Qualitative Validation of Radial Basis Function Networks

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

Abstract — In system identification applications of neural networks, the aim is usually to obtain a dynamically valid model of the system which can be used for system analysis and for controller design. In the present study, a cell to cell mapping procedure is adopted for the global analysis of nonlinear systems and the qualitative validation of radial basis function networks. The method is used to graphically display the dynamic properties of nonlinear systems in a cell state space, including the fixed points, periodic and aperiodic solutions or chaotic behaviour and the corresponding stability properties. The orthogonal least squares algorithm (OLS) is then used to train a radial basis function network and the trained network is analyzed using the cell mapping framework. In this way the dynamical properties of the trained network can be qualitatively compared with those of the original system. The effects of overparameterisation and output noise on the dynamic properties of the trained network are investigated using cell map analysis.

Keywords — Qualitative Validation, Radial Basis Function, Overparameterisation.

Research Report No. 659

October 1996

Qualitative Validation of Radial Basis Function Networks

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

October 1996

Abstract — In system identification applications of neural networks, the aim is usually to obtain a dynamically valid model of the system which can be used for system analysis and for controller design. In the present study, a cell to cell mapping procedure is adopted for the global analysis of nonlinear systems and the qualitative validation of radial basis function networks. The method is used to graphically display the dynamic properties of nonlinear systems in a cell state space, including the fixed points, periodic and aperiodic solutions or chaotic behaviour and the corresponding stability properties. The orthogonal least squares algorithm (OLS) is then used to train a radial basis function network and the trained network is analyzed using the cell mapping framework. In this way the dynamical properties of the trained network can be qualitatively compared with those of the original system. The effects of overparameterisation and output noise on the dynamic properties of the trained network are investigated using cell map analysis.

Keywords — Qualitative Validation, Radial Basis Function, Overparameterisation.

1 Introduction

Network validation constitutes an important aspect of neural network training. Statistical validation methods have been studied by many researchers (Bohlin, 1971; Leontaritis and Billings, 1987; Söderstrom and Stoica, 1990; Billings and Tao, 1991; Billings and Zhu, 1995) and are often used for validating a trained network. Since most statistical validation methods are based on a particular sequence of the output, they may be considered as 'local' validation methods. Therefore, a statistically valid network may not be dynamically valid (Haynes and Billings, 1994; Zheng and Billings, 1996) and may fail to produce the dynamical invariants of the original system. In system identification applications of neural networks the aim is usually to obtain not only a statistically but also a dynamically valid model of the system under study. The most important dynamical invariants of a nonlinear system are probably the equilibrium points or fixed points, periodic, aperiodic, or strange attractors and the domains of attraction associated with these attractors. 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 validatation of neural networks. With this approach the dynamical invariants of the network can be shown graphically in a cell state space and can be compared with those of the original system. Based on this analysis the aim of the validation presented in the paper is to determine which aspects of the network training algorithm result in a dynamically invalid network. The effects of overparameterisation and output noise on the dynamic properties of the radial basis function network are studied using cell map analysis and an **OLS** training algorithm.

The layout of the paper is as follows. The cell mapping analysis method is presented in section two and it is shown how the cell mapping algorithm can be used for the global analysis of nonlinear dynamical systems. Section three describes the **OLS** algorithm and outlines the qualitative validation procedures for radial basis function networks. Simulation examples are presented in section four to show the effects of overparameterisation and output noise on the trained network.

2 Cell to Cell Mapping

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

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

or a system of difference equations

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

where $\mathbf{x} \in \mathbf{R}^N$ is the state vector. $\mu \in \mathbf{R}^P$ is the parameter vector, $\mathbf{F} : \mathbf{R}^N \times \mathbf{R}^P \longrightarrow \mathbf{R}^N$, $\mathbf{G} : \mathbf{R}^N \times \mathbf{R}^P \longrightarrow \mathbf{R}^N$. The component of the state vector $\mathbf{x}$ is normally regarded as a continuum having an uncountable number of points in any interval. In practice, however the state variables are usually treated as collections of intervals because of the limited accuracy associated with measurements and computations. The state space can therefore be treated as a collection of N-dimensional cells. This state space is called a cell state space. A cell structure may be introduced into the cell state space in various ways. A simple structure can be constructed by dividing each state variable x_i $i = 1, 2, \ldots, N$ into N_i intervals. These intervals are identified by the integer values of the corresponding cell coordinates $z_i = 1, 2, \ldots, N_i$. The Cartesian product space of $z_1, z_2, \ldots, z_N$ is then 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 $i = 1, 2, \ldots, N$. 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

