 1/20f_m$ s.

$n_{y.min}$: The minimum number of degrees of freedom required to characterize the dynamics underlying the data can be obtained from estimates of the fractal dimension D_f [4].

[4]There are several kinds of fractal dimensions such as the information dimension, the correlation dimension, etc. (Grassberger 1986). It should be noted that D_f characterizes the attracting set (data)

If the system which produced the data has dynamical order n, then $D_f < n$ (Moon 1987). Methods for the direct determination of the minimum embedding dimension are available in the literature (Broomhead *et al.* 1987, Aleksić 1991, Kennel *et al.* 1992). If known, the order of the discretised model can be used.

$n_{y.max}$: Apparently there is no way of determining an *a priori* value for this limit. A theorem due to Takens (1981) states that the reconstructed attractor will retain the original characteristics if the embedding dimension [5] is greater than or equal to $2D_f + 1$. Although this provides a sufficient lower bound, in practice it has been found that this value may often be larger than is required (Broomhead *et al.* 1987, Marteau and Abarbanel 1991, Kennel *et al.* 1992) and therefore could be used as an upper bound.

$n_{p.min}$: This limit should mark the minimum number of terms for which some validation criterion is satisfied, for instance the correlation functions of equation (21). If known, the number of terms of the discretised model can be used.

$n_{p.max}$: There seems to be no way of determining an *a priori* value for this upper bound. Consequently, in a practical situation the searching subspace would be as shown in figure 19 where this subspace may extend along the n_p axis for thousands of units. This emphasizes the need for efficient and reliable algorithms for term selection. Thus in practice information criteria may be computed as the search advances along the n_p direction (note that in this direction, due to the properties of the term selection algorithm of §2.4, the structures are organised in a *hierarchical* or *nested* way) and the search may be halted not immediately but shortly after any of the criteria reaches a minimum.

These considerations help to reduce the search space. This is very desirable in practice since term selection is usually very time consuming. Different model structures, such as neural networks, would be difficult to represent in the MSS. However it is believed that for neural networks the 'subspace' in which good models are likely to be estimated is also limited.

It is noted that figure 18 is just a qualitative representation of some of the results described in the paper. No attempt has been made to define or quantify the shape and dimensions of the subspace $\mathcal{Q}$. Nevertheless, this figure helps to illustrate the following

---

and not necessarily the system which produced the data.

[5] See footnote 2

ideas

(a) The parameters n_y and n_p depend on T_s, in such a way that the former tend to decrease as the sampling period is shortened (this tendency is contrary to a widespread belief that n_y increases as T_s is made smaller).

(b) The subspace $\mathcal{Q}$ is *thicker* in the middle part corresponding to an *optimal* range of sampling periods. This is analogous to the information in figure 11.

(c) The lower end of $\mathcal{Q}$ coincides with the discretization subspace $\mathcal{D}$ for small values of T_s. This is consistent with the fact that the discretised models are more accurate as $T_s \to 0$.

(d) The upper end of $\mathcal{Q}$ is limited by the sampling theorem but good models tend to be found for sampling periods which are shorter than the upper limit imposed by this theorem.

(e) The shape and dimensions of $\mathcal{Q}$ are not as important as the fact that this subspace is *limited*. This is in accordance to experimental results which show that valid models tend to be found in a bounded region of the MSS, and that increasing n_y and/or n_p beyond certain limits deteriorates the quality of the identified models.

7 Conclusions

This paper has been concerned with the effect of some aspects of the identification of nonlinear systems on the dynamical characteristics of the estimated models.

The first aspect which has been investigated was input design. In linear systems the choice of the input signal is almost invariably specified based on frequency considerations in such a way as to guarantee that the signal is *persistently exciting*. For nonlinear systems not only the *frequency content* but also the *amplitude profile* of the input should be carefully chosen. Because of the energy transfer among different frequencies in nonlinear systems, the requirements on the frequency content of the input may be somewhat relaxed. On the other hand, because different dynamical regimes may be induced by simply altering

the amplitude of a sinusoidal input, additional excitation can be attained by specifying a particular amplitude profile. This has been shown by way of an example.

This paper has also investigated how several parameters which characterize the structure of a nonlinear polynomial affect the dynamical behaviour of this model. The parameters considered were the sampling rate, the number of process terms and the order or maximum lag of the model. Furthermore, comparing models with roughly the same quality, it was possible to verify how these parameters affect each other. It has been shown that a decrease in the sampling period is associated with a decrease in the number of terms and in the dynamical order of the model. This leads to the conclusion that using slightly oversampled data enables the estimation of models with simpler structure.

In order to gain further insight into the dependence of the various parameters characterizing the structure, the estimated models have been compared to the discretised counterparts obtained analytically from the original system. This has revealed that the process of including terms in a model may be viewed, to a certain extent, as a *compensation* for the accuracy lost in the sampling and digital representation of the data. The quality of this compensation is directly related to the sampling period and to the ability of choosing the best terms to include in the model. Furthermore, results have been reported which suggest that there is a limited range of sampling periods for which this compensation is possible.

Regarding model complexity, it has been argued that if a nonlinear model is unnecessarily complex, that is if the model has more terms or if it is of higher order than the *optimum*, such a model is prone to induce spurious dynamical regimes. This deleterious aspect of model overparametrization is in contrast to the enhanced predictive ability that such models exhibit over the estimation data. Examples have been provided to illustrate that models which fit the estimation data better are not necessarily the models which capture the underlying dynamics adequately. This is highly relevant because as a consequence of the ever increasing computational power available, there seems to be a *natural* tendency to overparameterise nonlinear models.

The performance of four information criteria (FPE, AIC(4), LILC and BIC) in selecting the best model from within a family has also been investigated. In one case the four criteria indicated a model which was also found to be the best based on bifurcation

analysis. In another example, however, three criteria indicated a twelve-term model as the optimum. Bifurcation analysis revealed that the thirteen-term model was the most accurate and that the twelve-term model exhibited spurious dynamical regimes. In both cases, these criteria only had *local minima* within the considered range as opposed to the global minimum usually verified in linear systems. Apparently a good choice (not necessarily the best, as one of the examples has revealed) would be the model corresponding to the first local minimum. This was found to be in accordance with the results of (Kortmann and Unbehauen 1988).

An additional pitfall is that most of the harmful effects of overparametrization are not revealed by classical tools used in signal processing. In this paper bifurcation diagrams and Poincaré sections have been used for this purpose. It is interesting to note that the former diagrams were particularly useful in revealing the dependence of the dynamics on the number of terms and the order of the models while Poincaré sections were found to be important in disclosing how the sampling rate affects the dynamics.

Finally, some of the estimated models considered in the paper have been mapped in what has been called the *model structure space* (MSS). In this space each different model structure can be represented by a point. This has been done only for NARMAX polynomial models though similar spaces could be defined for some different structures.

Although the definition of the MSS was made on empirical grounds, representing the models in such a space has been useful for illustrating many of the aspects investigated in the paper. Some aspects are

(a) The relationship among the sampling period, the number of process terms and the dynamical order of the models

(b) The structure of the discretised models appear as a limit subspace in the MSS

(c) Considerations on the bounds which limit the search subspace in the MSS illustrate the need for algorithms which, in a viable time, should be able to select the most important terms to be included in a model

(d) The subspace of the MSS where valid models are likely to be found is limited, thus highlighting the dangers of overparametrization

The field of nonlinear dynamics has experienced a great deal of excitement in the last few years as a consequence of the *chaos* advent. The main reason for this was that very complicated dynamics, even unpredictable motions, could be produced and therefore modelled by simple deterministic equations. If on the one hand the *chaos revolution* was disclosing the fact that apparently random dynamical regimes could be characterised without resorting to stochastic theory, on the other hand it was reminding practitioners that simple models would be enough to model complicated dynamics. Thus it seems paradoxical that when identifying nonlinear (perhaps chaotic) systems large models and networks tend to be used almost as a default.

ACKNOWLEDGMENTS

LAA gratefully acknowledges financial support from the Brazilian Council of Scientific and Technological Development - CNPq, under grant 200597/90-6. SAB gratefully acknowledges that this work was supported by SERC under grant GR/H 35286.

References

ADOMAITIS, R.A., FARBER, R.M., HUDSON, J.L., KEVREKIDIS, I.G., KUBE, M. and LAPEDES, A.S., 1990. Application of neural nets to system identification and bifurcation analysis of real world experimental data, in *Neural Networks: Biological computers or electronic brains*, Springer Verlag: Paris, 87–97.

AGUIRRE, L.A. and BILLINGS, S.A., 1992. Digital simulation and discrete modelling of a chaotic system, Research report no. 447, Dept. of Automatic Control and Systems Engineering, University of Sheffield, Sheffield, UK.

AGUIRRE, L.A. and BILLINGS, S.A., 1993. Validating identified nonlinear models with chaotic dynamics, Research report no. 469, Dept. of Automatic Control and Systems Engineering, University of Sheffield, Sheffield, UK.

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

ALEKSIĆ, Z. 1991. Estimating the embedding dimension, *Physica D*, **52**, 362–368.

BILLINGS, S.A., 1980. Identification of nonlinear systems — a survey, *IEE Proceedings*, **127 D**(6), 272–285.

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

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

BILLINGS, S.A., 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 LEONTARITIS, I.J. 1981. Identification of nonlinear systems using parametric estimation techniques, *IEE Conf. Control and its applications*, Warwick, 123–127.

BILLINGS, S.A. and VOON, W.S.F., 1983. Structure detection and model validity tests in the identification of nonlinear systems, *IEE Proceedings*, **130 D**(4), 193–199.

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., 1986a. A prediction-error and stepwise-regression estimation algorithm for non-linear systems, *Int. J. Control*, **44**(3), 803–822.

BILLINGS, S.A. and VOON, W.S.F., 1986b. A prediction-error and stepwise-regression estimation algorithm for non-linear systems, *Int. J. Control*, **44**(3), 803–822.

BROOMHEAD, D.S., JONES, R. and KING, G.P., 1987. Topological dimension and local coordinates for time series data, *J. Phys. A:Math. Gen.*, **20**, L563–L569.

CASDAGLI, M., 1989. Nonlinear prediction of chaotic time series, *Physica D*, **35**, 335–356.

CHEN, S. and BILLINGS, S.A., 1989. Representations of nonlinear systems: the NARMAX model, *Int. J. Control*, **49**(3), 1013–1032.

FARMER, J.D. and SIDOROWICH, J.J., 1988. Exploiting chaos to predict the future and reduce noise, in *Evolution, Learning and Cognition*, Y.C. Lee (Ed.), World Scientific.

FEIGENBAUM, M.J., 1983. Universal behaviour in nonlinear systems, *Physica D*, **7**, 16–39.

FRASER, A.M., 1989. Information and entropy in strange attractors, *IEEE Trans. Information Theory*, **35**(2), 245–262.

GIONA, M., LENTINI, F. and CIMAGALLI, V. 1991. Functional reconstruction and local prediction of chaotic time series, *Phy. Rev. A*, **44**(6), 3496–3502.

GOOIJER, J.G., ABRAHAM, B., GOULD, A. and ROBINSON, L., 1985. Methods for determining the order of an autoregressive-moving average process: a survey. *Int. Statistical Review*, **53**(3), 301–329.

GRASSBERGER, P., 1986. Estimating the fractal dimensions and entropies of strange attractors, in *Chaos*, A.V. Holden (Ed.), Manchester University Press, Manchester, 291–311.

GRASSBERGER, P., SCHREIBER, J. and SCHAFFRATH, C., 1991. Nonlinear time sequence analysis, *Int. J. Bifurcation and Chaos*, **1**(3), 521–547.

GUCKENHEIMER, J. and HOLMES P., 1983. *Nonlinear oscillations, dynamical systems, and bifurcation of vector fields*, Springer-Verlag, New York.

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

HANNAN, E.J. and QUINN, B.G., 1979. The determination of the order of an autoregression, *J. Royal Statist. Soc. B*, **41**(2), 190–195.

