



This is a repository copy of *Qualitative Validation and Generalization in Nonlinear System Identification*.

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/80504/>

Monograph:

Zheng, G.L. and Billings, S.A. (1996) *Qualitative Validation and Generalization in Nonlinear System Identification*. Research Report. ACSE Research Report 612 .
Department of Automatic Control and Systems Engineering

Reuse

Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher's website.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

Qualitative Validation and Generalization in Nonlinear System Identification

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

January 1996

Abstract — Cell to cell mapping for global analysis of nonlinear systems is adopted to enable the qualitative validation of identified nonlinear systems. The method provides a framework for the global analysis of a diverse range of nonlinear systems, including continuous and discrete time systems and nonlinear identified models. In the present study, the method is used to reveal the dynamic properties of a nonlinear system, including the fixed points, periodic, aperiodic solutions or chaotic behaviour and the corresponding stability properties. The orthogonal least squares algorithm (OLS) is then used to identify a parametric NARMAX model of the system. The resulting model is analyzed using the same framework and the dynamic properties of the model are qualitatively compared with those of the original system. Based on the results of the validation, a modified selection criterion for the OLS algorithm is proposed, which incorporates the nonlinear degree of the terms in the model complexity. The effectiveness of the new algorithm is demonstrated using examples.

Keywords — Qualitative Validation, Model Complexity, Generalization, System Identification.

Research Report No. 612

January 1996

200328356



Qualitative Validation and Generalization in Nonlinear System Identification

G. L. Zheng and S. A. Billings

Department of Automatic Control and Systems Engineering,
University of Sheffield, Mappin Street, Sheffield S1 3JD

January 1996

Abstract — Cell to cell mapping for global analysis of nonlinear systems is adopted to enable the qualitative validation of identified nonlinear systems. The method provides a framework for the global analysis of a diverse range of nonlinear systems, including continuous and discrete time systems and nonlinear identified models. In the present study, the method is used to reveal the dynamic properties of a nonlinear system, including the fixed points, periodic, aperiodic solutions or chaotic behaviour and the corresponding stability properties. The orthogonal least squares algorithm (OLS) is then used to identify a parametric NARMAX model of the system. The resulting model is analyzed using the same framework and the dynamic properties of the model are qualitatively compared with those of the original system. Based on the results of the validation, a modified selection criterion for the OLS algorithm is proposed, which incorporates the nonlinear degree of the terms in the model complexity. The effectiveness of the new algorithm is demonstrated using examples.

Keywords — Qualitative Validation, Model Complexity, Generalization, System Identification.

1 Introduction

Model validation constitutes an important aspect of any system identification study. In most applications, it is necessary to obtain not only a statistically but also a dynamically valid model of the system under study. Statistical validation methods have been studied by many researchers (Bohlin, 1971; Leontaritis and Billings, 1987; Billings and Tao, 1991; Billings and Zhu, 1994; Billings and Zhu, 1995). However, a statistically valid model may not be dynamically valid (Haynes and Billings, 1994) and may fail to produce the dynamical invariants of the original system. The most important dynamical invariants of a nonlinear system are probably the equilibrium points or fixed points, periodic, aperiodic, or strange attractors and domains of attraction of these attractors. Bifurcation theory and singularity theory are usually used for the analysis of parameterised nonlinear systems. But these

methods require *a priori* knowledge about the solution structure of the system and can usually only be applied to systems with simple specific structures. In the present paper, the cell to cell mapping method (Hsu, 1980) is applied to the global analysis of parameterised nonlinear systems and is adopted for the qualitative validation of identified nonlinear models. With this approach the dynamical invariants of the identified model can be shown graphically in a cell state space and can be compared with those of the original system. The ultimate aim of the validation presented in the paper is to find out which aspects of the system identification algorithm result in a dynamically invalid model. The OLS algorithm was applied to identify a parametric NARMAX model of a known system. The drawback of the selection criterion of the algorithm was identified as one cause of model deficiency. A modified selection criterion was therefore proposed and improved results were obtained. The prediction performance is often used as a measure of the validity of the identified model. Such a measurement was also investigated over a validation set and the pitfalls in applying such a measure for validation are investigated.

The layout of the paper is as follows. A cell mapping analysis method is presented in section two and a simple example is used to demonstrate the application of the method for the global analysis of nonlinear dynamical systems. Section three describes the OLS algorithm and associated selection criterion. A modified selection criterion is proposed which incorporates both the approximation capability and the complexity of the terms to be included in the model. Section four outlines the procedure of qualitative validation of identified nonlinear models. Simulation examples are presented in section five to demonstrate the usage of the qualitative validation and the effectiveness of the newly proposed model selection algorithm.

2 Cell to Cell Mapping

Most physical systems can be represented by a system of ordinary differential equations

$$\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}, \mu) \quad (1)$$

or a system of difference equations

$$\mathbf{x}(n+1) = \mathbf{G}(\mathbf{x}(n), \mu) \quad (2)$$

where $\mathbf{x} \in \mathbf{R}^N$ is the state vector, $\mu \in \mathbf{R}^P$ is the parameter vector, $\mathbf{F} : \mathbf{R}^N \times \mathbf{R}^P \rightarrow \mathbf{R}^N$, $\mathbf{G} : \mathbf{R}^N \times \mathbf{R}^P \rightarrow \mathbf{R}^N$. The component of the state vector \mathbf{x} is 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 limited accuracy of measurement and computation. The state space is 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 $i = 1, 2, \dots, N$ into N_i intervals. These intervals are identified by the integer values of the corresponding cell coordinates $z_i = 1, 2, \dots, N_i$. The Cartesian product space of z_1, z_2, \dots, z_N is then 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. An integer valued vector function $\mathbf{F}(\mathbf{z})$ can then be defined over the cell state space and is called a cell function.

Let x_i , $i \in \{N\}$, be the state variables, where $\{N\}$ denotes a set of positive integers $i = 1, 2, \dots, N$. Let the coordinate axis of a state variable x_i be divided into N_i uniform intervals of size h_i . The interval z_i along the x_i -axis is defined such that it contains all x_i satisfying

$$(z_i - \frac{1}{2})h_i \leq x_i < (z_i + \frac{1}{2})h_i \quad (3)$$

This defines a cell vector \mathbf{z} of an N -tuple of integers z_i , $i \in \{N\}$. A point \mathbf{x} in the N -dimensional state space with components x_i corresponds to a cell \mathbf{z} in the cell space with components z_i if and only if x_i and z_i satisfy equation (3) for all $i \in \{N\}$.

Having defined a cell state space, the discrete time evolution of the dynamical system described by equation (1) or (2) can now be put in the form of a simple cell mapping

$$\mathbf{z}(n+1) = \mathbf{C}(\mathbf{z}(n), \mu), \quad \mathbf{z} \in \mathbf{Z}^N, \mu \in \mathbf{R}^P \quad (4)$$

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

A singular cell or an equilibrium cell \mathbf{z}^* is a cell satisfying the relation

$$\mathbf{z}^* = \mathbf{C}(\mathbf{z}^*, \mu) \quad (5)$$

It is an analogue of the equilibrium point of equation (1) or the fixed point of equation (2).

Let \mathbf{C}^m denote the cell mapping \mathbf{C} applied m times, and \mathbf{C}^0 be the identify mapping. A sequence of K distinct cells $\mathbf{z}^*(j)$, $j \in \{K\}$, which satisfies

$$\begin{aligned} \mathbf{z}^*(m+1) &= \mathbf{C}^m(\mathbf{z}^*(1), \mu), \quad m \in \{K-1\}, \\ \mathbf{z}^*(1) &= \mathbf{C}^K(\mathbf{z}^*(1), \mu) \end{aligned} \quad (6)$$

is said to constitute a periodic solution (or motion) of period K . Each of its elements $\mathbf{z}^*(j)$ is called a periodic cell of period K . For ease of reference, such a motion 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. Obviously, a P-K solution is an analogue of the periodic solution of equation (1) or equation (2).

Accordingly, the domains of attraction for simple cell mapping systems can also be defined. If a cell \mathbf{z} is mapped in r steps into one of the P-K cells $\mathbf{z}^*(j)$ of a P-K solution, the cell \mathbf{z} is said to be r steps removed from the P-K solution, here r is the minimum positive integer such that $\mathbf{C}^r(\mathbf{z}) = \mathbf{z}^*(j)$. The set of all cells which are r steps or less removed from a P-K solution is called the r -step domain of attraction of that P-K solution. The total domain of attraction (or simply domain of attraction) of a P-K solution is the r -step domain of attraction with $r \rightarrow \infty$. The cells in the domain of attraction of a P-K solution are denoted as DOA-K cells.

Associated with the cell mapping, a cell function is defined as

$$\mathbf{F}(\mathbf{z}, \mathbf{C}, \mu) = \mathbf{C}(\mathbf{z}, \mu) - \mathbf{z} \quad (7)$$

and a k -step ahead cell function is defined as

$$\mathbf{F}(\mathbf{z}, \mathbf{C}^K, \mu) = \mathbf{C}^K(\mathbf{z}, \mu) - \mathbf{z} \quad (8)$$

By dividing the cell state space into multiplets, the singularities of the cell function and k -step ahead cell function can be determined by using their values at the vertices of the multiplets. For a comprehensive description of the cell mapping, the reader is referred to (Hsu, 1980).

For most practical problems there is only a finite region of the state space beyond which the further evolution of the system is no longer the concern of our study. Similarly, for a dynamical system governed by a cell mapping there is only a finite region of cell state space \mathbf{Z}^N which is of interest to us. The total number of cells in this region is N_c and the cells are called regular cells. A sink cell is then defined to encompass all the cells outside the region of interest. The total number of cells is therefore $N_c + 1$. 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. Let the cells be labelled using the following transform

$$\begin{aligned} l(\mathbf{z}) &= z_1 + (z_2 - 1)N_1 + (z_2 - 1)(z_3 - 1)N_1N_2 + \dots, \quad \forall z_i \in \{N_i\}, i \in \{N\} \\ l(\mathbf{z}) &= 0, \quad \exists i, z_i \notin \{N_i\} \end{aligned} \quad (9)$$

Such that the regular cells are labelled by positive integers $1, 2, \dots, N_c$, where $N_c = \prod_1^N N_i$, and the sink cell is labelled by 0, the zero cell. The N_c regular cells and the sink cell are then denoted by a set $S = \{N_c + \}$. This set is closed under the cell mapping

$$\begin{aligned} \mathbf{z}(n+1) &= \mathbf{C}(\mathbf{z}(n), \mu), \quad l(\mathbf{z}(n)) \in \{N_c\} \\ \mathbf{z}(n+1) &= \mathbf{C}(\mathbf{z}(n), \mu) = \mathbf{z}(n), \quad l(\mathbf{z}(n)) = 0 \end{aligned} \quad (10)$$

The set of regular cells within the influence of the sink cell forms the domain of attraction of the sink cell, and is labelled DOA-Sink. These cells will eventually be mapped into the sink cell and are denoted DOA-Sink cells.

