Effects of Overparametrisation in Nonlinear System Identification and Neural Networks

G. L. Zheng and S. A. Billings
Department of Automatic Control and Systems Engineering,
University of Sheffield, Mappin Street, Sheffield S1 3JD

Abstract — The effects of overparametrisation in nonlinear system identification and neural networks are often presented as overfitting to a particular data set. An overparametrised model tends to interpolate the data set and can fail to capture the underlying dynamics represented by the data. Past studies have usually concentrated on overparametrisation in the number of model terms or the number of hidden layer nodes. In the present study, the effects of several forms of overparametrisation on the model dynamics are compared using qualitative validation methods. The dynamic properties of the identified model or the trained network are graphically displayed and compared with those of the real system. It is shown that overparametrisation in the number of time lags or input nodes has the worst effect on the dynamic properties of the resulting model.

Keywords — Overparametrisation, Qualitative Validation, Nonlinear System Identification, Radial Basis Function, Neural Networks.

Research Report No. 707

February 1998

Effects of Overparametrisation in Nonlinear System Identification and Neural Networks

G. L. Zheng and S. A. Billings
Department of Automatic Control and Systems Engineering,
University of Sheffield, Mappin Street, Sheffield S1 3JD

February 1998

Abstract — The effects of overparametrisation in nonlinear system identification and neural networks are often presented as overfitting to a particular data set. An overparametrised model tends to interpolate the data set and can fail to capture the underlying dynamics represented by the data. Past studies have usually concentrated on overparametrisation in the number of model terms or the number of hidden layer nodes. In the present study, the effects of several forms of overparametrisation on the model dynamics are compared using qualitative validation methods. The dynamic properties of the identified model or the trained network are graphically displayed and compared with those of the real system. It is shown that overparametrisation in the number of time lags or input nodes has the worst effect on the dynamic properties of the resulting model.

Keywords — Overparametrisation, Qualitative Validation, Nonlinear System Identification, Radial Basis Function, Neural Networks.

1 Introduction

The effect of overparametrisation in nonlinear system identification and neural networks is usually presented as overfitting to a particular data set. Overparametrisation is therefore often measured based on predictive performance of the model or the network on the data set. But an overparametrised model will tend to interpolate the training data set. It will often fail to capture the underlying dynamics and spurious dynamics may be induced as a result of overfitting. It is therefore important to investigate the effects of overparametrisation on model dynamics. In this paper, three qualitative validation methods, the Poincaré section, the bifurcation diagram and the cell to cell map are used to graphically illustrate the dynamic properties of the identified model or trained network and to compare these to those of the real system. The effects of overparametrisation are qualitatively measured by the difference in the dynamic properties between the model or the network and the real system.

In nonlinear system identification and neural network training, it is usually assumed that the time lags or the number of input layer nodes are known a priori. Most system

identification and neural network training algorithms were developed for selecting the number of model terms or hidden layer nodes (Chen et al., 1989; Holcomb and Morari, 1991; Musavi et al., 1992; Judd and Mees, 1994; Billings and Zheng, 1995). The effects of overparametrisation in the time lags or the input layer nodes are therefore often ignored but the present study shows that this can have a severe effect on the dynamics of the identified model or trained network. Overparametrisation in the time lags can reduce the range of the number of terms in dynamically valid models and therefore reduces the possibility of identifying a dynamically valid model. In neural network training, overparametrisation in the number of input layer nodes may result in a network with an excessive number of weights which can induce spurious dynamics. It is therefore important to investigating methods for selecting time lags or input layer nodes in nonlinear system identification and neural network training.

The layout of the paper is as follows. Section two briefly describes several forms of overparametrisation in nonlinear system identification and neural networks. Three qualitative validation methods for nonlinear system identification and neural networks are presented in section three. Simulation results are shown in section four. The effects of different forms of overparametrisation on the model dynamics are compared by using the qualitative validation methods. Conclusions are given in section four.

2 Overparametrisation in Nonlinear System Identification and Neural Networks

In this section, three types of overparametrisation in nonlinear system identification and neural networks are briefly described. The nonlinear autoregressive moving average model with exogenous inputs (NARMAX) and radial basis function (RBF) networks are used as examples. Overparametrisation in other model types or neural networks is likely to have a similar form.

Consider a single input single output continuous time system described by the equations

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \mathbf{f}(\mathbf{x}(t)) + \mathbf{g}(\mathbf{x}(t))u(t) \\ y(t) &= \mathbf{h}(\mathbf{x}(t)) \end{aligned} \tag{1}$$

where $\mathbf{x}(t) \in R^n$, $y(t), u(t) \in R$, $\mathbf{f}, \mathbf{g}$ and $\mathbf{h}$ are analytic functions of appropriate dimensions. The assumption of linearity in $u(t)$ is usually made to simplify control analysis and is often verified in practice. Although the state $\mathbf{x}(t)$ is linear in $u(t)$, the output $y(t)$ is nonlinear in $u(t)$. A discretised model of the system can be obtained by representing the solution of the differential equation (1) in terms of a formal Lie exponential (Barbot, 1989).

$$\begin{aligned} \mathbf{x}(k+1) &= e^{(T_s L(\mathbf{f}+\mathbf{g}u))} Id = \sum_{I=0}^{\infty} \frac{T_s^I}{I!} L^I(\mathbf{f}+\mathbf{g}u) Id|_{\mathbf{x}(kT_s)} \\ y(k) &= \mathbf{h}(\mathbf{x}(k)) \end{aligned} \tag{2}$$

where the input u is assumed to be constant on the interval $[kT_s \quad (k+1)T_s]$, T_s is the sample interval and the I^{th} iterated composition of Lie derivatives $L^I(\mathbf{f}+\mathbf{g}u)Id|_{\mathbf{x}(kT_s)}$ are defined recursively as

$$\begin{aligned} L^0(\mathbf{f}+\mathbf{g}u)Id|_{\mathbf{x}(kT_s)} &= Id|_{\mathbf{x}(kT_s)} = \mathbf{x}(kT_s) \\ L^I(\mathbf{f}+\mathbf{g}u)Id|_{\mathbf{x}(kT_s)} &= L(\mathbf{f}+\mathbf{g}u)_{\bullet}L^{I-1}(\mathbf{f}+\mathbf{g}u)Id|_{\mathbf{x}(kT_s)} \end{aligned} \tag{3}$$

The Lie derivative $L\mathbf{f}$ acts on an analytic function $\mathbf{h} : R^n \rightarrow R$ as follows

$$L\mathbf{f}_{\bullet}\mathbf{h}|_{\mathbf{x}} = \sum_{i=1}^{n} \frac{\partial \mathbf{h}}{\partial x_i} f_i(\mathbf{x}) \tag{4}$$

where $f_i(\mathbf{x})$ and x_i are the i^{th} component of $\mathbf{f}(\mathbf{x})$ and $\mathbf{x}$ respectively.

The discrete time system is equal to the continuous time system at each sampling time kT_s. The discrete time system may have an infinite number of terms and the degree of nonlinearity may be infinity. In applications, the discrete time system is usually truncated to a finite order and a finite number of model terms are used to approximate the continuous time system at the sampling times kT_s since the coefficients of higher order terms are usually small and negligible. The discrete time system truncated at order k is the same as the Taylor expansion of the function defining the Runge-Kutta solution truncated at the k^{th} order (Barbot, 1989).

In system identification and neural network training, a model or network with limited complexity is usually used to approximate the continuous time system at the sampling times kT_s. Since only a finite number of data samples are available and the data are usually contaminated with noise, a model or network with too high a complexity may fail to capture the underlying dynamics of the system and could result in overfitting on the training data set. The model or network with too high a complexity is said to be overparametrised. To separate the effects of overparametrisation and noise on the dynamics of the model or the network, a finite number of clean data samples will be considered in the present study. For the discrete time state space system, there may be two types of overparametrisation, overparametrisation in the number of terms or hidden layer nodes and overparametrisation in the degree of nonlinearity. Note that the equilibrium points of the continuous time system will always be the fixed points of the discrete time system truncated at order k. However the discrete time system truncated at order k with $k > 1$ will always introduce extra fixed points. These fixed points should not have too much effect on the system dynamics.

A wide class of nonlinear discrete time systems can be represented by the NARMAX model (Leontaritis and Billings, 1985)

$$y(k) = F^*(y(k-1), \ldots, y(k-n_y), u(k-1), \ldots, u(k-n_u)) \tag{5}$$

where F^* is some nonlinear function. Expanding the nonlinear function F^*, the model may be written as

$$y(k) = \sum_{j=1}^{M} \theta_j \phi_j + \theta_0 \tag{6}$$

where θ_0 is a constant term and $\mathcal{M}$ is the total number of terms in the model, θ_j ($j = 1, \ldots, \mathcal{M}$) are the coefficients and ϕ_j ($j = 1, \ldots, \mathcal{M}$) are the product combinations of $y(k-1), \ldots, y(k-n_y), u(k-1), \ldots, u(k-n_u)$ from degree one up to n_l, and n_l is the degree of nonlinearity. In system identification, the model terms and model structure are unknown and must be identified from the input and output signals.

It can be seen from the model that there may be three types of overparametrisation, overparametrisation in the number of time lags n_y and n_u, overparametrisation in the number of model terms $\mathcal{M}$ and overparametrisation in the degree of nonlinearity n_l.