HAYNES, B.R., 1989. A qualitative approach to the global analysis of nonlinear systems with application to system identification, PhD thesis Univ. of Sheffield, Sheffield, UK.

HÉNON, M., 1982. On the numerical computation of Poincaré maps, *Physica D*, **5**, 412–414.

KASHYAP, R.L., 1977. A bayesian comparison of different classes of dynamica models using empirical data, *IEEE Trans. Automat. Contr.*, **22**(5), 715–727.

KENNEL, M.B., BROWN, R. and ABARBANEL, H.D.I., 1992. Determining embedding dimensions for phase-space reconstruction using geometrical construction, *Phys. Rev. A*, **45**(6), 3403–3411.

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.

KORENBERG, M.J. and PAARMANN, L.D., 1991. Orthogonal approaches to time-series analysis and system identification, *IEEE Signal Processing Magazine*, 8(3), 29–43.

KORTMANN, M. and UNBENHAUEN, H., 1988. Two algorithms for model structure determination of nonlinear dynamic systems with applications to industrial processes, *IFAC Symposium on Identification and System Parameter Estimation*, Beijing, 939–946.

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.

LEONTARITIS, I.J. and BILLINGS, S.A., 1987. Experimental design and identifiability for nonlinear systems, *Int. J. Systems Sci.*, **18**(1), 189–202.

LEONTARITIS, I.J. and BILLINGS, S.A., 1988. Prediction error estimator for non-linear stochastic systems, *Int. J. Systems Sci.*, , **19**(4), 519–536.

MARTEAU, P.F. and ABARBANEL, H.D.I., 1991. Noise reduction in chaotic time series using scaled probabilistic methods, *J. Nonlinear Sci.*, **1**, 313–343.

MASRI, S.F., CHASSIAKOS, A.G. and CAUGHEY, T.K. 1993. Identification of nonlinear dynamic systems using neural networks, *Transactions of the ASME, J. Appl. Mech.*, **60**, 123–133.

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

PACKARD, N.H., CRUTCHFILED, J.P., FARMER, J.D. and SHAW, R.S., 1980. Geometry from a time series, *Phys. Rev. Lett.*, **45**(9), 712–716.

PARKER, T.S. and CHUA, L.O. 1989. *Practical numerical algorithms for chaotic systems*, Springer Verlag, Berlin.

POTTMANN, M., UNBEHAUEN, H. and SEBORG, D.E., 1993. Application of a general multi-model approach for identification of highly nonlinear processes — a case study. *Int. J. Control*, **57** (1), 97–120.

SÖDERSTRÖM, T. and STOICA, P., 1989. *System Identification*, Prentice Hall, London.

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

UEDA, Y., 1980. Steady motions exhibited by Duffing's equation: A picture book of regular and chaotic motions, in *New approaches to nonlinear problems in dynamics*, P.J. Holmes (Ed.), SIAM, 311–322.

WOLF, A., 1986. Quantifying chaos with Lyapunov exponents, in *Chaos*, A.V. Holden (Ed.), Manchester University Press, 273–290.

Captions

Figure 1. (a) bifurcation diagram and (b) Poincaré section of the attractor at $A=11$ for the Duffing-Ueda equation.

Figure 2. Input/output data for $u_2(t)=11cos(t)$.

Figure 3. (a) bifurcation diagram and (c) Poincaré section for the model estimated from the data of figure 2. (b) bifurcation diagram and (d) Poincaré section for the model estimated from the data of figure 2 when corrupted by very low variance gaussian noise.

Figure 4. Input/output data when the input is a square wave superimposed on top of a gaussian component.

Figure 5. Bifurcation pattern of a family of models estimated from the data of figure 4.

Figure 6. Input/output data when the input is a square wave of increasing amplitude superimposed on top of a gaussian component.

Figure 7. Bifurcation pattern of a family of models estimated from the data of figure 6.

Figure 8. Dependence of the number of process terms, n_p, and the model order, n_y, on the sampling period. Each pair (o,*) represents the best model estimated from data sampled at the respective sampling rate.

Figure 9. Inaccurate bifurcation diagrams of comparable characteristics for the models (a) $T_s=\pi/30$ s, $n_y=3$, $n_p=11$ and (b) $T_s=\pi/60$ s, $n_y=3$, $n_p=7$.

Figure 10. Bifurcation diagrams for the models discretised with (a) $T_s=\pi/100$ s, (c) $T_s=\pi/60$ s, and (e) $T_s=\pi/30$ s. Bifurcation diagrams for the models estimated from data sampled at (b) $T_s=\pi/100$ s, equation (27), (d) $T_s=\pi/60$ s, equation (26), and (f) $T_s=\pi/30$ s, equation (25). Poincaré sections of the attractor at $A=11$ for the models discretised with (g) $T_s=\pi/100$ s, (i) $T_s=\pi/60$ s, and (k) $T_s=\pi/30$ s. Poincaré sections of the attractor at $A=11$ for the models estimated from data sampled at (h) $T_s=\pi/100$ s, equation (27), (j) $T_s=\pi/60$ s, equation (26), and (l) $T_s=\pi/30$ s, equation (25).

Figure 11. Comparison of the performance of the discretised and estimated models of figure 10. (*) discretised models, (+) estimated models.

Figure 12. Bifurcation pattern of a family of models estimated from the data of figure 6 sampled at $T_s=\pi/60$ s.

Figure 13. Bifurcation quality, J_b, for models with different number of terms and order.

Figure 14. Effects of overparametrization on a model estimated from data sampled at $T_s=\pi/30$ s (a) bifurcation diagram, Poincaré sections of the attractors at (b) $A=11$ and (c) $A=8.85$. The latter attractor is spurious.

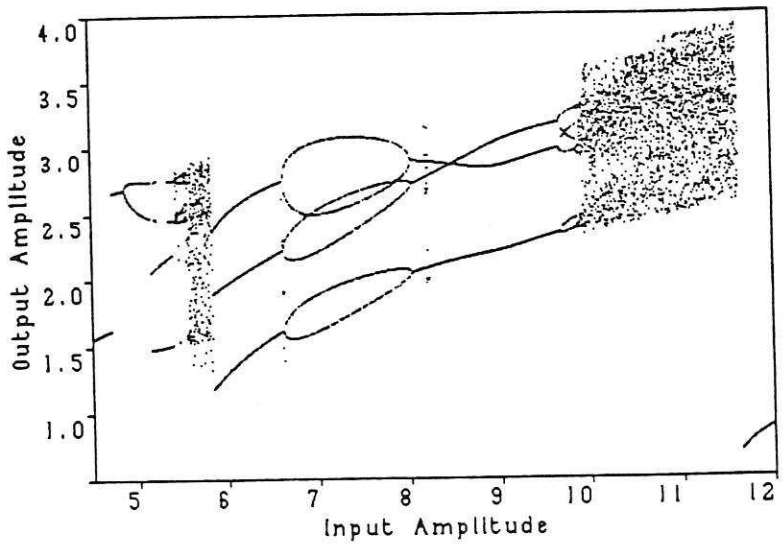
Figure 15. Effects of overparametrization on a model estimated from data sampled at $T_s=\pi/60$ s (a) bifurcation diagram, Poincaré sections of the attractors at (b) $A=11$ and (c) $A=8.85$. The latter attractor is spurious.

Figure 16. Variance of residuals for models with increasing numbers of terms.

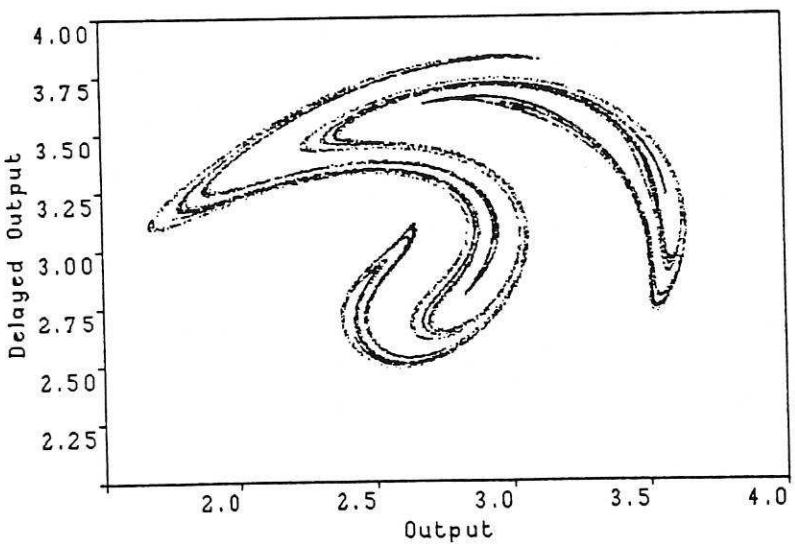
Figure 17. (- -) measured output and (—) predicted output using (a) the best model estimated, equation (26), and (b) the overparameterised thirteen-term model.

Figure 18. Representation of some models in the *model structure space* (MSS). The three black dots are the best models for the respective sampling rate, equations (25)-(27). The white dots are some of the reported inaccurate models. The ovals are the projections and the shaded regions is where the accurate models of figure 13 were found.

Figure 19. Search subspace represented in the *model structure space* (MSS).


4.0
3.5
3.0
2.5
2.0
1.5
1.0
Output Amplitude
5
6
7
8
9
10
11
12
Input Amplitude

(a)


4.00
3.75
3.50
3.25
3.00
2.75
2.50
2.25
Delayed Output
2.0
2.5
3.0
3.5
4.0
Output

(b)