The cell mapping given above represents the system for a given point in the parameter space \mathbf{R}^P . The cell mapping can be extended to enable the qualitative analysis of parameterised nonlinear systems by combining it with bifurcation theory. For a discrete system represented by the difference equations

$$x_i(n+1) = g_i(\mathbf{x}(n), \mu) \quad i \in \{N\} \quad (11)$$

Let the region of interest of the state variables be

$$\mathbf{x}^{(l)} \leq \mathbf{x} \leq \mathbf{x}^{(h)} \quad (12)$$

and the region of interest of the parameters be

$$\mu^{(l)} \leq \mu \leq \mu^{(h)} \quad (13)$$

Introduce an extended state variable $\bar{\mathbf{x}}$ as

$$\begin{aligned} \bar{\mathbf{x}} &= (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_{N+P})^T = (\mathbf{x}^T \mu^T)^T \\ &= (x_1, x_2, \dots, x_N, \mu_1, \mu_2, \dots, \mu_P)^T \end{aligned} \quad (14)$$

The point mapping may be rewritten as

$$\begin{aligned}\bar{x}_i(n+1) &= g_i(\mathbf{x}(n), \mu(n)) \quad i = 1, 2, \dots, N \\ \bar{x}_i(n+1) &= \bar{x}_i(n) = \mu_{i-N}(n) \quad i = N+1, N+2, \dots, N+P\end{aligned}\quad (15)$$

Introduce cell structures into both the state space and the parameter space by dividing \bar{x}_i into N_i uniform intervals of size h_i , such that z_i along the \bar{x}_i -axis contains all \bar{x}_i satisfying

$$\begin{aligned}(z_i - \frac{1}{2})h_i &\leq \bar{x}_i(=x_i) < (z_i + \frac{1}{2})h_i \quad i = 1, \dots, N \\ (z_i - \frac{1}{2})h_i &\leq \bar{x}_i(=\mu_{i-N}) < (z_i + \frac{1}{2})h_i \quad i = N+1, \dots, N+P\end{aligned}\quad (16)$$

Note the cell vector \mathbf{z} is an $N+P$ -tuple of integers z_i , $i \in \{N+P\}$. The centre point of each cell can be calculated as

$$\begin{aligned}\bar{x}_i^{(c)}(n) &= x_i^{(c)}(n) = x_i^{(l)} + h_i z_i(n) - \frac{h_i}{2} \\ &z_i(n) \in \{N_i\}, \quad i = 1, \dots, N \\ \bar{x}_i^{(c)}(n) &= \mu_{i-N}^{(c)}(n) = \mu_{i-N}^{(l)} + h_i z_i(n) - \frac{h_i}{2} \\ &z_i(n) \in \{N_i\}, \quad i = N+1, \dots, N+P\end{aligned}\quad (17)$$

The image of the centre point $\bar{x}^{(c)}(n)$ is then calculated using equation (15). The cell map $\mathbf{C}(\mathbf{z})$ is constructed by determining the image of the centre point of each cell within \mathbf{S} by

$$\begin{aligned}z_i(n+1) &= C_i(\mathbf{z}(n)) = \text{INT} \left[\frac{x_i^{(c)}(n+1) - x_i^{(l)}}{h_i} + 1 \right], \\ &i = 1, 2, \dots, N \\ z_i(n+1) &= C_i(\mathbf{z}(n)) = z_i(n) = \mu_{i-N}(n), \\ &i = N+1, N+2, \dots, N+p\end{aligned}\quad (18)$$

Note that the resulting cell mapping contains $\prod_{j=1}^P N_{N+j}$ slices of N dimensional cell mappings in the parameter space \mathbf{R}^P . Having defined a cell mapping over the Cartesian product of the state space and the parameter space, the dynamical behaviour of the system can then be revealed by computing the mapping images of all the regular cells within the set \mathbf{S} . For a system described by the difference equations (2), the cell map given in (15) and (18) is used to compute the images of all the regular cells. For a system described by the differential equations (1), the mapping images of the regular cell is computed by integrating the equations (1) at a fixed time interval T with the cell centre as the initial conditions. Equation (18) is then used to decide the image cells. The time interval T should be selected such that there are not too many cells being mapped into themselves or into the sink cell.

After the image of all the regular cells have been found, the regular cells within \mathbf{S} can be classified using the unravelling algorithm (Hsu, 1980). The algorithm calls all the regular cells within \mathbf{S} 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

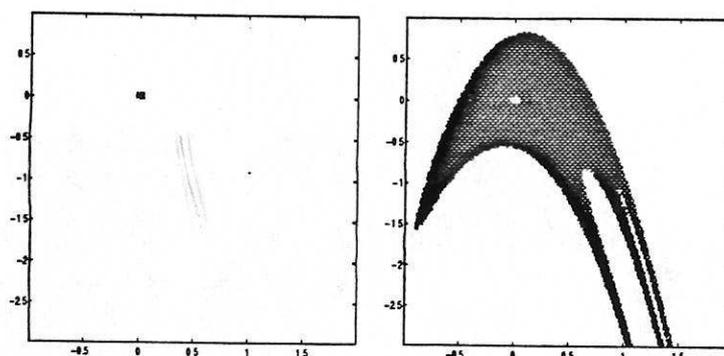


Figure 1: left: A core of period four cells correspond to the stable focus at $(0, 0)$ and a singular cell corresponding to the saddle point at $(1, -0.9)$. right: Regular cells in the domain of attraction of the period four cells, the regular cells in the blank region are mapped into the sink cell

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.

A Simple Cell Mapping Example. The dynamical properties of a simple point mapping system is analyzed here using the cell mapping algorithm. The system is

$$\begin{aligned} x_1(n+1) &= (1-\mu)x_2(n) + (2-\mu+\mu^2)x_1^2(n) \\ x_2(n+1) &= -(1-\mu)x_1(n) \end{aligned} \quad (19)$$

This system has a stable focus at $(x_1, x_2) = (0, 0)$ and a saddle point at $(x_1, x_2) = (1, -(1-\mu))$. Set up a cell structure for $\mu = 0.1$ in the region $-1 \leq x_1 \leq 2$, $-3 \leq x_2 \leq 1$. Divide each coordinate direction into 200 uniform intervals. There are 40000 regular cells in the cell state space. The stable focus was identified by the cell mapping analysis as a core of period 4 cells near the origin and the saddle point at $(1, -0.9)$ was revealed to be a period 1 cell as shown in Fig 1. The cells in the domain of attraction of the PK - 4 cells are also shown in the same figure. The cells in the blank region were mapped into the sink cell.

3 NARMAX Representation and the Orthogonal Least Squares Algorithm

Consider the nonlinear discrete time system

$$\begin{aligned} \mathbf{x}(k+1) &= \mathbf{g}(\mathbf{x}(k), u(k)) \\ y(k) &= h(\mathbf{x}(k), u(k)), \quad \mathbf{x}(0) = \mathbf{x}_0 \end{aligned} \quad (20)$$

where $\mathbf{x}(k) \in \mathbf{R}^N$ is the state at k , $k = 1, 2, \dots$, \mathbf{x}_0 is the initial state, $u(k)$, $y(k) \in \mathbf{R}$ are input and output signals respectively. $\mathbf{g} : \mathbf{R}^N \times \mathbf{R} \rightarrow \mathbf{R}^N$ is the one step ahead transition

function and $h : \mathbf{R}^N \times \mathbf{R} \rightarrow \mathbf{R}$ is the output function. A NARMAX model of the system (20) can be obtained (Leontaritis and Billings, 1985) as

$$y(k) = F^*(y(k-1), \dots, y(k-n_y), u(k-1), \dots, u(k-n_u)) \quad (21)$$

where F^* is a nonlinear function. The model is valid in a region around the equilibrium point of the system and is independent of the initial condition \mathbf{x}_0 . Assume the output $y(k)$ is corrupted by an additive noise $\xi(k)$, the NARMAX model is then in the form

$$y(k) = F(y(k-1), \dots, y(k-n_y), u(k-1), \dots, u(k-n_u), \xi(k-1), \dots, \xi(k-n_\xi)) + \xi(k) \quad (22)$$

Expanding the nonlinear function F , the model may be written as

$$y(k) = \sum_{j=1}^{n_\theta} \theta_j \phi_j + \theta_0 + \xi(k) \quad (23)$$

where θ_0 is a constant term and n_θ is the total number of terms in the model, θ_j ($j = 1, \dots, n_\theta$) are the coefficients and ϕ_j ($j = 1, \dots, n_\theta$) are the product combinations of $y(k-1), \dots, y(k-n_y), u(k-1), \dots, u(k-n_u), \xi(k-1), \dots, \xi(k-n_\xi)$ from one upto n_l . In system identification, the model terms and model structure (*i.e.* n_θ and n_l) are unknown and must be identified from the input and output signals. The OLS method was developed for selecting the model terms from the following full model

$$y(k) = \sum_{j=1}^M \theta_j \phi_j + \theta_0 + \xi(k) \quad (24)$$

where ϕ_j , $j = 1, \dots, M$ are all the product combinations of $y(k-1), \dots, y(k-n_y), u(k-1), \dots, u(k-n_u), \xi(k-1), \dots, \xi(k-n_\xi)$ from order one to a predetermined upper limit n_l . The number of terms in the full model (24) amounts to $C_{n_y+n_u+n_\xi+n_l}^{n_l}$. For a set of input output signals ($y(k)$ $u(k)$, $k=1, 2, \dots, \mathcal{N}$), the input output relation may be written in vector form as

$$\mathbf{Y} = \Phi \Theta + \Xi \quad (25)$$

An orthogonal decomposition of Φ is given as

$$\Phi = \mathbf{P} \mathbf{A} \quad (26)$$

where \mathbf{A} is an $(M+1) \times (M+1)$ unit upper triangular matrix and \mathbf{P} is an $\mathcal{N} \times (M+1)$ matrix with orthogonal columns that satisfy

$$\mathbf{P}^T \mathbf{P} = \mathbf{D} = \text{diag}\{d_0, d_1, \dots, d_M\} \quad (27)$$

with

$$d_j = \langle \mathbf{p}_j, \mathbf{p}_j \rangle, \quad j = 0, 1, \dots, M \quad (28)$$

where $\langle \bullet, \bullet \rangle$ denotes the inner product. Rearranging equation (25) yields

$$\mathbf{Y} = (\Phi \mathbf{A}^{-1})(\mathbf{A} \Theta) + \Xi = \mathbf{P} \mathbf{q} + \Theta \quad (29)$$