The input-output relation can also be approximated using a radial basis function network as

$$y(k) = \sum_{j=1}^{\mathcal{N}} \theta_j^{\phi} \phi_j + \sum_{j=1}^{n_y} \theta_j^{y} y(k-j) + \sum_{j=1}^{n_u} \theta_j^{u} u(k-j) + \theta_0 \tag{7}$$

where ϕ_j, $j = 1, \ldots, \mathcal{N}$ are the basis functions, $y(k-j)$, $j = 1, \ldots, n_y$ are the delayed outputs, $u(k-j)$, $j = 1, \ldots, n_u$ are the delayed inputs and θ_0 represents the bias term. Equation (7) can be rewritten in the following form

$$y(k) = \sum_{j=1}^{\mathcal{M}} \theta_j \phi_j + \theta_0 \tag{8}$$

where ϕ_j, $j = 1, \ldots, \mathcal{M}$ are the basis functions or the linear terms $y(k-1), \ldots, y(k-n_y), u(k-1), \ldots, u(k-n_u)$. It may be seen that there are two types of overparametrisation in RBF networks, overparametrisation in the number of input nodes n_y or n_u and overparametrisation in the number of hidden layer nodes or basis functions $\mathcal{M}$. If the input output relation of the discrete time system has only a finite number of terms, depending on the type of nonlinearity of the system (1), the RBF network may be overparametrised in the degree of nonlinearity. As a result, a large number of radial basis function nodes may be needed for the network to approximate the underlying dynamics in the data.

It is well known in linear system identification that model overfitting often induces extra poles and zeros which tend to cancel each other. While poles and zeros cannot easily be defined for nonlinear models our results suggest that an inspection of the poles and zeros of the linear model terms from a nonlinear **NARMAX** model are a useful indication of model overfitting. If the nonlinear model is overparametrised in the time lags n_y and n_u, the extra poles and zeros of the linear part of the nonlinear model tend to cancel each other. Our experience suggests that a similar phenomenon occurs in the RBF networks, where pole-zero cancellation in the linear part of the network architecture is a useful indication of model overfitting. This issue will be studied further in the future.

In linear system identification pole-zero cancellation will result in instability if any of the extra poles is unstable. In nonlinear system identification the effects of spurious unstable poles of the linear part of the nonlinear model on system dynamics are much more severe. While all these observations regarding the linear part of a nonlinear model are difficult to support in theory they do provide considerable insight during practical model fitting.

The generalization error of a model can be partitioned into a bias error and a variance error. The variance error will increase as the number of model parameters increases. Consequently, when the model is overparametrised, spurious dynamics will be induced and

the model or network will tend to have a worse generalization capability due to increased variance error.

When the model is overparametrised in the degree of nonlinearity, extra fixed points will be introduced. If the extra fixed points are within the range of the data set, spurious dynamics will tend to be induced. When the extra fixed points are outside the range of the data set or the range of interest of the application, the effects on the system dynamics may be neglected. The effect of the extra fixed points on the system dynamics is very complicated and will depend on the relative position of these to the fixed points of the system and the relative position of the domain of attraction compared to those of the system. Overparametrisation in the degree of nonlinearity also considerably increases the number of candidate terms in the NARMAX model. This may both reduce the possibility of identifying a dynamically valid model and the identification may become numerically ill-conditioned.

3 Qualitative Validation Methods

In system identification and neural network training, predictive performance and statistical methods are usually used for validation (Billings and Zhu, 1995; Söderstrom and Stoica, 1990). Both these methods are based on a particular data set and can be regarded as a local measure of the validity of the model or network. But in system identification it is also important to demonstrate that the estimation procedure can capture the correct underlying system dynamics. When the system is linear it is therefore common practice to benchtest algorithms by comparing the estimated model parameters with the true values using known simulated system models. This is a form of qualitative model validation. When the system is nonlinear it may no longer be possible to compare parameter values. But it is still important to use a form of qualitative model validation to determine if the model can capture the correct underlying system dynamics. Dynamic and global validation methods have therefore been studied in recent years (Haynes and Billings, 1994; Casdagli, 1989; Aguirre and Billings, 1994; Zheng and Billings, 1996). This can be considered as equivalent to the linear practice of comparing estimated parameters to known values. In the present study three qualitative model validation methods, which are introduced below, are used to measure the effectiveness of nonlinear modelling procedures.

3.1 Poincaré Sections

Given a non-autonomous system represented by the ordinary differential equation

$$\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}, t) \qquad \mathbf{x}(t_0) = \mathbf{x}_0 \tag{9}$$

The solution of the system passing through x_0 at time t_0 is denoted by $\Phi_t(\mathbf{x}_0, t_0)$. If $\mathbf{F}(\mathbf{x}, t) = \mathbf{F}(\mathbf{x}, t+T)$ for all x and t, the system is said to be time periodic with period T, the smallest such T is called the minimum period. Sampling the trajectory of the system every T seconds, a map $P_N(\mathbf{x}) : R^n \rightarrow R^n$ is defined as

$$P_N(\mathbf{x}) = \Phi_{t_0+T}(\mathbf{x}, t_0) \tag{10}$$

P_N is called the Poincaré map for the non-autonomous system (Parker and Chua, 1989). The orbit $\{P_N^k(\mathbf{x})\}_{k=1}^{\infty}$ of the map is

$$P_N^k(\mathbf{x}) = \Phi_{t_0+kT}(\mathbf{x}, t_0), \qquad k = 0, 1, \ldots \tag{11}$$

In validation applications, the output of the system at the sampling times kT, $y(t_0+kT)$ is usually plotted against $y(t_0 + kT + T_d)$ in a pseudo-phase plane, where $0 < T_d < T$ is selected by the user. The resulting plot is called a Poincaré section. The limit set of the solution constitutes the most important dynamic invariants of the system. The dynamic invariants of the original system and the models of various representations can be graphically shown in the pseudo-phase plane and hence can be compared to each other and used to qualitatively validate an identified model or a trained neural network. The limit set of a period one solution of the system will appear as an isolated point in the Poincaré section. The limit set of a K^{th} order subharmonic of the system becomes a period K closed orbit. The limit set of a quasi-periodic solution of the system becomes an embedded circle. The limit set of a chaotic solution of the system does not lie on a simple geometrical object and shows a fine structure in the Poincaré section.

3.2 Bifurcation Diagrams

Consider again the system given in equation (9) with a parameter $\mu \in R$

$$\dot{\mathbf{x}} = \mathbf{F}_1(\mathbf{x}, t, \mu) \tag{12}$$

As the parameter μ changes, the limit sets of the system also change. Typically, a small change in μ produces small quantitative changes in the limit set. Examples are a change in the position of a limit set, change in shape or size of a limit set. However, there is also the possibility that a small change in μ may cause a qualitative change. Such a qualitative change is called a bifurcation. Examples are the disappearance or creation of a limit set, a change in the stability type of a limit set.

The bifurcation diagram is a plot of the position of the limit sets versus the bifurcation parameter μ. The computation of the bifurcation diagram involves finding all the steady state solutions of the system for various values of the parameter μ. In the present study, the brute force method (parameter stepping) (Parker and Chua, 1989) will be used to find the steady state solutions.

Given a map $P_b : R^n \times R \to R^n$, which depends on a parameter $\mu \in R$, to calculate the bifurcation diagram of P_b over the parameter range $\mu_{min} \leq \mu \leq \mu_{max}$, the system is simulated for K evenly spaced parameter values between μ_{min} and μ_{max},

$$\mu_k = \mu_{min} + (k-1)\frac{\mu_{max} - \mu_{min}}{K-1}, \qquad k = 1, \ldots, K \tag{13}$$

For the kth simulation, the system is simulated a fixed number of iterations. The first N_t iterations are assumed to be the transient and are discarded. The following N_s iterations are plotted on the bifurcation diagram. For a non-autonomous system, P_b is usually chosen to be the Poincaré map. In this study, the map P_b is chosen to be the Poincaré section and the input magnitude is chosen as a bifurcation parameter. For each value of the input magnitude, the system is simulated a fixed number of iterations and the steady state output

of the system is sampled at every T seconds, where T is the period of the sinusoidal input. The samples of the output are then plotted against the magnitude values of the input.

In the bifurcation diagram, a stable equilibrium point of the system corresponds to a single point for each bifurcation parameter. Because of continuity, the equilibrium points form a curve in the bifurcation diagram. A stable periodic solution of the system corresponds to a single point for each bifurcation parameter. Therefore, stable period solutions of the system form a curve in the bifurcation diagram. A Kth order subharmonic of the system is represented by K curves in the bifurcation diagram. The quasi-periodic solutions appear as smears of points. It is impossible to distinguish the order of the quasi-periodic solutions from the bifurcation diagram alone. Chaotic solutions appear much the same as quasi-periodic solutions and there is no robust way to distinguish chaotic from quasi-periodic solutions on the basis of the bifurcation diagram alone. To distinguish chaotic from quasi-periodic solutions, the Poincaré sections can be applied. Note that the brute force method only shows the stable solutions of the system.

3.3 Cell to Cell Maps

Consider once again the system described by the equation (12). The component of the state vector $\mathbf{x}$ is normally regarded as a continuum having an uncountable number of points in any interval. In practice, however the state variables are usually treated as collections of intervals because of the limited accuracy associated with measurements and computations. The state space can therefore be treated as a collection of N-dimensional cells. This state space is called a cell state space. A cell structure may be introduced into the cell state space in various ways. A simple structure can be constructed by dividing each state variable x_i $i = 1, 2, \ldots, N$ into N_i intervals. These intervals are identified by the integer values of the corresponding cell coordinates $z_i = 1, 2, \ldots, N_i$. The Cartesian product space of $z_1, z_2, \ldots, z_N$ is then called a cell state space. Each element $\mathbf{z}$ of the space is an N-tuple of integers and is called a cell vector or simply a cell. The discrete time evolution of the dynamical system described by equation (12) can then be put in the form of a simple cell mapping

$$\mathbf{z}(n+1) = \mathbf{C}(\mathbf{z}(n), \mu), \quad \mathbf{z} \in \mathbf{Z}^N, \ \mu \in R \tag{14}$$

where $\mathbf{C} : \mathbf{Z}^N \times R \longrightarrow \mathbf{Z}^N$ is called a cell map.