FIG 1

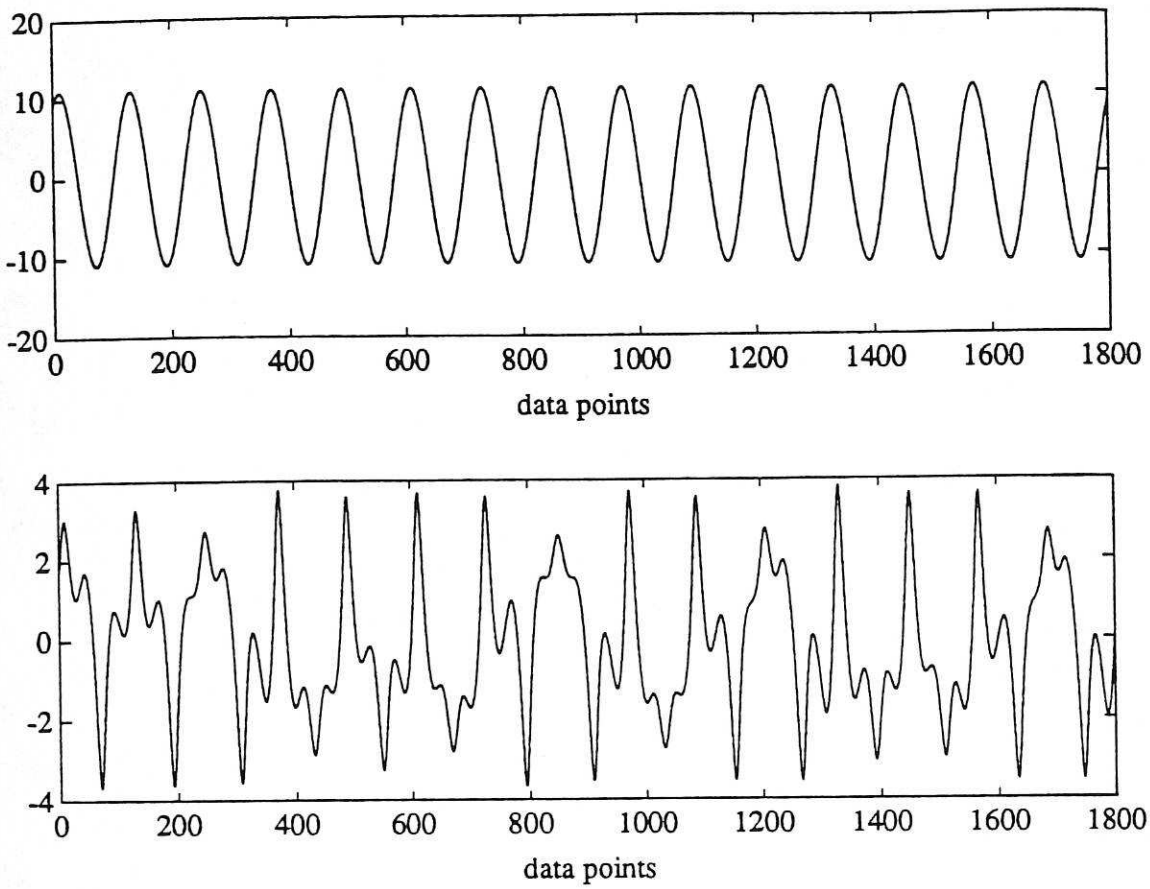
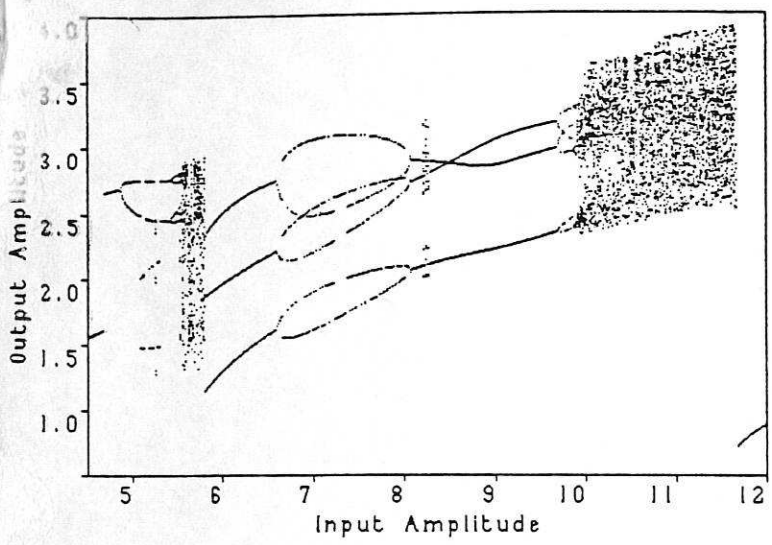
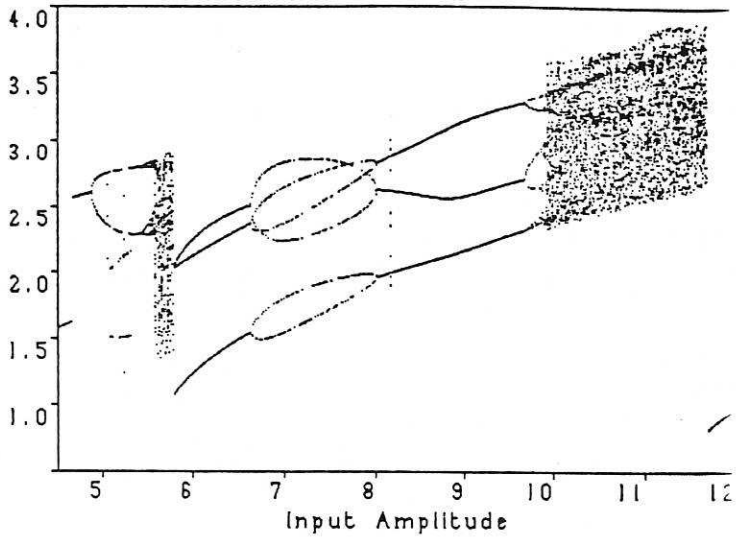

20
10
0
-10
-20
0
200
400
600
800
1000
1200
1400
1600
1800
data points
4
2
0
-2
-4
0
200
400
600
800
1000
1200
1400
1600
1800
data points

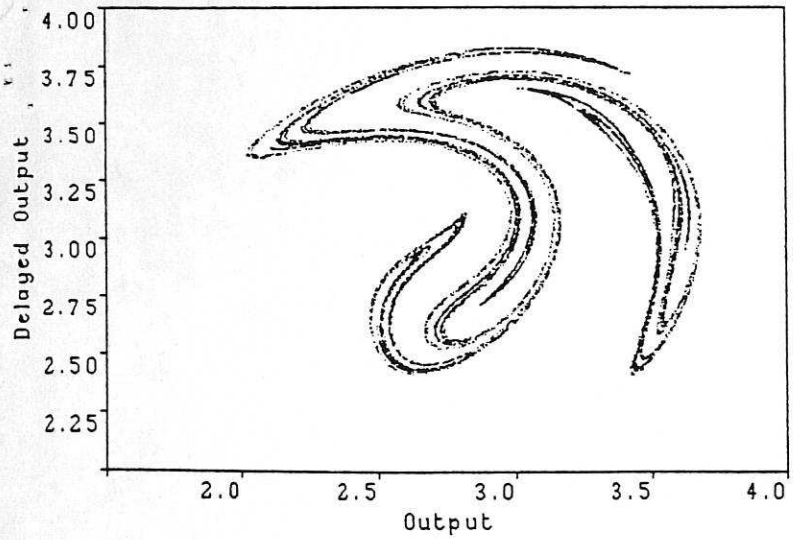
FIG. 2


Output Amplitude
4.0
3.5
3.0
2.5
2.0
1.5
1.0
5
6
7
8
9
10
11
12
Input Amplitude

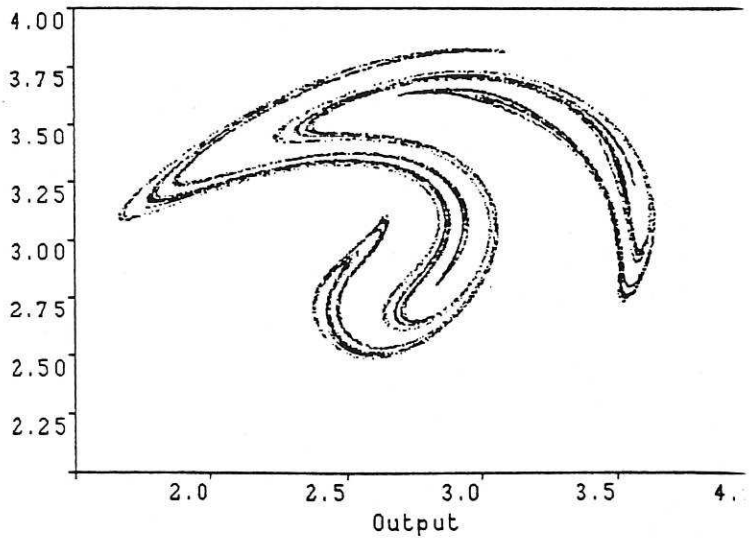
(a)


4.0
3.5
3.0
2.5
2.0
1.5
1.0
5
6
7
8
9
10
11
12
Input Amplitude

(b)


Delayed Output
4.00
3.75
3.50
3.25
3.00
2.75
2.50
2.25
2.0
2.5
3.0
3.5
4.0
Output

(c)


4.00
3.75
3.50
3.25
3.00
2.75
2.50
2.25
2.0
2.5
3.0
3.5
Output

(d)

FIG. 3

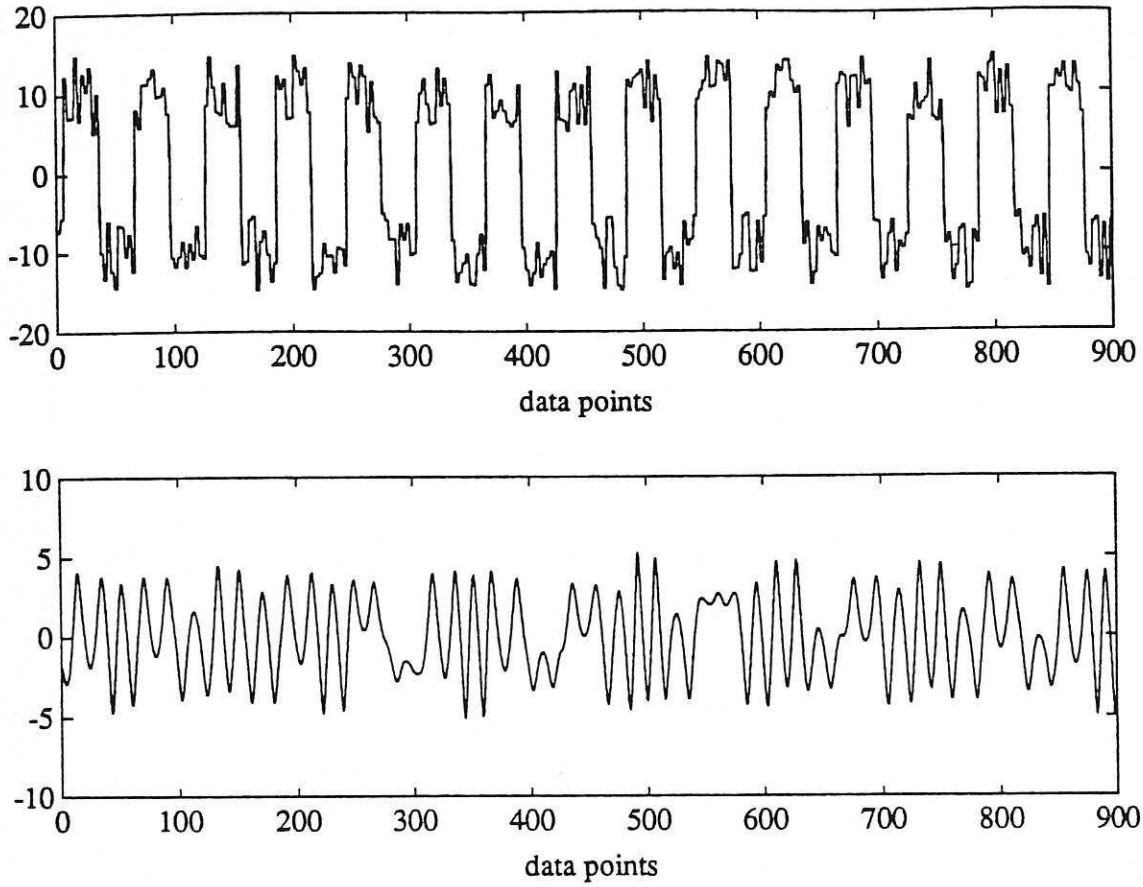

20
10
0
-10
-20
0
100
200
300
400
500
600
700
800
900
data points
10
5
0
-5
-10
0
100
200
300
400
500
600
700
800
900
data points