Because $\xi(k)$ is uncorrelated with the past input and output signals, it may be shown (Chen, Billings and Luo, 1989) that

$$q_i = \frac{\langle \mathbf{p}_j, \mathbf{Y} \rangle}{\langle \mathbf{p}_j, \mathbf{p}_j \rangle}, \quad j = 0, 1, \dots, M \quad (30)$$

The number of candidate terms M may be very large, but a small number of terms may be adequate to approximate the underlying system dynamics. These terms can be identified using an efficient forward selection procedure derived in (Billings, Korenberg and Chen, 1988; Billings, Chen and Korenberg, 1989; Billings and Chen, 1989). The principle of the method is shown below. From equation (29), the sum of the squares of the output is

$$\langle \mathbf{Y}, \mathbf{Y} \rangle = \sum_{j=0}^M q_j^2 \langle \mathbf{p}_j, \mathbf{p}_j \rangle + \langle \boldsymbol{\Xi}, \boldsymbol{\Xi} \rangle \quad (31)$$

where the errors are assumed to be uncorrelated with past input and output. The error reduction ratio (Billings et al., 1988) due to \mathbf{p}_j may be expressed as

$$\text{err}_j = \frac{q_j^2 \langle \mathbf{p}_j, \mathbf{p}_j \rangle}{\langle \mathbf{Y}, \mathbf{Y} \rangle} \quad (32)$$

The best candidate term at each step is the one which achieves the largest error reduction ratio err_j . The selection procedure may be terminated when a desired error tolerance ρ ($0 < \rho < 1$) is achieved.

$$1 - \sum_{j=0}^{n_\xi} \text{err}_j < \rho \quad (33)$$

Note that the constant term may or may not be selected for a particular application. The tolerance ρ will affect both the approximation accuracy and the complexity of the model. The criterion emphasizes the approximation accuracy of the model only. The resulting model may tend to interpolate the particular data set and thus lead to overfitting if an inappropriate ρ value is used. A compromise may be achieved by using Akaike's information criterion (AIC) to terminate the selection procedure (Chen et al., 1989). AIC is equivalent to the expected test set error or the prediction risk for model selection. The prediction risk is a measure of the generalization capability of the resulting model on the future data. It has the general form (Akaike, 1970; Barron, 1984; Murata and Yoshizawa, 1994; Liu, 1995)

$$\epsilon_{val}^2 = \epsilon_{est}^2 + \sigma^2 \frac{2n_{eff}}{\mathcal{N}} \quad (34)$$

where ϵ_{val}^2 and ϵ_{est}^2 are the sum of squared error on the validation set and the estimation set respectively, σ^2 is the variance of the noise and is usually replaced by ϵ_{est}^2 in practical applications, \mathcal{N} is the number of the data samples in the estimation set and n_{eff} is the effective number of parameters in the model. For a model which is linear in the parameters, the effective number of parameters is the same as the number of parameters in the model. AIC or similar model selection criteria are theoretically significant, they do, however seem to ignore the term complexity in the particular application of NARMAX model selection. As

implied in the criteria, all the terms are treated equivalently and are counted as one effective parameter no matter how complex they might be. The selection procedure described above selects the term which approximates the data set most effectively at each step and terminates when a minimum of the criterion is reached. One drawback of this selection procedure is that the complexity of the terms is not incorporated into the criterion. The higher degree terms usually have higher error reduction ratios and therefore are often included in the model. As shown in section five, higher degree terms can have a deleterious effect on the resulting model and often induce spurious limit cycles. It is therefore desirable to exclude the higher degree terms from the model.

In the present work, a new approach is introduced whereby the selection procedure is modified to reflect both the approximation capability on the estimation set and the complexity of the term. The degree of nonlinearity is used as a measure of complexity of the term. At each step k , the term which minimizes the following criterion is selected

$$\mathbf{J}(k) = \epsilon_{est}^2(k) \times \left(1.0 + \frac{2n_{eff}(k)}{\mathcal{N}}\right) \quad (35)$$

where $n_{eff}(k)$ has a different meaning to that given in equation (34) and the degree of nonlinearity n_l is defined as the effective number of parameter for each term. The selection terminates when the criterion reaches a minimum. The selection procedure is therefore equivalent to selecting a model which minimizes the following criterion

$$\mathbf{J} = \epsilon_{est}^2 \times \left(1.0 + \frac{2n_{eff}}{\mathcal{N}}\right) \quad (36)$$

In section five, both selection algorithms will be used to identify the NARMAX model of a given system. The resulting models are analyzed using the cell mapping algorithm and compared. It will be demonstrated in section five that the new selection algorithm achieves improved results.

4 Qualitative Validation of Identified NARMAX Models

The cell mapping method provides a useful framework for global analysis of a diverse range of nonlinear systems, such as the NARMAX model and feedforward neural networks. For a given system, the system dynamics can be shown graphically in a cell state space using the cell mapping algorithm. The dynamical properties of the identified nonlinear model of the system can then be analyzed in the same framework. The deficiency of the identified model is thus graphically revealed. To perform cell mapping analysis on a NARMAX model, a cell structure must be introduced. The validity of the identified model should be independent of any specific form of $u(k)$, the input signal $u(k)$ is therefore selected as a bifurcation parameter and the dependence on time is dropped in model validation. A state space representation of the identified NARMAX model is obtained and cell structures are introduced into both the state space and parameter space. In the validation, only the process model is considered, the noise terms $\xi(k)$ which are included in the model to avoid bias are neglected. The model then becomes

$$y(k) = F(y(k-1), \dots, y(k-n_y), \mu) \quad (37)$$

where $u(k-1), \dots, u(k-n_u) = \mu$. A state space representation of this model is

$$\begin{aligned} x_1(k+1) &= F(x_1(k), \dots, x_{n_y}(k), \mu) \\ x_2(k+1) &= x_1(k) \\ &\dots \\ x_{n_y}(k+1) &= x_{n_y-1}(k) \end{aligned} \quad (38)$$

Note the singular points of the system are all on the hyperplane $x_1 = x_2 = \dots = x_{n_y}$. These singular points correspond to all the fixed points of the identified NARMAX model and can be easily computed on this hyperplane. The limit cycles and domains of attraction however should be computed in the whole cell state space. The validation procedure may be summarized as follows.

1. Compute the cell to cell mapping of the true system. Unravel the cells to reveal the singular cells, the P-K cells and the corresponding domains of attraction.
2. Identify a NARMAX model of the system. Represent the model in state space form. Construct a cell state space in the state space and the parameter space (the input u). Compute the cell to cell mapping of the model and unravel the cells.
3. Find the singular points (P-1 cells) of the identified model on the hyperplane $x_1 = x_2 = \dots = x_{n_y}$, compare the singular points with those of the true system.
4. Find the limit cycles (P-K cells, $K > 1$) of the identified model, compare the P-K cells with those of the true system.
5. Compute the domains of attraction of the singular cells and the P-K cells. Compare them with those of the true system

The dimension of the cell state space is determined by the number of lags in the output signal $y(k)$. For a high dimensional problem (with larger time lags), the cell mapping analysis may become computationally expensive. If a model is inadequate, the validation based on the singular points in step 3 is often found to be sufficient to reveal the inadequacy. If a model satisfies the validation on all the singular cells and the P-K cells, the domains of attraction of these cells should be analyzed and the dynamical properties of the model can be graphically shown. Since the state variables of the model and the true system are usually not the same, care must be taken in explaining the domains of attraction of the P-K cells. The domains of attraction may not be exactly the same as those of the true system. However, the validation procedure should be able to pin point any inadequacy in the identified model and reveal in which region of the state space the model is valid. The validation will always be limited by the domain of the data used in the identification and will not in general be sufficient to justify that the model is valid in the whole space. In practice, the true system will be unknown, the main purpose of the qualitative validation therefore is to investigate which aspects of model identification affect the identified model and how these can be improved.

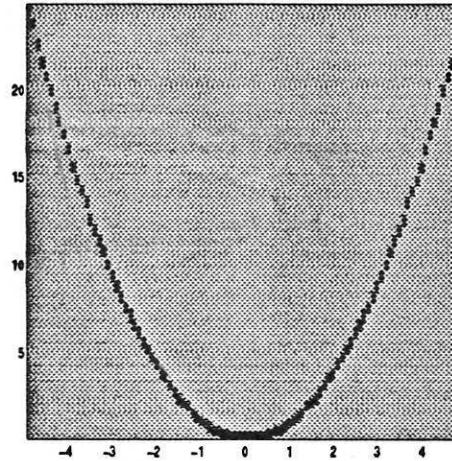


Figure 2: Cell mapping analysis of the differential equations (39) shows the P-1 cells (dark) and the DOA-1 cells (grey). vertical axis - y , horizontal axis - u

5 Simulation Results

Consider a nonlinear system represented by the following differential equations

$$\begin{aligned}\dot{x}_1(k) &= -2.0x_1(k) + 2.0x_2^2(k) \\ \dot{x}_2(k) &= -10.0x_1(k) + 10.0u(k) \\ y(k) &= x_1(k)\end{aligned}\quad (39)$$

Select the input $u(k)$ as a bifurcation parameter μ . Introduce a cell structure in the state space and the parameter space by dividing the the region $x_1 \in [0 \ 25]$, $x_2 \in [-5 \ 5]$, $\mu \in [-5 \ 5]$ into $125 \times 125 \times 100$ cells. By integrating the differential equations at an interval of 0.2 second and unravelling the cells, the cell diagram of the system projected onto the $y - u$ plane was obtained as shown in **Fig 2**. The equilibrium points of the system are identified as p-1 cells which form the curve $y = u^2$. All the cells in the region of interest are mapped to the equilibrium points as shown in the cell diagram. For this simple example, the singular points are computed in the whole cell state space instead of the hyperplane $x_1 = x_2 = \dots = x_{n_y}$ in all the simulations presented in this paper.