Similar to an ordinary map, singular cells or equilibrium cells, and periodic solutions of period K can all be defined for the cell to cell map. For ease of reference, a solution of period K is called a P-K solution and each of its elements a P-K cell. According to this definition, a singular cell is a P-1 cell. A P-K solution may represent the periodic solution of the system, it may merely represent the equilibrium point of the system. Accordingly, the domains of attraction for simple cell mapping systems can also be defined. The cells in the domain of attraction of a P-K solution are denoted as DOA-K cells.

For a dynamical system governed by a cell mapping there is only a finite region of cell state space $\mathbf{Z}^N$ which will be of interest. The cells in this region are called regular cells. A sink cell is then defined to encompass all the cells outside the region of interest. If the mapping image of a regular cell is outside the region of interest, the cell is then said to be mapped into the sink cell. By definition, the sink cell is a P-1 cell and the mapping image of the sink cell is itself. The set of regular cells within the influence of the sink cell forms

the domain of attraction of the sink cell, and is labelled the DOA-Sink. These cells will eventually be mapped into the sink cell and are denoted DOA-Sink cells.

The dynamical behaviour of the system can be revealed by computing the mapping images of all the regular cells. For a system described by the differential equation (12), the mapping images of the regular cells are computed by integrating equation (12) at a fixed time interval T_i with the cell centre as the initial conditions. The time interval T_i should be selected such that there are not too many cells which become mapped into themselves or into the sink cell. For a system described by difference equations, the corresponding cell map is used to compute the images of all the regular cells.

After the images of all the regular cells have been found, the regular cells can be classified using the unravelling algorithm (Hsu, 1980). The algorithm calls all the regular cells one by one and classifies each cell into one of the following four categories, a singular cell (or a P-1 cell), a P-K cell, a DOA-K cell or a DOA-Sink cell. The singular cells correspond to the fixed points or the equilibrium points and the periodic cells to the limit cycles of the system. They constitute the invariant stable and unstable orbits of the system within $\mathbf{Z}^N$. The dynamical behaviour of the system is thus revealed graphically in the cell state space.

For validation applications, the cell to cell map of the identified model or trained network can be plotted and compared with that of the original system. The deficiency of the model or the network can therefore be shown graphically in the cell state space.

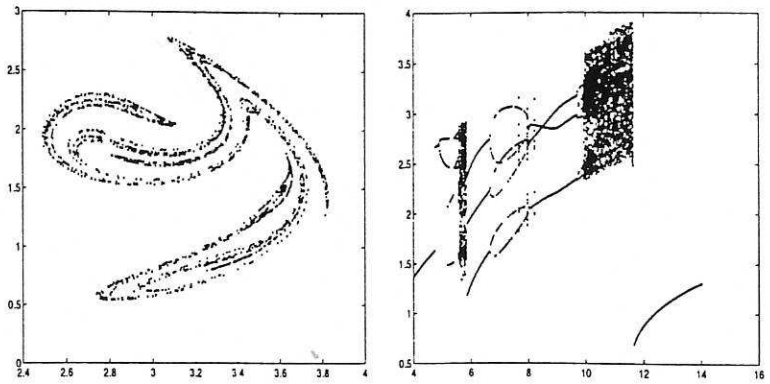
4 Simulation Results

4.1 Effects of Overparametrisation in NARMAX Model Identification

Consider the Duffing-Ueda equation expressed as (Ueda, 1980)

$$\begin{aligned} \frac{dx_1(t)}{dt} &= x_2(t) \\ \frac{dx_2(t)}{dt} &= -\delta x_2(t) - x_1^3(t) + Au(t) \\ y(t) &= x_1(t) \end{aligned} \tag{15}$$

For $\delta = 0.1$ and $u(t) = \cos(t)$, a variable order Runge-Kutta algorithm was used to integrate the equation. A sinusoidal input was applied to the system and the input was kept constant on the interval $[k\pi/1200s, (k+1)\pi/1200s]$. The Poincaré section and the bifurcation diagram of the system are shown in Figure 1. The Poincaré section was computed for $A = 11.0$ and shows that the output is chaotic. From the bifurcation diagram, some of the steady states of the system can be roughly classified as follows. From $A = 4.0$ to $A = 4.86$, the system settles in to a period one solution, from $A = 4.86$ to $A = 5.41$ a period two solution and from $A = 5.41$ to $A = 5.55$ a period four solution. The system then undergoes a series of period doublings and is driven into chaos. The chaos returns to a period three solution at $A = 5.82$. The system undergoes another series of period doublings and displays chaos at about $A = 9.94$. The chaos returns to a period one solution at $A = 11.64$. For validation, the Poincaré sections and the bifurcation diagrams of identified models will be qualitatively compared with those of the real system Figures 1a and 1b.



(a) (b)

Figure 1: (a) Poincaré section and (b) bifurcation diagram of the Duffing Ueda system. The integration interval was $\pi/1200$, the sample interval was $\pi/60$.

A discretised model of the system truncated at second order is

$$\begin{pmatrix} x_1(t+1) \\ x_2(t+1) \end{pmatrix} = \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} + T_s \begin{pmatrix} x_2(t) \\ -\delta x_2(t) - x_1^3(t) + Au(t) \end{pmatrix} + \frac{T_s^2}{2} \begin{pmatrix} 0 \\ -3x_1^2(t) \end{pmatrix} x_2(t) + \frac{T_s^2}{2} \begin{pmatrix} 1 \\ -\delta \end{pmatrix} \Big(-\delta x_2(t) - x_1^3(t) + Au(t) \Big) \tag{16}$$

An input-output expression of the second order discretised model is

$$\begin{aligned} y(t) = {} & (2 - T_s\delta + \frac{T_s^2}{2}\delta^2)y(t-1) + (-1 + T_s\delta - \frac{T_s^2}{2}\delta^2)y(t-2) \\ & + \frac{T_s^2}{2}Au(t-1) + (\frac{T_s^2}{2} - \frac{T_s^3}{2}\delta)Au(t-2) - \frac{T_s^2}{2}y^3(t-1) \\ & + (-\frac{T_s^2}{2} + \frac{T_s^3}{2}\delta)y^3(t-2) + \frac{3T_s^2}{2}y^4(t-2) - \frac{3T_s^2}{2}y(t-1)y^3(t-2) \\ & + \frac{3T_s^4}{4}y^3(t-2)Au(t-2) - \frac{3T_s^4}{4}y^6(t-2) \end{aligned} \tag{17}$$

For $\delta = 0.1$ and $Au(t) = 11.0\cos(t)$, a variable order Runge-Kutta algorithm was used to integrate the system (15). The input was kept as constant in an interval of $\pi/1200s$ and the resulting data was then sampled at an interval of $\pi/60s$. The output signal is plotted in Figure 2. The discretised model given in equation (17), which was obtained under the condition that the input was kept constant during each sample interval $\pi/60s$ rather than being constant during each integration interval of $\pi/1200s$ in the simulation, will be used to judge the complexity of the identified models.

To investigate the dynamic effects of the three types of overparametrisation in NARMAX model identification, models with different time lags, various numbers of model terms and degrees of nonlinearity were identified using the orthogonal least squares algorithm

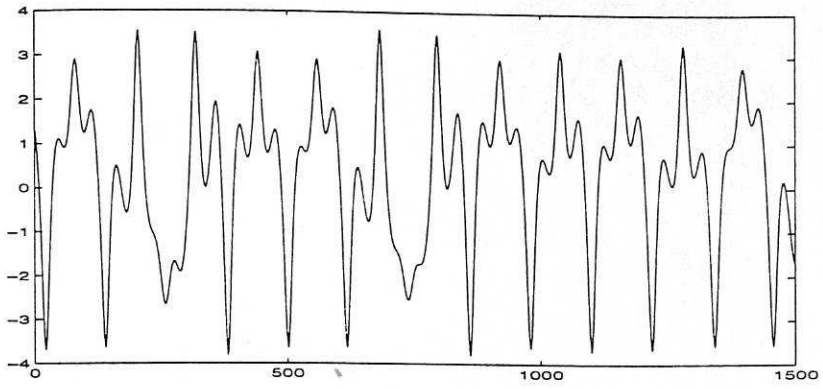


Figure 2: Output signal of the Duffing Ueda equation for the input $11.0\cos(t)$. The equation was integrated with a variable order Runge-Kutta algorithm and the input was kept constant in each interval of $\pi/1200s$. The data was then sampled at an interval of $\pi/60s$.

(OLS) (Chen et al., 1989). In all the simulations presented in section 4.1.1 and 4.1.2, the stopping criterion for the OLS algorithm was the number of terms in the model. Over-parametrised models were therefore deliberately obtained by including more terms than necessary to illustrate the problems which can occur unless the structure of the model is carefully selected. The OLS algorithm however can be used to identify the correct model with appropriate stopping criteria (Chen et al., 1989).