FIG. 4

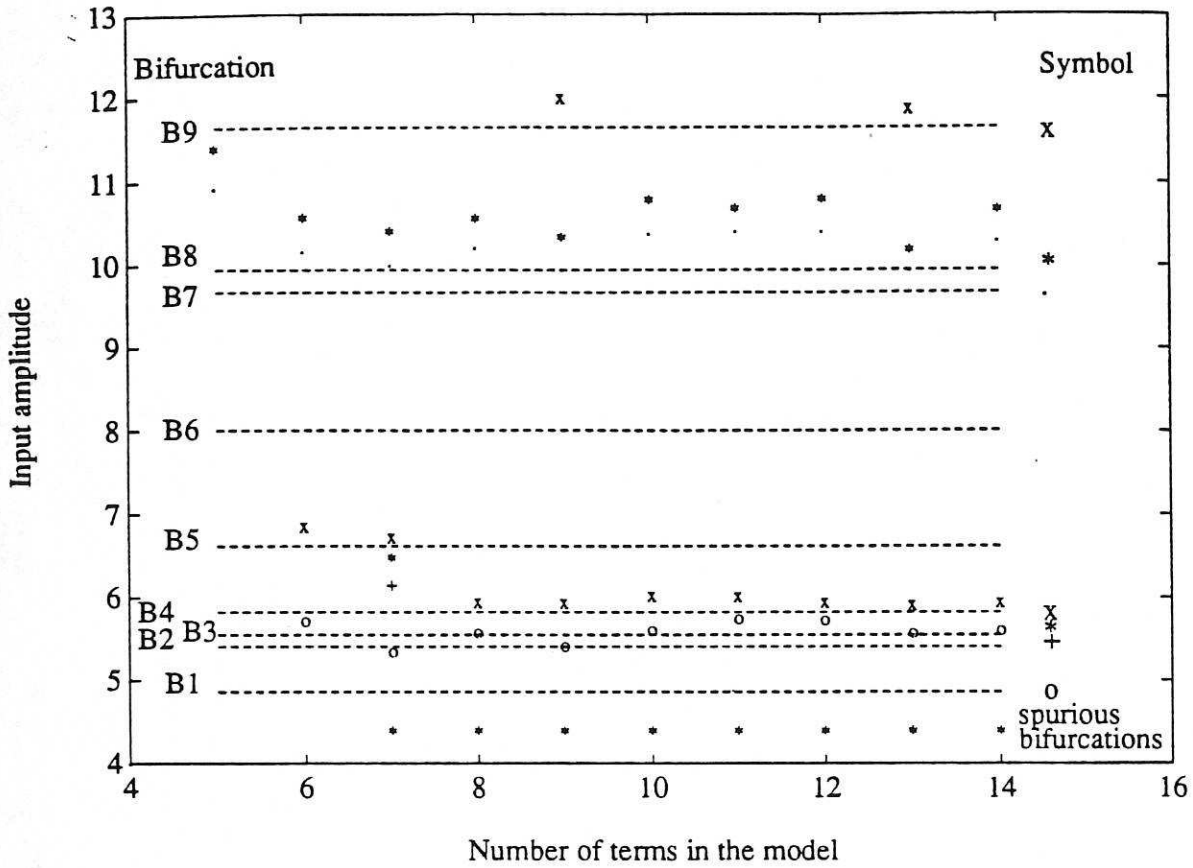

Bifurcation
Symbol
B9
B8
B7
B6
B5
B4
B2
B3
B1
x
*
o
spurious
bifurcations
Input amplitude
Number of terms in the model
13
12
11
10
9
8
7
6
5
4
4
6
8
10
12
14
16

FIG. 5

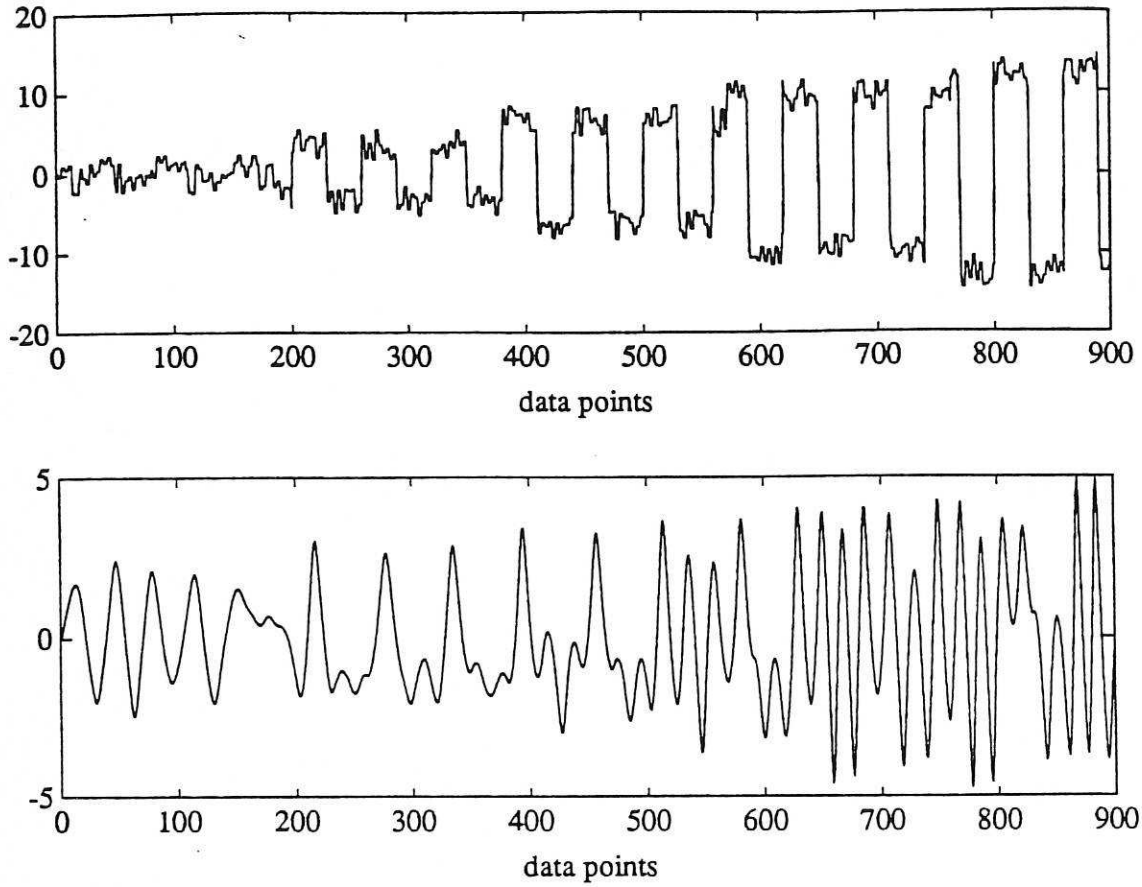

20
10
0
-10
-20
0
100
200
300
400
500
600
700
800
900
data points
5
0
-5
0
100
200
300
400
500
600
700
800
900
data points

FIG. 6

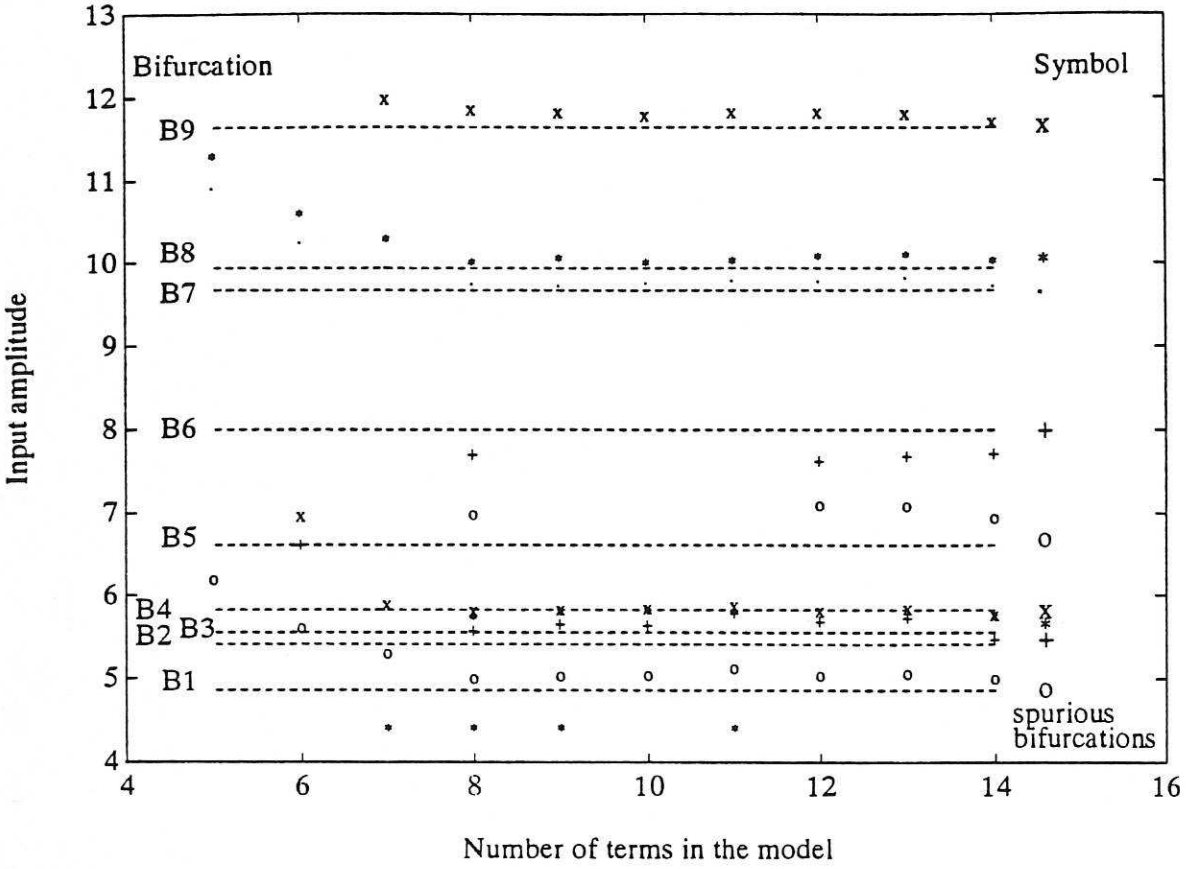

Bifurcation
Symbol
B9
B8
B7
B6
B5
B4
B2
B3
B1
spurious
bifurcations
Input amplitude
Number of terms in the model
13
12
11
10
9
8
7
6
5
4
4
6
8
10
12
14
16

Fig. 7

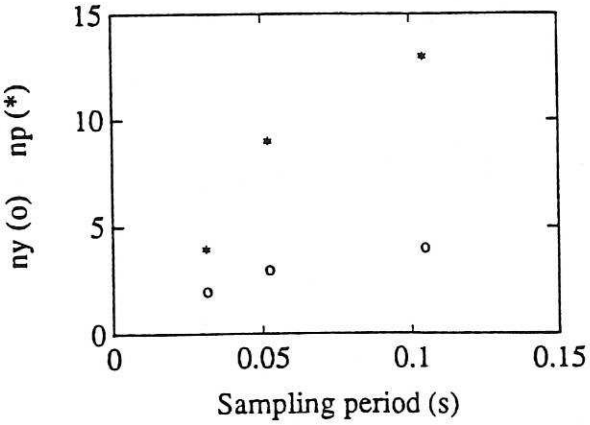
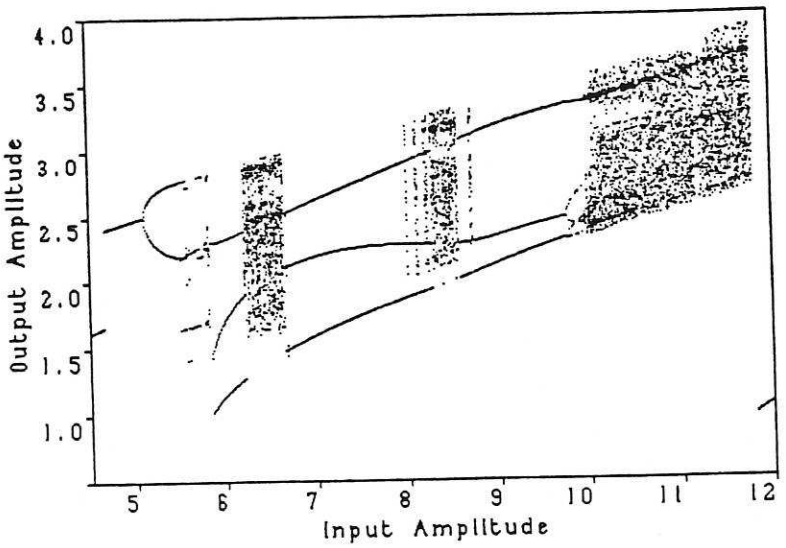
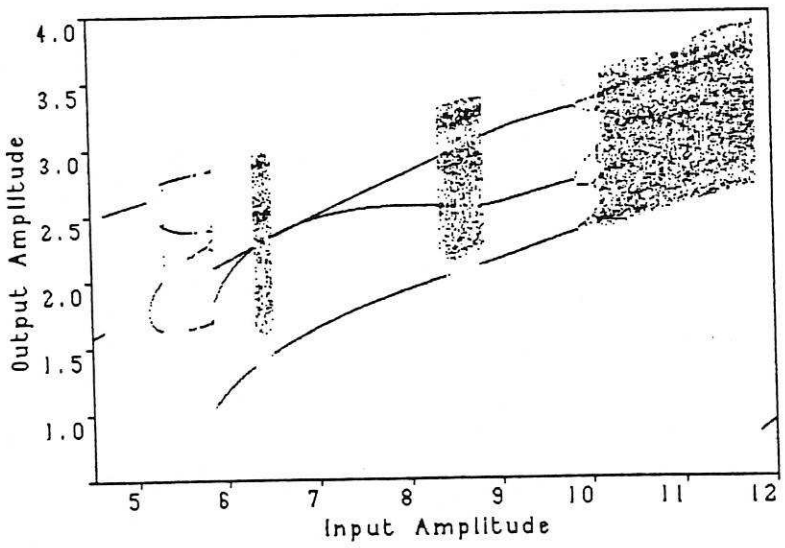