To identify a NARMAX model of the system, an input sequence uniformly distributed in the interval $(-0.5 \ 1.5)$ was applied to the system and an output sequence was obtained by integrating the differential equations and sampling at 0.2 second intervals. The 800 samples shown in **Fig 3** are used for identification. Two selection algorithms are used for model identification, one selects the terms with the ERR criterion and terminates the selection when the criterion in equation (34) reaches a minimum, the other selects the terms using the modified criterion (35) and terminates when the criterion reaches a minimum. For ease of reference, the former will be called the OLS algorithm and the later the modified OLS algorithm. For comparison, NARMAX models of the system for various values of n_y , n_u and n_l with linear noise terms were identified using both the algorithms.

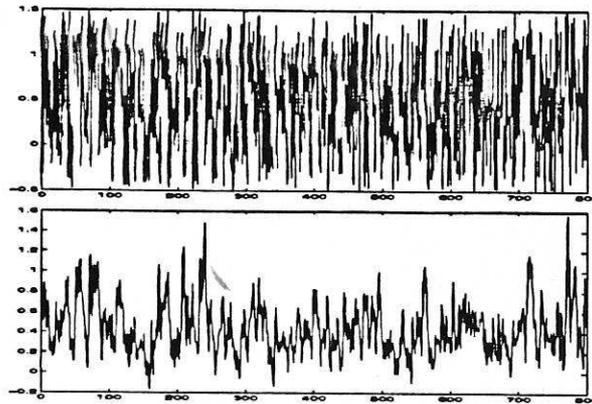


Figure 3: Input (above) and output signals of the system given in equation (39) sampled at 0.2 second intervals. The input signal is uniformly distributed in the region (-0.5 1.5). A Gaussian noise was added to the output. The resulting signal to noise ratio was 20 DB

Table 1: Selected terms using the OLS algorithm for $n_y = n_u = 1$, $n_l = 2, \dots, 5$ and $n_\xi = 5$

$n_l = 2$ $\frac{\sum \xi_{op}^2(k)}{N} = 3.1288e-2$ $\frac{\sum \xi_{mp}^2(k)}{N} = 1.7494e-1$	$n_l = 3$ $\frac{\sum \xi_{op}^2(k)}{N} = 3.4381e-2$ $\frac{\sum \xi_{mp}^2(k)}{N} = \infty$	$n_l = 4$ $\frac{\sum \xi_{op}^2(k)}{N} = 3.8951e-2$ $\frac{\sum \xi_{mp}^2(k)}{N} = \infty$	$n_l = 5$ $\frac{\sum \xi_{op}^2(k)}{N} = 3.2227e-2$ $\frac{\sum \xi_{mp}^2(k)}{N} = 4.8367e-2$
$y(k-1)$ $u^2(k-1)$ $u(k-1)$ $u(k-1)y(k-1)$ constant $y^2(k-1)$ $\xi(k-1)$ $\xi(k-3)$ $\xi(k-2)$	$y(k-1)$ $u^2(k-1)$ $u(k-1)$ $u(k-1)y^2(k-1)$ constant $\xi(k-1)$ $\xi(k-3)$ $\xi(k-2)$	$y(k-1)$ $u^2(k-1)$ $u(k-1)$ $u^2(k-1)y^2(k-1)$ constant $\xi(k-1)$ $\xi(k-3)$ $\xi(k-2)$	$y(k-1)$ $u^2(k-1)$ $u(k-1)$ $u^3(k-1)y^2(k-1)$ $u^3(k-1)y(k-1)$ $u^2(k-1)y^3(k-1)$ $u^4(k-1)y(k-1)$ $\xi(k-1)$ $\xi(k-3)$ $\xi(k-2)$

Consider initially constraining the model set by the selection $n_y = n_u = 1$, $n_l = 2, \dots, 5$ and $n_\xi = 5$. The terms selected using the OLS algorithm and the mean square errors are given in Table 1. The terms are presented in the order of selection. A cell state space was introduced by dividing the region $y \in [0 \ 25]$ and $u \in [-5 \ 5]$ into 200×100 uniform intervals. The cell mapping diagrams of the models and the DOA-Sink cells are presented in Fig 4 and Fig 5 respectively. The resulting models are only valid in a small region of the cell state space and most of the regular cells are mapped outside the region considered which may result in instability. For $n_l = 3$ and 5, spurious limit cycles are also induced. The generalization capability of the models was tested on the data set shown in Fig 6. The input is a Gaussian signal superimposed on a square wave. Notice that the range of the input is larger than the input which was used in the model identification. A Gaussian noise was added to the output and the signal to noise ratio was 20 DB. The covariance of the output noise was $1.6504e-2$. The prediction error and the model prediction error were computed as follows

$$\begin{aligned} y_{op}(k) &= F(y(k-1), \dots, y(k-n_y), u(k-1), \dots, u(k-n_u), \xi_{op}(k-1), \dots, \xi_{op}(k-n_\xi)) \\ \xi_{op}(k) &= y(k) - y_{op}(k) \end{aligned} \quad (40)$$

$$\begin{aligned} y_{mp}(k) &= F(y_{mp}(k-1), \dots, y_{mp}(k-n_y), u(k-1), \dots, u(k-n_u)) \\ \xi_{mp}(k) &= y(k) - y_{mp}(k) \end{aligned} \quad (41)$$

The mean square errors for the models are also listed in Table 1. The model prediction error is a much more strict measure than the prediction error.

When the modified selection algorithm was applied to the estimation set, the following NARMAX model was identified for $n_y = n_u = 1$, $n_l = 2, \dots, 5$ and $n_\xi = 5$.

$$\begin{aligned} y(k) &= 0.800375y(k-1) + 0.105117u^2(k-1) + 0.065003u(k-1) - 0.35377\xi(k-1) \\ &\quad - 0.127177\xi(k-3) - 0.065083\xi(k-2) - 0.052298\xi(k-4) + \xi(k) \end{aligned} \quad (42)$$

Note that the same model was identified for different degrees of nonlinearity n_l . The selection algorithm works in favour of the simpler terms. The cell mapping of the model is shown in Fig 7. The mean square prediction error and the mean square model prediction error on the validation set are $3.0077e-2$ and $4.3194e-2$ respectively. It can be seen that the resulting model is valid in a slightly larger region of the cell state space and achieved better generalization on the validation set. The results are still far from adequate however and suggest that the constraint $n_y = n_u = 1$ which was deliberately imposed has caused the system to be undermodelled.

Increasing the size of the model set by defining $n_y = n_u = 2$, $n_l = 2, \dots, 5$ and $n_\xi = 5$, the terms identified using the OLS selection algorithm are listed in Table 2.

A cell state space was introduced by dividing the region $y(k) \in [0 \ 25]$, $y(k-1) \in [0 \ 25]$ and $u \in [-5 \ 5]$ into $125 \times 125 \times 100$ uniform intervals. The cell mapping diagrams of the models are presented in Fig 8. The resulting model for $n_l = 2$ is valid in a large region of the cell state space considered. The resulting models for $n_l = 3, 4$ and 5 are only valid in a much smaller region of the cell state space and spurious limit cycles are also induced. Most of the regular cells are mapped outside the region considered. The mean square errors of the models when tested on the validation set given in Fig 6 are also listed in Table 2. Note

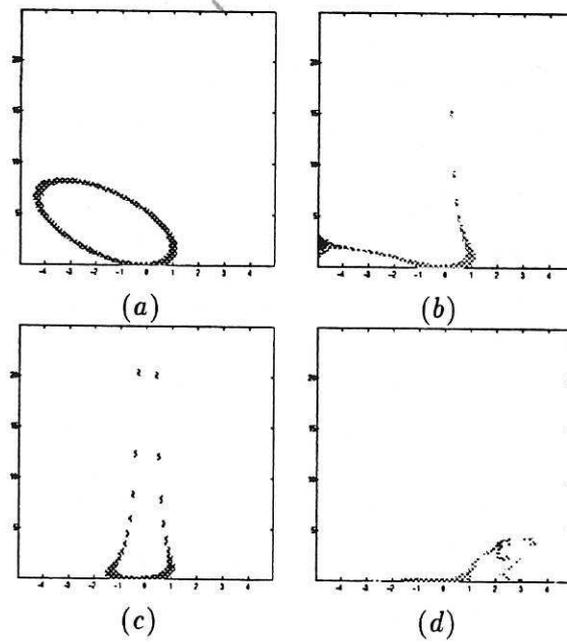


Figure 4: Cell mapping analysis of the models selected using the error reduction ratio. a). P-1 cells, $n_y=n_u=1$, $n_l=2$, $n_e=5$. b). P-1 and P-2 cells, $n_y=n_u=1$, $n_l=3$, $n_e=5$. c). P-1 cells $n_y=n_u=1$, $n_l=4$, $n_e=5$. d). P-1, P-2, P-3 and P-6 cells $n_y=n_u=1$, $n_l=5$, $n_e=5$

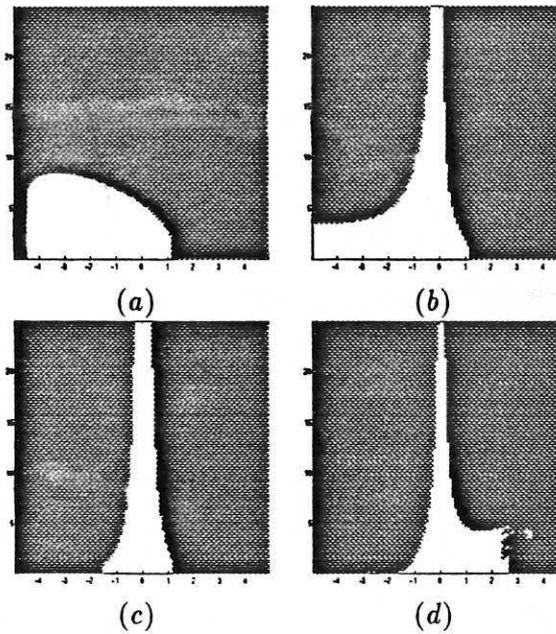


Figure 5: Cell mapping analysis of the models selected using the error reduction ratio. a). DOA-Sink cells, $n_y=n_u=1$, $n_l=2$, $n_e=5$. b). DOA-Sink cells, $n_y=n_u=1$, $n_l=3$, $n_e=5$. c). DOA-Sink cells, $n_y=n_u=1$, $n_l=4$, $n_e=5$. d). DOA-Sink cells, $n_y=n_u=1$, $n_l=5$, $n_e=5$