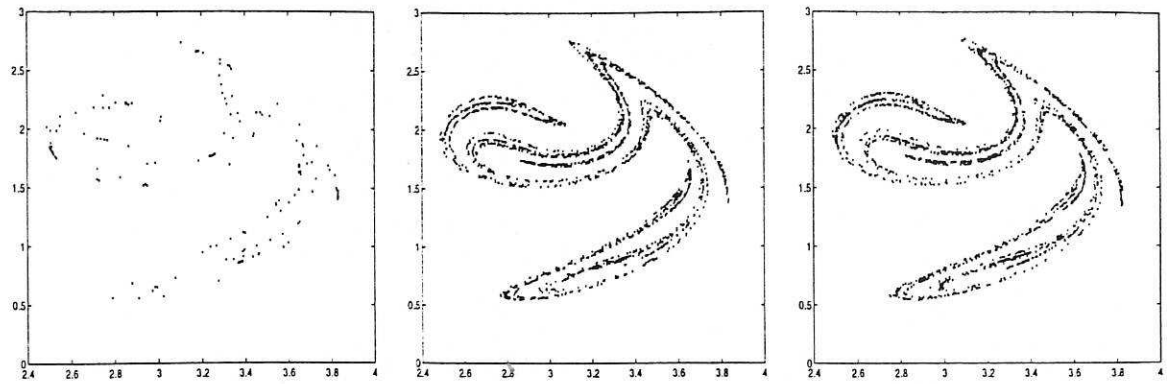
4.1.1 Effects of Overparametrisation in the Time Lags and the Number of Model Terms

The time lags were initially selected to be their correct values as $n_y = 2$, and $n_u = 2$ and the degree of nonlinearity was chosen as $n_l = 3$. The total number of possible candidate terms was 35. The OLS algorithm was then used to identify a group of models with a preset number of model terms. The Poincaré sections and the bifurcation diagrams of the identified models are shown in Figure 3 and 4 respectively.

A model with 5 terms was obtained as

$$\begin{aligned} y(t) &= 1.9946y(t-1) - 9.9466e^{-1}y(t-2) - 2.6999e^{-3}y^3(t-1) \\ &\quad +2.7345e^{-3}u(t-1) - 1.9836e^{-6}u^2(t-1)y(t-1) \end{aligned} \tag{18}$$

Each of the 5 terms in the model (18) was selected in such a way that the approximation error of the resulting model was most effectively reduced at each selection step. Figure 3a and 4a show that the Poincaré section of the model is very different from that of the original system (Figure 1a) and the bifurcation diagram of the model is different from that of the system in the range of $A = 6.5$ to 8. The bifurcation point near $A = 6.5$ has been shifted to the right and that near $A = 8$ has been shifted to the left. The model has therefore clearly failed to capture the system dynamics. This indicates that a model with less than 5 terms may not be able to capture the underlying dynamics of the data. This is evident from equation (17) which shows that for a model to approximate the underlying dynamics



(a) (b) (c)

Figure 3: Poincaré sections of identified models of the Duffing Ueda system with $n_l = 3$, $n_y = 2$, $n_u = 2$. (a) 5 terms, (b) 7 terms, (c) 30 terms. The sample interval was $\pi/60$.

more terms than equation (18) will be required. Increasing the number of model terms to 7, produced the following model

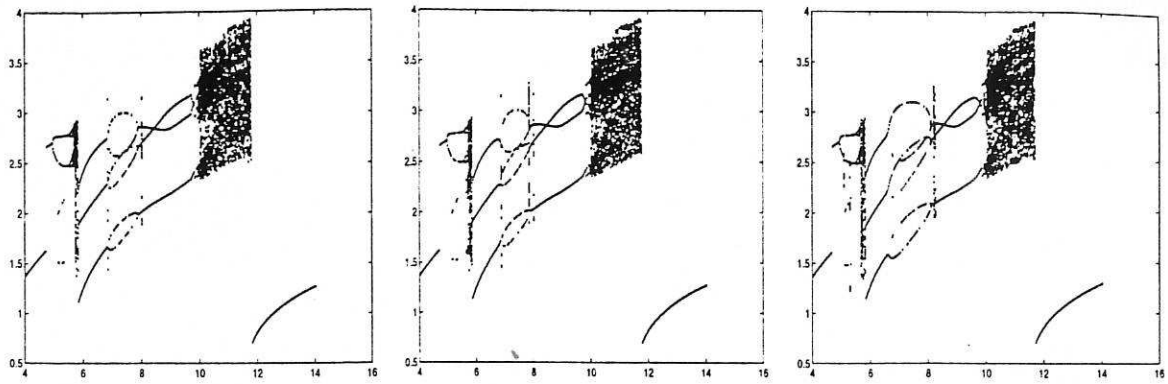
$$\begin{aligned} y(t) &= 1.9946y(t-1) - 9.9467e^{-1}y(t-2) - 2.68599e^{-3}y^3(t-1) \\ &\quad +2.7383e^{-3}u(t-1) - 1.1970e^{-6}u^2(t-1)y(t-1) + 4.7794e^{-7}u^2(t-2) \\ &\quad -5.9609e^{-6}u(t-1)y^2(t-1) \end{aligned} \tag{19}$$

This model had almost the same Poincaré section as the real system compare Figure 3b with 1a. This suggests that the resulting model has captured the local invariants of the system. However, the bifurcation diagram of the model (Figure 4b) in the range of $A = 6.5$ to 8 is still different from that of the real system Figure 1b and the model is therefore not globally valid. Increasing the number of model terms produced globally valid models. The number of terms can be increased above 30. However the output of the model with 34 terms diverged to infinity. This simulation suggests that overparametrisation in the number of model terms has only a marginal effect on model dynamics when the time lags are not overspecified and the data is noise free. When the data is contaminated with noise, the effect of overparametrisation in the number of model terms on the model dynamics can be very different. This will be investigated in a future study.

Allowing overparametrisation in the time lags by setting $n_y = 3$ and $n_u = 3$, the effects of overparametrisation become more obvious. For models with less than 6 terms, no input model terms were selected and the outputs diverged to infinity. The best model with 6 terms was identified as

$$\begin{aligned} y(t) &= 1.9623y(t-1) - 9.3051e^{-1}y(t-2) - 3.1811e^{-2}y(t-3) \\ &\quad -5.9662e^{-5}y^3(t-3) - 2.7471e^{-3}y^3(t-1) + 2.8138e^{-3}u(t-1) \end{aligned} \tag{20}$$

Figure 5a shows that this model has the same Poincaré section as that of the real system (Figure 1a) and captures the local invariants of the system. However, the model fails to



(a) (b) (c)

Figure 4: Bifurcation diagrams of the identified models of the Duffing Ueda system with $n_l = 3$, $n_y = 2$, $n_u = 2$. (a) 5 terms, (b) 7 terms, (c) 30 terms. The sample interval was $\pi/60$.

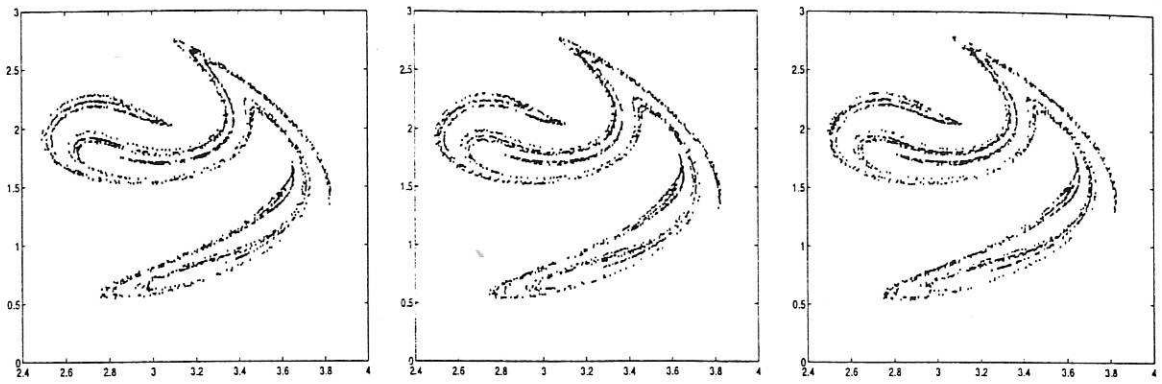
capture the global dynamic invariants of the system and has a bifurcation diagram Figure 6a which is different from that of the system Figure 1b in the range of $A = 6.5$ to 8. A further increase in the number of model terms was sufficient to achieve a globally valid model provided the number of terms was less than 19. Note that the range of the number of terms, 7 to 18, in the globally valid models was much smaller when compared with the previous simulations, 8 to 30 where the maximum time lags allowed was set correctly. The model with 7 terms was

$$\begin{aligned} y(t) &= 2.0163y(t-1) - 1.0381y(t-2) - 2.1716e^{-2}y(t-3) - 1.1612e^{-4}y^3(t-3) \\ &\quad -3.3396e^{-3}y^3(t-1) + 2.6732e^{-3}u(t-1) + 7.9070e^{-4}y(t-2)y^2(t-1) \end{aligned} \quad (21)$$

A comparison of Figures 5b and 6b with Figures 1a and 1b show that the model captures the local and global invariants of the system. The linear component of the model with 19 terms had an unstable pole at 15.339 and the model output diverged to infinity. An overparametrisation of just one in the maximum number of time lags clearly produced effects on the model dynamics which were much worse than in the simulations where the time lags were correctly selected.