FIG. 8

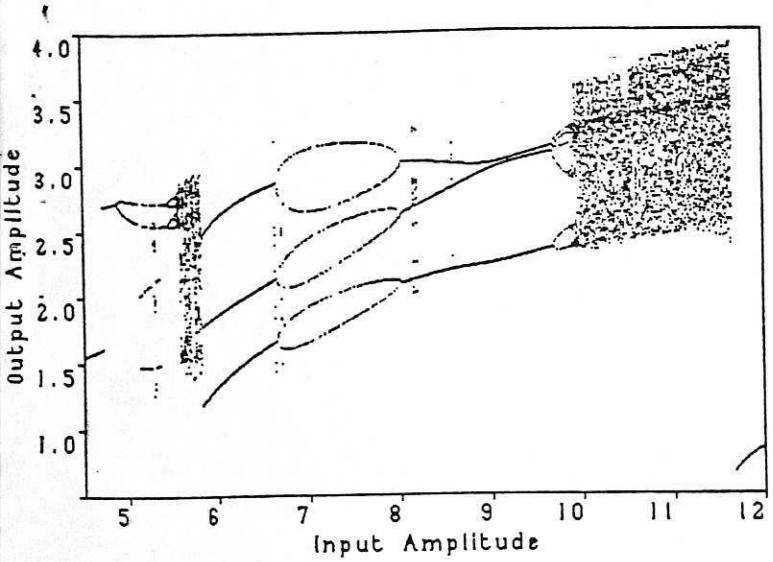

4.0
3.5
3.0
2.5
2.0
1.5
1.0
Output Amplitude
5
6
7
8
9
10
11
12
Input Amplitude

(a)

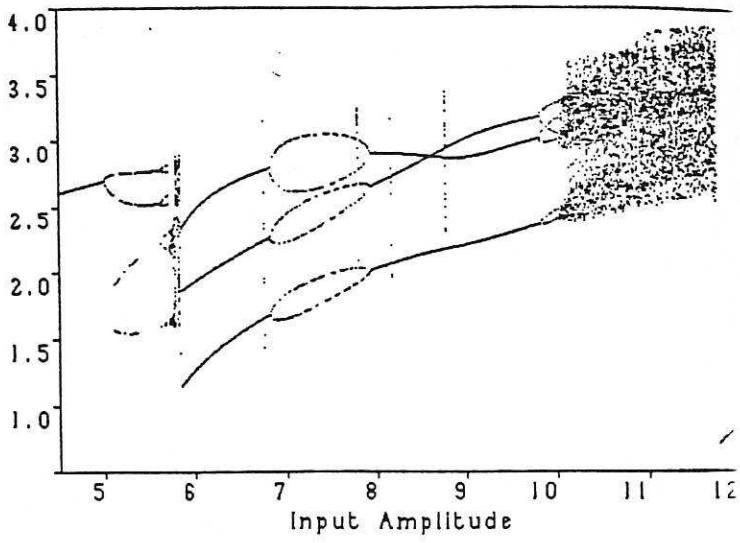

4.0
3.5
3.0
2.5
2.0
1.5
1.0
Output Amplitude
5
6
7
8
9
10
11
12
Input Amplitude

(b)