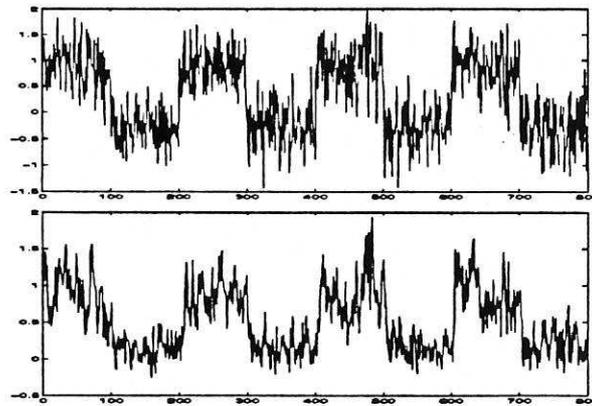


Figure 6: Input (above) and output signals of the system given in equation (39) sampled at 0.2 second intervals. A Gaussian random signal with a standard deviation of 0.4 was added to the square wave with a magnitude of 0.6. A Gaussian noise was added to the output. The signal to noise ratio was 20 DB

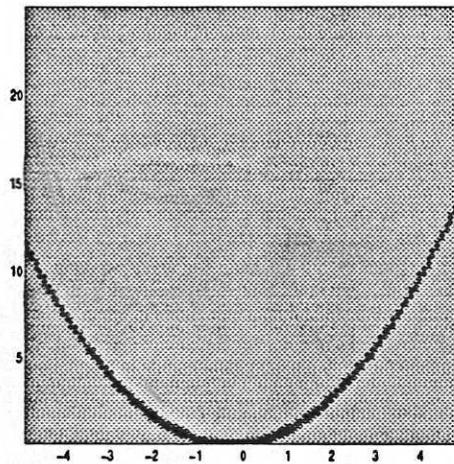


Figure 7: Cell mapping analysis of the models identified using the modified OLS algorithm. P-1 cells (dark) and DOA-1 cells (grey) $n_y=n_u=1, n_l=2, \dots, 5, n_e=5$

Table 2: Selected terms using the OLS algorithm for $n_y = n_u = 2, n_l = 2, \dots, 5$ and $n_\xi = 5$

$n_l = 2$ $\sum \frac{\xi_{op}^2(k)}{N} = 1.7407e-2$ $\sum \frac{\xi_{mp}^2(k)}{N} = 1.7365e-2$	$n_l = 3$ $\sum \frac{\xi_{op}^2(k)}{N} = 2.7422e-2$ $\sum \frac{\xi_{mp}^2(k)}{N} = 3.2781e-2$	$n_l = 4$ $\sum \frac{\xi_{op}^2(k)}{N} = 2.2844e-2$ $\sum \frac{\xi_{mp}^2(k)}{N} = 2.5068e-2$	$n_l = 5$ $\sum \frac{\xi_{op}^2(k)}{N} = 2.8602e-2$ $\sum \frac{\xi_{mp}^2(k)}{N} = 5.1434e-2$
$y(k-1)$	$y(k-1)$	$y(k-1)$	$y(k-1)$
$u^2(k-1)$	$u^2(k-1)$	$u^2(k-1)$	$u^2(k-1)$
$u(k-2)$	$u(k-2)$	$u(k-2)$	$u(k-2)$
$u(k-2)u(k-1)$	$u^2(k-2)u(k-1)$	$u^2(k-2)u(k-1)$	$u^2(k-2)u(k-1)$
$y(k-2)$	$y(k-2)$	$y(k-2)$	$y(k-2)$
$u^2(k-2)$	$u^2(k-2)y(k-1)$	$u^2(k-2)y(k-1)$	$u^2(k-2)y(k-1)$
$u(k-2)y(k-1)$	$u(k-2)u(k-1)$	$u^3(k-2)y(k-2)$	$u^3(k-2)y(k-2)$
$u(k-2)y(k-2)$	$u(k-2)y(k-2)$	$u(k-2)u(k-1)$	$u(k-2)u(k-1)$
$\xi(k-1)$	$u(k-2)u^2(k-1)$	$u(k-2)y^3(k-1)$	$y^5(k-1)$
$\xi(k-2)$	$u(k-1)$	$u(k-2)y^3(k-2)$	$y^3(k-2)y^2(k-1)$
	$u(k-2)y^2(k-2)$	$u(k-2)y(k-2)$	constant
	$u(k-2)y(k-2)y(k-1)$	$u(k-2)y^2(k-1)$	$y^2(k-1)$
	$u(k-2)y(k-1)$	$u(k-2)y(k-1)$	$u(k-2)y(k-1)$
	$\xi(k-1)$	$u(k-2)u^3(k-1)$	$u(k-2)y^4(k-2)$
	$\xi(k-2)$	$u(k-1)$	$y(k-2)y^2(k-1)$
		$u^2(k-2)$	$u(k-2)u^3(k-1)$
		$u^2(k-2)y^2(k-1)$	$u(k-1)$
		$u(k-2)y(k-2)y(k-1)$	$\xi(k-1)$
		$\xi(k-1)$	$\xi(k-2)$
		$\xi(k-2)$	

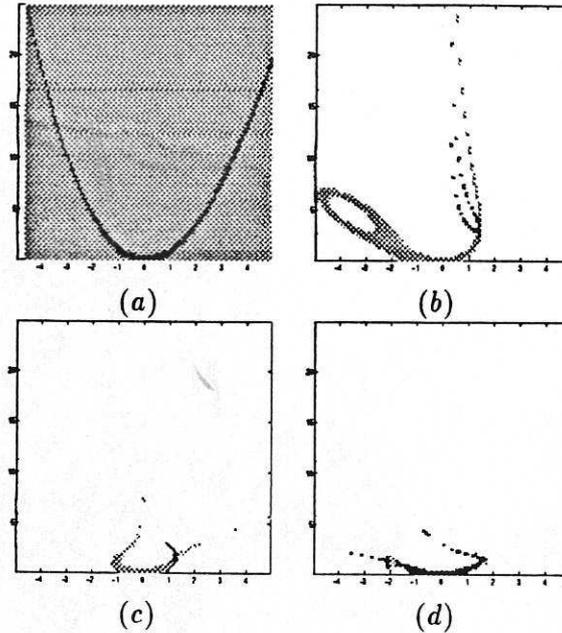


Figure 8: Cell mapping analysis of the models selected using the error reduction ratio. a). P-1 cells and DOA-1 cells, $n_y=n_u=2$, $n_l=2$, $n_e=5$. b). P-1 and P-2 cells, $n_y=n_u=2$, $n_l=3$, $n_e=5$. c). P-1 and P-2 cells, $n_y=n_u=2$, $n_l=4$, $n_e=5$. d). P-1, P-2, P-3, P-4, P-5, P-6 and P-9 cells $n_y=n_u=2$, $n_l=5$, $n_e=5$

that the model for $n_l = 2$ achieved much better predictions on the validation data set than the other three models. However, the model is not valid in the whole region of interest.

When the modified OLS algorithm was applied to the estimation data set, the following models were identified.

$$\begin{aligned}
 y(k) = & 0.832588y(k-1) + 0.131462u^2(k-1) + 0.006892u(k-2) \\
 & + 0.106844u(k-2)u(k-1) - 0.11554y(k-2) + 0.030943u^2(k-2) \\
 & - 0.830768\xi(k-1) + 0.089671\xi(k-2) + \xi(k) \quad n_l = 2
 \end{aligned} \tag{43}$$

$$\begin{aligned}
 y(k) = & 0.823393y(k-1) + 0.130232u^2(k-1) + 0.039562u(k-2) \\
 & + .021112u^2(k-2)u(k-1) - .102947y(k-2) + .027454u^2(k-2)y(k-1) \\
 & + 0.086467u(k-2)u(k-1) - 0.04051u(k-2)y(k-2) - 0.838699\xi(k-1) \\
 & + 0.104357\xi(k-2) + \xi(k) \quad n_l = 3, 4, 5
 \end{aligned} \tag{44}$$

The cell mapping diagrams of the models are shown in **Fig 9**. The model for $n_l = 2$ is valid in the whole region considered and the models for higher nonlinearity are only valid in a smaller region. Note that improved results were achieved when the modified algorithm was used to identify the models for $n_l = 3, 4$ and 5 . The same model was identified for higher nonlinearity and no spurious limit cycles were induced. The mean square errors on the validation set for the model with $n_l = 2$ are $1.6931e-2$ and $1.6911e-2$ respectively.

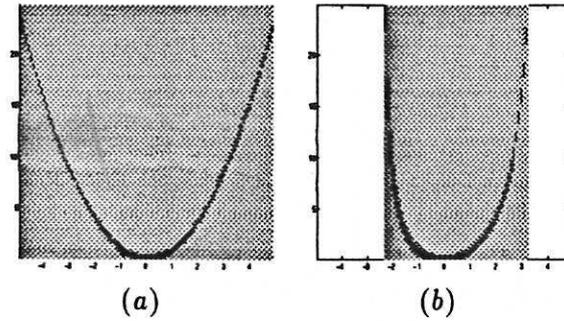


Figure 9: Cell mapping analysis of the models identified using the modified OLS algorithm. a). P-1 cells (dark) and DOA-1 cells (grey), $n_y=n_u=2$, $n_l=2$, $n_e=5$. b). P-1 cells (dark) and DOA-1 cells (grey), $n_y=n_u=2$, $n_l=3, 4, 5$, $n_e=5$

The mean square errors for the model with higher nonlinearity are $2.0805e-2$ and $2.0971e-2$ respectively. The model with $n_l = 2$ achieved the best prediction performance on the validation set compared with the other models identified so far. The values of the mean square errors are very close to the covariance of the output noise ($1.6504e-2$).

