Global Analysis and Model Validation in Nonlinear System Identification

B.R. Haynes †

S.A. Billings ††

† Department of Electronic and Electrical Engineering

University of Leeds

Leeds LS2 9JT, U.K.

†† Department of Automatic Control and Systems Engineering

University of Sheffield

Sheffield S1 3JD, U.K.

Research Report No. 429

April 1991

Global Analysis and Qualitative Model Validation

in

Nonlinear System Identification

Abstract

The analysis of parameterised nonlinear models is considered. In particular the emphasis is on models produced as a result of applying nonlinear system identification techniques. A qualitative approach is taken which calls upon ideas from bifurcation theory, which when combined with an existing numerical method enables the analysis of a wide variety of nonlinear model forms. The combination of qualitative and numerical aspects provides a flexible framework for analysis at the same time as providing a global perspective that is difficult to achieve using analytical methods alone. When combined with the well tried methods of system identification the approach enables the validation of nonlinear models from a qualitative viewpoint. This contrasts well with the statistical model validation techniques used in identification. The technique is applied to a number of examples to illustrate the effect of input design and signal to noise ratio on the modelling process.

Introduction

Traditionally the analysis of parameterised nonlinear systems has called upon ideas and methods from bifurcation and singularity theory. The so called dynamical systems approach is attractive, at least at first sight, in that it provides information on the very type of qualitative behaviour the nonlinear model was constructed to emulate. In addition the methods become of use just at the point where traditional linear theory breaks down. That is when one or more linearised eigenvalues become degenerate.

Unfortunately these methods have two drawbacks. First, being generally analytically based, they depend on a good deal of a priori knowledge of the solution structure of the system. In a general parameterised nonlinear model such information will not be available and indeed may be difficult to obtain. Secondly, most of the analytical approaches require the model structure to take a particular form.

In contrast the techniques of nonlinear system identification developed in last ten years place few such restrictions on the form of the nonlinear model. Adopting the model structure introduced by Billings and Leontaritis, 1981, allows the construction of a wide class of parametric nonlinear input output models. These recursive models map past inputs and outputs to the present output. The question then becomes, how to assess the validity of such representations.

One answer to this question which is proposed in the present study is to suspend the parametric models in a discrete, or *cellular*, state space. It is shown how this allows the application of a particularly attractive and simple numerical algorithm which has the advantage of providing

information on both the stationary and periodic nature of the models characteristic along with local and global stability information. As an added bonus this new approach places few restrictions on the form the model may take.

This paper is organized as follows: Section I of this paper provides a brief description of the basis and general procedure behind fitting nonlinear models using system identification techniques. Section II describes the cell map approach applied to parameterised models produced as a result of applying the nonlinear system identification technique. Section III introduces via a number of examples the concept of the *qualitative validation* of nonlinear models. In particular evidence of the existence of bifurcation effects introduced in the course of nonlinear system identification is provided and then used to construct nonlinear models that are both statistically and qualitatively valid.

1. System Identification

The development and use of system identification techniques has received much attention over recent years. This paper does not attempt to review or extend these ideas, for background material see in the first instance Ljung, 1987, or Billings, 1985. It is important however to understand the context which system identification plays in this work and to understand the procedure for fitting a nonlinear model using measurements of only the system input and output.

The present study focuses on one particular useful nonlinear model structure, the so called NARMAX model [Billings and Leontaritis, 1981]. This model structure can be used to represent all the commonly occurring nonlinear systems, including Volterra series, block structured, bilinear, state-affine, output-affine and rational models. However, the approach developed here is by no means constrained to the analysis of NARMAX models alone [Haynes and Billings, 1991].

Consider the general nonlinear discrete time system represented by

$$\begin{aligned} \mathbf{x}(t+1) &= g(\mathbf{x}(t),\mathbf{u}(t)) \qquad \mathbf{x}\in\mathbf{X} \quad \mathbf{u}\in\mathbf{U} \quad \mathbf{y}\in\mathbf{Y} \quad \mathbf{x}(t)=(x_1(t),..,x_n(t))^T \\ \mathbf{y}(t) &= h(\mathbf{x}(t),\mathbf{u}(t)) \qquad \mathbf{u}(t)=(u_1(t),..,u_r(t))^T \quad \mathbf{y}(t)=(y_1(t),..,y_m(t))^T \end{aligned} \tag{1.1}$$

where $t\in\mathbf{Z}$ is the set of integers, $\mathbf{X}$ is the state set of dimension n, $\mathbf{U}$ the input set of dimension r, $\mathbf{Y}$ the output set of dimension m, $\mathbf{x}_0$ is the initial state, $g : \mathbf{X}\times\mathbf{U}\rightarrow\mathbf{X}$ represents the one step ahead state transition function and $h : \mathbf{X}\times\mathbf{U}\rightarrow\mathbf{Y}$ is the output function.

Let the vector of all inputs from 1 to t be $\mathbf{u}^t = (\mathbf{u}(t),\mathbf{u}(t-1),..,\mathbf{u}(1))^T$. The function that describes the input output behaviour of a system is of primary importance in systems theory and may be utilised to describe the experimental observations made by an observer. When the system is initially at the zero equilibrium $\mathbf{x}_0$ the *zero-state response* function for an input sequence $\mathbf{u}^t$ of length t is

$$\mathbf{y}(t) = \mathbf{f}_{\mathbf{x}_0|t}(\mathbf{u}^t) = \mathbf{f}_{\mathbf{x}_0|t}(\mathbf{u}(t),\mathbf{u}(t-1),..,\mathbf{u}(1)) \qquad \mathbf{f}_{\mathbf{x}_0} : \mathbf{U}^t\rightarrow\mathbf{Y} \tag{1.2}$$

The response function $\mathbf{f}_{\mathbf{x}_0|t}$ is a different function for every $t = 1,2,..$ since the domain of $\mathbf{u}^t$ differs. The response function is said to be finitely realisable if and only if the dimension of the state vector in (1.1) is finite. The validity of the input output representation (1.2) depends upon the form of $\mathbf{f}_{\mathbf{x}_0|t}$ and two rather mild conditions [Leontaritis and Billings, 1985]. Firstly, the response function of the system should have a finite dimensional state space. Secondly, a rank condition is satisfied requiring the state space of the linearised system to be the same as the true system. Then, defining the observability indices of the linearised system as $n_1,n_2,..,n_m$ Padulo and Arbib, 1974, then the following recursive model describes the system around $\mathbf{x}_0$

$$y_i(t+n) = q_i\,(y_1(t+n_1-1),..,y_1(t),..,y_m(t+n_m-1),..,y_m(t),$$
$$u_1(t+p),..,u_1(t),\ ,..,\ u_r(t+p),..,u_r(t)\,) \qquad i=1,2,..,m \tag{1.3}$$

For the case when $r=1$ and $m=1$ the system is SISO. Introducing the delay, d, gives

$$y(t) = F^l(\,y(t-1),..,y(t-n_y)\,,\,u(t-d),..,u(t-d-n_u)\,) \tag{1.4}$$

where $F^l(.)$ is some nonlinear function. The input output description (1.4) expands the current output in terms of past inputs and outputs providing a model structure which will represent a broad class of nonlinear systems. Unlike the response function (1.2) the model (1.3) is independent of the initial condition, and is valid in a region about the equilibrium point. The recursive input output model (1.4) is a valid realisation of the general state space representation (1.1). Of course in practical system identification the state space model is unknown and hence the response function $\mathbf{f}_{\mathbf{x}_0|t}$ is unavailable. However, this does not prevent the construction of models of this form using nonlinear system identification techniques.

As part of the process of system identification is is necessary to construct estimates of the parameters, or coefficients, in (1.4). Expanding $F^l(.)$ as a polynomial and defining

$$V_1 = y(t-1),..,\ V_{n_y} = y(t-n_y)\,,\ V_{n_y+1} = u(t-d),..,\ V_S = u(t-d-n_u)$$

where $S = n_y + n_u$, enables (1.4) to be expressed in the form

$$y(t) = F^l(\,V_1\,,\,V_2,..,\,V_S\,) \tag{1.5}$$

which when expanded yields

$$y(t) = \sum_{i_1=1}^{S}\theta_{i_1}V_{i_1} + \sum_{i_1=1}^{S}\sum_{i_2=1}^{i_1}\theta_{i_1i_2}V_{i_1}V_{i_2} + \cdots + \sum_{i_1=1}^{S}\cdots\sum_{i_l=1}^{i_{l-1}}\theta_{i_1\cdots i_l}V_{i_1}\cdots V_{i_l} \tag{1.6}$$

where θ_i represents the coefficients of the linear, quadratic, cubic ... terms. Assuming that the system output $y(t)$ is corrupted by zero mean additive noise $e(t)$ such that $y(t)$ becomes $y(t) + e(t)$ then the input output model (1.6) is then written as

$$y(t) = \sum_{i_1=1}^{n_y}\theta_{i_1}(\,y(t-i_1)-e(t-i_1)\,) + \sum_{i_1=1}^{n_u}\theta_{n_y+i_1}(\,u(t-d-i_1+1)\,)$$
$$+ \sum_{i_1=1}^{n_y}\sum_{i_2=1}^{i_1}\theta_{i_1i_2}(\,y(t-i_1)-e(t-i_1)\,)(\,y(t-i_2)-e(t-i_2)\,)$$
$$+ \sum_{i_1=1}^{n_y}\sum_{i_2=1}^{n_u}\theta_{i_1n_y+i_2}(\,y(t-i_1)-e(t-i_1)\,)(\,u(t-d-i_2+1)\,)$$

$$+ \sum_{i_1=1}^{n_u} \sum_{i_2=1}^{i_1} \theta_{n_y+i_1 n_y+i_2} (u(t-d-i_1+1)(u(t-d-i_2+1)) + e(t) + \text{HOT...} \qquad (1.7)$$

Inspection of (1.7) shows that even though the noise enters the process as an additive signal at the output it appears in the model as multiplicative terms with the systems input and output. Although the model is linear in the parameters the inclusion of lagged process outputs introduces cross product terms between the noise and the process input and output signals. Rewriting (1.7) gives

$$y(t) = F^l(y(t-1),..,y(t-n_y) , u(t-d),..,u(t-d-n_u) , e(t-1),..,e(t-n_e)) + e(t) \qquad (1.8)$$

where n_e is the order of the additive noise $e(t)$. This model is referred to as a Nonlinear AutoRegressive Moving Average model with eXogenous inputs, or NARMAX model. A NARMAX model with first order dynamics expanded as a second order polynomial nonlinearity would for example be represented as

$$\begin{aligned} y(t) &= \theta_1 y(t-1) + \theta_2 u(t-1) + \theta_{11} y^2(t-1) + \theta_{12} y(t-1)u(t-1) + \theta_{22} u^2(t-1) \\ &\quad + e(t) - \theta_3 e(t-1) - \theta_{13} y(t-1)e(t-1) - \theta_{23} e(t-1)u(t-1) + \theta_{33} e^2(t-1) \end{aligned}$$

Much has been achieved in developing parameter estimation techniques for the NARMAX and other models. In this paper we are interested only in the general model fitting procedure. It is important to understand the basic ideas that lie behind the process of constructing a parametric model before we consider the analysis of such models. The NARMAX model representation has been chosen in order to illustrate this. Details of the algorithms used may be found in Billings and Voon, 1984, 1986, and Korenberg et al, 1988. Identification techniques making use of different model structures can be found in Chen and Billings, 1988a, 1988b, and Billings and Chen, 1989,

1.1. Input Output Data Collection

The process of fitting a model of the above form begins with data collection. In carrying this out the input signal should be designed so as to persistently excite the system. This requires selection of the mean, μ, variance, σ^2, and bandwidth of the input carefully. The input should be selected to excite all the systems dynamic modes over the complete amplitude range of interest. Correct input design is of primary importance if the model fitting exercise is to be successful. Indeed optimal input design for the purpose of identification has been attempted [Kabola and Springer, 1982], though this can only be achieved if the system model structure is known a priori.

Traditionally in system identification work pseudo-random binary sequences, or PRBS, are used. Unfortunately PRBS signals may not be persistently exciting for nonlinear systems. Additionally using a PRBS biases the estimates in the vicinity of those binary levels chosen for the PRBS. This may result in an effective linearisation of the nonlinear system. One approach used with some success is to construct a compound PRBS input [Billings and Fakhouri, 1982]. A better idea is to adopt a gaussian distributed input over the required amplitude range [Leontaritis and Billings, 1987.]

1.2. Nonlinear Structure Detection

The *Structure Detection* test, **STD**, can be used to determine whether the data collected is representative of a linear or nonlinear system. Assume that the input $u(t)=u'(t)+b$ and output noise $e(t)$ are independent, $e(t)$ and $u'(t)$ are zero mean processes, $b \neq 0$, all odd order moments of $u'(t)$ and $e(t)$ are zero mean and all even order moments are assumed to exist.

Then it can be shown that the process is linear iff

$$\phi_{y'y'^2}(\tau) = E[y'(t+\tau)y'^2(t)] = 0 \qquad \forall \tau \tag{1.9}$$

where $y'(t)$ is the system output, with mean level removed [Billings and Voon, 1983]. Note that the test distinguishes between linear additive noise corruption of the measurements and distortion due to nonlinear effects. The input conditions are satisfied by selecting an input such as a sine wave, gaussian signal, independent uniformly distributed process or PRBS.

1.3. Parameter Estimation

This involves estimation of the model coefficients θ_i in the NARMAX model expansion given values for the model parameters n_u, n_y, n_e, d and l in (1.4). When the system under test is nonlinear it will in general be impossible to solve (1.8) for $e(t)$ and consequently the noise source $e(t)$ and the prediction error will not be equal. Reformulating (1.8) into a *Prediction Error* model by defining $\hat{y}(t)$ as the one step ahead prediction such that

$$\xi(t) = y(t) - \hat{y}(t) \tag{1.10}$$

gives

$$y(t) = F^l(y(t-1),..,y(t-n_y), u(t-d),..,u(t-d-n_u), \xi(t-1),..,\xi(t-n_\xi)) + \xi(t) \tag{1.11}$$

where

$$\hat{y}(t) = F^l(y(t-1),..u(t-d),..)$$

$$E[\xi(t) \mid y(t-1),..,u(t-d),..] = 0 \tag{1.12}$$

and n_ξ is the order of the prediction error $\xi(t)$. In general the noise source $e(t)$ can neither be measured nor computed from (1.8), all the parameter estimation algorithms must therefore be developed based on (1.11). Direct application of least squares yields biased estimates in nonlinear systems and alternative algorithms must be developed to overcome this difficulty. The *Maximum Likelihood* method can be applied directly when it is known that a system is linear and the innovations or prediction errors have a gaussian distribution. Again when the system is nonlinear this is not possible [Billings and Voon, 1986]. The *Prediction Error* method [Billings and Voon, 1986], can however be applied without any knowledge of the distribution of the innovations. For gaussian innovations the Prediction Error method is equivalent to the Maximum Likelihood method. Alternatively the *Orthogonal Estimation* algorithm by Korenberg et al, 1988, can be used. This method is used in this work as it enables both parameter estimation and model structure selection to be carried out simultaneously.

1.4. Orthogonal Parameter Estimation

Estimation of the model coefficients, θ_i in (1.11) is achieved by constructing an auxiliary orthogonal parameter set, w_i, such that the linear in the parameters expansion of (1.11)

$$y(t) = \sum_{i=1}^{n_\theta} \theta_i p_i(t) + \xi(t) \tag{1.13}$$

where n_θ denotes the number of parameters in the model and $p_i(t)$ represents a term in the expansion of a NARMAX model with no two $p_i(t)$s are identical, eg. $p_1(t)=y(t-1)$, $p_2(t)=u(t-1)$.... The orthogonal algorithm involves transforming (1.13) into an equivalent auxiliary model

$$y(t) = \sum_{i=1}^{n_\theta} g_i w_i(t) + \xi(t) \tag{1.14}$$

where the g_i's are the orthogonal coefficients and $w_i(t)$, $i = 1,..,n_\theta$ are constructed to be orthogonal over the data records such that $\overline{w_i(t)\, w_{j+1}(t)} = 0$ for $i = 1,..,j$. The orthogonal estimation procedure obtains estimates of the g_i's and transforms them back to the system parameters θ_i's [Korenberg et al, 1988].

1.5. Model Structure Selection

The successful application of a parameterised model in nonlinear identification is critically dependent upon detecting which candidate terms should be included. This is carried out using the *Error Reduction Ratio*, or ERR test which is derived as a by-product of the orthogonal estimation algorithm. Consider the auxiliary model described by (1.14). Multiplying the model by itself and taking the time average gives

$$\overline{y^2(t)} = \sum_{i=1}^{n_\theta} g_i^2 \, \overline{w_i^2(t)} + \overline{\xi^2(t)} \tag{1.15}$$

Assuming that $\xi(t)$ is a zero mean white noise sequence and the orthogonal property holds. The maximum mean squared prediction error is achieved when no terms are included in the model, $n_\theta = 0$, such that $\overline{\xi^2(t)} \mid_{n_\theta=0} = \overline{y^2(t)}$. The ERR_i is defined as the reduction in mean square error by including a term $g_i w_i(t)$ in (1.15), equal to $g_i^2 \overline{w_i^2(t)}$, expressed as a percentage reduction in the maximum mean squared error, such that

$$ERR_i = \frac{g_i^2 \, \overline{w_i^2(t)}}{\overline{y^2(t)}} \times 100 \qquad i = 1,..,n_\theta \tag{1.16}$$

Insignificant terms in (1.14) can then be discarded by defining a threshold value of ERR below which terms are considered to contribute a negligible reduction in the mean squared error.

The great advantage of the ERR is that it does not require the estimation of a complete model in order to determine the significance of a candidate term. The ERR value is however dependent upon the order in which a term enters the auxiliary equation, for more detail on this see Billings et al, 1989.

1.6. Nonlinear Model Validation Test

The final stage in the model fitting process involves statistically checking the whiteness of modelling residues for unmodelled predictable behaviour. *Model Validity Tests*, MVT detect information in the residuals which, if neglected, will introduce bias in the parameter estimates. If the system under test is linear the residuals should be unpredictable from all past inputs and outputs. Assume that the input $u(t)$ and noise $e(t)$ are independent zero mean processes, and all odd order moments are zero, $e(t)$ is white and $u(t)$ may be white. The traditional MVT tests for linear systems require [Box and Jenkins, 1976]

$$\phi_{\xi\xi}(\tau) = \delta(\tau)$$

$$\phi_{u\xi}(\tau) = 0 \qquad \forall \tau \tag{1.17}$$

The use of these correlation techniques for model validation has been shown to be inappropriate when the system under test is nonlinear . In the general case when it is known that the system under test is nonlinear, as indicated by the **STD** test, then the residuals will be unpredictable from all past inputs and outputs if, in addition to (1.17), [Billings et al, 1989].

$$\begin{aligned} \phi_{\xi\xi u}(\tau) &= 0 \qquad && \forall \tau \geq 0 \\ \phi_{u^2\xi^2}(\tau) &= 0 \qquad && \forall \tau \\ \phi_{u^2\xi}(\tau) &= 0 \qquad && \forall \tau \end{aligned} \tag{1.18}$$

The identified model is then said to pass both the linear and nonlinear **MVT** tests.