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

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

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

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

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

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

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 identity mapping. A sequence of K distinct cells $\mathbf{z}^*(j), j \in \{K\}$, which satisfies

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

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. A P-K solution may represent the periodic solution of equation (1) or equation (2), it may merely represent the equilibrium point of equation (1) or the fixed point of 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 $r \longrightarrow \infty$. 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

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

and a k-step ahead cell function is defined as

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

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 refered 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. Label the cells using the following transform

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

Such that the regular cells are labelled by positive integers $1, 2, \ldots, N_c$, 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 $\mathbf{S} = \{N_c+\}$. This set is closed under the cell mapping

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

The set of regular cells within the influence of the sink cell forms the domain of attraction of the sink cell, and is labelled the 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 now be extended to enable the qualitative analysis of parameterised nonlinear systems by combining it with bifurcation theory. Consider a discrete system represented by the difference equations

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

Let the region of interest of the state variables be

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

and the region of interest of the parameters be

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

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

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

The point mapping may be rewritten as

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

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

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

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

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

The image of the centre point $\overline{\mathbf{x}}^{(c)}(n)$ 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

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

Note that the resulting cell mapping contains $\prod_{j=1}^{P} N_{N+j}$ slices of N dimensional cell mappings in the parameter space 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 determine the image cells. The time interval T should be selected such that there are not too many cells which become mapped into themselves or into the sink cell.

After the image of all the regular cells has 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 limit cycles of the system. They constitute the invariant stable and unstable orbits of the system within Z^N. The dynamical behaviour of the system is thus revealed graphically in the cell state space.

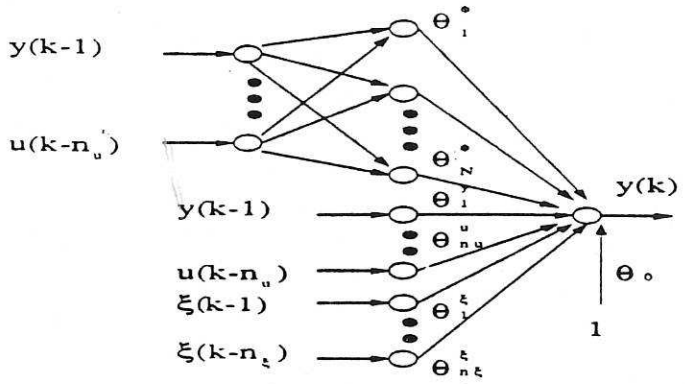

Figure 1: **RBF** network with linear inputs and linear noise inputs

3 Qualitative Validation of RBF Networks

In this study, the radial basis function network as shown in **Fig 1** will be used for modelling nonlinear dynamical systems. The output of the network may be represented as (Broomhead and Lowe, 1988; Moody and Darken, 1989; Poggio and Girosi, 1990)

$$y(k) = \sum_{j=1}^{\mathcal{N}} \theta_j^{\phi} \phi_j + \sum_{j=1}^{n_y} \theta_j^{y} y(k-j) + \sum_{j=1}^{n_u} \theta_j^{u} u(k-j) + \theta_0 + \sum_{j=1}^{n_\xi} \theta_j^{\xi} \xi(k-j) + \xi(k) \qquad (19)$$

where ϕ_j, $j = 1, \ldots, \mathcal{N}$ are the basis functions, $y(k-j)$, $j = 1, \ldots, n_y$ are the delayed outputs, $u(k-j)$, $j = 1, \ldots, n_u$ are the delayed inputs, $\xi(k-j)$, $j = 1, \ldots, n_\xi$ are the noise sequences and θ_0 represents the bias term. Equation (19) can be rewritten in the following form

$$y(k) = \sum_{j=1}^{M} \theta_j \phi_j + \theta_0 + \xi(k) \qquad (20)$$

where ϕ_j, $j = 1, \ldots, M$ are the basis functions or the linear terms $y(k-1), \ldots, y(k-n_y), u(k-1), \ldots, u(k-n_u), \xi(k-1), \ldots, \xi(k-n_\xi)$. The number of hidden layer nodes M is usually very large. However the **OLS** algorithm (Billings, Korenberg and Chen, 1988) can be applied to select a much smaller subset of hidden layer nodes for the network. For a set of input/output signals ($y(k)$ $u(k)$, $k = 1, 2, \ldots, \mathcal{N}$), the input/output relation of the network may be written in vector form as