Increasing the size of the model set further by selecting $n_y = n_u = 3$, $n_l = 2, \dots, 5$ and $n_e = 5$, the terms identified using the OLS algorithm and the mean square errors are given in Table 3. A cell state space was introduced by dividing the region $y(k) \in [0 \ 25]$, $y(k-1) \in [0 \ 25]$, $y(k-2) \in [0 \ 25]$ and $u \in [-5 \ 5]$ into $40 \times 40 \times 40 \times 40$ uniform intervals. The cell mapping diagrams of the models are presented in Fig 10. Note that spurious limit cycles are induced in all the models. When the modified OLS algorithm was applied to the estimation set, the following models were identified.

$$\begin{aligned}
 y(k) = & .000595y(k-1) + .134213u^2(k-1) + .009833u(k-2) \\
 & +.104572u(k-2)u(k-1) + .015736u(k-3)y(k-2) + .141978u^2(k-2) \\
 & +.578331y(k-2) + .084522u(k-3)u(k-2) + .02636u^2(k-3) - .117612 \\
 & \times y(k-3) + .004987 - .596641\xi(k-2) + .09078\xi(k-3) + \xi(k) \quad n_l = 2 \quad (45)
 \end{aligned}$$

$$\begin{aligned}
 y(k) = & -0.000307y(k-1) + 0.132861u^2(k-1) + 0.023262u(k-2) \\
 & +.019459u^2(k-2)u(k-1) + .012893u(k-3)y(k-2) + .575979y(k-2) \\
 & +0.009543u(k-3)u^2(k-2) + 0.028084u^2(k-3) + 0.12788u^2(k-2) \\
 & 0.085708u(k-2)u(k-1) - 0.113304y(k-3) + 0.0754u(k-3)u(k-2) \\
 & +0.00516 - 0.590164\xi(k-2) + 0.083561\xi(k-3) + \xi(k) \quad n_l = 3 \quad (46)
 \end{aligned}$$

$$\begin{aligned}
 y(k) = & 0.003354y(k-1) + 0.133032u^2(k-1) + 0.023856u(k-2) \\
 & +.018813u^2(k-2)u(k-1) + .054936u(k-3)y(k-2) + .490672y(k-2) \\
 & +0.009271u(k-3)u^2(k-2) + 0.039565u^2(k-3) + 0.12849u^2(k-2) \\
 & 0.08512u(k-2)u(k-1) - 0.043131y(k-3) + 0.071711u(k-3)u(k-2) \\
 & -0.027585u^3(k-3)y(k-3) - 0.525433\xi(k-2) + \xi(k) \quad n_l = 4, 5 \quad (47)
 \end{aligned}$$

The cell diagrams of the models are shown in Fig 11. It may be appreciated from

Table 3: Selected terms using the OLS algorithm for $n_y = n_u = 3$, $n_l = 2, \dots, 5$ and $n_\xi = 5$

$n_l = 2$ $\frac{\sum \xi_{op}^2(k)}{N} = 1.7154e-2$ $\frac{\sum \xi_{mp}^2(k)}{N} = 1.7124e-2$	$n_l = 3$ $\frac{\sum \xi_{op}^2(k)}{N} = 2.0559e-2$ $\frac{\sum \xi_{mp}^2(k)}{N} = 2.0717e-2$	$n_l = 4$ $\frac{\sum \xi_{op}^2(k)}{N} = 1.9318e-2$ $\frac{\sum \xi_{mp}^2(k)}{N} = 1.8836e-2$	$n_l = 5$ $\frac{\sum \xi_{op}^2(k)}{N} = 2.2656e-2$ $\frac{\sum \xi_{mp}^2(k)}{N} = 2.0707e-2$
$y(k-1)$	$y(k-1)$	$y(k-1)$	$y(k-1)$
$u^2(k-1)$	$u^2(k-1)$	$u^2(k-1)$	$u^2(k-1)$
$u(k-2)$	$u(k-2)$	$u(k-2)$	$u(k-2)$
$u(k-2)u(k-1)$	$u^2(k-2)u(k-1)$	$u^2(k-2)u(k-1)$	$u^2(k-2)u(k-1)$
$u(k-3)y(k-2)$	$u(k-3)y(k-2)$	$u(k-3)y(k-2)$	$u(k-3)y(k-2)$
$u^2(k-2)$	$y(k-2)$	$y(k-2)$	$y(k-2)$
$y(k-2)$	$u(k-3)u^2(k-2)$	$u(k-3)u^2(k-2)$	$u(k-3)u^2(k-2)$
$u(k-3)u(k-2)$	$u^2(k-3)$	$u^2(k-3)$	$u^2(k-3)$
$u^2(k-3)$	$u^2(k-2)$	$u^2(k-2)$	$u^2(k-2)$
$y(k-3)$	$u(k-2)u(k-1)$	$u(k-2)u(k-1)$	$u(k-2)u(k-1)$
$u(k-2)y(k-2)$	$y(k-3)$	$y(k-3)$	$y(k-3)$
$u(k-2)y(k-3)$	$u(k-3)u(k-2)$	$u^3(k-3)y(k-3)$	$u^4(k-3)y(k-3)$
$\xi(k-1)$	$u^2(k-3)y(k-3)$	$u(k-3)u(k-2)$	$u(k-3)u(k-2)$
$\xi(k-2)$	$u(k-2)y(k-2)$	$u^2(k-3)y^2(k-2)$	$u^2(k-3)y^2(k-2)$
	$u(k-2)y(k-3)$	$y^4(k-1)$	$y(k-2)y^4(k-1)$
	$\times y(k-2)$		
	$u(k-3)u(k-1)$	$u(k-3)u(k-1)y(k-3)$	$u^2(k-2)y(k-3)$
	$\times y(k-3)$		$\times y^2(k-1)$
	$u^2(k-1)y(k-2)$	$u^2(k-1)y(k-2)$	$u^2(k-2)y(k-2)$
	$u(k-2)u^2(k-1)$	$u^2(k-1)y(k-2)y(k-1)$	$u(k-3)y^2(k-1)$
	$u(k-1)y^2(k-2)$	$u^2(k-2)y(k-3)y(k-1)$	$u^2(k-2)y(k-3)$
			$\times y(k-1)$
	$u(k-3)y^2(k-3)$	$u^2(k-2)y(k-2)$	$u(k-2)y(k-1)$
	$u^2(k-2)y(k-3)$	$u(k-3)y^2(k-1)$	$u(k-2)y^2(k-1)$
	$\xi(k-2)$	$u(k-1)y(k-3)y(k-2)$	$u^3(k-2)u^2(k-1)$
		$u(k-1)y(k-3)y(k-2)$	$u(k-1)$
		$\times y(k-1)$	
		$u(k-2)u^2(k-1)$	$\xi(k-2)$
		$\xi(k-2)$	$\xi(k-1)$

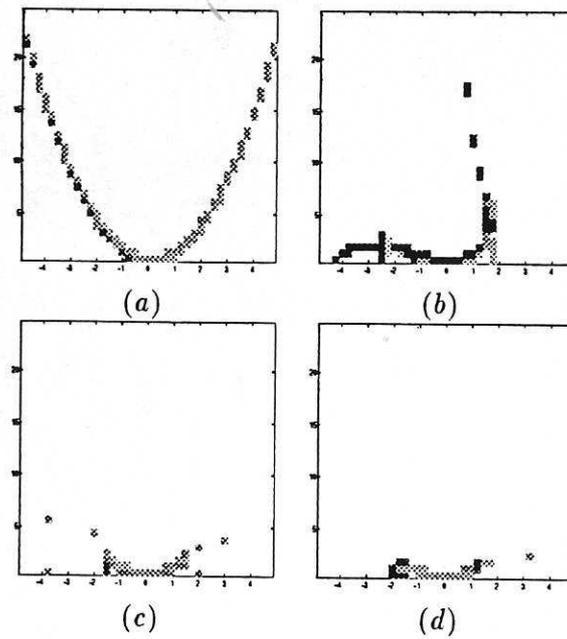


Figure 10: Cell mapping analysis of the models selected using the error reduction ratio. a). P-1 and P-2 cells, $n_y=n_u=3$, $n_l=2$, $n_e=5$. b). P-1, P-2, P-3 and P-10 cells, $n_y=n_u=3$, $n_l=3$, $n_e=5$. c). P-1, P-2, P-3 and P-5 cells, $n_y=n_u=3$, $n_l=4$, $n_e=5$. d). P-1 and P-2 cells, $n_y=n_u=3$, $n_l=5$, $n_e=5$

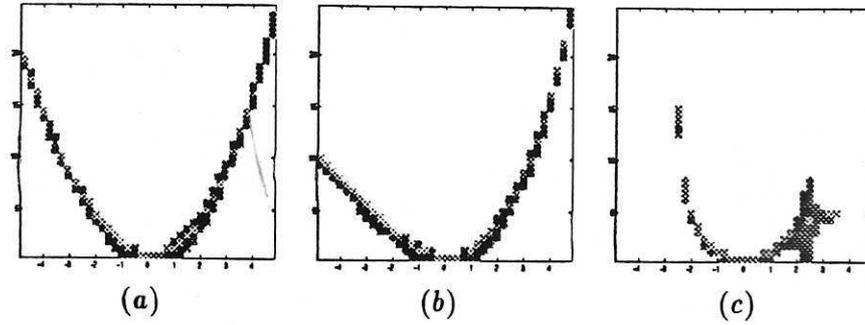


Figure 11: Cell mapping analysis of the models identified using the modified OLS algorithm. a). P-1 and P-2 cells, $n_y=n_u=3$, $n_l=2$, $n_e=5$. b). P-1 and P-2 cells, $n_y=n_u=3$, $n_l=3$, $n_e=5$. c). P-1, P-2 and P-3 cells, $n_y=n_u=3$, $n_l=4$, $n_e=5$.

Table 4: MSEs on validation set for the models identified using the modified OLS algorithm ($n_y = n_u = 3$)

	$n_l = 2$	$n_l = 3$	$n_l = 4, 5$
$\frac{\sum \xi_{op}^2(k)}{N}$	1.7057e-2	1.7890e-2	1.8549e-2
$\frac{\sum \xi_{mp}^2(k)}{N}$	1.7025e-2	1.7780e-2	1.8040e-2

the diagrams that slightly better results were obtained when compared with the models identified using the OLS algorithm for larger value of n_l . However, spurious limit cycles still exist in all the models and the models are only valid in a small region of the cell state space. The deficiency in the models is probably caused by inadequate excitation of the system dynamics coupled with incorrect selection of the time lags. It may be seen that the model complexity increases with increased time lags for both selection algorithms. It is therefore necessary to select the time lags appropriately. The input selection algorithms presented in (Zheng and Billings, 1996) may be used for this purpose. This issue will be studied in further publications. In the following examples, the range of the input signal is increased to improve the estimation results. Good prediction performance was achieved for the models with $n_y = n_u = 3$, $n_l = 2$. If the MSE is used as a measure of goodness of fit, the models identified from the settings $n_y = n_u = 3$, $n_l = 2$ would be considered as valid models although these models contain many spurious limit cycles. This indicates that the MSEs obtained on a single data set may be not sufficient to validate an identified model.