1.7. Example

The NARMAX model orthogonal estimator coupled with the nonlinear structure detection test and model validity tests are combined together. The arrangement of the candidate terms, $p_i(t)$, in (1.13) is accomplished using a *Forward Regression* algorithm applied to the estimated prediction error $\xi(t)$. To illustrate the procedure for fitting a NARMAX model consider the system described by

$$y(t) = 0.8\,y(t-1) + 0.5\,u(t-1) + 0.2\,u(t-1)y(t-1) + e(t) \tag{1.19}$$

The input $u(t)$ is composed of a uniformly distributed signal $u(t)\in[\pm 1.0]$ superimposed on a mean level $b = 0.2$. The additive output noise signal, $e(t)$, is a gaussian white noise process N(0.0,0.2). Input output data generated is shown in Fig. (1a), along with the results of the **STD** test, (1.9), Fig. (1b), that clearly indicate the data to be from a nonlinear system. For this trivial system selecting $n_u = 1$, $n_y = 1$, $n_e = 3$, $d = 1$ and $l = 1$ gave the linear model with coefficients

Terms	Estimates	ERR	Standard Deviation
y(t- 1)=	0.841e+00	(0.863e+02)	(0.662e-02)
u(t- 1)=	0.699e+00	(0.101e+02)	(0.137e-01)
e(t- 1)=	0.855e+00	(0.149e+01)	(0.428e-01)

e(t- 3)=	-0.328e+00	(0.193e+00)	(0.463e-01)

Note the total error sum contributed by the 4 terms due to (1.16) is $ERR_i = 98.1\%$. Rewriting this model in the more usual form gives

$$y_t = 0.841y_{t-1} + 0.699u_{t-1} + 0.855e_{t-1} - 0.328e_{t-3} \tag{1.20}$$

where $y_t = y(t)$, $u_t = u(t)$ and $e_t = e(t)$. The model above gives good predictions, Fig. (2a). and passes the linear **MVT** tests, (1.17), shown in the upper two plots in Fig. (2b). However, the deficiency of the linear model is highlighted by the nonlinear **MVT** tests, (1.18), which clearly lie outside the 95% confidence intervals, shown in the lower three plots in Fig. (2b). Setting $l = 2$ a quadratic nonlinear model is now fitted resulting in the nonlinear model

Terms	**Estimates**	**ERR**	**Standard Deviation**
y(t- 1)=	0.800e+00	(0.863e+02)	(0.125e-02)
u(t- 1)=	0.502e+00	(0.101e+02)	(0.284e-02)
y(t- 1)*u(t- 1)=	0.200e+00	(0.348e+01)	(0.125e-02)
e(t- 1)=	-.759e+00	(0.381e-01)	(0.427e-01)

The model produces excellent predictions, Fig. (3a), having a total error sum $ERR_i = 99.9\%$, and passes both the linear and nonlinear **MVT** test, Fig. (3b). Rewriting this model gives

$$y_t = 0.800y_{t-1} + 0.502u_{t-1} + 0.200u_{t-1}y_{t-1} - 0.759e_{t-1}$$

Notice that the inclusion of the nonlinear term improves the ERR sum in this case by about 2% over the best linear fit. It may then be argued that the best linear model would, for all intents and purposes be sufficient. We show later in this paper the folly in this reasoning.

2. Global Analysis

The stability of a model, both absolute and structural, is dependent on the location and distribution of degenerate singularities on the solution manifold of that system. To date little has been done to study the qualitative aspects of the type of nonlinear models produced as a result of the identification process. Analytical methods meet some of the requirements for certain classes of models. Such methods however, tend to be valid for only one particular model structure or type. For application to the type of nonlinear models produced in system identification, a flexible approach is essential that is not limited in this way. Furthermore, it is desirable to add a global aspect to the analysis to enable both the state space, and parameter space, of the model to be probed for *interesting* behaviour.

In order to achieve these aims a dual approach has been adopted. This combines the essentially qualitative ideas of Bifurcation theory and Singularity methods with a simple yet attractive numerical algorithm. The analysis of nonlinear systems using the *Cell Map* approximation was first carried out by Hsu, see for example Hsu and Guttalu, 1980. This method has been extended to the qualitative analysis of nonlinear parameterised models of the type commonly encountered in bifurcation studies [Haynes and Billings, 1991]. The approach proves attractive for a number of reasons. Firstly, it provides a method of enumerating both the stationary and periodic solution structure of the system over a given parameter range. Secondly, and more importantly, information of a global nature is provided on the extent of the systems stability domains. As a result of this both local and global bifurcation phenomena can be detected.

2.1. Cell Mappings

The process of analysing a system using Cell Map analysis comprises of a number of steps. The first is the suspension of the nonlinear system, say (1.1), in a *cell state space,* $\mathbf{Z}^n$. This is an n-dimensional space whose elements are n-tuples of integers. Each element is called a *cell vector,* or simply a *cell*, and is denoted by **z**.

There are many ways to obtain a cell structure over a given euclidean state space. The simplest way is to construct a cell structure consisting of rectangular parallelepipeds of uniform size (squares, cubes etc). Let x_i ,i = 1,..,n be the state variables and let each coordinate axis of the state variable be divided into a number N_i, of intervals of uniform size h_i. The interval z_i along the x_i axis is defined such that it covers all the x_i of interest.

$$(z_i - \frac{1}{2})\,h_i \le x_i \le (z_i + \frac{1}{2})\,h_i \qquad z_i = 1,2..,N_i \tag{2.1}$$

The n-tuple z_i ,$i = 1,..,n$ is then called the cell vector, denoted by **z**. A point **x** belongs to a cell **z** iff x_i and z_i satisfy (2.1) $\forall i \in [1,n]$. Each cell **z** is considered as an entity and the entire collection of cells as the *cell state space.* Consider now the mapping between two cells $z(j)$ and $z(j+1)$, where $j = 1,2,..$ is used to denote a sequence in the same manner as the iterates of a mapping. The *cell map,* **C(z)**, is a mapping of a set of integers $\{N+\}$ such that

$$\mathbf{z}(j+1) = \mathbf{C}(\mathbf{z}(j)) \qquad \mathbf{z}(j) \in \mathbf{Z}^n \subset \mathbf{S} \tag{2.2}$$

The *cell function,* **F(z,C)**, is then defined as

$$\mathbf{F}(\mathbf{z},\mathbf{C}) = \mathbf{C}(\mathbf{z}) - \mathbf{z} \tag{2.3}$$

If $\mathbf{C}^K(\mathbf{z})$ denotes **C(z)** applied K times, and the K-th step ahead *cell function is*

$$\mathbf{F}^K(\mathbf{z},\mathbf{C}^K) = \mathbf{C}^K(\mathbf{z}) - \mathbf{z} \tag{2.4}$$

A *singular cell* $\mathbf{z}^*$ as a cell satisfying the relationship

$$\mathbf{F}(\mathbf{z}^*,\mathbf{C}) = 0 \qquad \text{or} \qquad \mathbf{z}^* = \mathbf{C}(\mathbf{z}^*) \tag{2.5}$$

and a periodic cell cycle, given that $\mathbf{C}^0(\mathbf{z})$ is the identity mapping, is a sequence of K distinct cells $\mathbf{z}^*(j)$, $j = 1,..,K$, K being the minimum value which satisfies

$$\mathbf{z}^*(m+1) = \mathbf{C}^m(\mathbf{z}^*(1)) \qquad m = 1,..,K-1 \qquad \mathbf{z}^*(1) = \mathbf{C}^K(\mathbf{z}^*(1)) \tag{2.6}$$

Each element of the periodic cycle is a *periodic cell.* The complete cell cycle is labelled as **P-K.** Additionally those cells eventually mapped onto the **P-K** cycle by (2.2) are defined as within the *domain of attraction*, or **DOA**, of the cell cycle and labelled as the **DOA-K** cells.

The size of the cell state space is determined by the system itself. For most practical systems there are ranges of values of the state variable beyond which we are no longer interested. This means that there is only a finite region of the state space which is of concern. Similarly for a dynamical system governed by a cell mapping there is only a finite region of cell space of interest, and correspondingly a finite number, N_c, of cells, named the *regular cells.* The *sink cell* is used to encompasses all possible cells outside the region of interest. If the mapped image of a regular cell lies outside the region of interest it is then said to be mapped into the sink cell. The regular cells are labelled by positive integers $\{1,2..,N_c\}$. The sink cell is labelled as $\{0\}$, the zero cell. This makes the total number of cells N_c+1, such that $\mathbf{S}=\{N_c+\}$. The set, **S**, is closed under the mapping described by

$$\mathbf{z}(j+1) = \mathbf{C}(\mathbf{z}(j)) \qquad \mathbf{z}(j), \mathbf{z}(j+1) \in \mathbf{Z}^n \subset \mathbf{S}$$

$$\mathbf{C}(0) = 0 \qquad \mathbf{S} = \{N_c+\} \tag{2.7}$$

A number of consequences arise from the definitions above. The sink cell, **C**(0), is a **P-1** cell. The set of regular cells within the influence domain of the sink cell, these being eventually mapped to **C**(0), are in the domain of attraction of the sink cell, and labelled the **DOA-Sink** cell.

2.2. Cell Mapping Discretisation

The cell map, **C**(**z**), system may be considered as a discrete system similar to the point mapping

$$\mathbf{x}_{k+1} = \mathbf{g}(\mathbf{x}_k) \qquad \mathbf{x} \in \mathbf{R}^n \tag{2.8}$$

In order to make use of the qualitative ideas provided by dynamical systems theory it is necessary to extend the algorithm of Hsu to the more general parameterised nonlinear system

$$\mathbf{x}_{k+1} = \mathbf{g}(\mathbf{x}_k, \mu) \qquad \mathbf{x} \in \mathbf{R}^n \quad \mu \in \mathbf{R}^m \tag{2.9}$$

Applying the center point method of discretisation requires the division of $\mathbf{R}^n$ into a collection of cells according to (2.1) and the calculation of each cells center point $\mathbf{x}^{(d)}(j)$ such that

$$x_i^{(d)}(j) = x_i^{(l)} + h_i\, z_i(j) - \frac{h_i}{2} \qquad z_i(j) = 1,..,N_i \quad j = 1,..,N_i \quad i = 1,..,n \tag{2.10}$$

where h_i is the cell size and $x_i^{(l)}$ the lower bound defining the region of interest such that

$$\mathbf{x}^{(l)} \le \mathbf{x} \le \mathbf{x}^{(h)} \qquad \mathbf{x}^{(l)} = (x_1^{(l)},..,x_n^{(l)}) \quad \mathbf{x}^{(h)} = (x_1^{(h)},..,x_n^{(h)}) \tag{2.11}$$

Similarly $\mathbf{R}^m$ is discretised using

$$\mu_i^{(d)}(j) = \mu_i^{(l)} + g_i\, z_i(j) - \frac{g_i}{2} \qquad z_i(j) = 1,..,N_i \quad j = 1,..,N_i \quad i = n+1,..,n+m \tag{2.12}$$

where g_i is the cell size and the region of interest defined by

$$\mu^{(l)} \le \mu \le \mu^{(h)} \qquad \mu^{(l)} = (\mu_1^{(l)},..,\mu_m^{(l)}) \quad \mu^{(h)} = (\mu_1^{(h)},..,\mu_m^{(h)}) \tag{2.13}$$

The point mapping or image of the center point $\mathbf{x}^{(d)}(j)$ is then calculated using (2.9) such that

$$\mathbf{x}_{k+1}^{(d)} = \mathbf{g}(\mathbf{x}_k^{(d)}, \mu^{(d)}) \tag{2.14}$$

and the cell map **C**(**z**) is constructed by determining the image of each cell within **S** using

$$C_i(z_i(j)) = z_i(j+1) = \mathrm{INT}\left[\frac{x_i^{(d)}(k+1) - x_i^{(l)}}{h_i} + 1\right] \quad i = 1,..,n \quad j = 1,..,N_i$$

$$C_i(z_i(j)) = z_i(j) \qquad i = n+1,..,n+m \quad j = 1,..,N_i \tag{2.15}$$

Note we are in effect constructing m separate cell mappings of dimension n, each representing a

slice within the parameter space $\mu=(\mu_1,..,\mu_m)^T$.

2.3. Cell Map Analysis

Having constructed C(z) the classification of all cells within S is carried out using an *Unravelling Algorithm* [Hsu and Guttalu, 1980]. The algorithm involves calling up each cell in turn and processing it in order to determine its global characteristics. Each cell is then classified accordingly.

A cell may be a singular cell, **P-1**, or a periodic cell cycle, **P-K**, satisfying (2.5) or (2.6). The set of all such cells make up the invariant stable and unstable orbits within $\mathbf{Z}^n$. When $K=1$ a fixed point has been located, when $K>1$ a periodic solution or limit cycle has been detected.

Alternatively a cell may simply be a regular cell in the **DOA** of a **P-K** cell. Each such cell is then said to be in the same *group* and have the same *periodicity* as that cell cycle and is labelled accordingly.

Finally a cell may be mapped by (2.2) outside the region of interest into the Sink cell. Such a cell is then said to be in the domain of attraction of the sink cell, the **DOA-Sink**, and is labelled accordingly. For more detail on the Unravelling algorithm and is variants see Hsu, 1987 or Haynes and Billings, 1991.

2.4. Suspension of the NARMAX Model

To analyse a NARMAX model within the framework outlined above it is first necessary to suspend the model, appropriately parameterised in the cell state space. Rewriting (1.8) assuming the input $u(t)$ to be a quasi static parameter and neglecting the noise terms $e(t),..,e(t-n_e)$ gives

$$y(t) = F^l[y(t-1),..,y(t-n_y),\mu] \tag{2.16}$$

where $u(t)$ is assumed constant such that

$$\mu = u(t) = u(t-d) = u(t-d-1) = \vdots = u(t-d-n_u) \tag{2.17}$$

The NARMAX model can now be written in the form of a parameterised discrete dynamical system

$$\mathbf{x}_{k+1} = F(\mathbf{x}_k, \mu, \alpha) \qquad \mathbf{x}\in\mathbf{R}^n \quad \mu\in\mathbf{R} \quad \alpha\in\mathbf{R}^k \tag{2.18}$$

where $\mathbf{x}_{k+1}=y(t)$, $\mathbf{x}_k=(y(t-1),..,y(t-n))^T$, $\mu=u(t)$, $\alpha=(\theta_1,..,\theta_k)$, $n=n_y$ and $k=n_\theta$.

The convention adopted here when considering multiparameter problems is to label μ as the primary parameter and α as an auxiliary parameter vector. The auxiliary parameter vector α is only considered when taking into account perturbations to the basic problem. These are brought about by variations in the coefficients, α_i , $i = 1,..,k$ which in this parameterisation correspond to the identified model parameters θ_i , $i = 1,..,n_\theta$ in (1.13).

2.5. Example

Consider the simple difference equation or NARMAX model identified earlier. Applying (2.16-18) gives

$$x_{k+1} = 0.8\,x_k + 0.2\,\mu + 0.2\,\mu\,x_k \tag{2.19}$$

Assuming the parameter vector, α, in (2.18) is a constant, gives

$$\mathbf{x}_{k+1} = g\,(\,\mathbf{x}_k\,,\,\mu\,) \qquad \mathbf{x}\in\mathbf{R} \quad \mu\in\mathbf{R} \tag{2.20}$$

Suspending (2.21) in a cell state space gives

$$\mathbf{z}(j+1) = \mathbf{C}(\,\mathbf{z}(j)\,,\,\mu\,) \qquad \mathbf{z}\in\mathbf{Z} \quad \mu\in\mathbf{R} \tag{2.21}$$

Analysis of this example was carried out by defining a cell state space, such that $x\in[-20,20]$ and $\mu\in[-2,3]$ over a mesh of 125×125 cells and applying (2.9-15). Application of the Unravelling algorithm produced the Cell Map diagram, Fig. (4).

Here a continuous curve of **P-1** cells represents the variation of the stable fixed points in (2.19) for $\mu<1$. These points are globally stable in that they act as an attractor $\forall z\in Z$ or $\forall x\in R$. The shaded region represents the **DOA-Sink** cells for $\mu>1$, that is those cell mapped out of the region of interest into the sink cell.

A stability picture of the system is thus quickly built up. This example is of course trivial, the same results can be obtained analytically directly from (2.19)

$$\overline{x} - g(\overline{x},\mu) = 0 \qquad \text{such that} \qquad \overline{x} = \frac{5}{2}\,\frac{\mu}{1-\mu} \tag{2.22}$$

linearising about $\overline{x}$ gives

$$\lambda(\mu) = D_x\,g(\,x,\mu\,)\,\big|_{\,x=\overline{x}} = 0.8 + 0.2\mu \tag{2.23}$$

Hence for $\mu\leq 1$ the eigenvalue, $|\,\lambda(\mu)\,| < 1$, for all x, and the curve of fixed points is stable.

3. Qualitative Model Validation and Bifurcation

Section I served to illustrate how both the theory, and implementation, of an identification algorithm can have fundamental bearing on the resulting form of the nonlinear model. In order to access the qualitative characteristic of such models in the present study the parameterised equation is suspended in a cell map framework. This enables the characterisation of both stationary and periodic behaviour at the same time as providing global and local stability information. The framework of analysis described in previous section is now employed to carry out what can be thought of as the *qualitative validation of nonlinear models.* In so doing two problems are addressed which are not easy to answer using traditional techniques alone.

3.1. Input Experiment and Design

The importance of input experiment design and the collection of I/O data to be used in identification is well known [Goodwin and Payne, 1975] Criteria such as input amplitude, probability distribution and bandwidth play an important role. In this section it is demonstrated using the global analysis procedure how simple input design considerations can influence the results obtained in the identification process. More importantly, the *statistical* model validity tests do not always point out these deficiencies. The approach outlined comprises of an iterative sequence of input design; model identification; statistical then qualitative model validation.