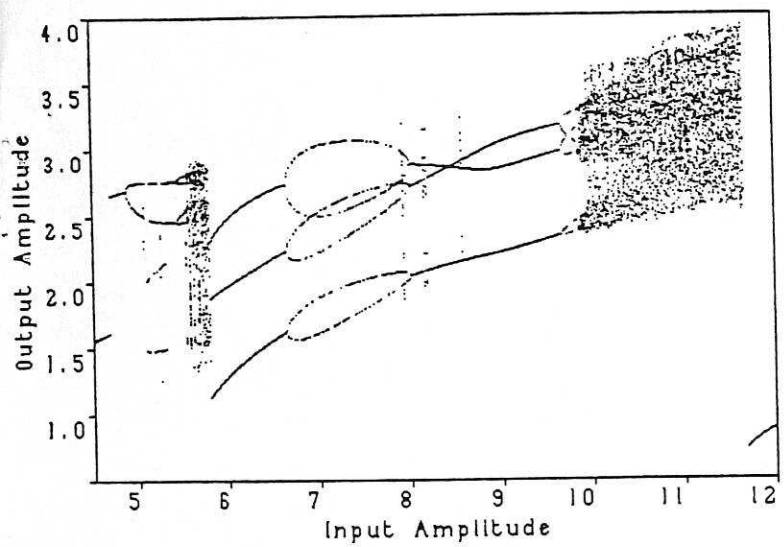
Fig. 9



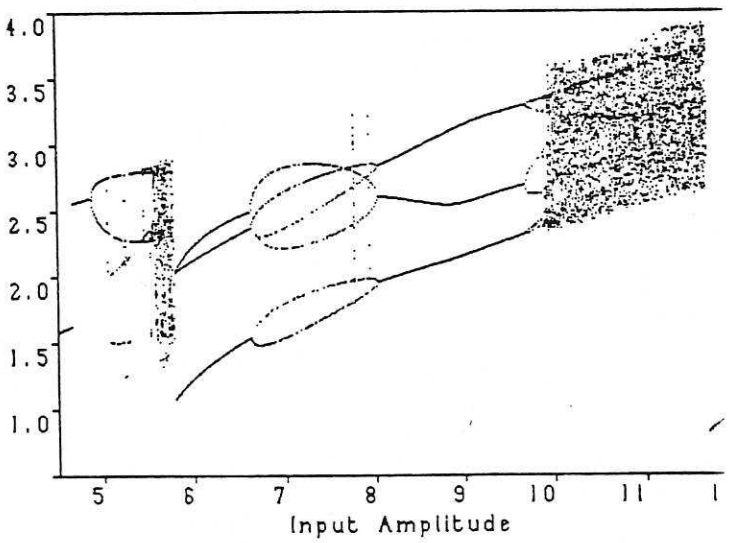
a



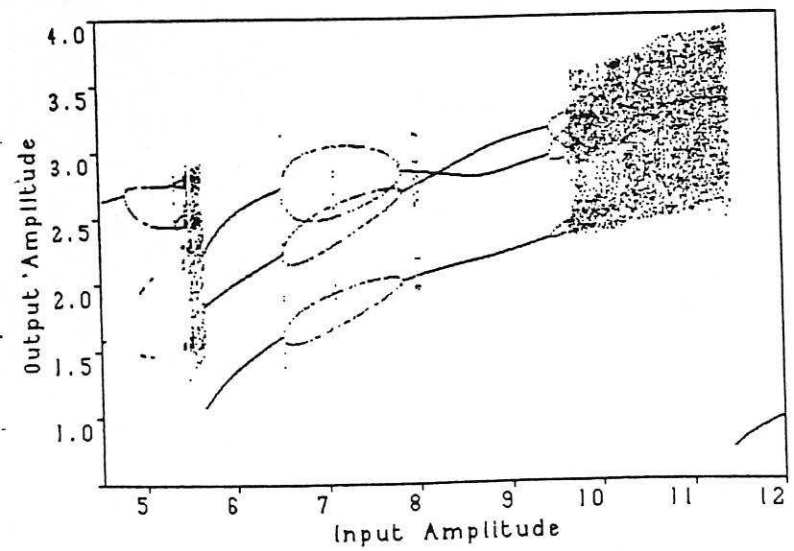
b



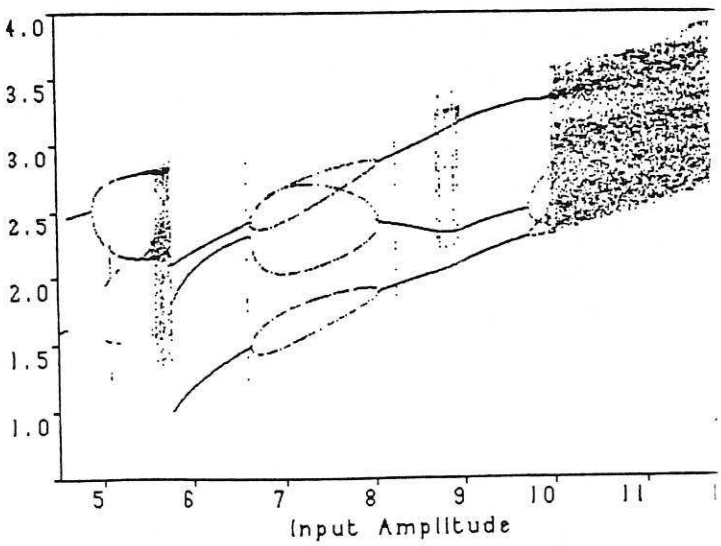
c



d

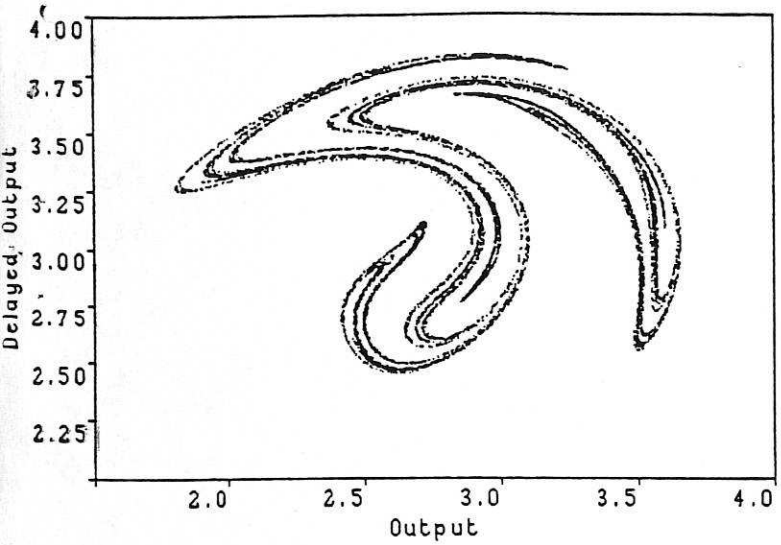


e

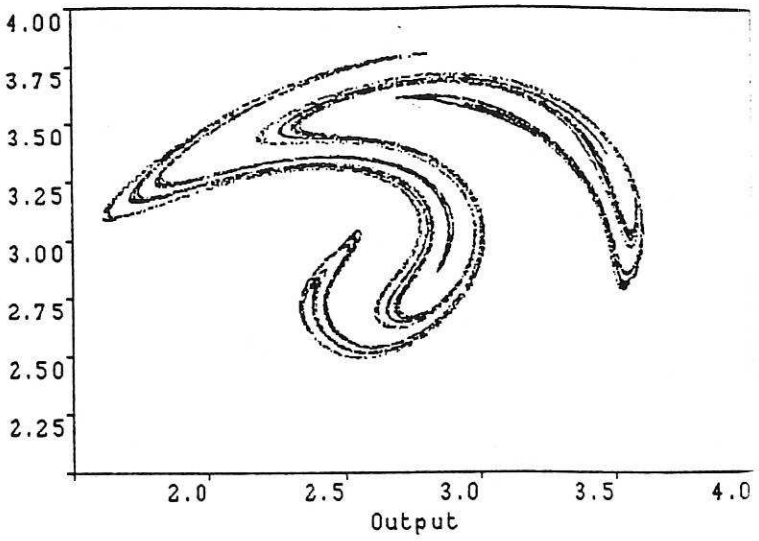


f

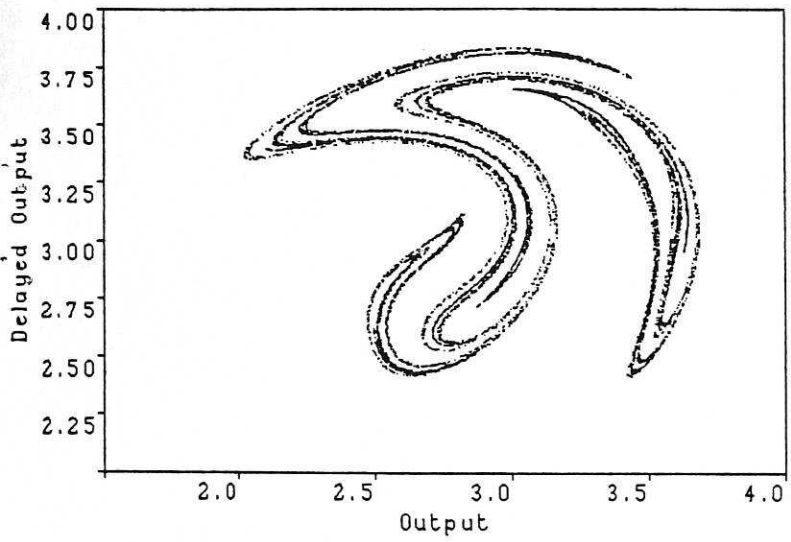
FIG. 10


4.00
3.75
3.50
3.25
3.00
2.75
2.50
2.25
Delayed Output
2.0
2.5
3.0
3.5
4.0
Output

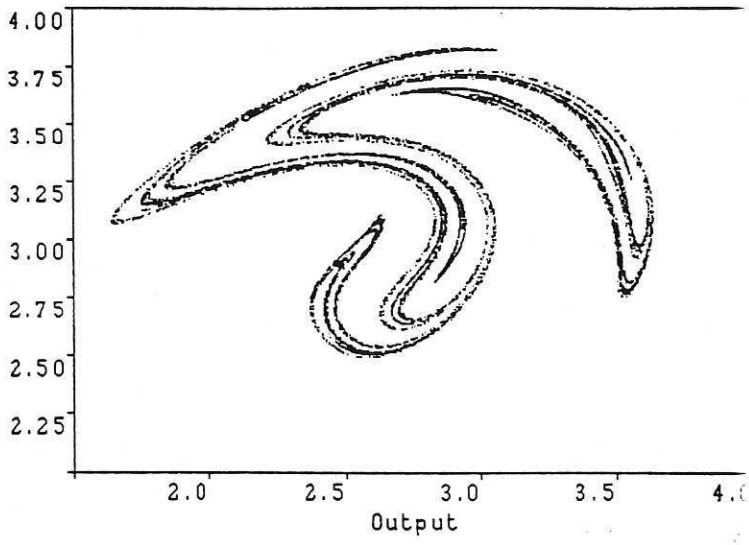
g


4.00
3.75
3.50
3.25
3.00
2.75
2.50
2.25
2.0
2.5
3.0
3.5
4.0
Output

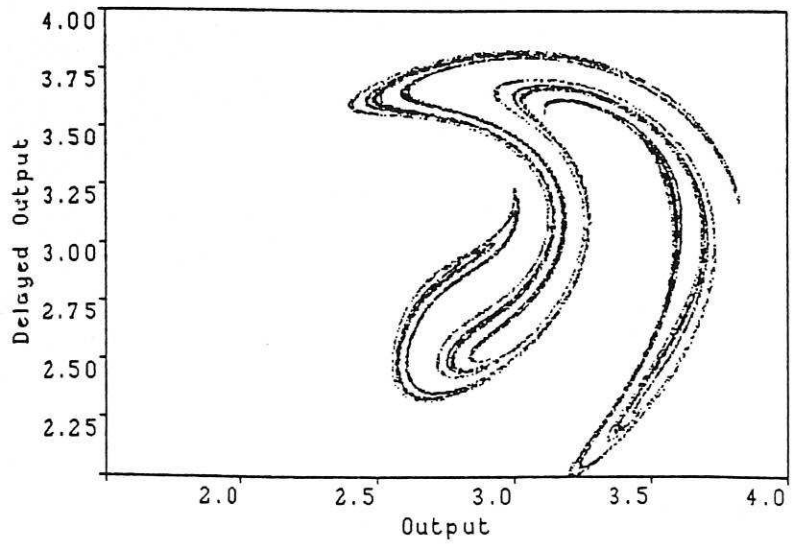
h


4.00
3.75
3.50
3.25
3.00
2.75
2.50
2.25
Delayed Output
2.0
2.5
3.0
3.5
4.0
Output

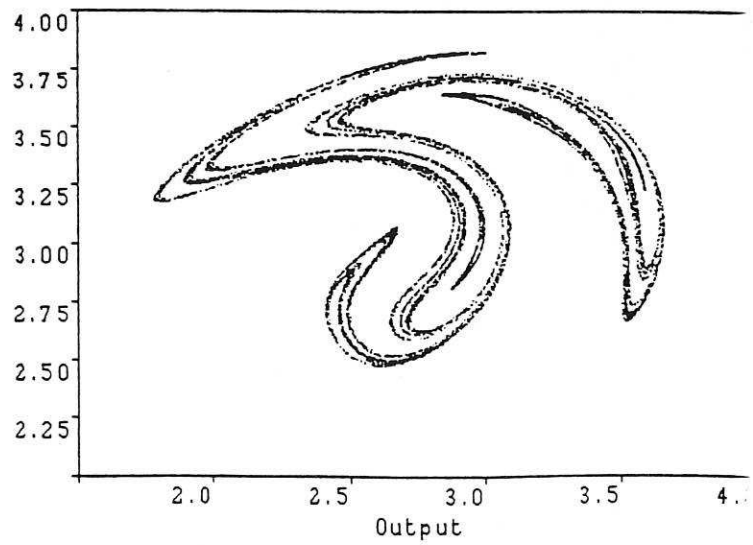
i


4.00
3.75
3.50
3.25
3.00
2.75
2.50
2.25
2.0
2.5
3.0
3.5
Output

j


4.00
3.75
3.50
3.25
3.00
2.75
2.50
2.25
Delayed Output
2.0
2.5
3.0
3.5
4.0
Output

k


4.00
3.75
3.50
3.25
3.00
2.75
2.50
2.25
2.0
2.5
3.0
3.5
Output

l

F 6. 10

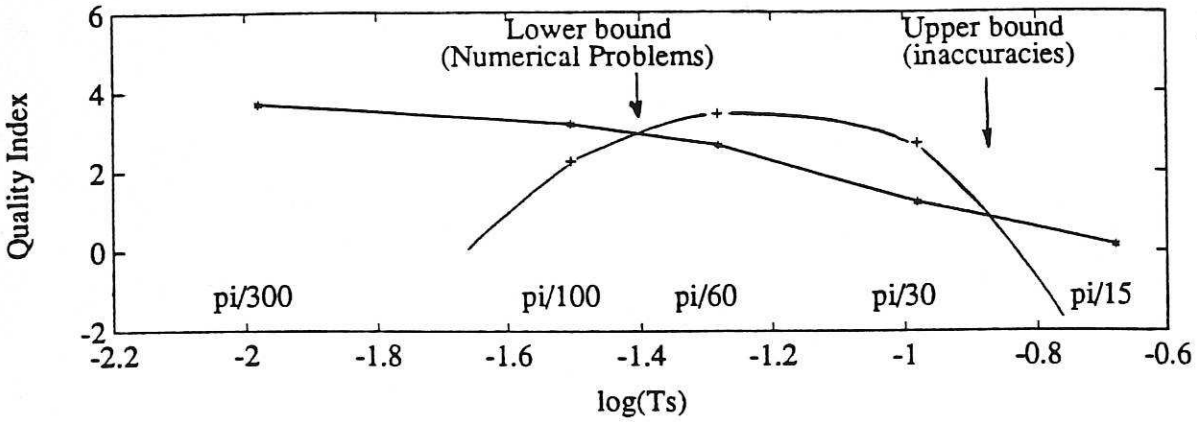

Lower bound
(Numerical Problems)
Upper bound
(inaccuracies)
Quality Index
6
4
2
0
-2
pi/300
pi/100
pi/60
pi/30
pi/15
-2.2
-2
-1.8
-1.6
-1.4
-1.2
-1
-0.8
-0.6
log(Ts)

Fig. 11

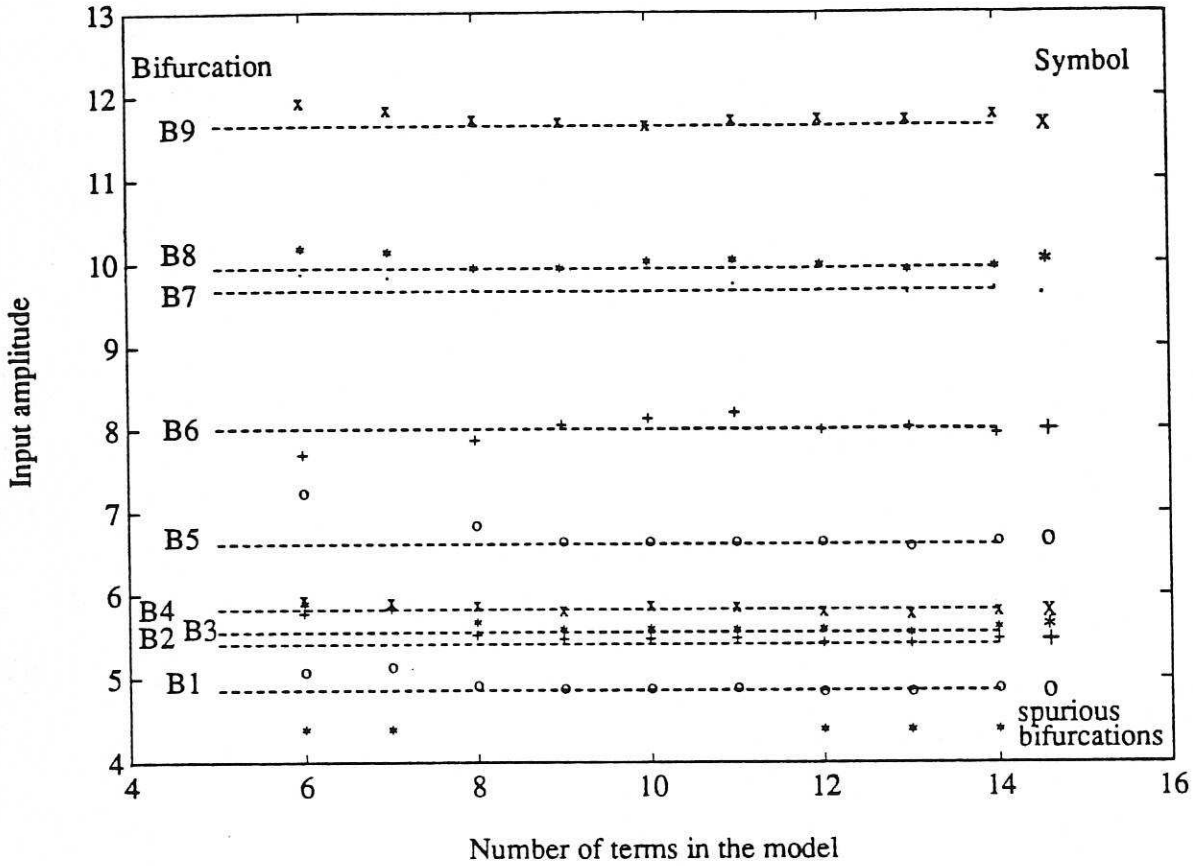

Bifurcation
Symbol
B9
B8
B7
B6
B5
B4
B3
B2
B1
spurious
bifurcations
Input amplitude
Number of terms in the model
13
12
11
10
9
8
7
6
5
4
4
6
8
10
12
14
16