In the previous examples, the system dynamics were excited in a small region. When incorrect model sets were used, spurious dynamics were introduced as limit cycles. In the following, the signal shown in Fig 12 was used to excite the system dynamics in a wider range. The input is a Gaussian sequence superimposed on a square wave. The magnitude

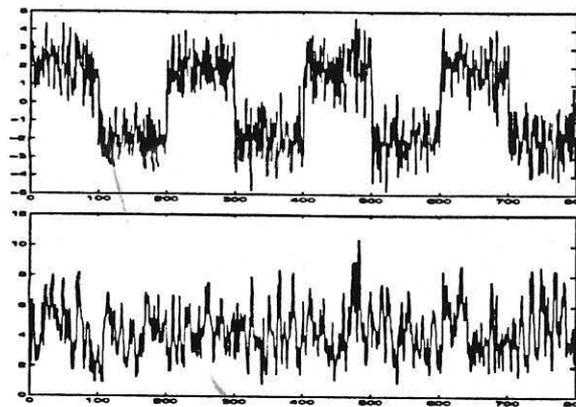


Figure 12: Input (above) and output signals of the system given in equation (39) sampled at 0.2 second intervals. A Gaussian random signal with a standard deviation of 1.0 was added to the square wave with a magnitude of 2.0. A Gaussian noise was added to the output. The signal to noise ratio was 20 DB

of the square wave was 2.0 and the Gaussian sequence had a zero mean and a standard deviation of 1.0. The output signal was obtained by integrating the differential equations (39) and sampling at 0.2 second intervals. A sequence of 800 samples were then generated for validation by applying an input sequence uniformly distributed in the interval $[-4 \ 4]$. The validation set is plotted in **Fig 13**. A Gaussian noise sequence was added to the output and the resulting signal to noise ratio was 20 DB. The variance of output noise was $1.824e-1$. For $n_y = n_u = 2$, $n_l = 2, \dots, 5$ and $n_\ell = 5$, the terms selected using the **OLS** algorithm and the mean square errors are listed in **Table 5**.

A cell state space was introduced by dividing the region $y(k) \in [0 \ 25]$, $y(k-1) \in [0 \ 25]$ and $u \in [-5 \ 5]$ into $125 \times 125 \times 100$ uniform intervals. The cell mapping diagrams of the models are presented in **Fig 14**. The results are much better when compared with those shown in **Fig 8** and valid models are obtained for the settings $n_l = 2$ and $n_l = 3$. Note that there are no spurious limit cycles, although the models for $n_l = 4, 5$ are not valid in the whole region considered.

For $n_y = n_u = 3$, $n_l = 2, \dots, 5$ and $n_\ell = 5$, the terms given in **Table 6** were selected using the **OLS** algorithm. The region $y(k) \in [0 \ 25]$, $y(k-1) \in [0 \ 25]$, $y(k-2) \in [0 \ 25]$ and $u \in [-5 \ 5]$ were divided into $50 \times 50 \times 50 \times 50$ uniform intervals. The cell mapping diagrams of the models are presented in **Fig 15**. The results are significantly improved as compared with those shown in **Fig 10**. The same model was identified for $n_l = 2, 3$ and 4 and it is valid in the whole region of interest. However, the algorithm failed to identify a valid model in the whole region of interest for $n_l=5$ and spurious dynamics were induced due to the inability to recognize the complexity of the terms.

When the modified **OLS** algorithm was applied to the data samples, the following two models were identified for $n_y = n_u = 2$ and $n_y = n_u = 3$ respectively. For each setting of the time lags, the same model was identified for different nonlinearities. The cell diagrams of the models are shown in **Fig 16**. It may be seen that the models are valid in the whole

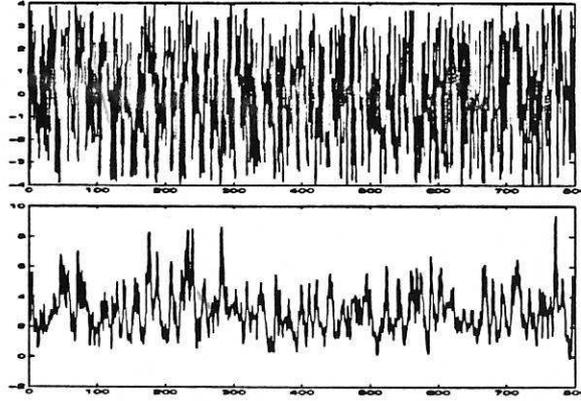


Figure 13: Input (above) and output signals of the system given in equation (39) sampled at 0.2 second intervals. The input was uniformly distributed in the region (-4.0 4.0). A Gaussian noise was added to the output. The signal to noise ratio was 20 DB

Table 5: Selected terms using the OLS algorithm for $n_y = n_u = 2$, $n_l = 2, \dots, 5$ and $n_\xi = 5$

$n_l = 2$ $\frac{\sum \xi_{op}^2(k)}{N} = 2.0824e-1$ $\frac{\sum \xi_{mp}^2(k)}{N} = 2.0725e-1$	$n_l = 3$ $\frac{\sum \xi_{op}^2(k)}{N} = 2.0948e-1$ $\frac{\sum \xi_{mp}^2(k)}{N} = 2.0901e-1$	$n_l = 4, 5$ $\frac{\sum \xi_{op}^2(k)}{N} = 2.1775e-1$ $\frac{\sum \xi_{mp}^2(k)}{N} = 2.0935e-1$
$y(k-1)$ $u(k-2)u(k-1)$ $u^2(k-1)$ $u^2(k-2)$ $y(k-2)$ $\xi(k-1)$	$y(k-1)$ $u(k-2)u(k-1)$ $u^2(k-1)$ $u^2(k-2)$ $y(k-2)$ $u^2(k-2)y(k-2)$ $\xi(k-1)$	$y(k-1)$ $u(k-2)u(k-1)$ $u^2(k-1)$ $u^2(k-2)$ $y(k-2)$ $u^2(k-2)u^2(k-1)$ $u(k-2)u(k-1)y^2(k-1)$ $\xi(k-1)$

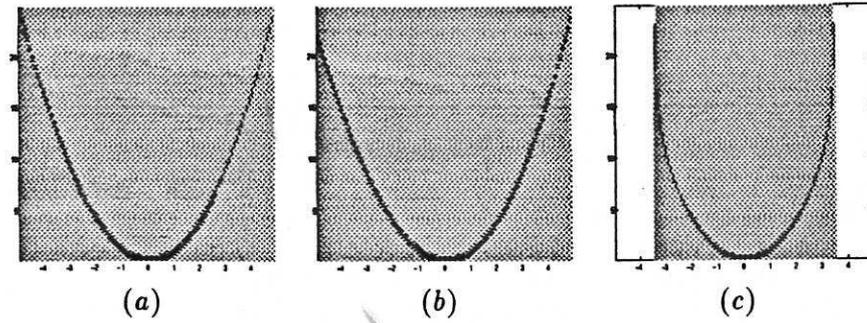


Figure 14: Cell mapping analysis of the models selected using the error reduction ratio. a). P-1 cells (dark) and DOA-1 cells (grey), $n_y=n_u=2$, $n_l=2$, $n_e=5$. b). P-1 cells (dark) and DOA-1 cells (grey), $n_y=n_u=2$, $n_l=3$, $n_e=5$. c). P-1 cells (dark) and DOA-1 cells (grey), $n_y=n_u=2$, $n_l=4, 5$, $n_e=5$.

Table 6: Selected terms using the OLS algorithm for $n_y = n_u = 3$, $n_l = 2, \dots, 5$ and $n_e = 5$

$n_l = 2, 3, 4$ $\frac{\sum \xi_{op}^2(k)}{N} = 2.1775e-1$ $\frac{\sum \xi_{mp}^2(k)}{N} = 2.0935e-1$	$n_l = 5$ $\frac{\sum \xi_{op}^2(k)}{N} = 3.6356e-1$ $\frac{\sum \xi_{mp}^2(k)}{N} = 3.4137e-1$
$y(k-1)$	$y(k-1)$
$u(k-2)u(k-1)$	$u(k-2)u(k-1)$
$u^2(k-1)$	$u^2(k-1)$
$u^2(k-2)$	$u^2(k-2)$
$y(k-2)$	$y(k-2)$
$u(k-3)u(k-2)$	$u(k-3)u(k-2)$
$u^2(k-3)$	$u^2(k-3)$
$y(k-3)$	$y(k-3)$
$u(k-3)u(k-1)$	$u(k-3)u(k-1)$
$\xi(k-2)$	$u^4(k-2)y(k-3)$
$\xi(k-1)$	$u^2(k-2)u^2(k-1)$
	$u(k-3)y^2(k-3)y^2(k-2)$
	$\xi(k-2)$
	$\xi(k-1)$
	$\xi(k-5)$

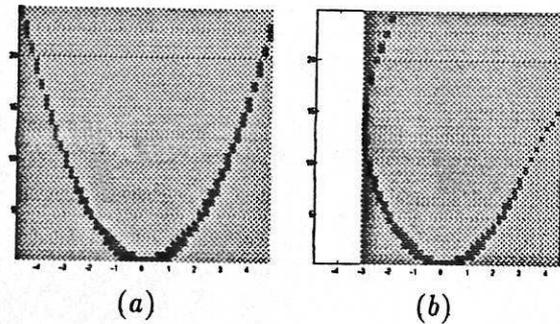


Figure 15: Cell mapping analysis of the models selected using the error reduction ratio. a). P-1 cells (dark) and DOA-1 cells (grey), $n_y=n_u=2$, $n_l=2, 3, 4$, $n_e=5$. b). P-1, P-2, P-6 cells and DOA-1, DOA-2, DOA-6 cells, $n_y=n_u=2$, $n_l=5$, $n_e=5$.

region of interest. The mean square prediction error and the mean square model prediction error for $n_y = n_u = 2$ are $2.0824e-1$ and $2.0725e-1$, and those for $n_y = n_u = 3$ are $2.1775e-1$ and $2.0935e-1$ respectively.

$$y(k) = 0.738605y(k-1) + 0.127458u(k-2)u(k-1) + 0.124555u^2(k-1) + 0.047947u^2(k-2) - 0.037582y(k-2) - 0.725945\xi(k-1) + \xi(k) \quad (48)$$

$$y(k) = 0.347498y(k-1) + 0.111462u(k-2)u(k-1) + 0.120715u^2(k-1) + 0.105012u^2(k-2) + 0.279564y(k-2) + 0.04671u(k-3)u(k-2) + 0.002749u^2(k-3) - 0.041016y(k-3) + 0.030807u(k-3)u(k-1) - 0.30343\xi(k-2) - 0.361127\xi(k-1) + \xi(k) \quad (49)$$

Note that all the valid models identified in the simulations achieved good predictions on the validation set. However good prediction performance does not necessarily mean the model is valid. When the model is valid in a smaller region, the model prediction could be as good as a valid model or comparable with those of the valid model.