To illustrate the importance of input design consider the identification of a general known block structured model, Fig. (5a) [Billings and Fakhouri, 1982]. Block structured continuous systems of this type can be analysed using the multidimensional Laplace transform by constructing a model of the form [George, 1959, Schetzen, 1980].

$$Y(s_1,..,s_n) = \sum_{i=1}^{n} G_i(s_1,..,s_i) \prod_{j=1}^{i} U(s_j) \tag{3.1}$$

where $G_i(\cdot)$ is the i-th order kernel transform. Applying the multidimensional Z-transform [Alper, 1964] to (3.1) yields

$$Y(z_1,..,z_n) = H_1(z_1) + \cdots + H(z_1,..,z_n) = \sum_{i=1}^{n} H_i(z_1,..,z_i) \tag{3.2}$$

where

$$H_i(z_1,..,z_i) = \mathbf{Z}\left[G_i(s_1,..,s_i) \prod_{j=1}^{i} U(s_j) \right] \tag{3.3}$$

An expression for $Y(z)$ may then be constructed using the process of *association of variables*

[Jagen and Reddy, 1972] given $U(s)$ explicitly and the position of any sample or zero order hold devices. Unfortunately a recursive input output model, as in (1.8), can not be constructed.

3.1.1. System Characteristic

The continuous system depicted in Fig. (5a) may be represented, using (3.1), as

$$Y(s_1,s_2) = \frac{200}{(s_1 + s_2 + 2)(s_1 + 10)(s_2 + 2)} \tag{3.4}$$

Alternatively, writing this system as a set of ODE's gives

$$\begin{aligned} \dot{x_1}(t) &= -2x_1(t) + 2x_2^2(t) \\ \dot{x_2}(t) &= -10x_2(t) + 10u(t) \end{aligned} \tag{3.5}$$

Taking the input u(t) as the parameter μ gives

$$\dot{\mathbf{x}}(t) = F(\mathbf{x},\mu) \qquad \mathbf{x}\in\mathbf{R}^2 \quad \mu\in\mathbf{R} \tag{3.6}$$

This parameterised continuous system may then be analysed directly using the framework outlined above, see also Haynes and Billings, 1991. Suspending (3.6) in a parameterised cell state space gives

$$\mathbf{z}(j+1) = \mathbf{C}(\mathbf{z}(j),\mu) \qquad \mathbf{z}\in\mathbf{Z}^2 \quad \mu\in\mathbf{R} \tag{3.7}$$

Defining the region of interest such that $x_1\in[-25,25]$, $x_2\in[-25,25]$, and $\mu\in[-5,5]$ over 125x125x40 cells and applying (2.9-15) gives $\mathbf{C}(\mathbf{z})$. Applying the Unravelling algorithm to $\mathbf{C}(\mathbf{z})$ results in the cell diagram, Fig. (5b). A stationary fixed point, represented by the **P-1** curve, varies smoothly with the parameter μ. Not unsurprisingly this system is globally stable for all input values. However, as the Cell Map algorithm always requires a finite range of μ over which to work Fig. (5b) does not truly reflect this global stability w.r.t μ. However, extending the range of $\mu\in\mathbf{R}$ can easily demonstrate this.

3.1.2. Identification and Cell Map Analysis

Proceeding now to generate some I/O data from the system. Linearisation suggests an input bandwidth of 1Hz will suffice and a sampling interval $f_s = 5Hz$ was chosen. The I/O data was generated by simulating the system using a Runge-Kutter 4th order integration with a small step size compared to f_s. The input was a uniformly distributed signal $u(t)\in[\pm 1.0]$ superimposed on a mean level, $b = 0.5$, Fig. (6a). The corresponding **STD** test for this data again clearly indicates the need for a nonlinear model, Fig. (6b).

Nonlinear identification of this data, setting $n_u = n_y = 1$, $n_e = 5$, $d = 0$ and $l = 2$ produced the quadratic NARMAX model with 1st order dynamics

$$\begin{aligned} y_t &= 0.660y_{t-1} + 0.030u_{t-1} + 0.237u_{t-1}^2 + 0.046u_{t-0}^2 + 0.034y_{t-1}\,u_{t-0} + 0.014y_{t-1}^2 \\ &\quad - 0.023u_{t-1}\,y_{t-1} - 0.641e_{t-1} - 0.098e_{t-5} + 0.002e_{t-2} + 0.046e_{t-4} \end{aligned} \tag{3.8}$$

The identified NARMAX model produces reasonable predictions, with $ERR_i = 97.2\%$, Fig. (7a), and passed all the **MVT** tests, Fig. (7b).

Over-fitting by setting $n_u = n_y = 2$, $d = 0$ and $l = 2$ produced a NARMAX model with 2nd order dynamics.

$$\begin{aligned} y_t &= 0.086y_{t-1} + 0.146u_{t-1}^2 + 0.352y_{t-2} + 0.215u_{t-2}^2 + 0.216u_{t-1}\,u_t + 0.049u_t - 0.062u_{t-1} \\ &\quad + 0.053u_{t-2} - 0.014u_{t-2}\,y_{t-1} + 0.022y_{t-2}^2 + 0.177y_{t-2}\,u_{t-1} - 0.044y_{t-1}\,u_t - 0.123u_{t-2}\,y_{t-2} \\ &\quad - 0.037u_t^2 - 0.055y_{t-2}\,u_t - 0.089y_{t-1}^2 + 0.114y_{t-2}\,y_{t-1} - 0.378e_{t-2} - 0.101e_{t-5} \end{aligned} \tag{3.9}$$

Clearly a greater number of terms have been included in this model. Again the identified NARMAX model produces good predictions, Fig. (8a), with $ERR_i = 97.3\%$ and passed all the **MVT** tests, Fig. (8b).

Looking at the results from the above identification, either model seems to give an acceptable fit to this I/O data. No indication is given, in either the model predictions or the **MVT** test, as to which is the better model. Indeed the only discriminating feature is the number of terms in the respective models.

Analysis of the two representations was carried out by suspending the each model in a parameterised cell state space in turn. Defining a cell state space for (3.8) using (2.9-15) with $x_1 \in [-25,25]$ and $\mu \in [-5,5]$ over a mesh of 125x125 cells and applying the Unravelling algorithm produced the cell diagram, Fig. (9a). Defining a similar cell state space for (3.9) with $x_1, x_2 \in [-25,25]$ and $\mu \in [-5,5]$ results in the cell diagram, Fig. (9b).

Comparing these two cell diagrams with Fig. (5b) highlights the inadequacy of both models. The shaded area represents the **DOA** of the attracting **P-1** cells. Notice for the 2nd order model the Unravelling algorithm detects a number of **P-2** cell cycles. Qualitative agreement between the either model and the true system characteristic is only achieved over a relatively narrow range of input parameter, μ. Also notice that the first order model (3.8) is stable over a larger domain.

The failure of these two models to capture the true characteristic of the original system, is due to the inability of the input to excite the system over a wide enough range. Indeed a rough

correspondence can be seen between size of the input used in the identification experiment and the domain of stability apparent in the identified models.

3.1.3. Input Design and Cell Map Analysis

In an attempt to improve the qualitative characteristic of the model the identification is now repeated, redesigning the input signal so as to excite the nonlinear system over a wider region.

Fig. (10a) shows the I/O data generated using a noise input $u(t) \in [+-1.0]$ superimposed upon a square wave of amplitude ± 1 along with the corresponding **STD** test, Fig. (10b) . Identification of this data, by setting $n_u = n_y = 1$, $n_e = 5$, $d = 0$ and $l = 2$, produced the 1st order model with $\sum \mathrm{ERR}_i = 99.4\%$.

$$y_t = 0.683y_{t-1} + 0.247u_{t-1}^2 + 0.065u_t^2 - 0.579e_{t-1} - 0.049e_{t-5} + 0.049e_{t-2} \tag{3.10}$$

Identification by setting $n_u = n_y = 2$, $n_e = 5$, $d = 0$ and $l = 2$, produced the 2nd order model with $\sum \mathrm{ERR}_i = 99.5\%$.

$$\begin{aligned} y_t = {} & 0.026y_{t-1} + 0.256u_{t-1}^2 + 0.441y_{t-2} + 0.163u_{t-2}^2 + 0.063u_t^2 + 0.049u_{t-2}\,u_t \\ & - 0.393e_{t-2} - 0.071e_{t-5} \end{aligned} \tag{3.11}$$

Both the identified NARMAX models produced good predictions and passed the **MVT** tests.

Analysis of the models (3.10) and (3.11) using the Cell Map framework yields the cell diagrams Fig. (11a) and Fig. (11b). Comparing this to Fig. (5b) shows a clear qualitative agreement between system and model characteristics over the entire range of input considered for both the above models. The only noticeable difference being a small core of **P-2** cells detected around the origin for the second model. This effect can be attributed to the coarse cell size used in the analysis.

An important point has arisen in the process of developing the iterative procedure outlined above. Statistical model validation techniques cannot always be relied upon to produce qualitatively correct results over a wide range of operation. They cannot detect poor experimental design for example. Input design is of primary importance in the identification of nonlinear systems, both the deterministic and stochastic elements need to be chosen correctly.

3.2. Model Structure Selection

As briefly discussed above the process of obtaining a parsimonious model from the full model representation (1.11), uses the method of Forward Regression to delete insignificant terms from the auxiliary model, (1.14), according to the value of the ERR_i calculated using (1.16). This problem is magnified in nonlinear systems due to the large number of candidate terms the algorithm is required to consider. Consider the effect of S/N ratio on the qualitative behaviour of a model. It is well known that parameter estimates obtained using virtually any algorithm deteriorate as the S/N is decreased. However the approach described in this work can be used to good effect in order to gain a deeper understand of this problem.

3.2.1. System Definition

To illustrate this problem consider the system depicted in Fig. (12a). This system is to be studied in a number of differing S/N environments. Assume now that the output of this system is corrupted by an additive white noise signal such that

$$z(t) = y(t) + e(t) \tag{3.12}$$

and the S/N ratio for the system is defined as

$$S/N = \frac{Var(y(t))}{Var(e(t))} \tag{3.13}$$

Note from the discussion earlier this will induce multiplicative correlated noise terms in the NARMAX model.

Writing this system as a set of ODE's and analysing it directly using the Cell Map algorithm results in the cell diagram, Fig. (12b). Once again the input $u(t)$ was chosen as the parameter μ and the cell state space was defined as $x_1, x_2 \in [-125, 125]$ and $\mu \in [-5, 5]$ over 125x125x40 cells. Again this simple system exhibits global stability for all input values and has a fixed point, delineated by the curve of **P-1** cells, that varies smoothly with μ in a cubic fashion.

I/O data from the system was generated using an input with mean $\mu = 1.0$, RMS $\sigma = 1.0$ and bandwidth, $f_i = 1.0$Hz. The sampling interval was set at $f_s = 6.0$Hz over 1000 samples. A number of different I/O datasets were constructed such that the signal to noise ratio varied from ∞ and 1 by adjusting $e(t)$ the output noise appropriately. Consider just the cases where S/N=∞, 1, 3, and 6 . Fig. (13a) shows the noise free signal, S/N=∞ and Fig. (13b) the noisy data, when S/N$\approx$1. The other cases lie in between these extreme values.

3.2.2. Identification

A 2nd order cubic NARMAX model with 3rd order noise model was fitted to each of the datasets. Fig. (14a) and Fig. (14b) show the model predictions, $\hat{y}(t)$, obtained using this model structure for S/N=∞ and S/N$\approx$1. The identified models pass all the **MVT** tests and produce reasonable predictions for all the S/N ratios considered. Although the performance of the model predictions seems to significantly deteriorate as the S/N ratio decreases. If however the predicted output, $\hat{y}(t)$, is compared with the true noise free output $y(t)$, that is when S/N=∞, very little difference in the outputs is noticeable, Fig. (15). Indeed, the identification algorithm is remarkably successfully in fitting an accurate noise model to the I/O data despite the adverse S/N conditions. The fitted models were:-

NARMAX Model **S/N**=∞, $\sum \mathrm{ERR}_i = 99.99\%$

$$y_t = 0.720y_{t-1} + 0.039u_{t-1}^3 + 0.172u_{t-1}^2 u_t + 0.065u_{t-2}^2 u_{t-1}$$

NARMAX Model **S/N**$\approx$1, $\sum \mathrm{ERR}_i = 46.14\%$

$$\begin{aligned} y_t = {} & 0.298u_{t-2}^2 + 0.100y_{t-1} + 1.230u_t + 0.017y_{t-2}^2 - 0.173u_{t-2}^3 - 0.007y_{t-2}^2 u_{t-1} - 0.001y_{t-2} u_t^2 \\ & + 0.216u_t^2 - 0.742u_{t-1}^2 + 0.562u_{t-2} u_{t-1}^2 + 0.014u_{t-2} y_{t-1} - 0.001y_{t-2} y_{t-1}^2 \\ & - 0.133u_{t-1}^2 y_{t-1} + 0.016y_{t-2} y_{t-1} + 0.115u_{t-2}^2 y_{t-1} + 1.420u_{t-2} - 1.420u_{t-1} \\ & - 0.137y_{t-1} u_t + 0.048y_{t-1} u_t^2 + 0.178u_{t-1} y_{t-1} - 0.029u_{t-2} y_{t-2} + 0.028y_{t-2} u_{t-1}^2 + 0.001y_{t-1}^3 \\ & - 0.018u_{t-2} y_{t-1}^2 + 0.008u_{t-1} y_{t-1}^2 + 0.046u_t^3 - 0.091e_{t-2} + 0.038e_{t-3} - 0.130e_{t-1} \end{aligned}$$

NARMAX Model **S/N**$\approx$3, $\sum \mathrm{ERR}_i = 93.41\%$

$$\begin{aligned} y_t = {} & +0.2535e+00y_{t-1} + 0.2524e+00u_{t-1}^3 + 0.3154e+00y_{t-2} + 0.8977e-01u_{t-2}^3 + 0.2304e+00u_t^2 \\ & + 0.9628e-02y_{t-2}^2 - 0.6897e-03y_{t-2}^3 + 0.2613e+00u_{t-2}^2 - 0.4792e+00u_{t-1}^2 + 0.1815e+00u_{t-2} u_t \\ & - 0.6192e-01u_{t-2} y_{t-2} + 0.4616e-02u_{t-2} y_{t-2}^2 - 0.9311e-02u_{t-2} - 0.3338e+00e_{t-2} - 0.1815e+00e_{t-1} \end{aligned}$$

NARMAX Model **S/N**$\approx$6, $\sum \mathrm{ERR}_i = 98.86\%$

$$y_t = +0.089y_{t-1} + 0.007u_{t-1}^3 + 0.467y_{t-2} + 0.108u_{t-2}^3 + 0.201u_{t-1}^2 u_t + 0.121u_{t-2}^2 u_{t-1} - 0.479e_{t-2}$$

Based on these results it would be tempting to conclude that all these models provide adequate prediction over the range of S/N environments considered. However, inspection of the model structures above highlights a noticeable difference in the number of parameters included in each. This is due to the combined effect of the output noise and the thresholds set for including terms in the model, (1.16). The later being set to the same value in each case to show the influence of S/N

ratio and emphasize the danger of over parameterising the model. Intuitively the percentage reduction in the mean square error in (1.16), for any one term in the auxiliary model, (1.14), will be significantly altered by noise in the measurement and otherwise insignificant terms can be wrongly included in the final model. In the worst case, here when S/N≈1, a model with 29 terms has been identified. It is for these reasons that intelligent structure detection has been introduced [Billings et al, 1989].

3.2.3. Cell Map Analysis

Analysis of these models using the Cell Map framework is achieved simply by suspending each model turn in the cell state space defining such that $x_1 \in [-125, 125]$, $x_2 \in [-125, 125]$ and $\mu \in [-5, 5]$ over a mesh of 125×125×50 cells. Applying the Unravelling algorithm produces the cell diagrams shown in Fig. (16a-c). In each diagram the shaded region denotes the **DOA** of all the **P-K** cell cycles detected, that is the entire stable domain within **Z**.

For S/N≈1 a very poor stability characteristic exhibited, the models stable **DOA** covers only a small fraction of the region of interest considered, Fig. (16a). However, the stability of the identified model is seen to improve as the S/N increases. For S/N≈3 the nonlinear discrete model is stable over the same range of input used to generate the I/O data, $u \in [-4, 4]$. But additional induced bifurcations are present outside this parameter region, the model still being only locally valid, Fig. (16b). For S/N ratios greater than this the model is stable over the entire input parameter range, see for example Fig. (16c), where S/N≈6. The Cell Map characteristic is then qualitatively the same as that of the original system. Additional bifurcations introduced, by the modelling process, are no longer evident and the model is globally valid and indeed globally stable.

Note also how an associated reduction in model complexity accompanies the improvement in model validity, demonstrating the need for intelligent structure detection [Billings et al, 1989]. For lower values of S/N ratio the qualitative information provided by the analysis above proves to be very useful in accessing both model validity and model stability. In addition the analysis clearly shows the folly of fitting an overly complex nonlinear parametric model.

3.3. Conclusions

In this work we have attempted to illustrate the importance of model validation from the qualitative viewpoint. In the process of constructing a nonlinear model a number of choices have to be made. Broadly these can be viewed as deciding what generic model type is to be adopted, for example NARMAX, experiment design for data collection and choosing a mechanism for structure selection. In system identification studies the later depends to a large extent on what statistical methods are used.

It is obviously important to validate the model both from the statistical and behavioural viewpoints. This means checking not only that the modelling residuals are white, indicating that all predictable dynamics has been captured. But that also the model exhibits the expected qualitative characteristics.

The approach introduced in this paper makes use of an iterative framework of experiment design, system identification and model validation in order to achieve this. Use is made of the parameterised cell state space in order to provide both local and global information on a wide variety of different model types and at the same time provide an iterative aspect to the analysis.

In the process of achieving this two important aspects of the modelling process have been addressed. The first concerns input design and illustrates the importance of choosing the correct input if locally valid models are to be avoided. The second illustrates the dramatic effect low S/N ratios can have on the model and its stability. Here a trade off exists between model complexity, S/N ratio and qualitative model validity. The later being measured in this study mainly in terms of the size of the stability domain, or **DOA**, and to a lesser extent by the presence of induced bifurcations. Most of all this work demonstrates the need for feedback of both statistical and qualitative information in the sequence of events comprising input design, identification and cell map analysis.

Acknowledgment

The authors gratefully acknowledge that this work has been supported by the SERC.

References

ALPER P., 1964, Higer-Dimensional Z-Transforms and Non-linear Discrete Systems, *Revue A*, **VI**, 199-212.