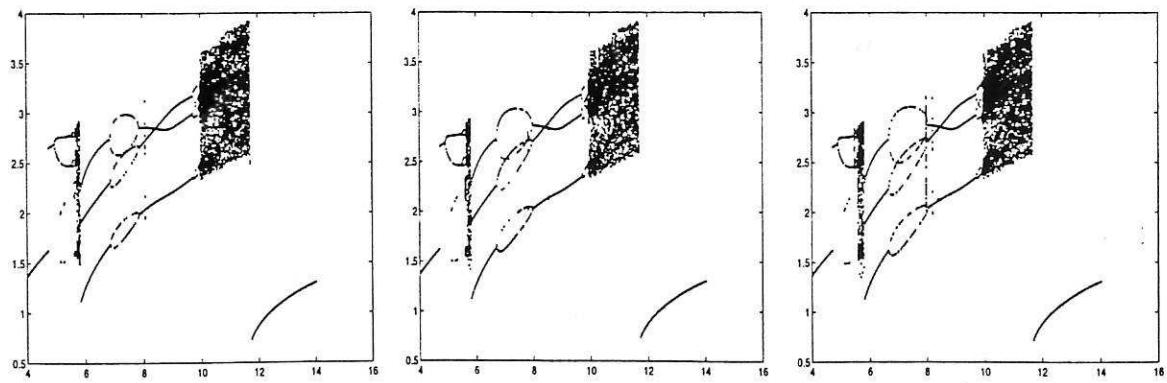
Increasing the maximum number of time lags further to $n_y = 4$ and $n_u = 4$ further reduced the range of the number of terms which produced globally valid models. For models with less than 7 terms, the outputs diverged. A comparison of Figures 7a and 8a with Figures 1a and 1b shows that the model with 7 terms was globally valid. The model was

$$\begin{aligned} y(t) &= 1.8440y(t-1) - 6.0717e^{-1}y(t-2) - 3.2368e^{-1}y(t-3) \\ &\quad +8.6821\epsilon^{-2}y(t-4) + 1.7083e^{-5}y(t-4)y(t-2)y(t-1) \\ &\quad -2.9218\epsilon^{-3}y^3(t-1) + 2.9070\epsilon^{-3}u(t-1) \end{aligned} \quad (22)$$



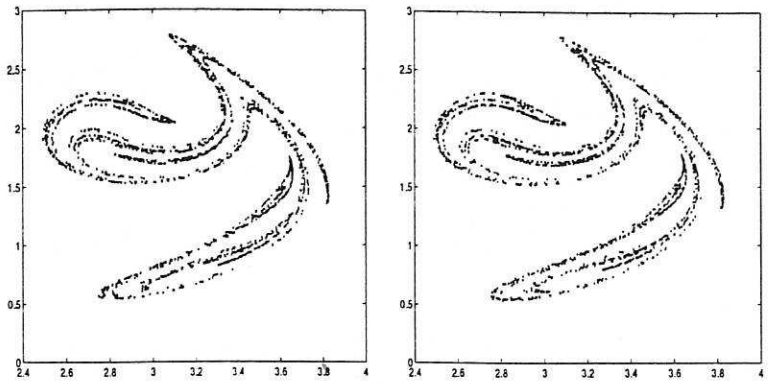
(a) (b) (c)

Figure 5: Poincaré sections of the identified models of the Duffing Ueda system with $n_l = 3$, $n_y = 3$, $n_u = 3$. (a) 6 terms, (b) 7 terms, (c) 18 terms. The sample interval was $\pi/60$.



(a) (b) (c)

Figure 6: Bifurcation diagrams of the identified models of the Duffing Ueda system with $n_l = 3$, $n_y = 3$, $n_u = 3$. (a) 6 terms, (b) 7 terms, (c) 18 terms. The sample interval was $\pi/60$.



(a) (b)

Figure 7: Poincaré sections of the identified models of the Duffing Ueda system with $n_l = 3$, $n_y = 4$, $n_u = 4$. (a) 7 terms, (b) 11 terms. The sample interval was $\pi/60$.

Globally valid models were obtained until the number of terms was increased to 11, the results are shown in Figures 7 and 8. Models with more than 11 terms had infinite outputs and the effects of overparametrisation were obvious. When the models were overparametrised in both the number of time lags and the number of terms the linear model terms produced unstable poles and the magnitude of the unstable pole increased rapidly as the number of terms was increased. For the models with 12, 14, 16 and 18 terms, the largest poles of the models were -1.4026, -3.4598, -12.1995 and -22.1832 respectively.

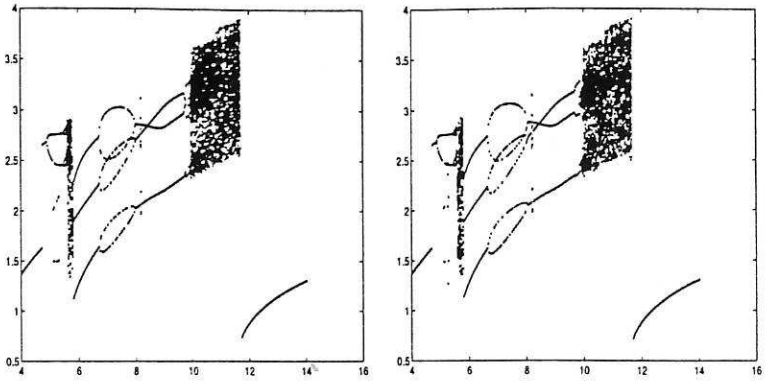
4.1.2 Combined Effects of Three Types of Overparametrisation

In the following simulations, the combined effects of overparametrisation in the number of time lags, the number of terms and the degree of nonlinearity will be investigated. Initially the maximum time lags were set as $n_y = 3$ and $n_u = 3$ and the degree of nonlinearity as $n_l = 5$. All models with less than 7 terms or more than 18 terms diverged. The linear part of the model with 6 terms had a pole at 1.0616. The model with 7 terms captured both the local and the global invariants of the system and the Poincaré section and the bifurcation diagram of the model resembled those of the real system, compare Figures 9a and 10a with Figures 1a and 1b. The model was identified as

$$\begin{aligned} y(t) \;=\; & 1.9877y(t-1) - 9.8084e^{-1}y(t-2) - 6.8760e^{-3}y(t-3) - 1.0757e^{-6}y^5(t-3) \\ & +3.8115e^{-6}y^5(t-1) - 2.7734e^{-3}y^3(t-1) + 2.7464e^{-3}u(t-1) \end{aligned} \tag{23}$$

Both the models with 12 terms and 18 terms had Poincaré sections and bifurcation diagrams close to those of the real system. Compared to the simulations with the same time lags and lower degree of nonlinearity, overparametrisation in the degree of nonlinearity appears only to have a marginal effect on the model dynamics.

Increasing the maximum time lags to $n_y = 4$ and $n_u = 4$, no valid model was obtained. Many models with different numbers of terms were obtained, but in all cases the linear



(a) (b)

Figure 8: Bifurcation diagrams of the identified models of the Duffing Ueda system with $n_l = 3$, $n_y = 4$, $n_u = 4$. (a) 7 terms, (b) 11 terms. The sample interval was $\pi/60$.

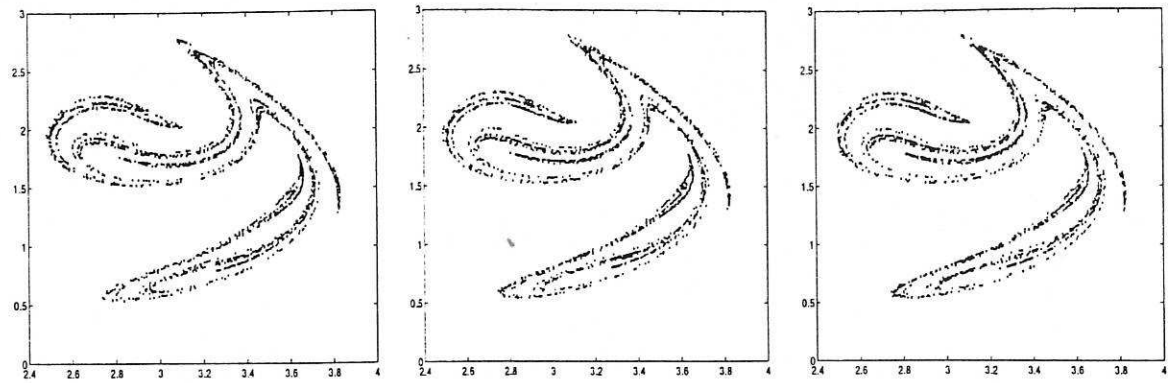
components of the models had unstable poles and the output diverged. The divergence of the models therefore appears to be mainly caused by overparametrisation in the time lags.

The simulation results presented in sections 4.1.1 and 4.1.2, suggest that incorrect settings in the time lags n_y and n_u considerably reduces the probability of identifying a globally valid model. For $n_l = 3$, the range of the number of terms in the globally valid models was 8 to 30 when the time lags were correctly selected as $n_y = 2$ and $n_u = 2$, this was reduced to 7 to 18 when the time lags were overparametrised by one to $n_y = 3$ and $n_u = 3$ and was further reduced to 7 to 11 when the time lags were overparametrised by two to $n_y = 4$ and $n_u = 4$. Notice that the total number of candidate terms for the corresponding settings are 35, 84 and 164 respectively. As the degree of nonlinearity was increased to $n_l = 5$, the effects of overparametrisation in the time lags became more obvious. For $n_y = 3$ and $n_u = 3$, the range of the number of terms in the models which were globally valid was 7 to 18, however no valid model was obtained for settings $n_y = 4$ and $n_u = 4$. The number of candidate terms for the corresponding settings are 462 and 1287. Clearly the large number of candidate terms may also induce numerical ill conditioning since many of the terms are highly correlated.