FIG. 12

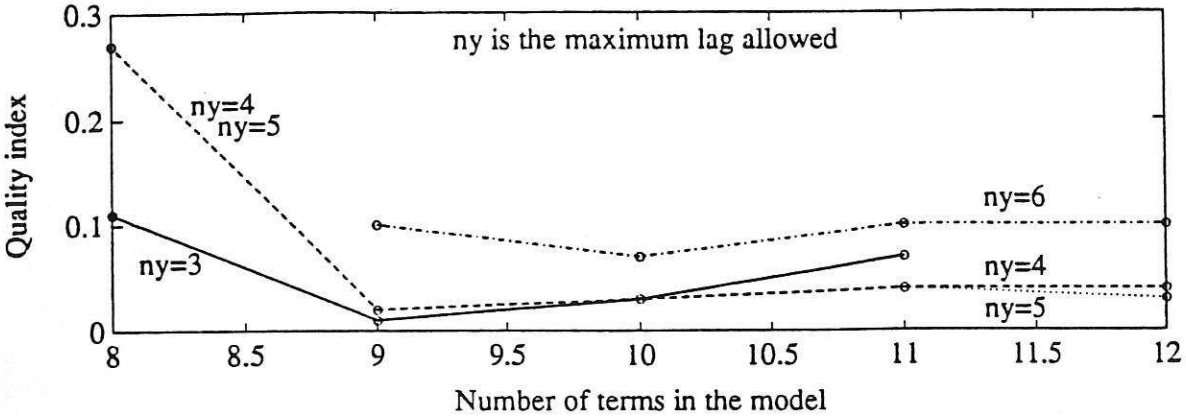
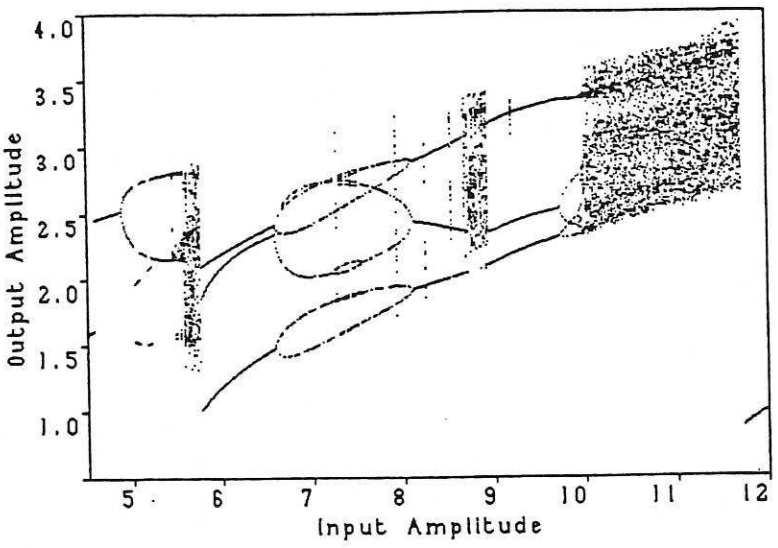
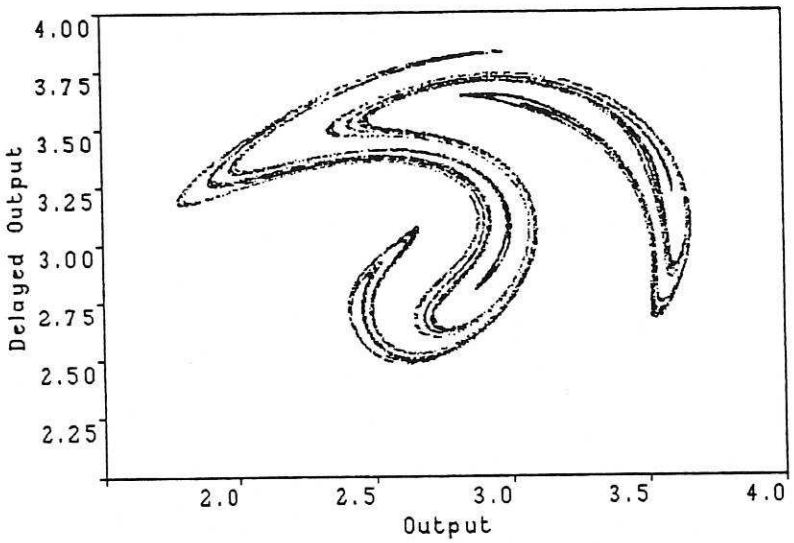

ny is the maximum lag allowed
Quality index
Number of terms in the model
0.3
0.2
0.1
0
8
8.5
9
9.5
10
10.5
11
11.5
12
ny=4
ny=5
ny=6
ny=3
ny=4
ny=5

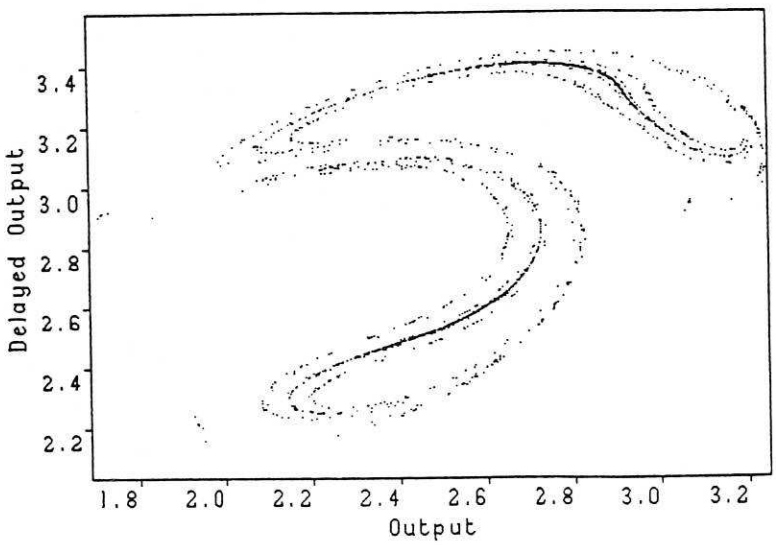
Fig. 13


Output Amplitude
4.0
3.5
3.0
2.5
2.0
1.5
1.0
5
6
7
8
9
10
11
12
Input Amplitude

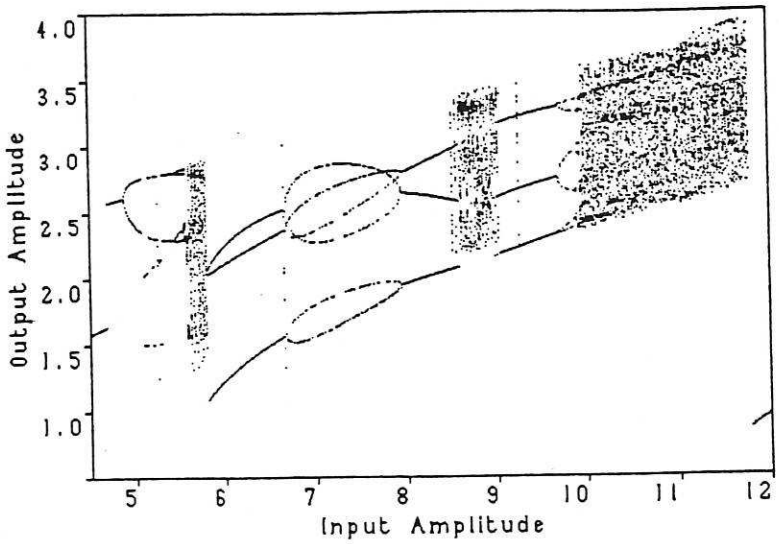
(a)


Delayed Output
4.00
3.75
3.50
3.25
3.00
2.75
2.50
2.25
2.0
2.5
3.0
3.5
4.0
Output

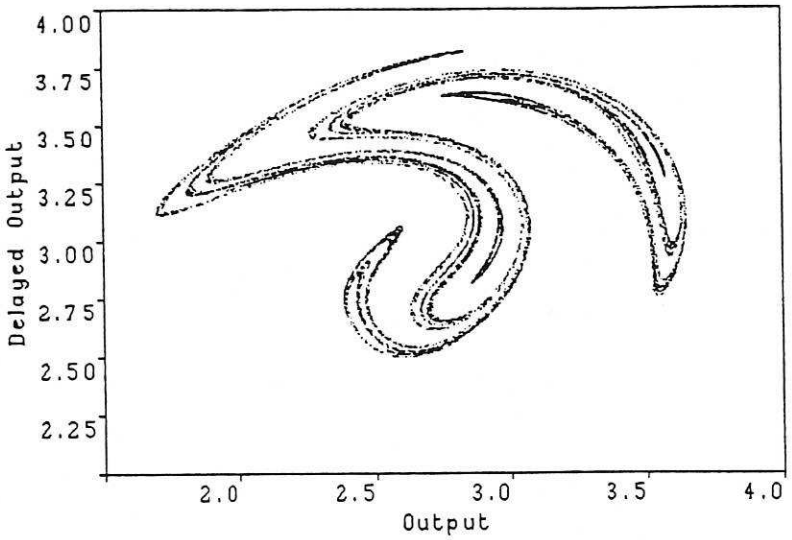
(b)


Delayed Output
3.4
3.2
3.0
2.8
2.6
2.4
2.2
1.8
2.0
2.2
2.4
2.6
2.8
3.0
3.2
Output

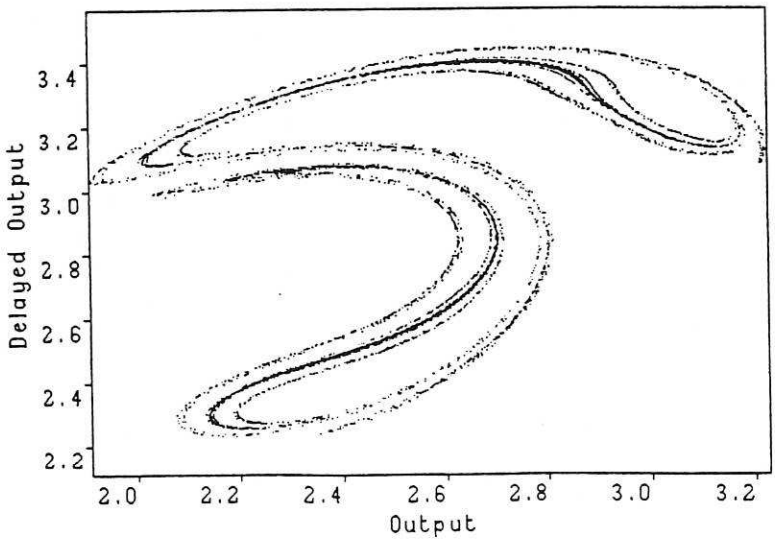
(C)


Output Amplitude
4.0
3.5
3.0
2.5
2.0
1.5
1.0
5
6
7
8
9
10
11
12
Input Amplitude

(a)


Delayed Output
4.00
3.75
3.50
3.25
3.00
2.75
2.50
2.25
2.0
2.5
3.0
3.5
4.0
Output

(b)


Delayed Output
3.4
3.2
3.0
2.8
2.6
2.4
2.2
2.0
2.2
2.4
2.6
2.8
3.0
3.2
Output

(c)