BILLINGS S.A., 1985, An overview of nonlinear systems identification, in, *Proc. 7th IFAC Symp. Identification and System Parameter Estimation, York UK.*

BILLINGS S.A., I. LEONTARITIS, 1981, Identification of nonlinear systems using parameter estimation techniques, in, *Proc. IEE Conference on Control and its Applications, Warwick Univ.*, 183-87.

BILLINGS S.A., S. CHEN, 1989, Identification of nonlinear rational systems using a prediction error estimation algorithm, *Int. J. Sys. Sci.*, **20**, 467-94.

BILLINGS S.A., S. CHEN, M. KORENBERG, 1989, Identification of MIMO nonlinear systems using a forward regression orthoganal estimator, *Int. J. Control*, **49**, 2157-89.

BILLINGS S.A., S.Y. FAKHOURI, 1982., Identification of nonlinear systems containing linear dynamic and static nonlinear elements, *Automatica*, **18**, 15-26.

BILLINGS S.A., W.S.F. VOON, 1983, Structure detection and model validity tests in the identification of nonlinear systems, *Proc. IEE, Part D*, **130**, 193-9.

BILLINGS S.A., W.S.F. VOON, 1984, Least squares parameter estimation algorithms for nonlinear systems, *Int. J. Sys. Sci.*, **15**, 601-615.

BILLINGS S.A., W.S.F. VOON, 1986, A prediction error and stepwise regression estimation algorithm for nonlinear systems, *Int. J. Control*, **44**, 803-822.

BOX G.E.P., G.M. JENKINS, 1976, Time Series Analysis, (Holden-Day: Oakland, California).

CHEN S., S.A. BILLINGS, 1988a, Prediction-error estimation algorithm for nonlinear output-affine systems, *Int. J. Control*, **47**, 309-332.

CHEN S., S.A. BILLINGS, 1988b, Recursive maximum likelihood identification of a nonlinear output-affine model, *Int. J. Control*, **48**, 1605-29.

GEORGE D.A., 1959, Continuous Nonlinear Systems, Tech. Report: 355, (MIT research lab.: California).

GOODWIN G.C., R.L. PAYNE, 1975, Dynamic System Identification: Experiment Design and Data Analysis, (Prentice Hall: Englewood Cliffs).

HAYNES B.R., S.A. BILLINGS, 1991, A Method for the Global Analysis of Nonlinear Systems, *Int. J. Control*, (Accepted for publication).

HSU C.S., 1987, Cell-to-Cell Mapping: A method for the global analysis of nonlinear systems, *Appl. Math. Sci.*, **64**, (Springer Verlag: New York, Berlin).

HSU C.S., R.S. GUTTALU, 1980, An unravelling algorithm for the global analysis of dynamical systems. An application of cell-to-cell mappings, *Trans. ASME J. Appl. Mech.*, **47**, 940-48.

JAGEN N.C., D.C. REDDY, 1972, Evaluation of the response of Nonlinear sampled data systems using multidimensional Z-transforms, *Proc. IEE* , **119**, 1521-25.

KABOLA R., K. SPRINGER, 1982, Control, Identification and Input Optimisation, (Plenum Press: New York).

KORENBERG M., S.A. BILLINGS, Y.P. LIU, P.J. MCILROY, 1988, An orthogonal parameter estimation algorithm for nonlinear stochastic systems, *Int. J. Control.*, **48**, 193-210.

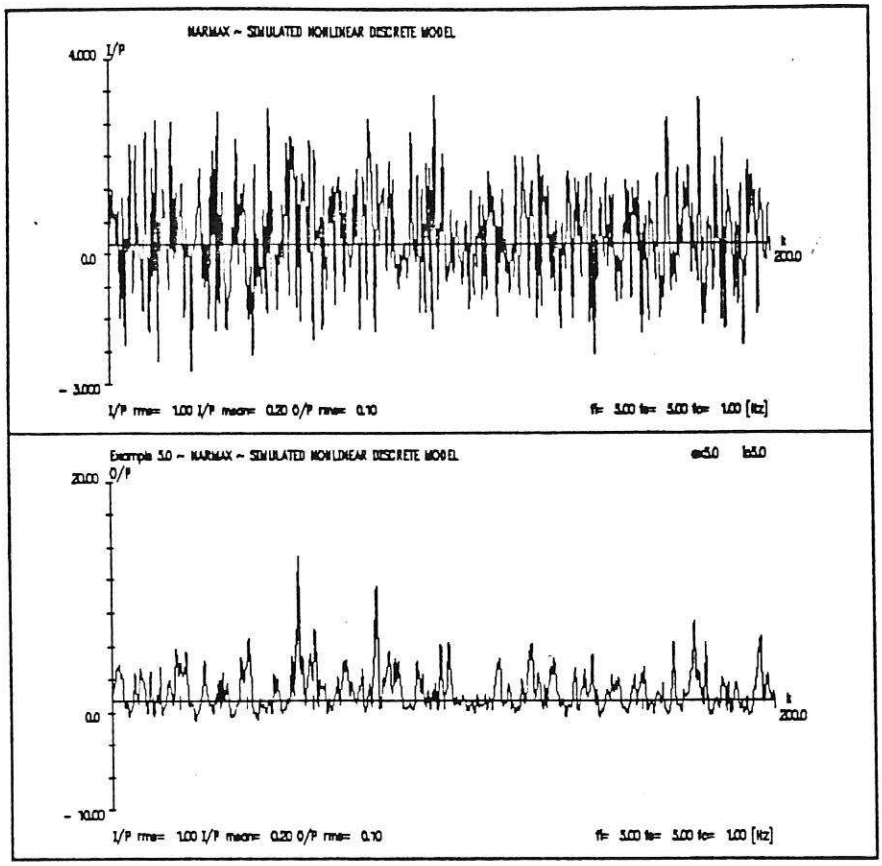
LEONTARITIS I., S.A. BILLINGS, 1985, Input-output parametric models for nonlinear systems, Parts I+II, *Int. J. Control*, **41**, 303-44.

LEONTARITIS I., S.A. BILLINGS, 1987, Experimental design and identifiability for nonlinear systems, *Int. J. Sys. Sci.*, **18**, 189-202.

LJUNG L., 1987, System Identification: Theory for the User, (Prentice Hall: Englewood Cliffs).

PADULO L., M. ARBIB, 1974, System Theory : A Unified State Space Approach, (Saunders: Philadelphia).

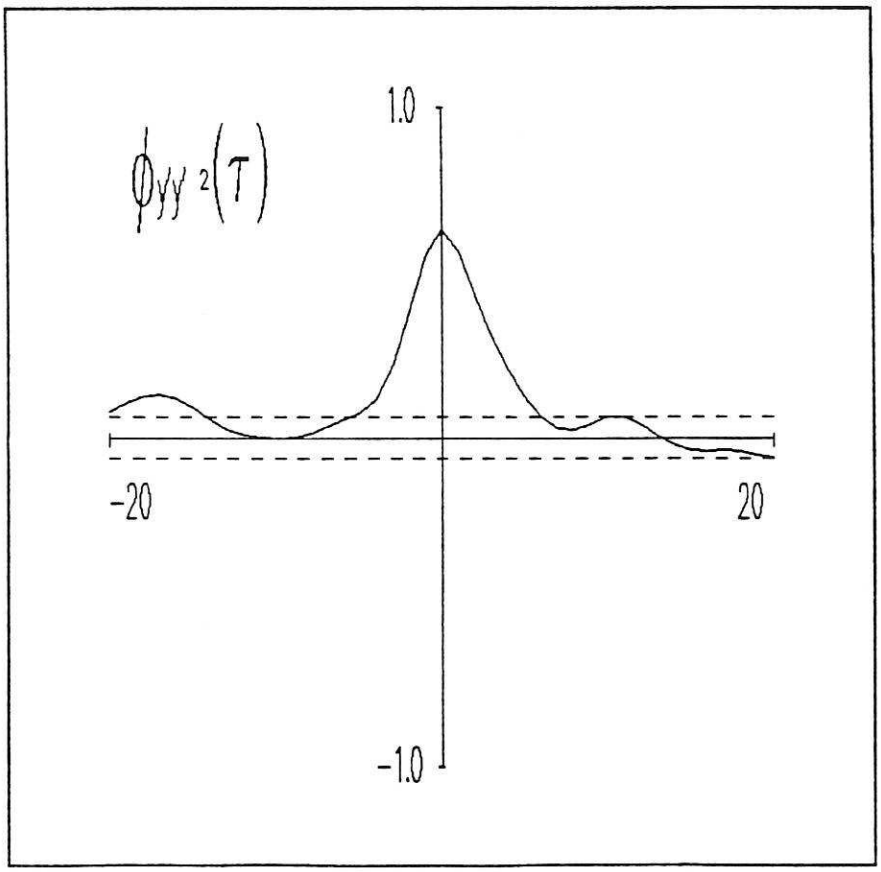
SCHETZEN M., 1980, The Volterra & Weiner Theories of Nonlinear Systems, (Wiley: New York).



Input

Output

Fig. (1a) - I/O data - Input $u(t) \in [\pm 1.0]$ with mean $\mu = 0.2$



Structure Detection

Fig. (1b) - Structure Detection Test

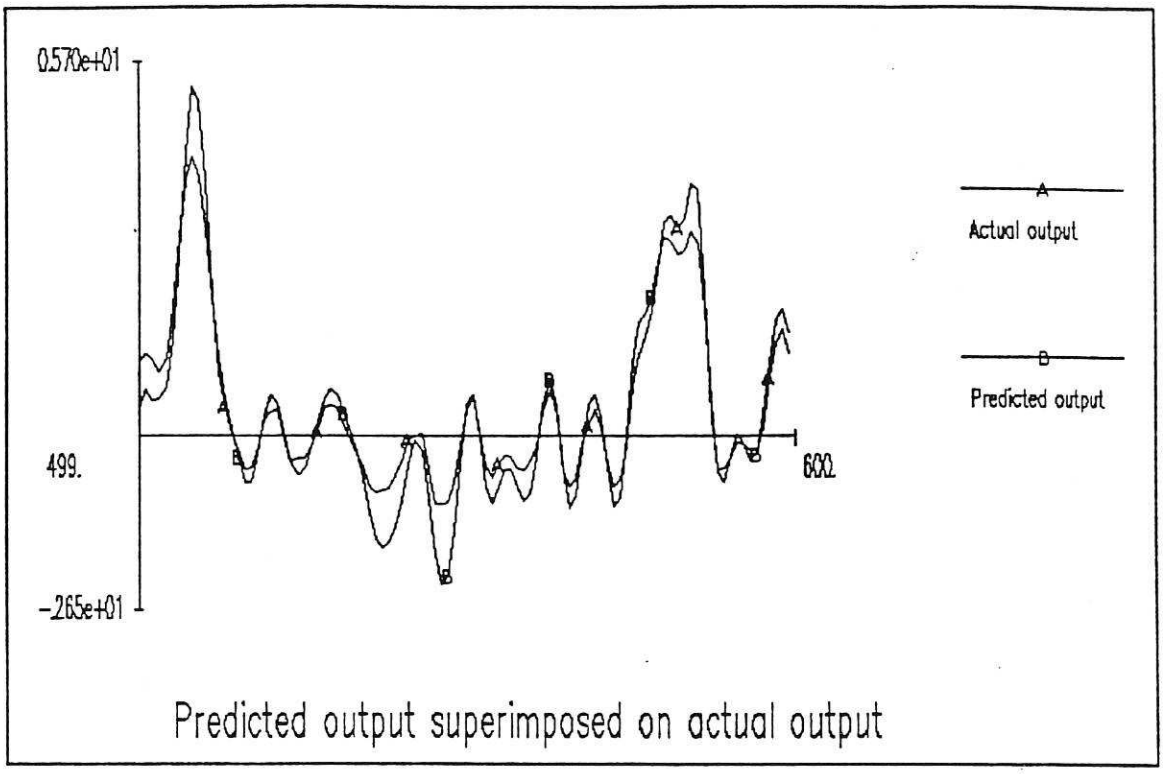
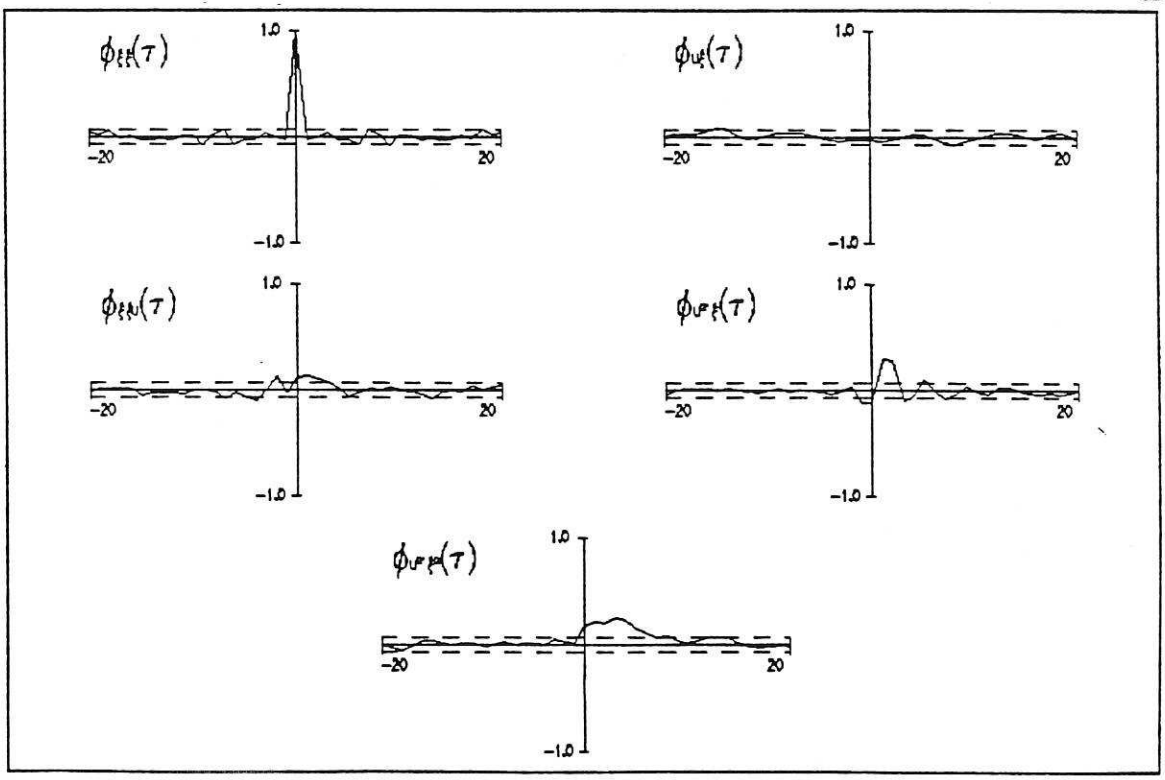


Fig. (2a) - Predicted + Measured Output - Linear Model



Linear and Nonlinear Model Validity Correlation Tests

Fig. (2b) - Model Validity Test - Linear model

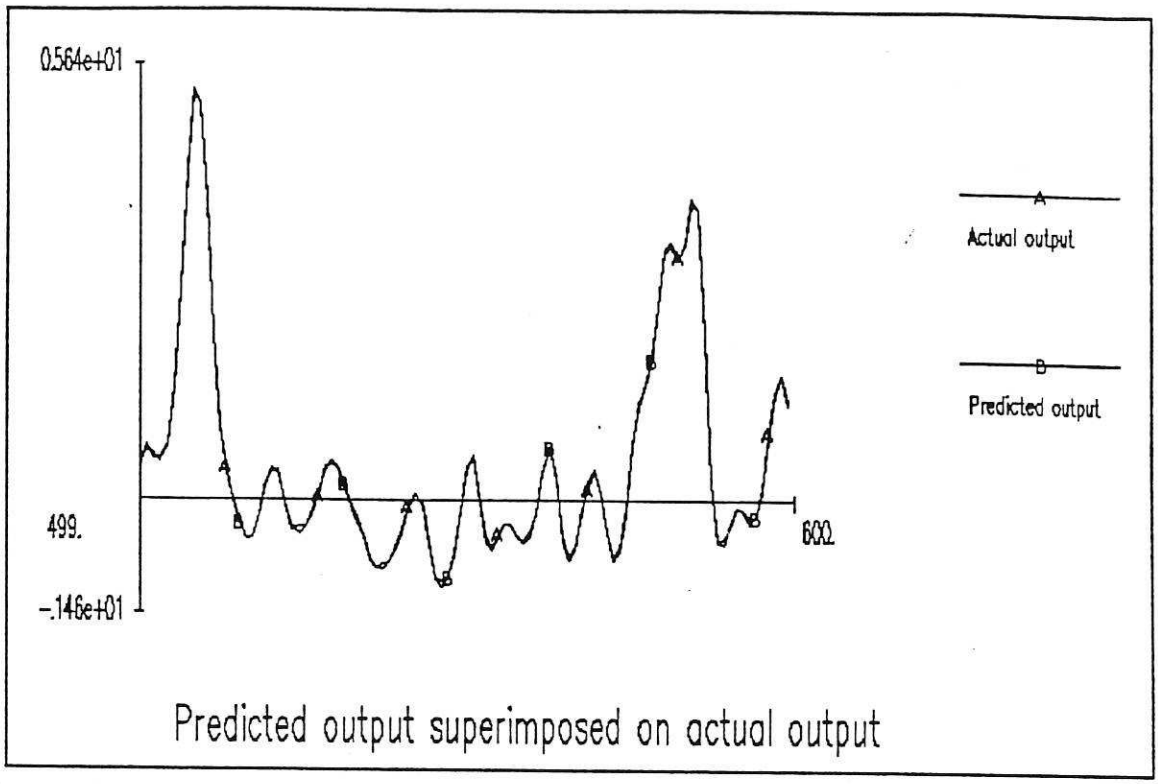
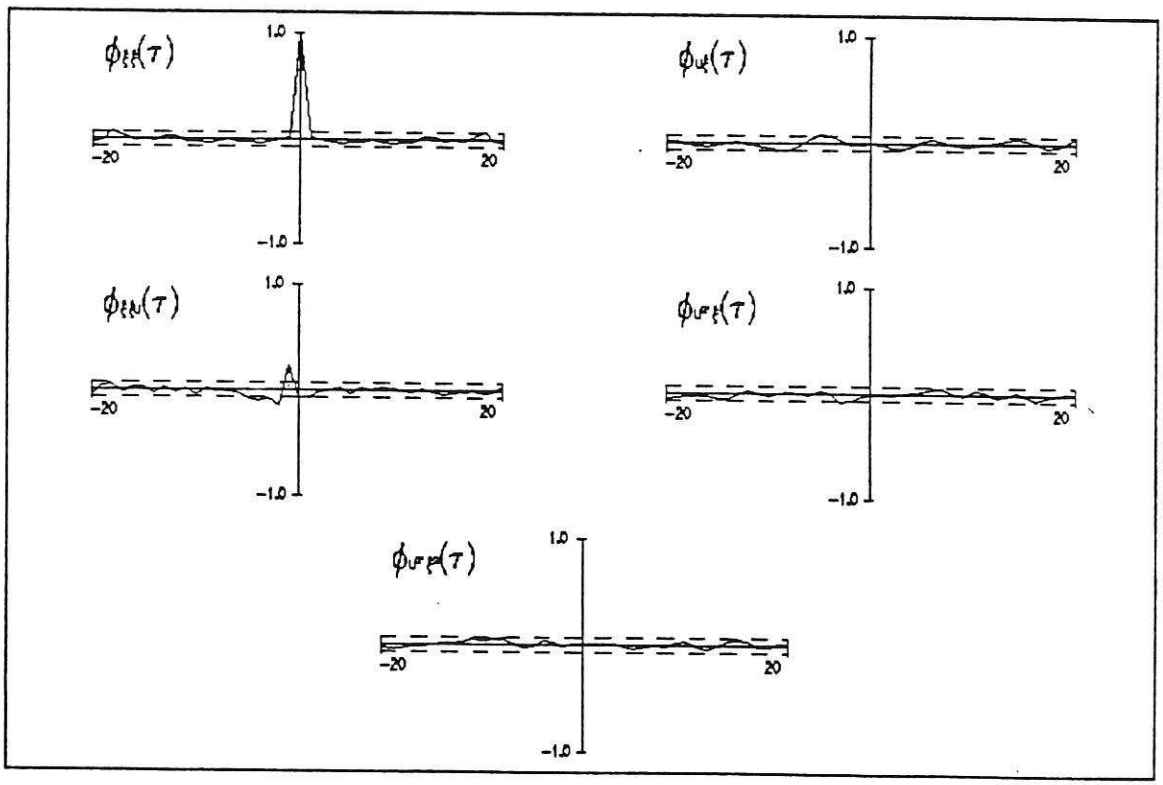