4.2 Effects of Overparametrisation in NARMAX Model Identification with a Square Wave Input of Increasing Amplitude

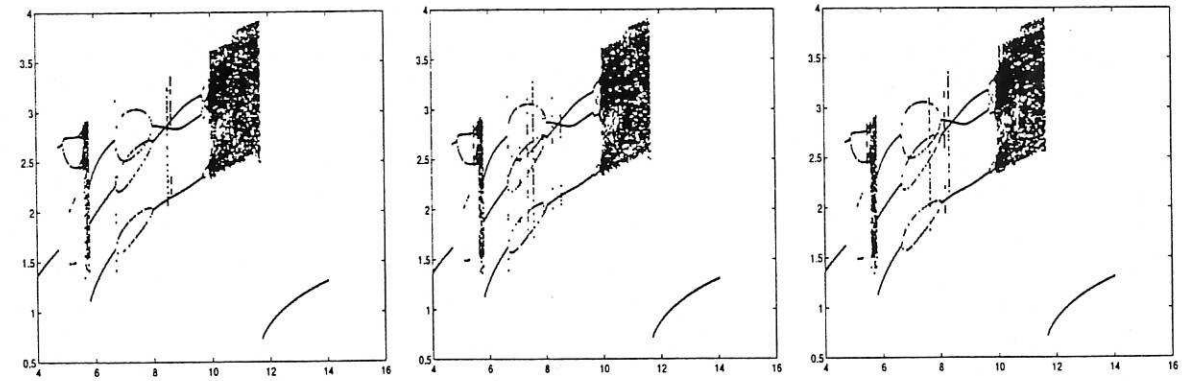
In the following simulations, the input to the system was changed to be a square wave with increasing amplitude with a Gaussian sequence superimposed. The amplitude of the square wave was gradually increased from 0.5 to 10 and the Gaussian sequence had a zero mean and a standard derivation of 1.0. The sample interval and the integration interval was both set to $\pi/60$. The input and output signals for identification are shown in Fig 11. The resulting discretised input-output model has the form given in equation (17). Models with various settings were identified. The results are summarised in Table 1.

For the setting $n_l = 3$ and the correct time lags $n_y = 2$, $n_u = 2$, all the identified



(a) (b) (c)

Figure 9: Poincaré sections of the identified models of the Duffing Ueda system with $n_l = 5$, $n_y = 3$, $n_u = 3$. (a) 7 terms, (b) 12 terms, (c) 18 terms. The sample interval was $\pi/60$.



(a) (b) (c)

Figure 10: Bifurcation diagrams of the identified models of the Duffing Ueda system with $n_l = 5$, $n_y = 3$, $n_u = 3$. (a) 7 terms, (b) 12 terms, (c) 18 terms. The sample interval was $\pi/60$.

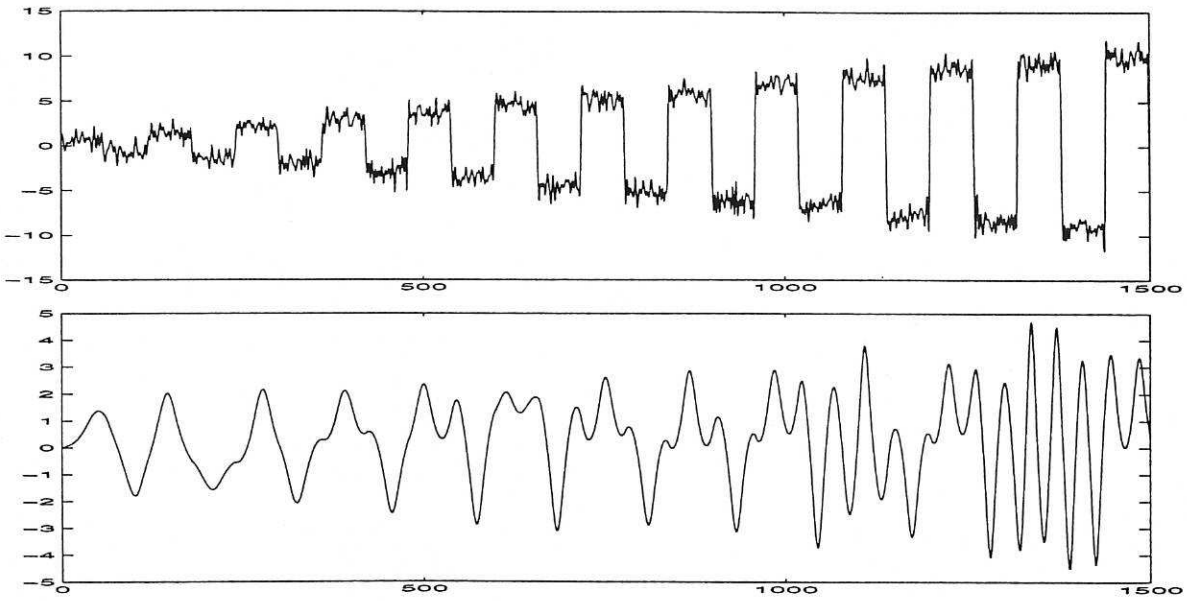


Figure 11: Input and output signals of the Duffing Ueda equation. The input was a square wave with increasing amplitude with a Gaussian sequence superimposed. The equation was integrated with a variable order Runge-Kutta algorithm and the input was kept as constant in each interval of $\pi/60s$. The data was sampled at an interval of $\pi/60s$.

models failed to capture the dynamic invariants of the system and the Poincaré sections and bifurcation diagrams were not correct. These results are much worse than those obtained when the input was a single frequency sinusoid. Because the amplitude was increased after every period of the square wave, the output was always in transition from one attractor to a nearby attractor and was not driven deep into chaos. In comparison for the sinusoidal input the output was driven into chaos by the single frequency input and the system dynamics were persistently excited.

Increasing the degree of nonlinearity n_l from 3 to 5, the resulting models had the correct Poincaré sections and bifurcation diagrams. As an extreme example, the number of terms was increased to 70. The model still captured the dynamic invariants of the real system. This example shows that overparametrisation in the number of terms appears to only have a marginal effect on the dynamics when the output is noise free. If the output is contaminated with noise such a model would certainly fail to capture the system dynamics.

Overparametrising the models by increasing the time lags to $n_y = 3$, $n_u = 3$ and setting the degree of nonlinearity n_l to 3 produced many valid models. However when the number of terms was less than 11, the model either diverged or did not have the correct Poincaré section and bifurcation diagram. The linear part of the model with 6 terms had an unstable pole and that of the model with 8 terms had a near pole-zero cancellation which was unstable. The linear components of all the models with more than 38 terms had an unstable pole-zero cancellation and diverged to infinity.

Overparametrising the time lags further by setting $n_y = 4$ and $n_u = 4$ resulted in a smaller range of the number of terms for valid models compared with the case of $n_y = 3$ and $n_u = 3$. The linear components of the models with less than 11 terms or more than 19 terms had one unstable and one stable pole-zero cancellation and the output diverged to infinity.

The simulation results above suggest that overparametrisation in the time lags has a severe effect on the resulting model dynamics. The linear component of the overparametrised models often had unstable poles and the magnitude of the unstable poles increased rapidly as the number of terms in the model was increased. Overparametrisation in both the time lags n_y and n_u may result in pole-zero cancellations in the linear components of the model and the cancellation may be stable or unstable. Clearly the unstable pole of the linear part of the model can lead to output instability. Although some of the models with near cancellation of stable poles and zeros captured the system dynamics, such a cancellation should be avoided when the data is noisy, since noise will introduce errors in the model coefficients and could induce unstable poles.

4.3 Overparametrisation in Radial Basis Function Networks

In a second set of simulations, a system represented by the nonlinear rational difference equation

$$y(k) = \frac{y(k-1)y(k-2)\left[y(k-1) + y(k-2) * u(k-1) - 2.8\right]}{1.0 + y^2(k-1) + y(k-1) * y(k-2) + y^2(k-2)} + u(k-1) \qquad (24)$$

Table 1: **Validation Results of the Duffing Ueda Models Identified from the Data with a Square Wave Input of Increasing Amplitude Superimposed with a Gaussian Sequence**

Settings	Number of Model Terms	Poincare Section	Bifurcation Diagram	unstable pole(p), pole(p) zero(z) cancellation
$n_y = 2,\ n_u = 2,\ n_l = 3$	6,8,10,12,24,30,35	nearly correct	incorrect	
$n_y = 2,\ n_u = 2,\ n_l = 5$	6,36,50,70	correct	correct	
$n_y = 3,\ n_u = 3,\ n_l = 3$	6	divergent	divergent	p: 2.9426
	8	divergent	divergent	p: 1.3490, z: 1.3805
	10	incorrect	incorrect	near cancellation,stable
	12,24,36,37	correct	correct	near cancellation,stable
	38	divergent	divergent	p: -1.3824, z: -1.3854
	40	divergent	divergent	p: -1.7324, z: -1.7345
	48	divergent	divergent	p: -2.5288, z: -2.5312
$n_y = 4,\ n_u = 4,\ n_l = 3$	10	divergent		p: -1.0186 .5048 z: -1.0258 .5053
	12,14,16,18,19	correct	correct	near cancelations,stable
	20	divergent	divergent	p: -5.6754 0.8491 z: -5.6739 0.8491
	24	divergent	divergent	p: -25.6158 -0.4521 z: -25.5947 -0.4520

was studied. This system can be expressed in state space form as

$$\begin{aligned} x_1(k+1) &= \frac{x_1(k)x_2(k)\left[x_1(k)+x_2(k)u(k)-2.8\right]}{1.0+x_1^2(k)+x_1(k)x_2(k)+x_2^2(k)} + u(k) \\ x_2(k+1) &= x_1(k) \\ y(k) &= x_1(k) \end{aligned} \tag{25}$$

Radial basis function neural networks were used to model the system. Cell to cell maps were then used to validate the trained networks and the effects of overparametrisation in the number of input nodes n_y and n_u and the number of radial basis function centres were investigated by comparing the cell to cell maps of the trained networks with that of equation (25).