6 Conclusions

In this paper, cell mapping analysis was used for qualitatively validating an identified NARMAX model. The cell mapping method can be applied to the global analysis of a diverse range of identified nonlinear systems such as the NARMAX model. The most important characteristics of the original nonlinear system and the identified model, including the equilibrium points, fixed points, limit cycles and the domains of attraction are revealed graphically in a cell state space. The method can pin point any deficiency in the identified model by comparing the resulting cell mapping diagrams with those of the original system. In practice, however the true system will be unknown *a priori*, the main purpose is therefore to investigate which aspects of identification can affect the resulting model and how.

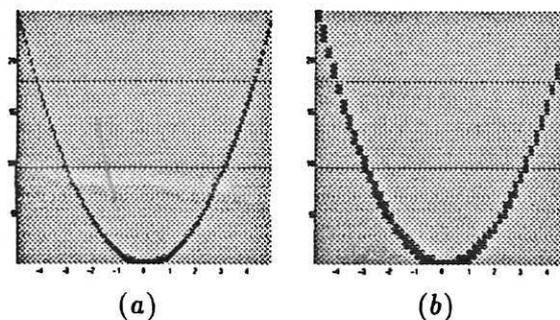


Figure 16: Cell mapping analysis of the models identified using the modified OLS algorithm. a). P-1 cells (dark) and DOA-1 cells (grey), $n_y = n_u = 2$, $n_l = 2, \dots, 5$, $n_e = 5$. b). P-1 cells (dark) and DOA-1 cells (grey), $n_y = n_u = 3$, $n_l = 2, \dots, 5$, $n_e = 5$

Based on the qualitative validation of models identified using the error reduction ratio, a modified selection criterion was proposed. The new method incorporates both the approximation capability and the complexity of the term (*i.e.* the degree of nonlinearity) into the selection criterion. The new criterion may be considered as a modified prediction risk on future data sets. The criterion is therefore closely related to the generalization capability of the terms.

A dynamically valid model enables high prediction accuracy over the estimation and validation data, but high prediction accuracy does not necessarily imply a dynamically valid model. The prediction accuracy on a particular validation data set, which is often used to select models, can therefore be misleading.

Acknowledgements

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

References

- Akaike, H. (1970). Statistical predictor identification. *Ann. Inst. Statist. Math.*, 22:203-217.
- Barron, A. R. (1984). Predicted squared error: A criterion for automatic model selection. In Stanley, J. F., editor, *Self-organizing Methods in Modelling GMDH Type Algorithms*, pages 87-103. Marcel Dekker, Inc. New York.
- Billings, S. A. and Chen, S. (1989). Extended model set, global data and threshold model identification of severely non-linear systems. *Int. J. Control*, 50(5):1897-1923.
- Billings, S. A., Chen, S., and Korenberg, M. J. (1989). Identification of mimo non-linear systems using a forward-regression orthogonal estimator. *Int. J. Control*, 49:2157-2189.
- 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 Science*, 19:1559-1568.
- Billings, S. A. and Tao, Q. H. (1991). Model validity tests for nonlinear signal processing applications. *Int. J. Control*, 54(1):157 - 194.
- Billings, S. A. and Zhu, Q. M. (1994). Nonlinear model validation using correlation tests. *Int. J. Control*, 60:1107-1120.
- Billings, S. A. and Zhu, Q. M. (1995). Model validity tests for multivariable nonlinear models including neural networks. *Int. J. Control*, 62:749-766.
- Bohlin, T. (1971). On the problem of ambiguities in maximum likelihood identification. *Automatica*, 7:199 - 210.
- 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.
- Hsu, C. S. (1980). A theory of cell-to-cell mapping for nonlinear dynamical systems. *J. Applied Mechanics*, 47:931 - 939.
- 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.
- Leontaritis, I. J. and Billings, S. A. (1987). Model selection and validation methods for nonlinear systems. *Int. J. Control*, 45(1):311 - 341.
- Liu, Y. (1995). Unbiased estimate of generalization error and model selection in neural network. *Neural Networks*, 8(2):215-219.
- Murata, N. and Yoshizawa, S. (1994). Network information criterion - determining the number of hidden units for an artificial neural network model. *IEEE. Trans. on Neural Networks*, 5(6):865-871.

Zheng, G. L. and Billings, S. A. (1996). Radial basis function network configuration using mutual information and the orthogonal least squares algorithm. *To appear in Neural Networks.*

List of Figures

- 1 left: A core of period four cells correspond to the stable focus at (0, 0) and a singular cell corresponding to the saddle point at (1, -0.9). right: Regular cells in the domain of attraction of the period four cells, the regular cells in the blank region are mapped into the sink cell 6
- 2 Cell mapping analysis of the differential equations (39) shows the P-1 cells (dark) and the DOA-1 cells (grey). vertical axis - y, horizontal axis - u . . . 11
- 3 Input (above) and output signals of the system given in equation (39) sampled at 0.2 second intervals. The input signal is uniformly distributed in the region (-0.5 1.5). A Gaussian noise was added to the output. The resulting signal to noise ratio was 20 DB 12
- 4 Cell mapping analysis of the models selected using the error reduction ratio. a). P-1 cells, $n_y=n_u=1, n_l=2, n_e=5$. b). P-1 and P-2 cells, $n_y=n_u=1, n_l=3, n_e=5$. c). P-1 cells $n_y=n_u=1, n_l=4, n_e=5$. d). P-1, P-2, P-3 and P-6 cells $n_y=n_u=1, n_l=5, n_e=5$ 14
- 5 Cell mapping analysis of the models selected using the error reduction ratio. a). DOA-Sink cells, $n_y=n_u=1, n_l=2, n_e=5$. b). DOA-Sink cells, $n_y=n_u=1, n_l=3, n_e=5$. c). DOA-Sink cells, $n_y=n_u=1, n_l=4, n_e=5$. d). DOA-Sink cells, $n_y=n_u=1, n_l=5, n_e=5$ 15
- 6 Input (above) and output signals of the system given in equation (39) sampled at 0.2 second intervals. A Gaussian random signal with a standard deviation of 0.4 was added to the square wave with a magnitude of 0.6. A Gaussian noise was added to the output. The signal to noise ratio was 20 DB 15
- 7 Cell mapping analysis of the models identified using the modified OLS algorithm. P-1 cells (dark) and DOA-1 cells (grey) $n_y=n_u=1, n_l=2, \dots, 5, n_e=5$ 16
- 8 Cell mapping analysis of the models selected using the error reduction ratio. a). P-1 cells and DOA-1 cells, $n_y=n_u=2, n_l=2, n_e=5$. b). P-1 and P-2 cells, $n_y=n_u=2, n_l=3, n_e=5$. c). P-1 and P-2 cells, $n_y=n_u=2, n_l=4, n_e=5$. d). P-1, P-2, P-3, P-4, P-5, P-6 and P-9 cells $n_y=n_u=2, n_l=5, n_e=5$ 17
- 9 Cell mapping analysis of the models identified using the modified OLS algorithm. a). P-1 cells (dark) and DOA-1 cells (grey), $n_y=n_u=2, n_l=2, n_e=5$. b). P-1 cells (dark) and DOA-1 cells (grey), $n_y=n_u=2, n_l=3, 4, 5, n_e=5$. . 18
- 10 Cell mapping analysis of the models selected using the error reduction ratio. a). P-1 and P-2 cells, $n_y=n_u=3, n_l=2, n_e=5$. b). P-1, P-2, P-3 and P-10 cells, $n_y=n_u=3, n_l=3, n_e=5$. c). P-1, P-2, P-3 and P-5 cells, $n_y=n_u=3, n_l=4, n_e=5$. d). P-1 and P-2 cells, $n_y=n_u=3, n_l=5, n_e=5$ 20
- 11 Cell mapping analysis of the models identified using the modified OLS algorithm. a). P-1 and P-2 cells, $n_y=n_u=3, n_l=2, n_e=5$. b). P-1 and P-2 cells, $n_y=n_u=3, n_l=3, n_e=5$. c). P-1, P-2 and P-3 cells, $n_y=n_u=3, n_l=4, n_e=5$. . 21
- 12 Input (above) and output signals of the system given in equation (39) sampled at 0.2 second intervals. A Gaussian random signal with a standard deviation of 1.0 was added to the square wave with a magnitude of 2.0. A Gaussian noise was added to the output. The signal to noise ratio was 20 DB 22

- 13 Input (above) and output signals of the system given in equation (39) sampled at 0.2 second intervals. The input was uniformly distributed in the region (-4.0 4.0). A Gaussian noise was added to the output. The signal to noise ratio was 20 DB 23
- 14 Cell mapping analysis of the models selected using the error reduction ratio. a). P-1 cells (dark) and DOA-1 cells (grey), $n_y=n_u=2, n_l=2, n_e=5$. b). P-1 cells (dark) and DOA-1 cells (grey), $n_y=n_u=2, n_l=3, n_e=5$. c). P-1 cells (dark) and DOA-1 cells (grey), $n_y=n_u=2, n_l=4, 5, n_e=5$ 24
- 15 Cell mapping analysis of the models selected using the error reduction ratio. a). P-1 cells (dark) and DOA-1 cells (grey), $n_y=n_u=2, n_l=2, 3, 4, n_e=5$. b). P-1, P-2, P-6 cells and DOA-1, DOA-2, DOA-6 cells, $n_y=n_u=2, n_l=5, n_e=5$ 25
- 16 Cell mapping analysis of the models identified using the modified OLS algorithm. a). P-1 cells (dark) and DOA-1 cells (grey), $n_y=n_u=2, n_l=2, \dots, 5, n_e=5$. b). P-1 cells (dark) and DOA-1 cells (grey), $n_y=n_u=3, n_l=2, \dots, 5, n_e=5$ 26