Fig. (3a) - Predicted + Measured Output



Linear and Nonlinear Model Validity Correlation Tests

Fig. (3b) - Model Validity Test - Nonlinear model

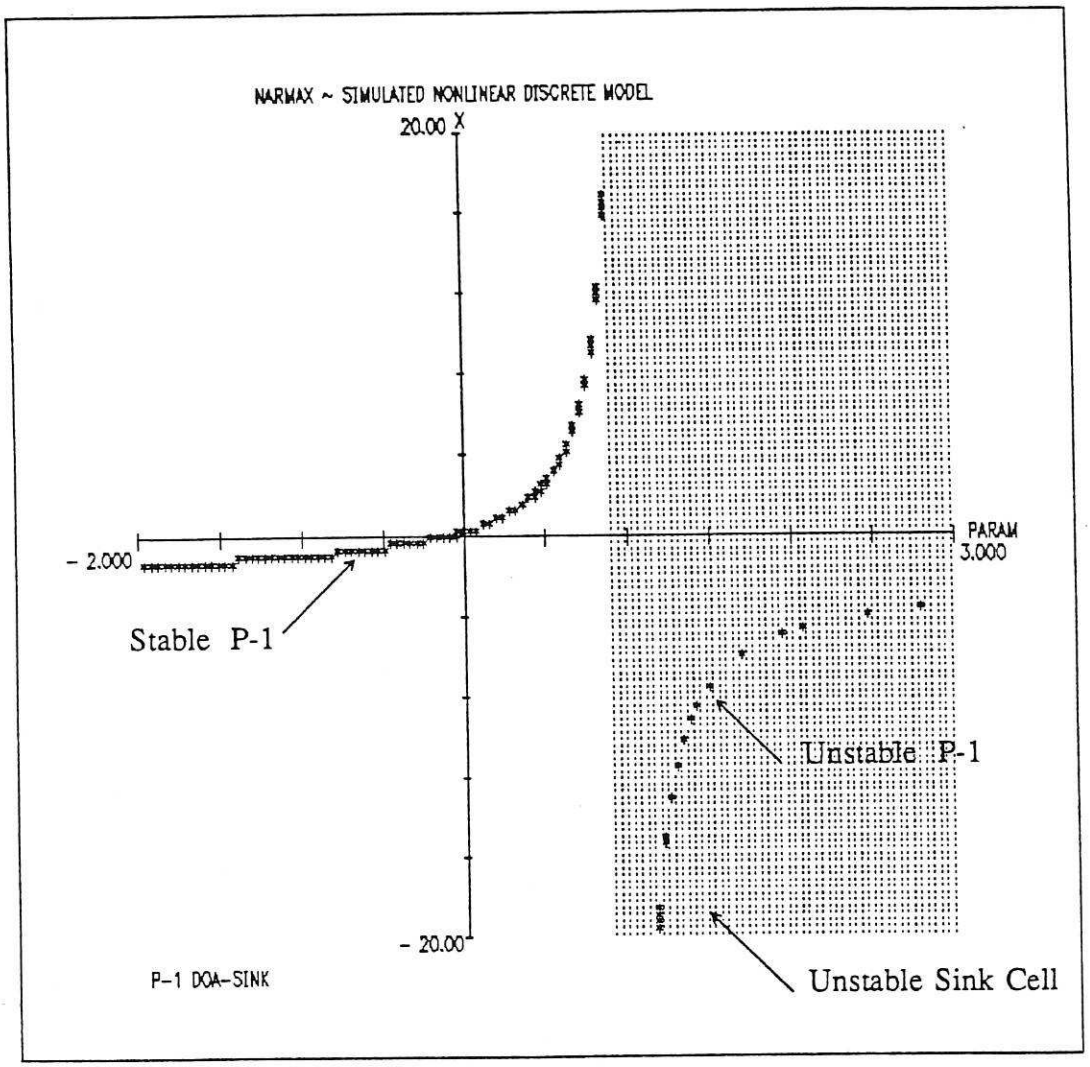


Fig. (4) - Cell Diagram - NARMAX Model

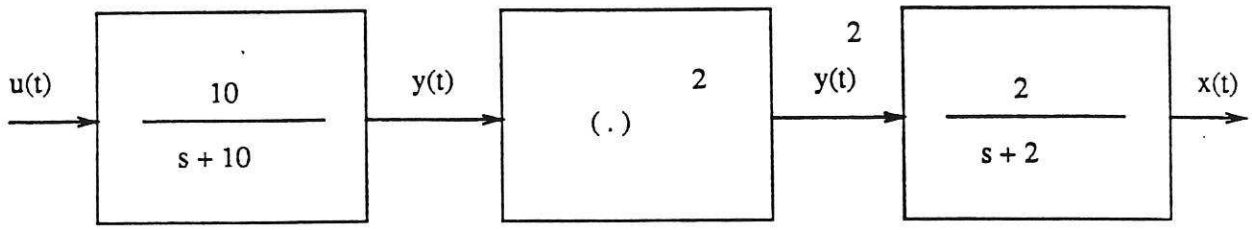


Fig. (5a) - Block Structured Quadratic System

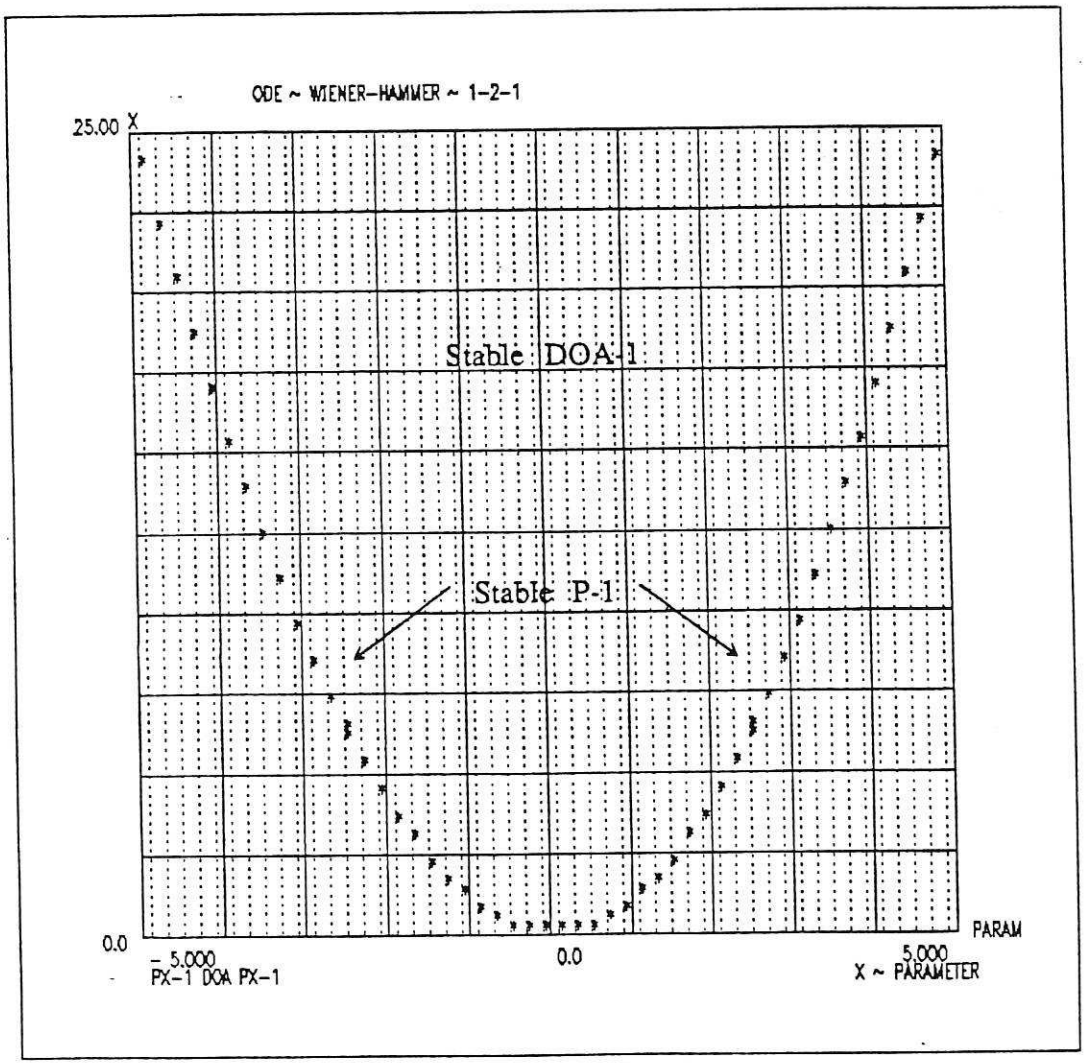
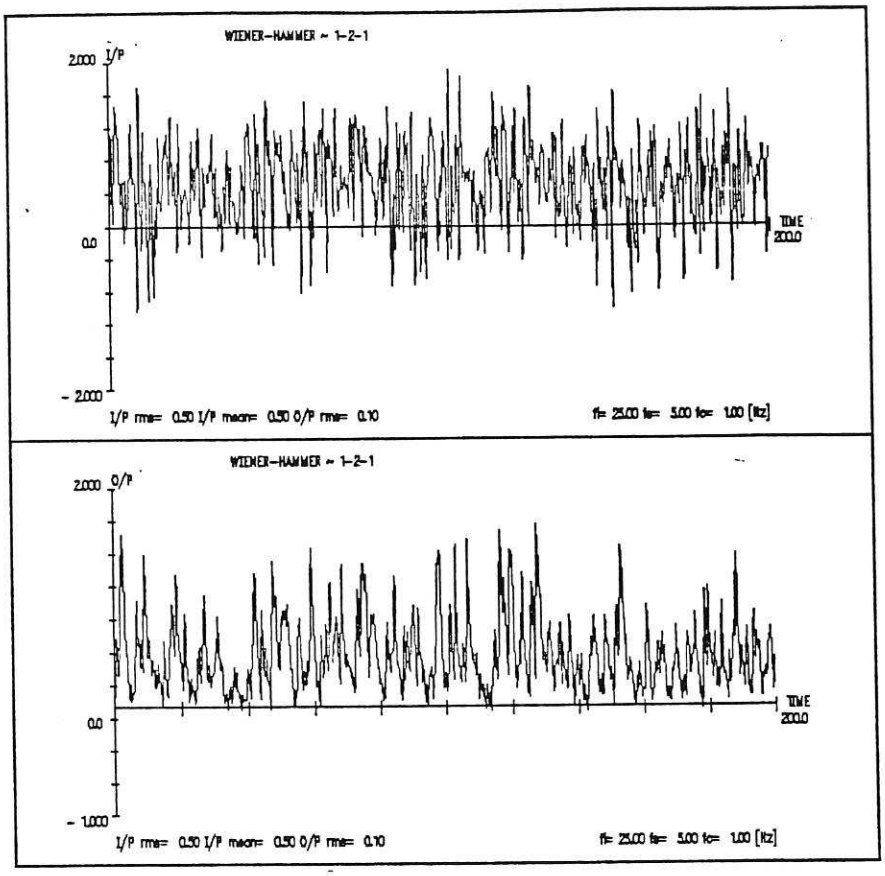


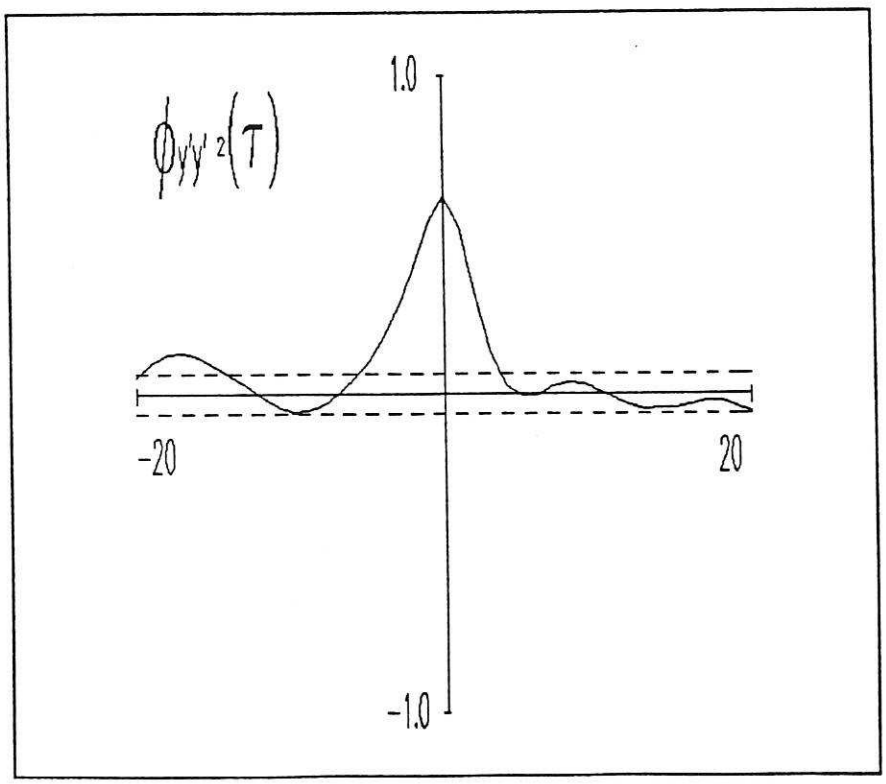
Fig. (5b) - Cell Diagram Block Structured Quadratic System