The input $u(k)$ was selected as the bifurcation parameter μ and a cell structure was introduced in the state space and the parameter space by dividing the region $x_1 \in [-2\ \ 0.5]$, $x_2 \in [-2\ \ 0.5]$, $\mu \in [-0.5\ \ 0.5]$ into $50 \times 50 \times 50$ cells. By evaluating the difference equation (25) in the given region and unravelling the cells, the cell diagram of the system projected onto the $y-u$ plane is shown in Figure 12. The fixed points of the system are identified as p-1 and p-3 cells which form the curve shown in Figure 12. As the input u increases from -0.5 to -0.1, the system bifurcates and two more fixed points are created. As the input u increases further from -0.1 to 0, the system bifurcates again and the number of fixed points decreases to one.

One advantage of cell mapping analysis is that both stable and unstable fixed points

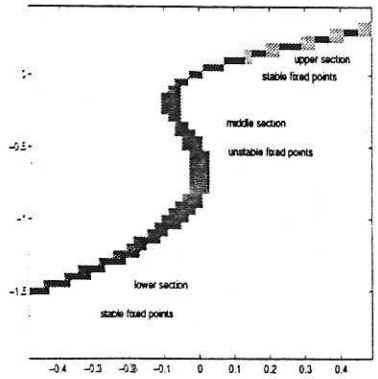


Figure 12: Cell mapping analysis of the difference equation (25) shows the P-1 cells (dark) and the P-3 cells (grey). Vertical axis - y, horizontal axis - u

of the system can be shown in the cell diagram. Because the system trajectory diverges from the unstable fixed point the unstable fixed points of the system can not be shown in the Poincaré map or the bifurcation diagram since they are based on the sampling of the system trajectory. In Figure 12, the lower and upper sections correspond to the stable fixed points, while the middle section, which connects the lower and upper sections, represents the unstable fixed points. The singular points are computed in the whole cell state space. The domains of attraction of the fixed point are not investigated in the present study. In all the simulations, the input signal to the system was a band limited uniformly distributed sequence in the range of [-0.8 0.8].

Initially the time lags were set to their correct values $n_y = 2$, $n_u = 1$, so that the network had 3 input nodes. Networks with various numbers of centres were trained and the cell maps of these networks are shown in Figure 13. It may be seen from the cell maps that the dynamics of the networks change slightly as the number of centres changes from 50 to 70. The fixed points of the networks with 20 and 40 centres were different from that of the true system near $u = -0.5$. It appears that a network with about 30 centres (Figure 13b) is sufficient to capture the underlying dynamics in the data.

Increasing the time lag of the output signal to $n_y = 3$, just one value larger than the correct value produced networks with very different cell maps as shown in Figure 14. For networks with a small number of centres, the fixed points near $u = -0.5$ were shifted downwards. The amount of shift decreased as the number of centres was increased. Notice that the cell map of the network with 70 centres, Figure 14f, for $n_y = 3$ and $n_u = 1$ is slightly worse than that of the network with 30 centres, Figure 13b, for $n_y = 2$ and $n_u = 1$. The increase in the number of centres appears to partly compensate for the incorrect setting in the number of input nodes or lagged variables. However, the number of weights in the former network is 280 compared with 90 in the latter.

Increasing further the number of input node assignments by setting the time lags to $n_y = 3$ and $n_u = 2$, so that the networks have 5 input nodes produced the cell maps of the trained networks with various numbers of centres shown in Figure 15. For networks with a small number of centres, the fixed points of the networks near $u = -0.5$ are again far away

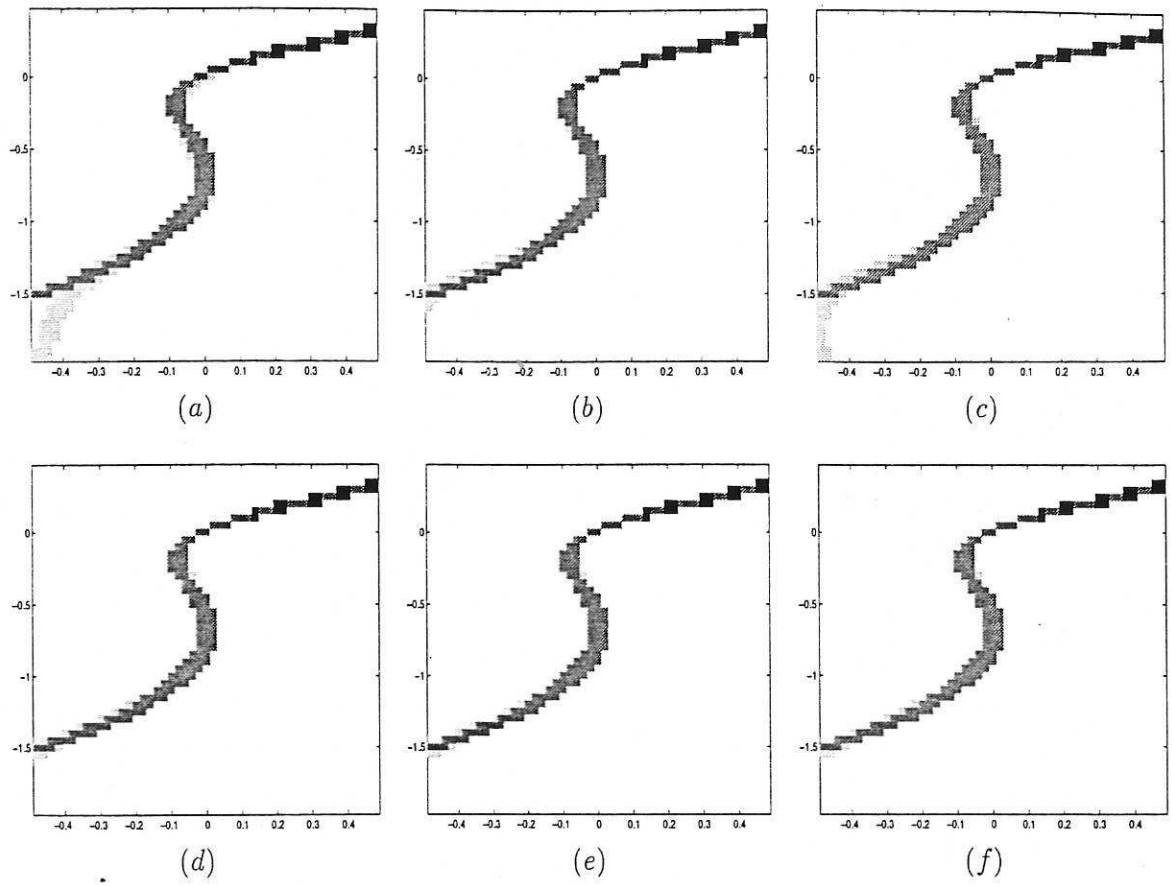


Figure 13: The cell maps of the system (dark) given in equation (25) and the trained RBF networks (grey) with $n_y = 2$, $n_u = 1$. (a) 20 centres, (b) 30 centres, (c) 40 centres, (d) 50 centres, (e) 60 centres, (f) 70 centres.

from the correct values. But as the number of centres is increased the differences becomes smaller.

The simulation results clearly show that overparametrisation in the number of input nodes can have deleterious effects on the resulting network. When the network was overparametrised in the number of input nodes, a large increase in the number of centres was necessary to partly counter the effect of overparametrisation in the input nodes. But this was achieved at the expense of a considerable increase in the number of weights in the network. If noisy data is used for the training the increase in the number of centres would be expected to induce large variance errors and result in overfitting but further work is required to study these effects.