FIG. 15

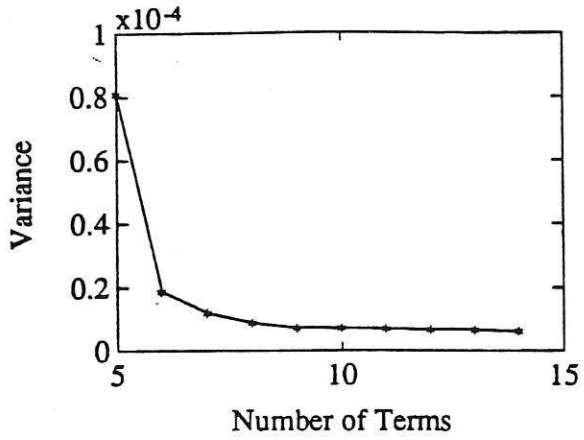
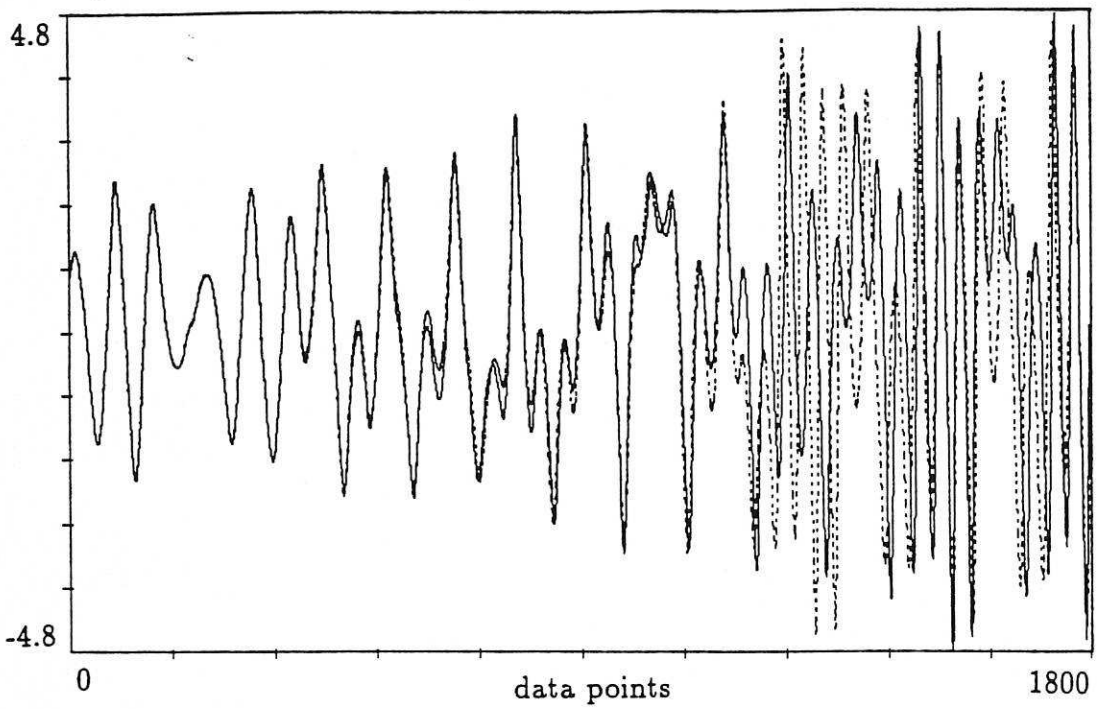
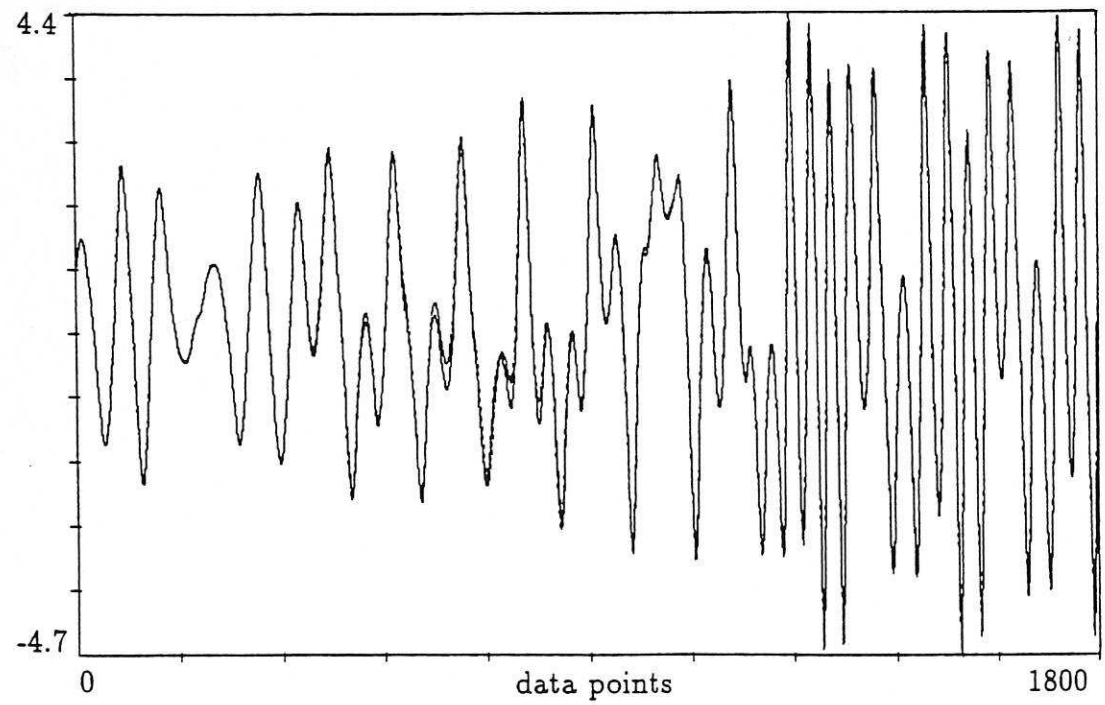


Figure 16



(a)



(b)

Figure 17

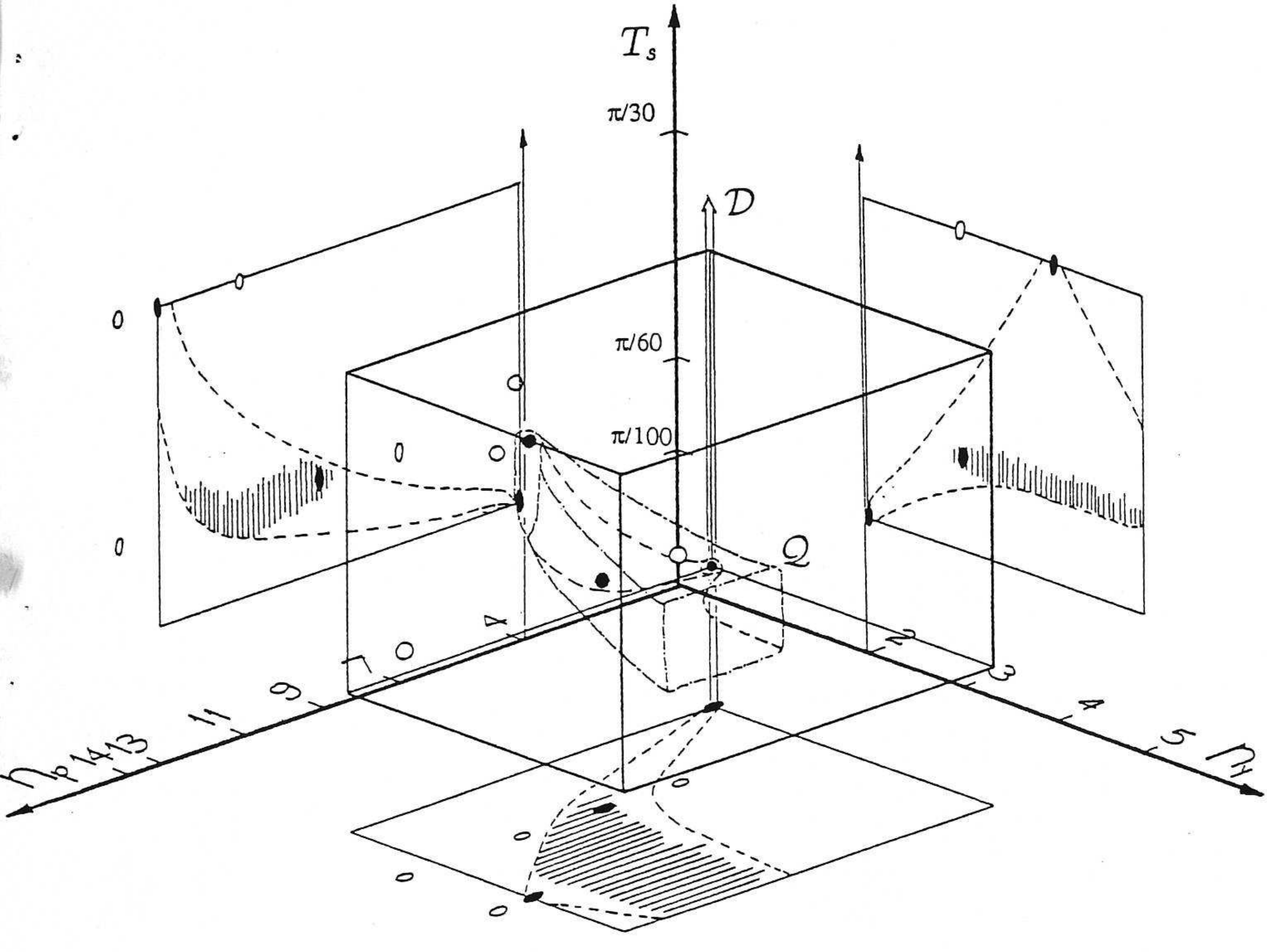


FIG. 18

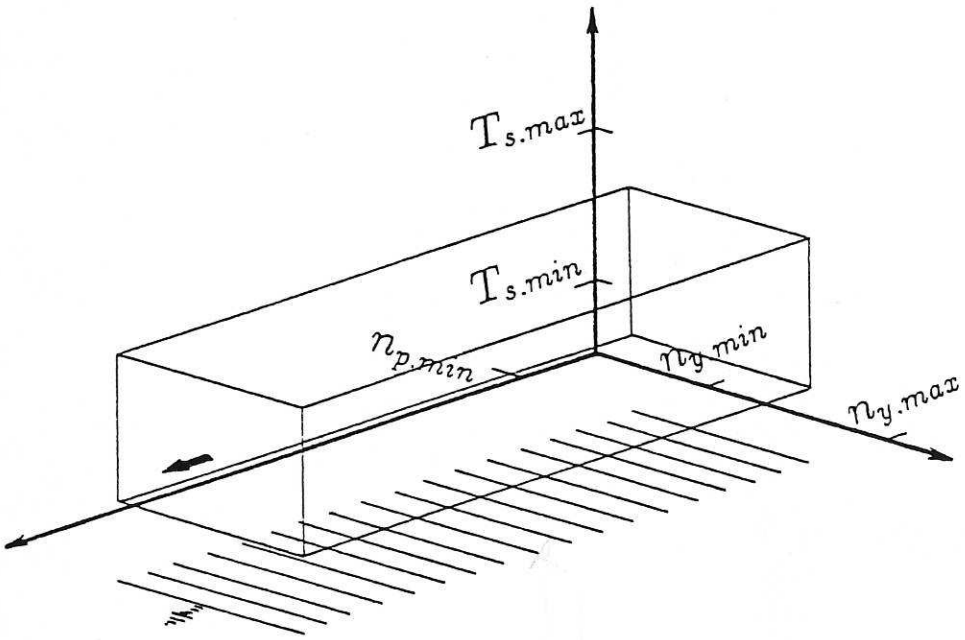


FIG. 19