Input

Output

Fig. (6a) - I/O data - Input $u(t) \in [\pm 1.0]$ with mean $\mu = 0.5$



Structure Detection

Fig. (6b) - Structure Detection Test

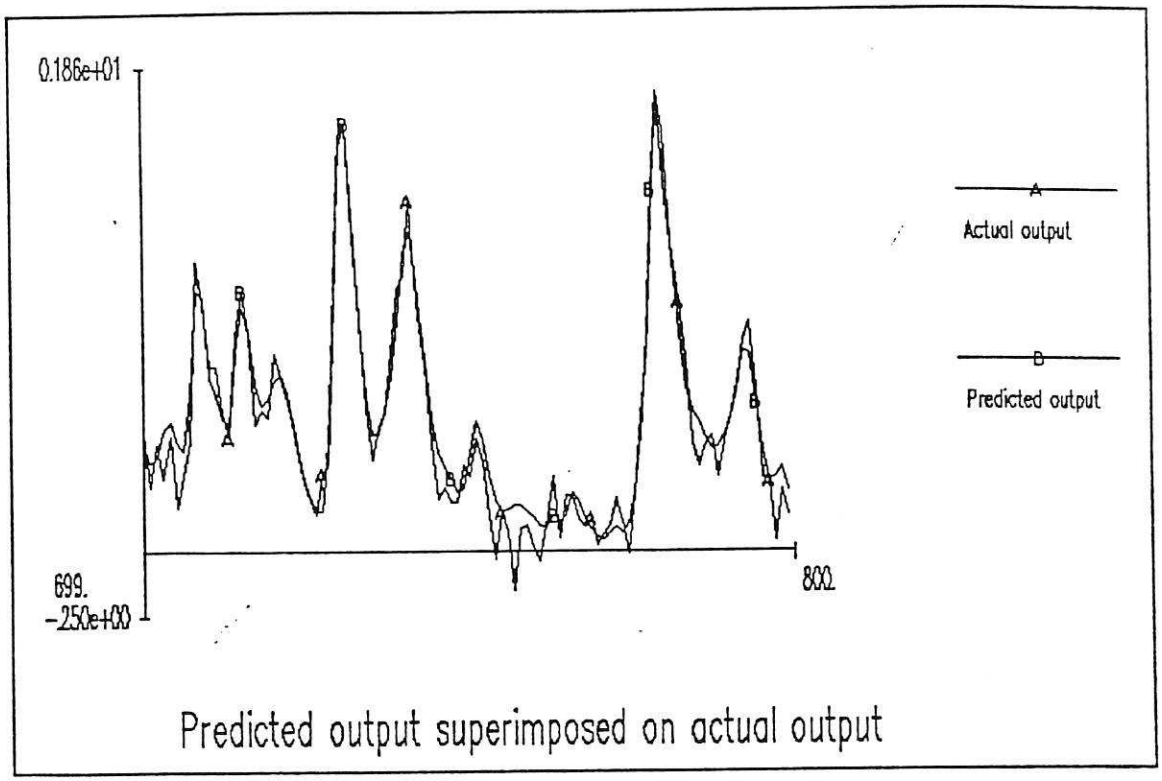


Fig. (7a) - Predicted + Measured Output - 1st Order Model

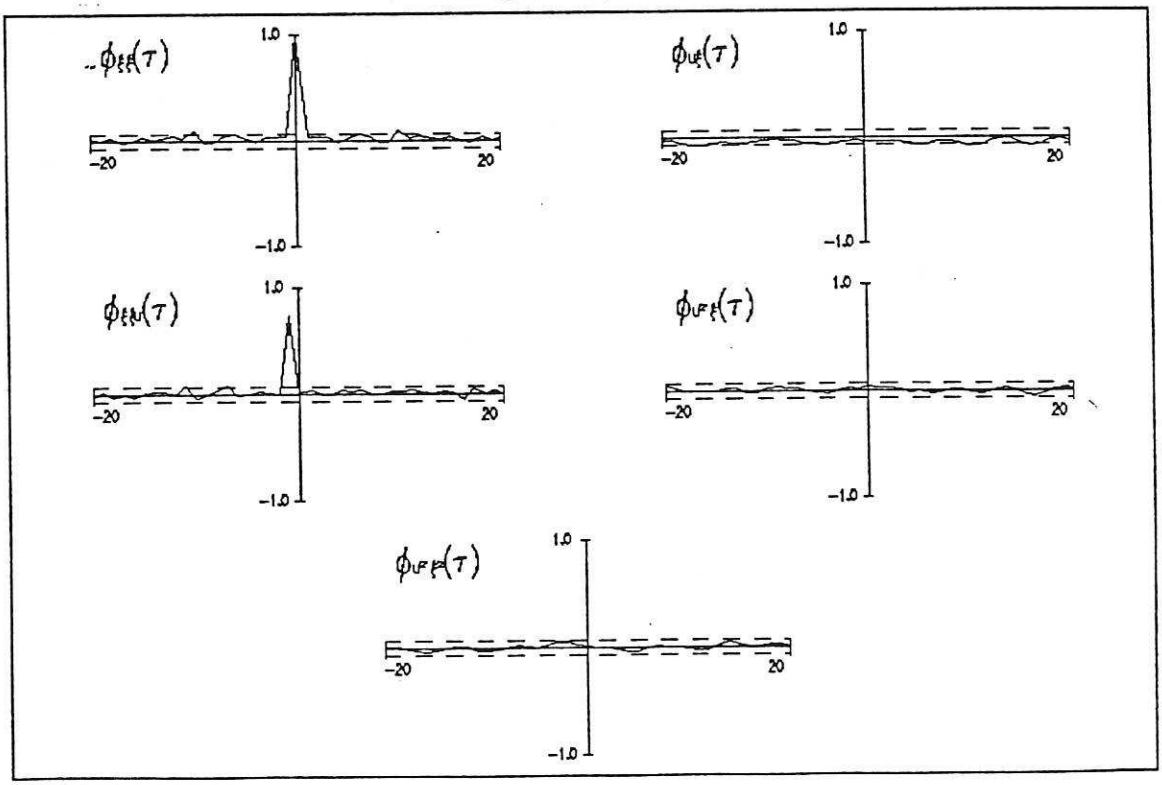


Fig. (7b) - Model Validity Test - 1st Order Model

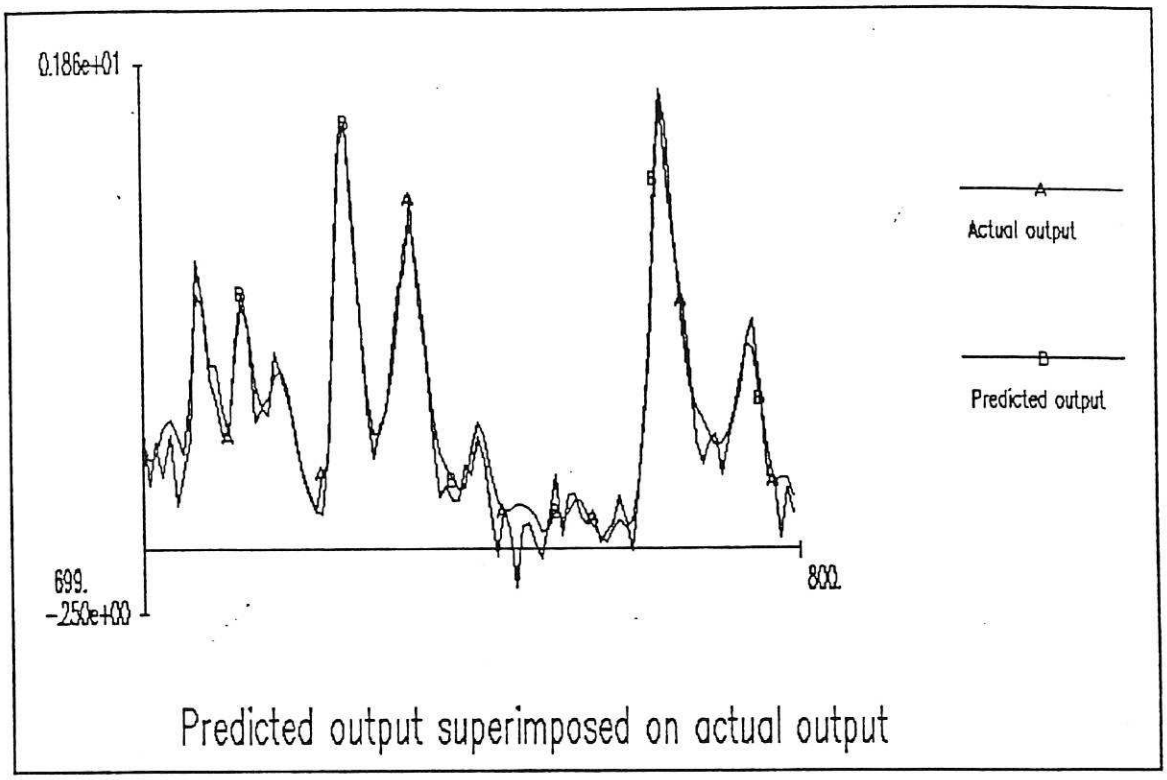
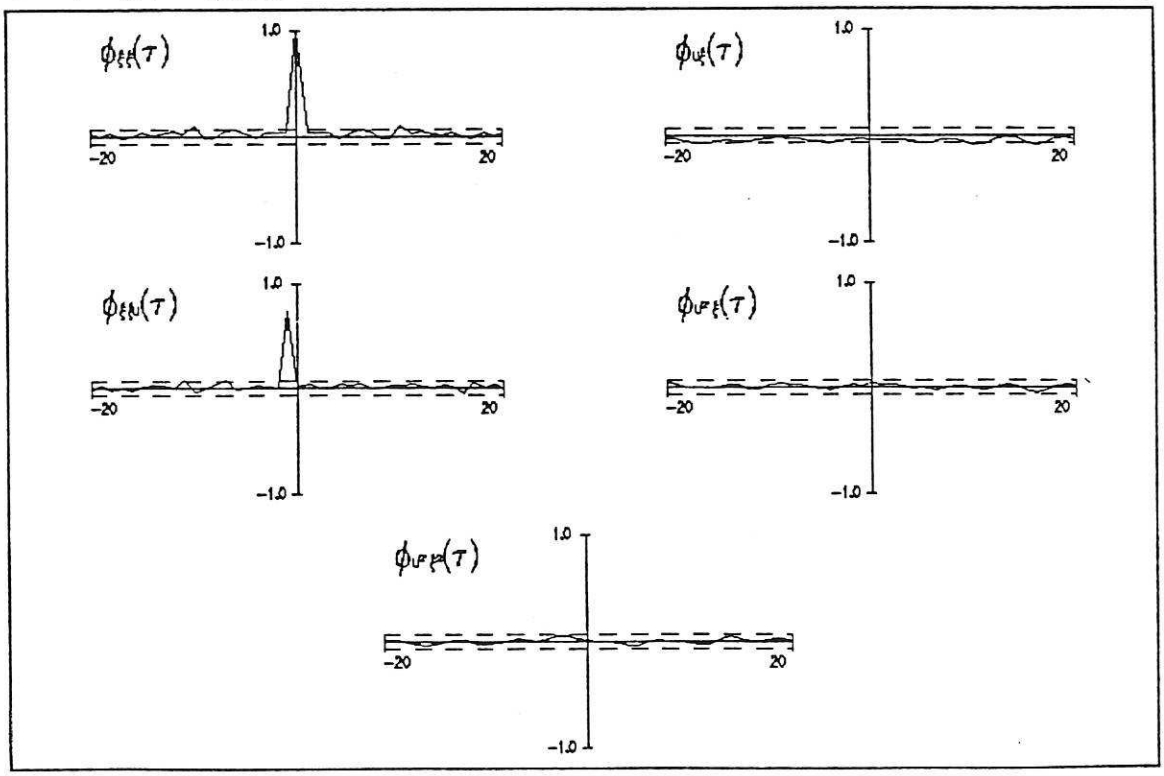


Fig. (8a) - Predicted + Measured Output - 2nd Order Model



Linear and Nonlinear Model Validity Correlation Tests

Fig. (8b) - Model Validity Test - 2nd Order Model

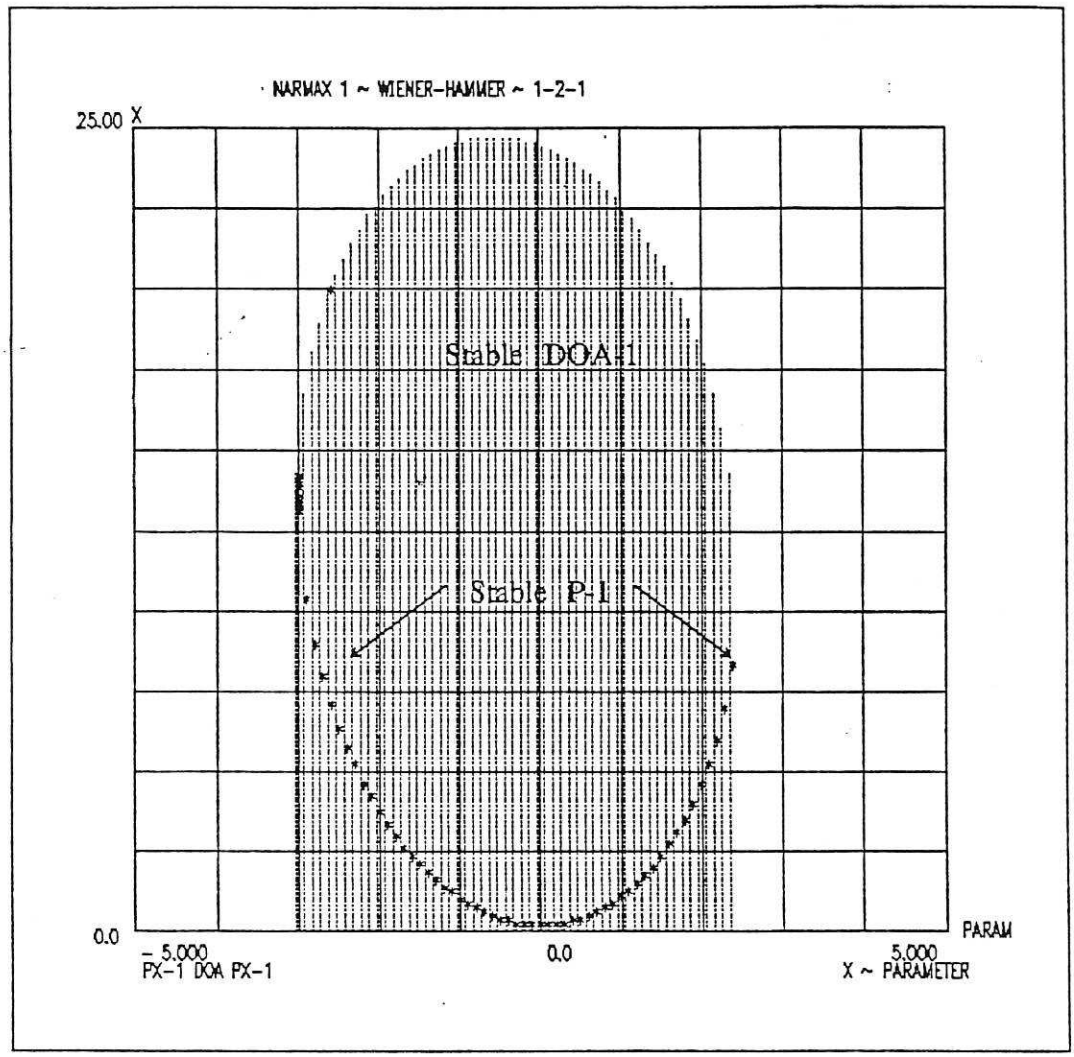


Fig. (9a) - Cell Diagram - 1st Order Model

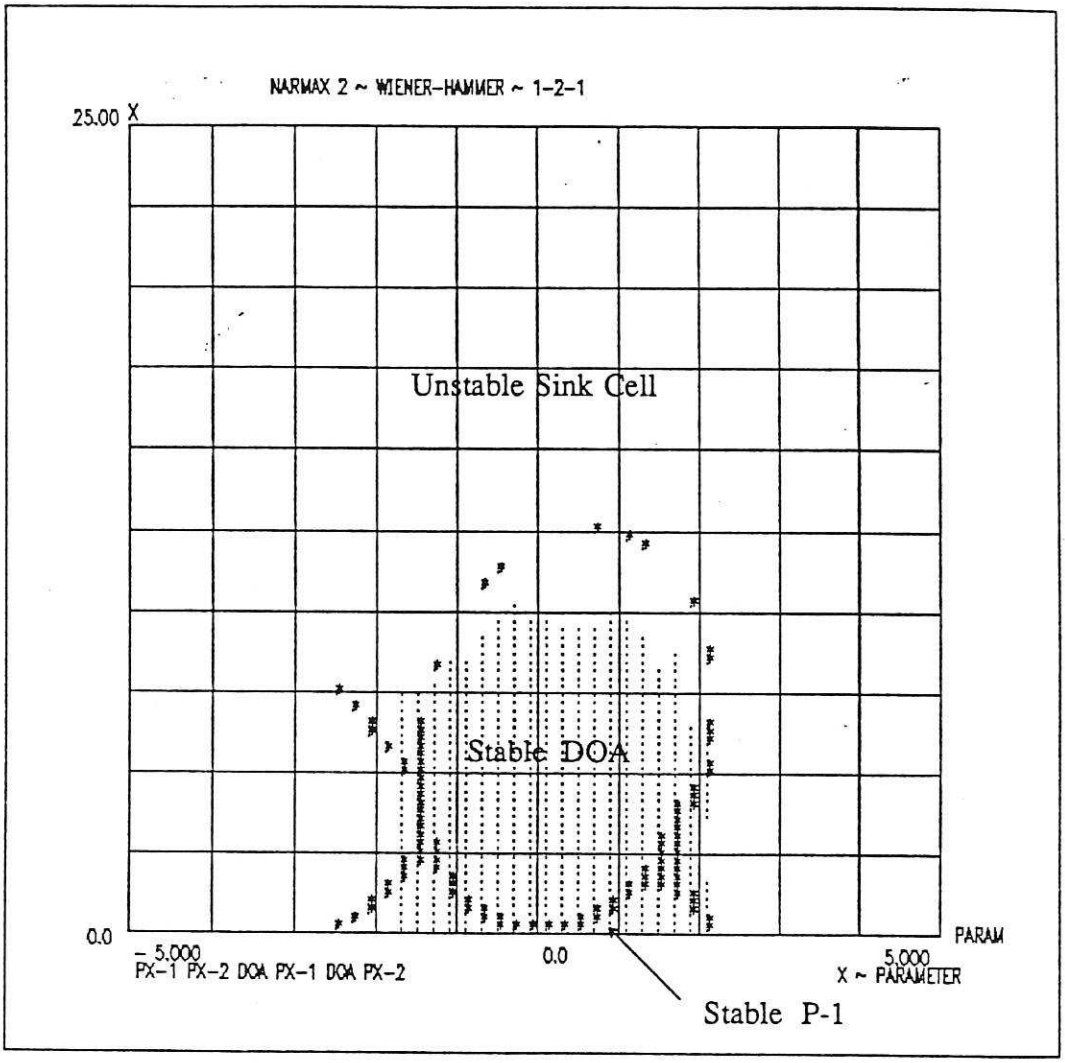
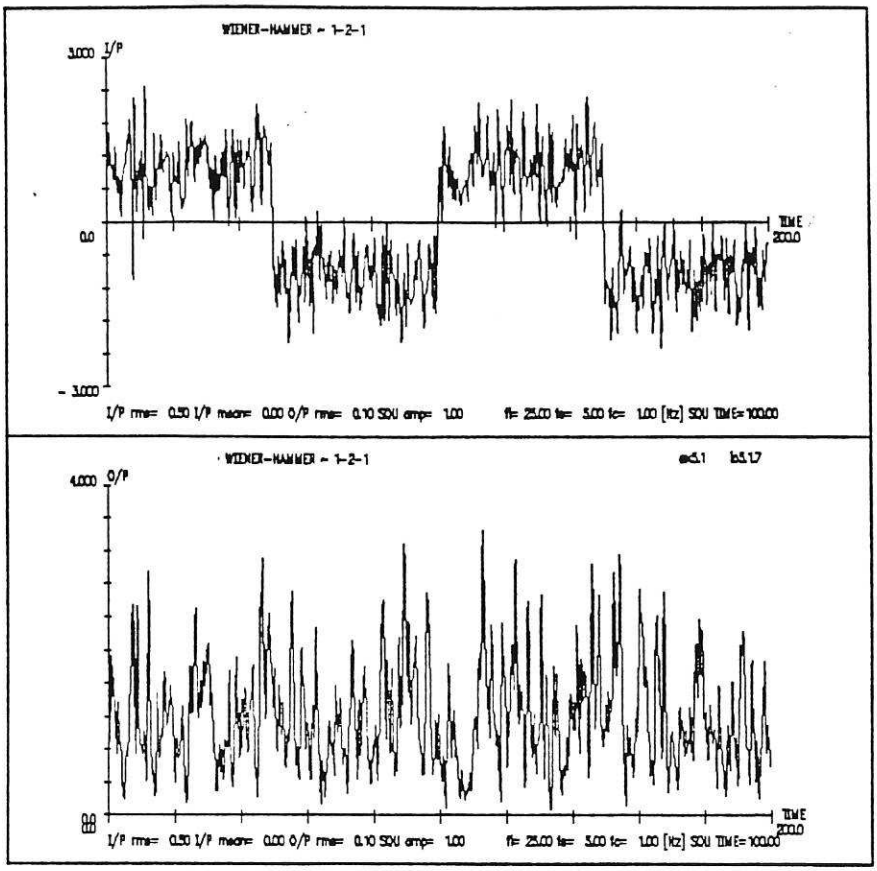