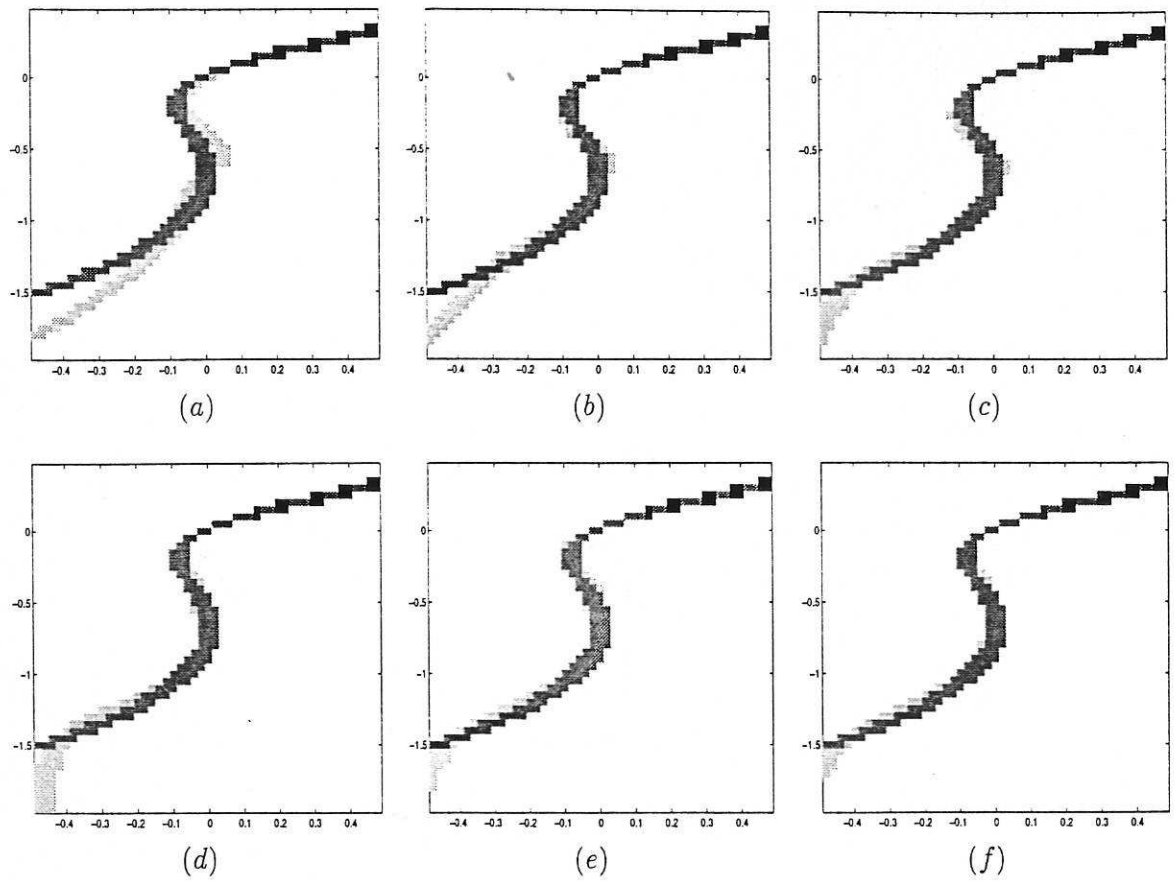


Figure 14: The cell maps of the system (dark) given in equation (25) and the trained RBF networks (grey) with $n_y = 3$, $n_u = 1$. (a) 20 centres, (b) 30 centres, (c) 40 centres, (d) 50 centres, (e) 60 centres, (f) 70 centres.

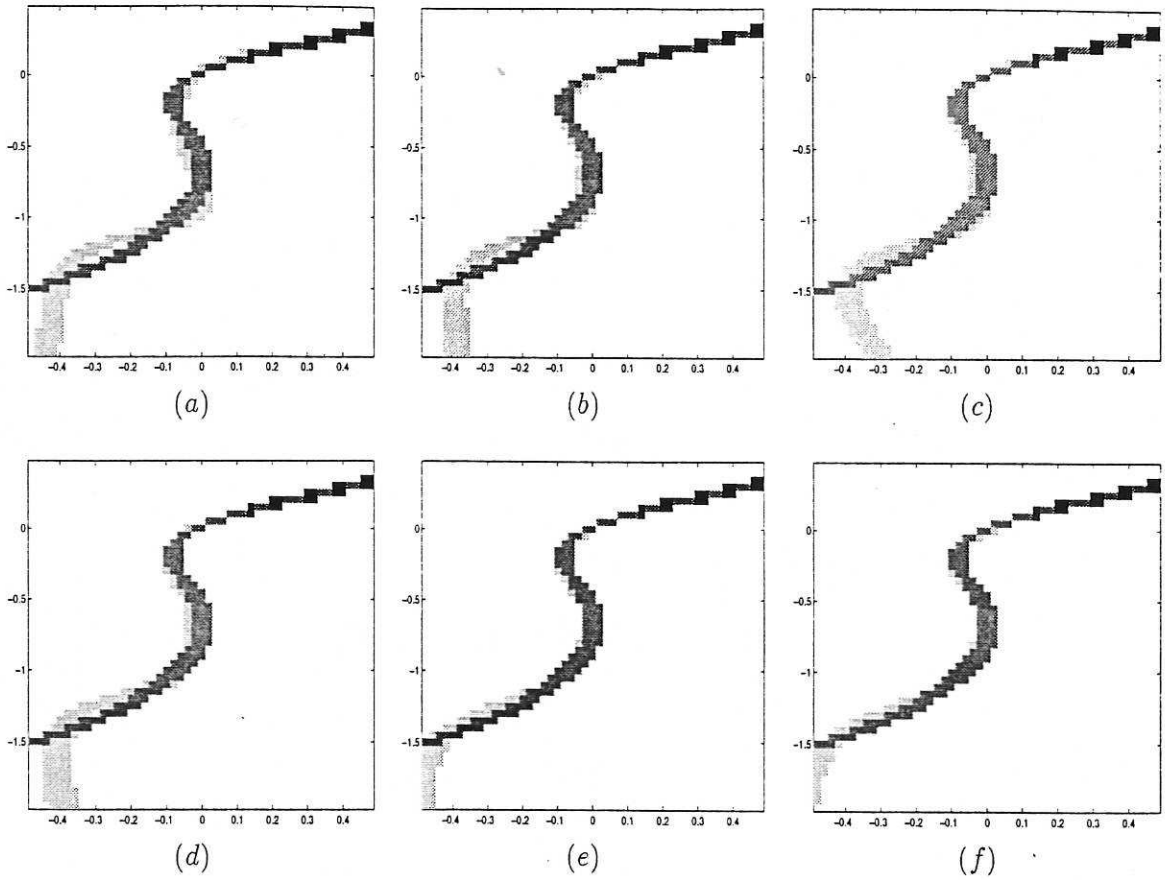


Figure 15: The cell maps of the system (dark) given in equation (25) and the trained RBF networks (grey) with $n_y = 3$, $n_u = 2$. (a) 20 centres, (b) 30 centres, (c) 40 centres, (d) 50 centres, (e) 60 centres, (f) 70 centres

5 Conclusions

The effect of overparametrisation on model dynamics has been investigated by qualitatively measuring the difference between the dynamic invariants of simulated systems with those of identified models or trained neural networks. The results suggest that overparametrisation in the number of time lags or the number of input layer nodes can have a much worse effect compared with overparametrisation in the number of hidden layer nodes or the degree of nonlinearity. The combined effects of overparametrisation in the number of time lags, the number of model terms and the degree of nonlinearity can induce unstable poles or lead to near pole-zero cancellations in the linear model terms of the NARMAX model. The magnitude of the unstable pole tends to increase rapidly as the number of terms is increased. Many simulations showed that similar phenomena can be observed in a wide range of systems.

A common practice in nonlinear system identification and neural network training is to assume that the time lags are known a priori and to optimise the degree of nonlinearity by varying the network architecture or order of nonlinear expansion. But the time lags will usually be unknown in practical applications and as the simulation results show a small increase in the number of time lags can have severe dynamic effects on the resulting models. Overparametrisation in the number of time lags can therefore considerably reduce the possibility of obtaining a globally valid model. It is therefore important to select the correct time lags in nonlinear system identification and neural network training.

Acknowledgements

The authors gratefully acknowledge that this work was supported by the EPSRC under contract GR/K71370.

References

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

Barbot, J.-P. (1989). A computer-aided design for sampling a nonlinear analytic system. In Mora, T., editor, *Applied Algebra, Algebraic Algorithms and Error Correcting Codes, Lecture Notes in Computer Science*, volume 357, pages 74–88. Springer-Verlag.

Billings, S. A. and Zheng, G. (1995). Radial basis function network configuration using genetic algorithms. *Neural Networks*, 8(6):877 – 890.

Billings, S. A. and Zhu, Q. M. (1995). Model validity tests for multivariable nonlinear models including neural networks. *Int. J. Control*, 62:749–766.

Casdagli, M. (1989). Nonlinear prediction of chaotic time series. *Physica D*, 20:335–356.

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

Haynes, B. R. and Billings, S. A. (1994). Global analysis and model validation in nonlinear system identification. *Nonlinear Dynamics*, 5:93–130.

Holcomb, T. and Morari, M. (1991). Local training for radial basis function networks: Towards solving the hidden unit problem. *Proc. American Control Conference*, pages 2331 – 2336.

Hsu, C. S. (1980). A theory of cell-to-cell mapping for nonlinear dynamical systems. *J. Applied Mechanics*, 47:931 – 939.

Judd, K. and Mees, A. I. (1994). On selecting models for nonlinear time series. Technical report, Department of Mathematics, University of Western Australia.

Leontaritis, I. J. and Billings, S. A. (1985). Input -output parametric models for nonlinear systems, part i + ii. *Int. J. Control*, 41(1):303 – 344.

Musavi, M. T., Ahmed, W., Chan, K. H., Faris, K. B., and Hummels, D. M. (1992). On the training of radial basis function classifiers. *Neural Networks*, 5:595 – 603.

Parker, T. S. and Chua, L. O. (1989). *Practical Numerical Algorithms for Chaotic Systems*. Springer Verlag.

Söderstrom, T. and Stoica, P. (1990). On covariance function tests used in system identification. *Automatica*, 26(1):125–133.

Ueda, Y. (1980). Steady motions exhibited by duffing's equation: A picture book of regular and chaotic motions. In Holmes, P. J., editor, *New approaches to nonlinear problems in dynamics*, pages 311–322. Springer-Verlag.

Zheng, G. L. and Billings, S. A. (1996). Qualitative validation and generalization in nonlinear system identification. *Submitted for Publication*.