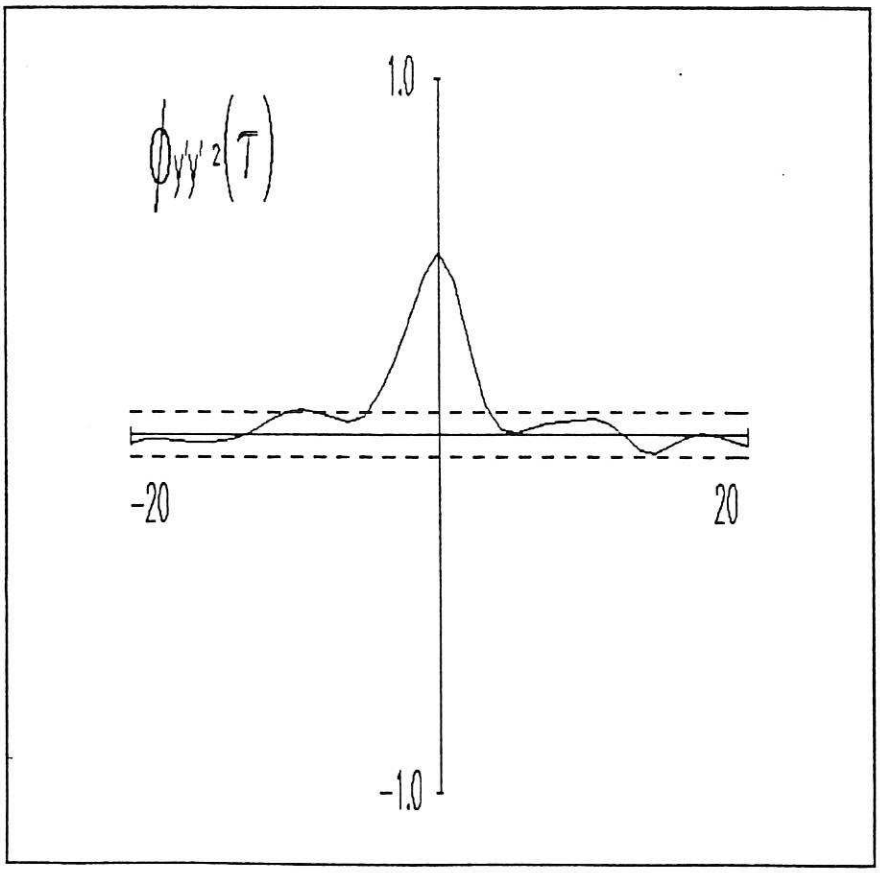
Fig. (9b) - Cell Diagram - 2nd Order Model



Input

Output

Fig. (10a) - I/O data - Input $u(t) \in [\pm 1.0]$ with mean $\mu = \pm 1.0$



Structure Detection

Fig. (10b) - Structure Detection Test

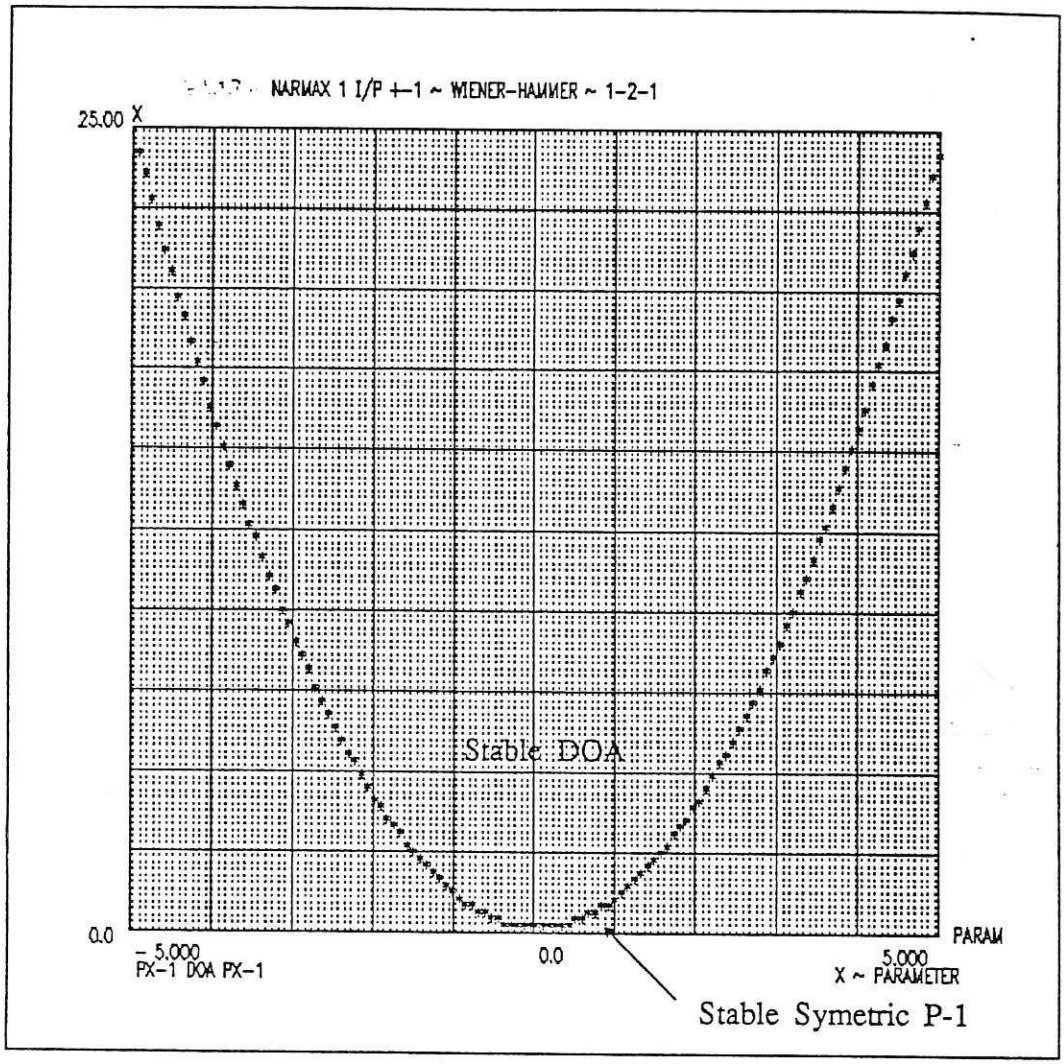


Fig. (11a) - Cell Diagrams - 1st Order Model

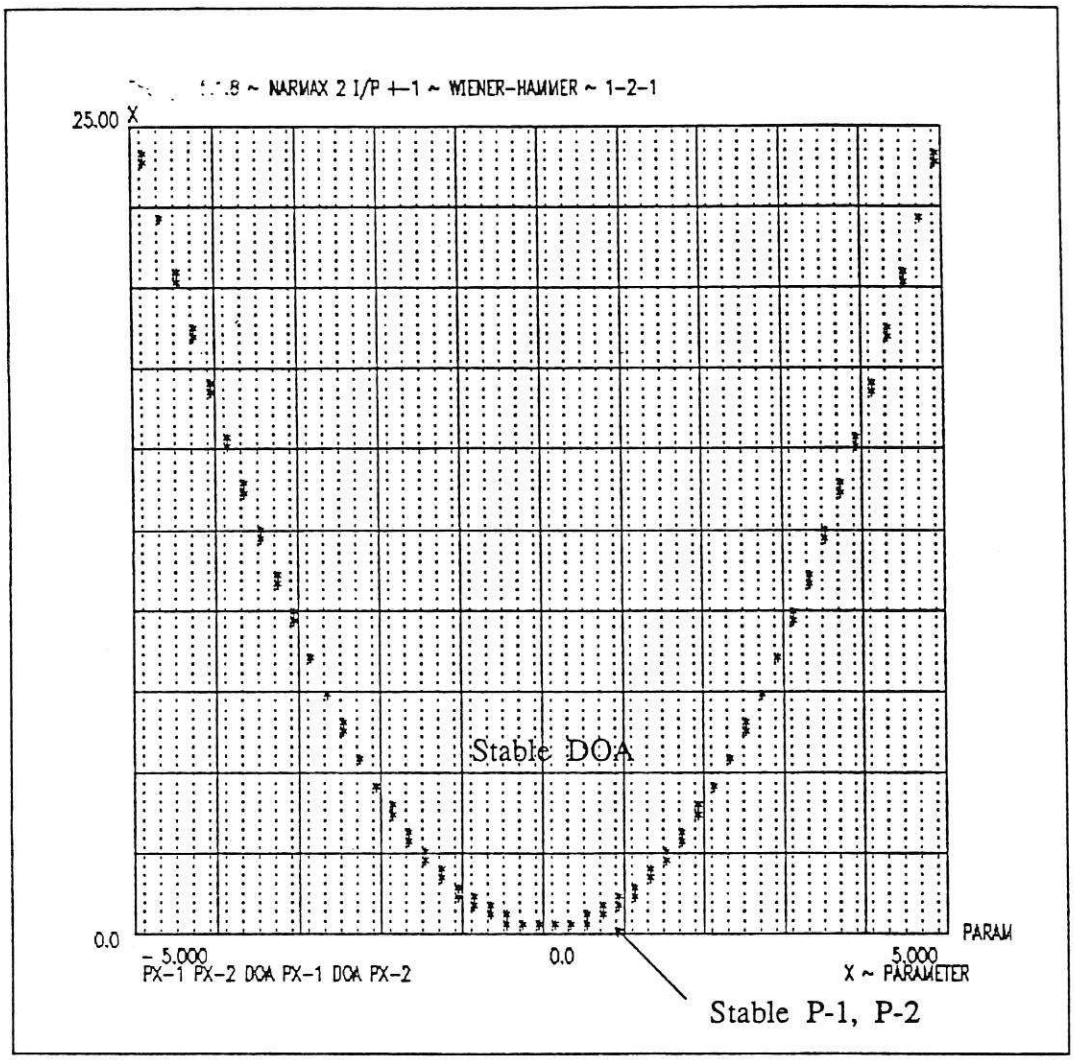


Fig. (11b) - Cell Diagrams - 2nd Order Model

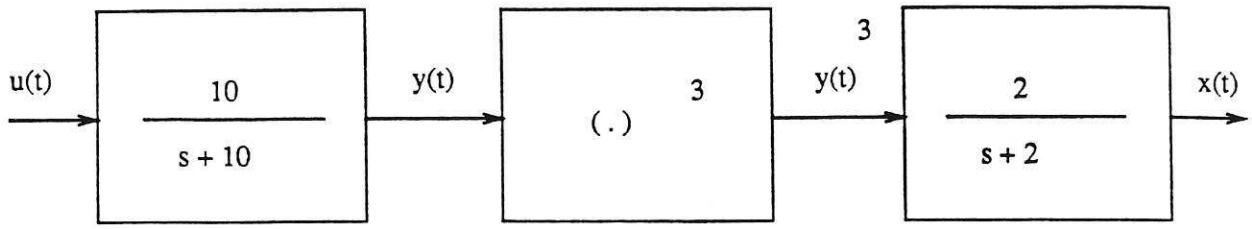


Fig. (12a) - Block Structured Cubic System

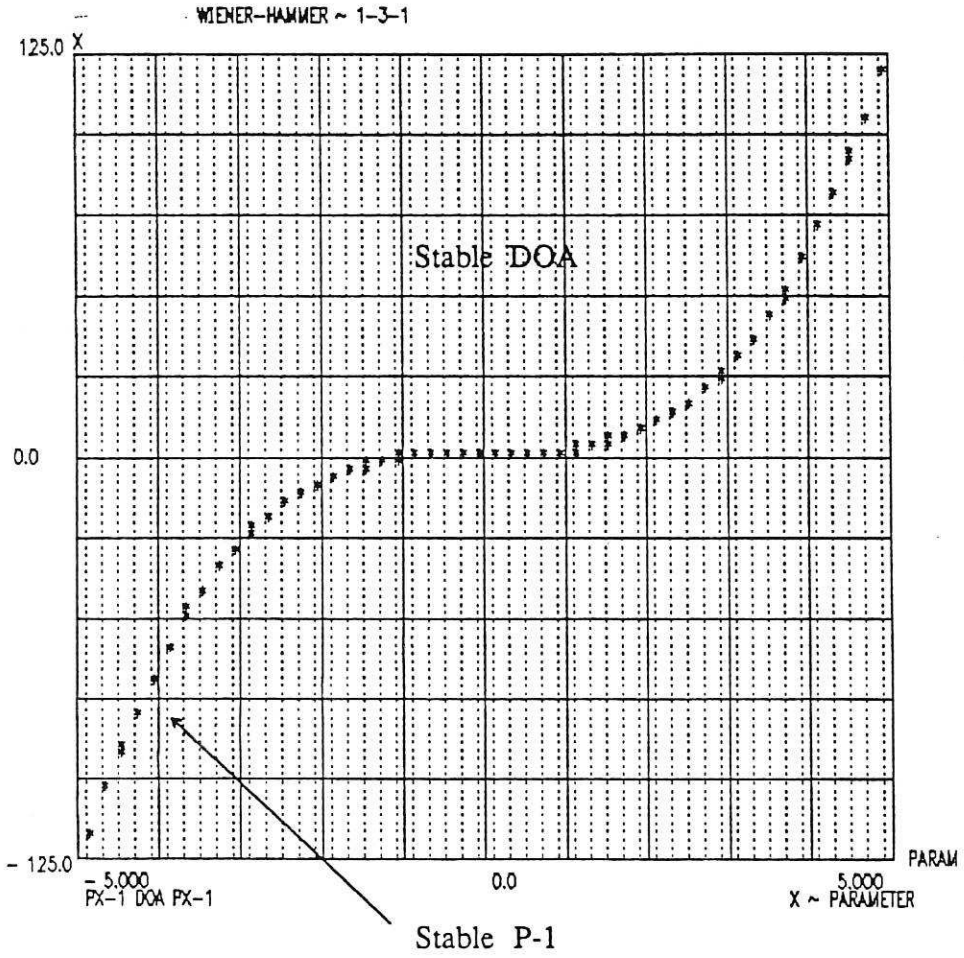
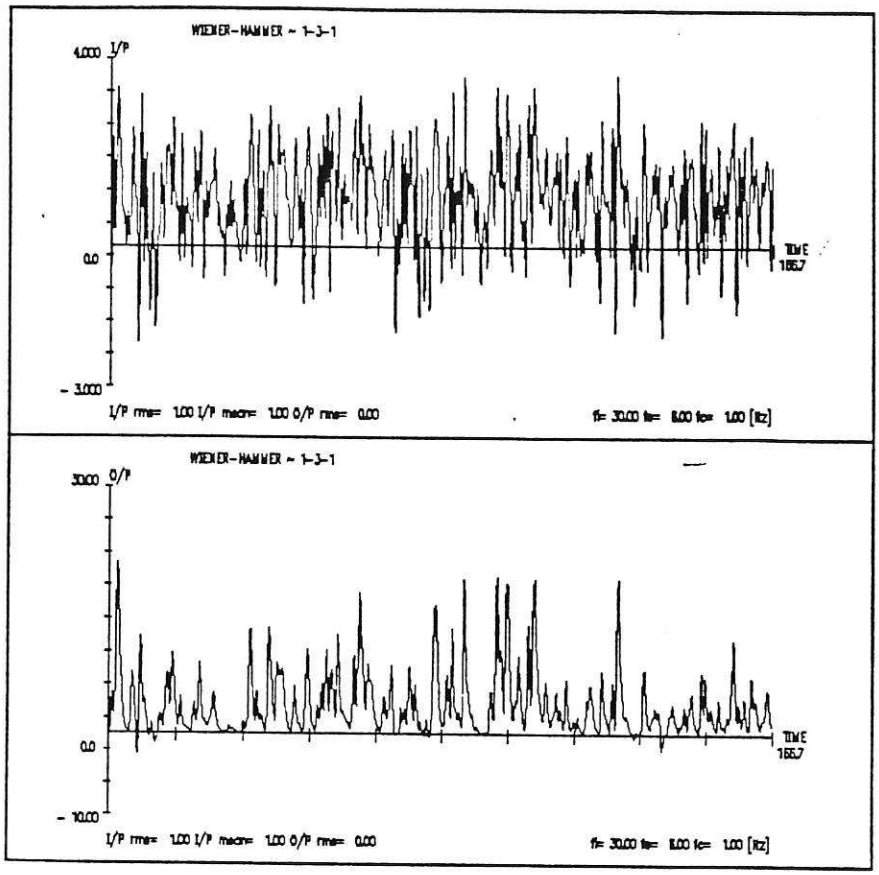


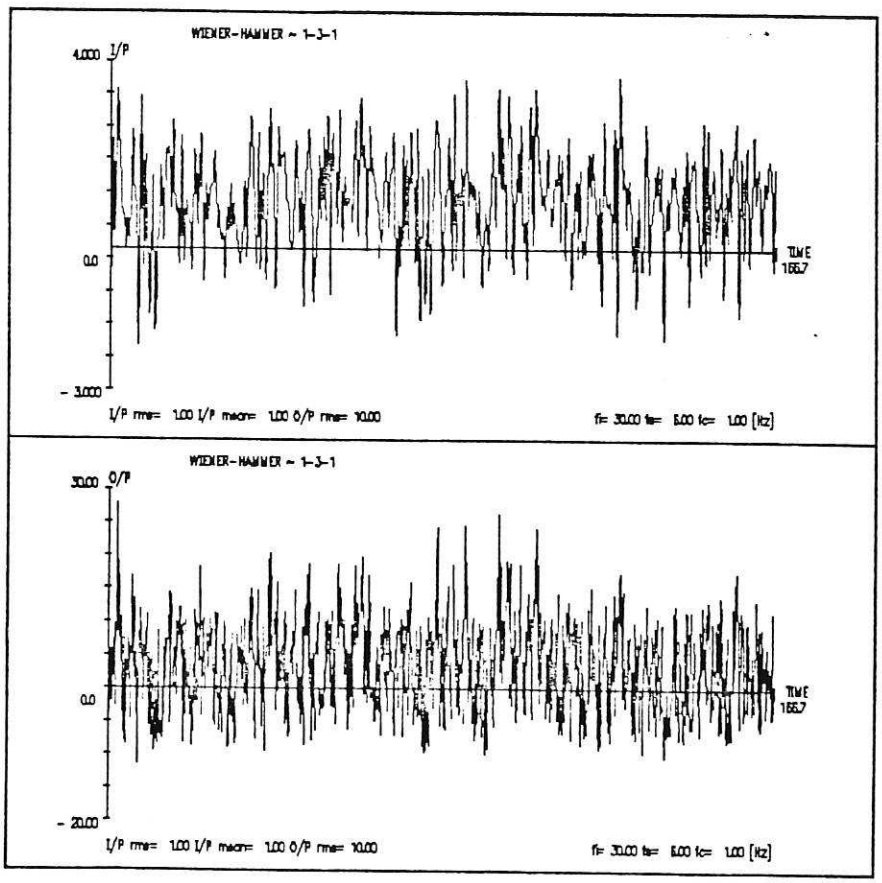
Fig. (12b) - Cell Diagram Block Structured Cubic System



Input

Output

Fig. (13a) - I/O data, Noise Free Data - S/N=∞



Input

Output

Fig. (13b) - I/O data, Noisy Corrupted Data - S/N≈1

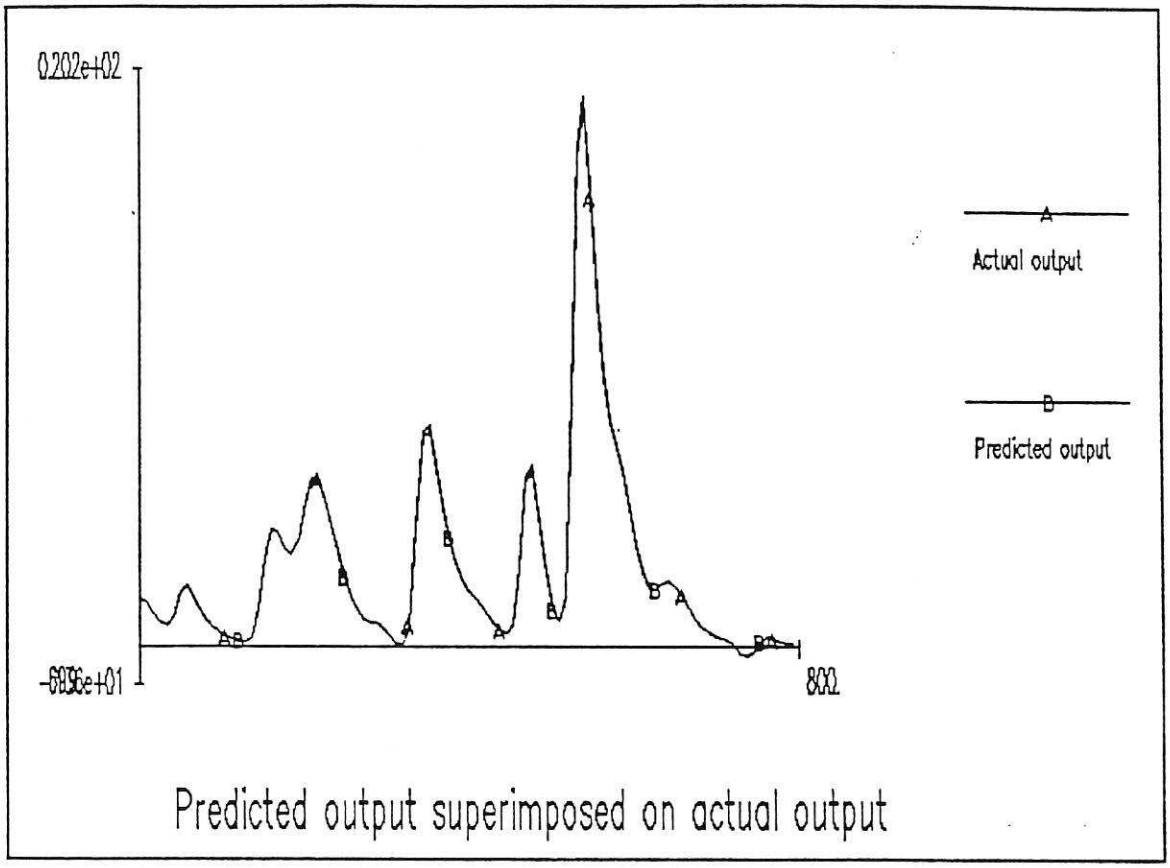


Fig. (14a) - Predicted + Measured Output - Noise Free S/N=∞

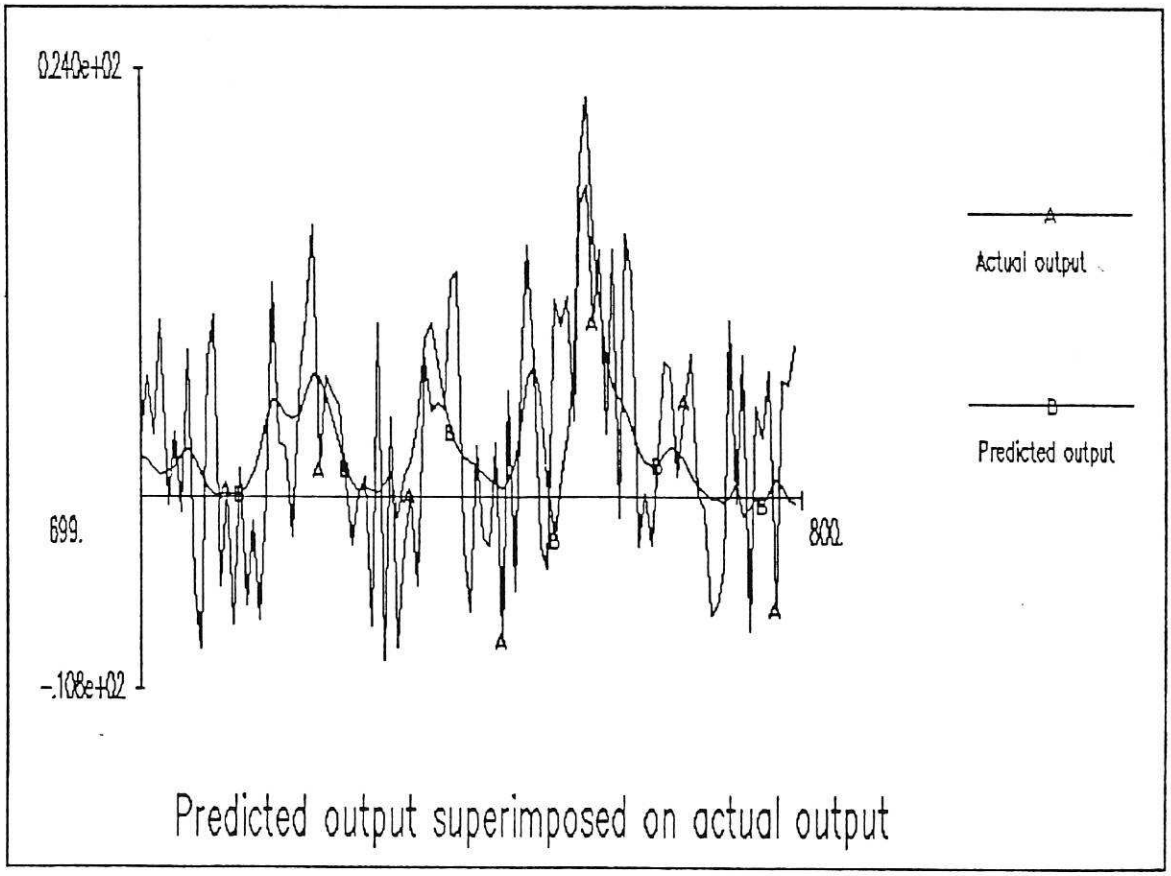


Fig. (14b) - Predicted + Measured Output - Noisy Data S/N≈1

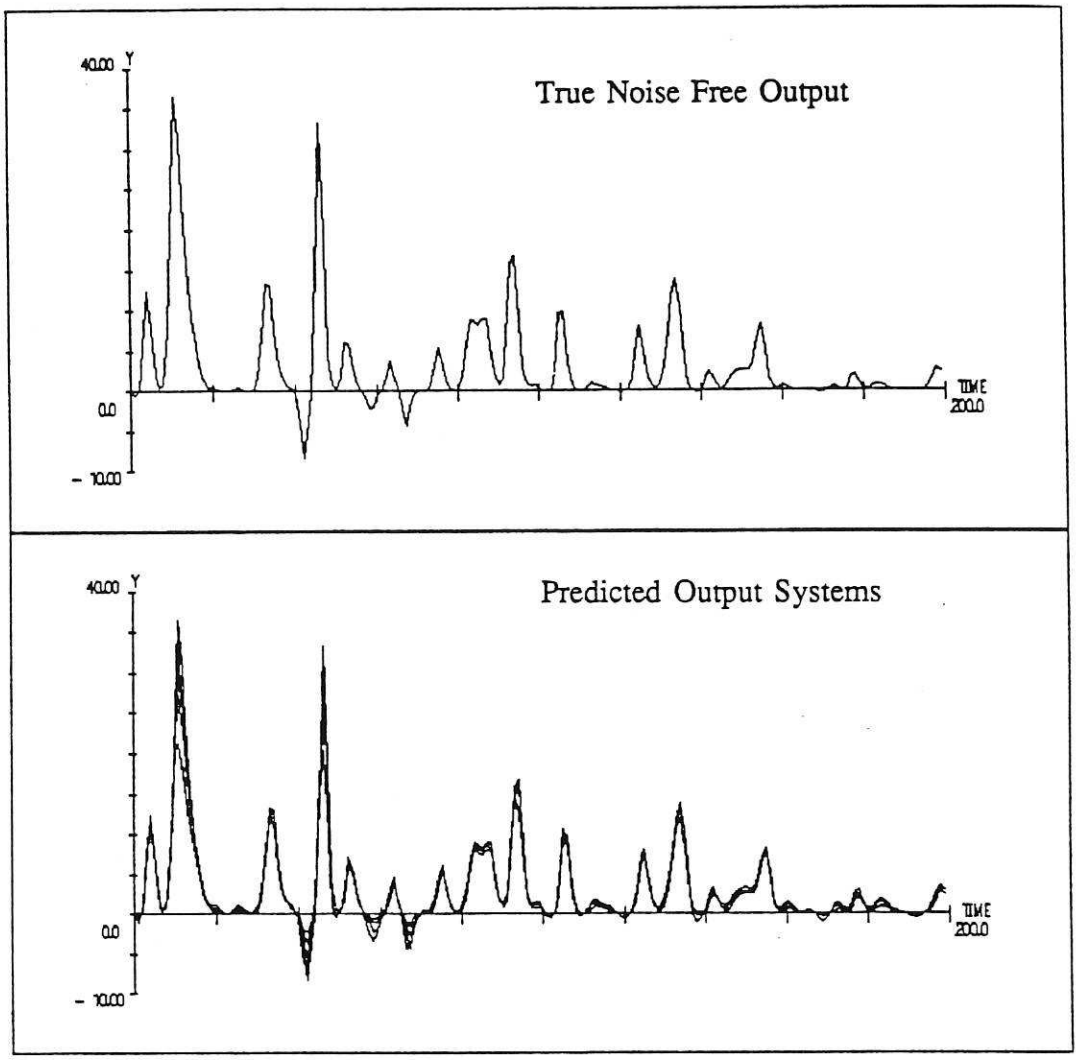


Fig. (15) - Noise Free Output $y(t)$- S/N=∞ + Noisy Predicted Output $\hat{y}(t)$ - S/N=1, 3, 6

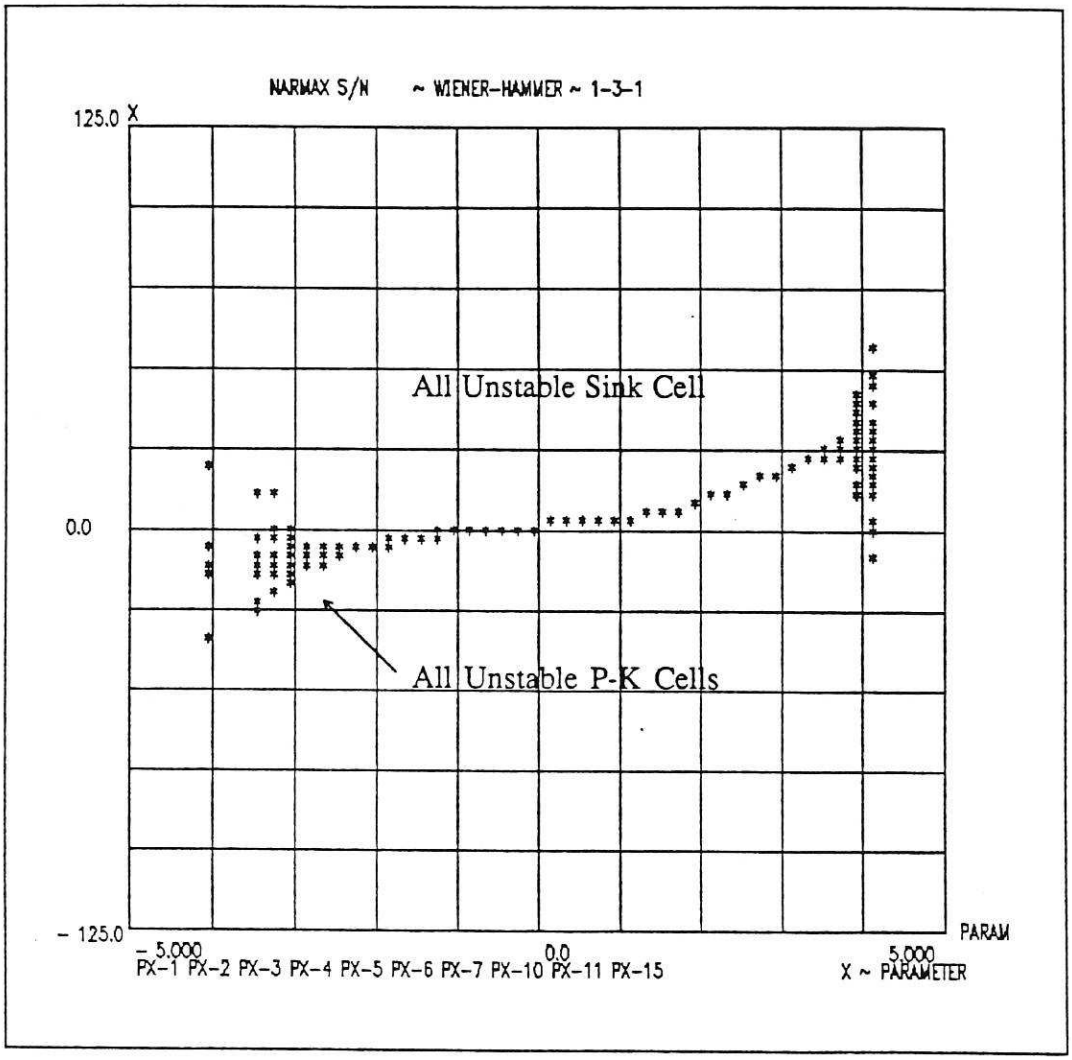


Fig. (16a) - Cell Diagram - S/N= 1

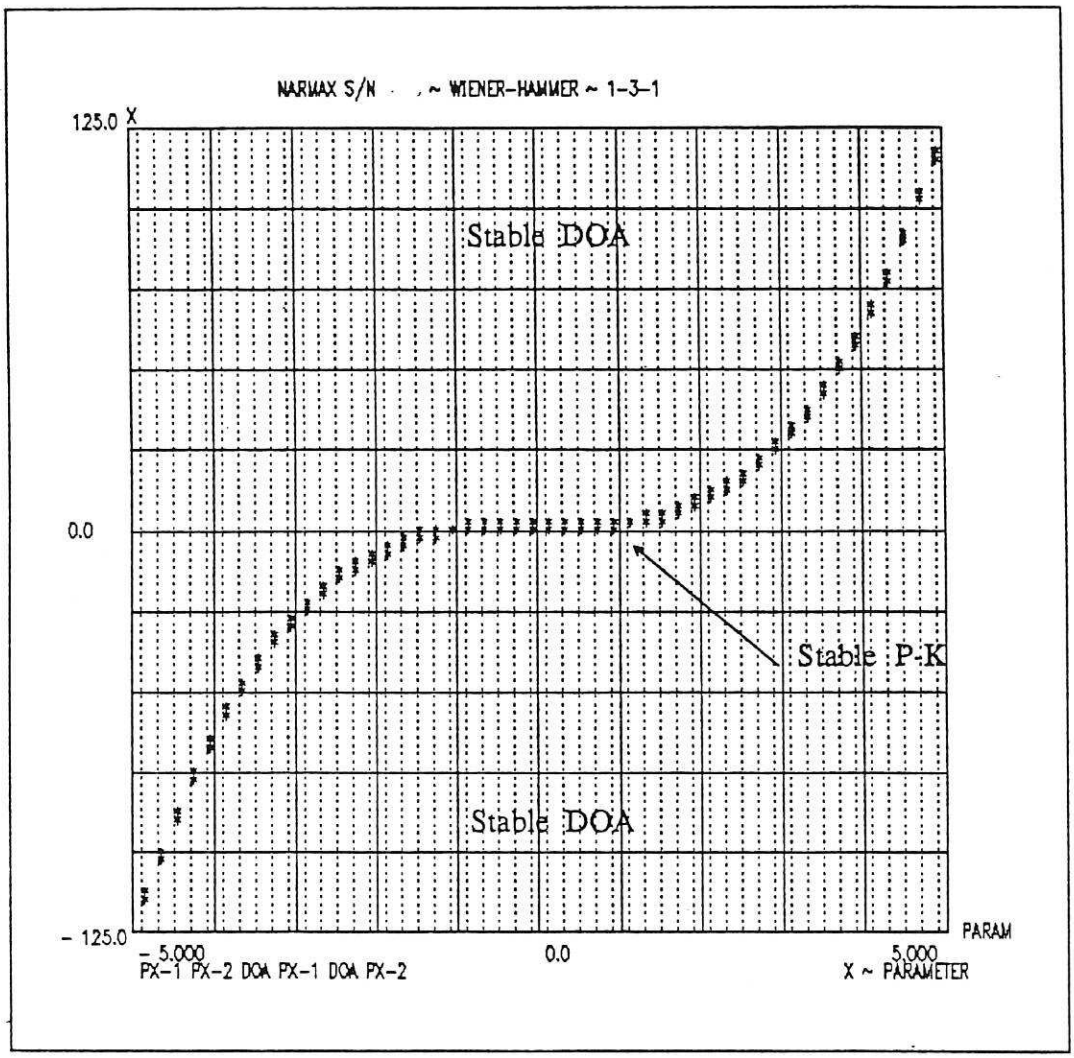


Fig. (16c) - Cell Diagram - S/N= 6