$$\mathbf{Y} = \Phi\Theta + \Xi \qquad (21)$$

An orthogonal decomposition of Φ is given as

$$\Phi = \mathbf{PA} \qquad (22)$$

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

$$\mathbf{P}^T\mathbf{P} = \mathbf{D} = diag\{d_0, d_1, \ldots, d_M\} \qquad (23)$$

with

$$d_j = < \mathbf{p}_j, \mathbf{p}_j >, \quad j = 0, 1, \ldots, M \tag{24}$$

where $< \bullet, \bullet >$ denotes the inner product. Rearranging equation (21) yields

$$\mathbf{Y} = (\Phi \mathbf{A}^{-1})(\mathbf{A}\Theta) + \Xi = \mathbf{P}\mathbf{q} + \Xi \tag{25}$$

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

$$q_i = \frac{< \mathbf{p}_j, \mathbf{Y} >}{< \mathbf{p}_j, \mathbf{p}_j >}, \qquad j = 0, 1, \ldots, M \tag{26}$$

The number of candidate nodes M may be very large, but a small number of nodes may be adequate to approximate the underlying system dynamics. These nodes can be identified using an efficient forward selection procedure derived in (Billings et al., 1988). The principle of the method is briefly shown below. From equation (25), the sum of the squares of the output is

$$< \mathbf{Y}, \mathbf{Y} >= \sum_{j=0}^{M} q_j^2 < \mathbf{p}_j, \mathbf{p}_j > + < \Xi, \Xi > \tag{27}$$

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

$$\mathbf{err}_j = \frac{q_j^2 < \mathbf{p}_j, \mathbf{p}_j >}{< \mathbf{Y}, \mathbf{Y} >} \tag{28}$$

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

$$1 - \sum_{j=0}^{n_\xi} \mathbf{err}_j < \rho \tag{29}$$

The tolerance ρ will affect both the approximation accuracy and the complexity of the network. The criterion emphasizes the approximation accuracy of the model only and 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). In the present study, the training is terminated when a given number of nodes are selected.

The cell mapping method provides a useful framework for the global analysis of a diverse range of nonlinear systems including 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 network can then be analyzed in the same framework. The deficiency of the network is thus graphically revealed. To perform a cell mapping analysis of a RBF network, a cell structure must be introduced. The validity of the network 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 the validation. A state space representation of the network is obtained and cell structures are introduced into

both the state space and the parameter space. During validation, only the process model is considered, the noise terms $\xi(k)$ which are included in the network to avoid bias are neglected. The network then becomes

$$y(k) = F(y(k-1), \ldots, y(k-n_y), \mu) \tag{30}$$

where F represents the nonlinear function on the right hand side of equation (20) with $\xi(k-j)$, $j = 1, \ldots, n_\xi$ set to zero and $u(k-1), \ldots, u(k-n_u)$ set to μ. A state space representation of the network is therefore

$$\begin{aligned} x_1(k+1) &= F(x_1(k), \ldots, x_{n_y}(k), \mu) \\ x_2(k+1) &= x_1(k) \\ &\ldots \\ x_{n_y}(k+1) &= x_{n_y-1}(k) \end{aligned} \tag{31}$$

Note the singular points of the system are all on the hyperplane $x_1 = x_2 = \ldots = x_{n_y}$. These singular points correspond to all the fixed points of the network and can easily be 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 summaried 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. Train a **RBF** network. Represent the network 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 network and unravel the cells.

3. Find the singular points (P-1 cells) of the network on the hyperplane $x_1 = x_2 = \ldots = x_{n_y}$, compare the singular points with those of the true system.

4. Find the limit cycles (P-K cells, $K > 1$) of the network, 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 will be 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 network is inadequate, the validation based on the singular points in step **3** is often found to be sufficient to reveal the inadequacy. If a network 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 network can be graphically displayed. Since the state variables of the network 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 inadquadacy in

the network and reveal in which region of the state space, if any, the network is valid. The validation will always be limited by the domain of the data used in the training and will not in general be sufficient to justify that the network is valid in the whole space. In practice, the true system will be unknown, but the main purpose of qualitative validation in the present study is to investigate which aspects of the training procedure affect the dynamical properties of the network and how these can be improved.

4 Simulation Results

Consider a nonlinear system represented by the following difference equations (Narendra and Parthasarathy, 1990)

$$y(k) = \frac{y(k-1)y(k-2)\left[y(k-1)+2.5\right]}{1.0+y^2(k-1)+y^2(k-2)} + u(k-1) \tag{32}$$

This system can be expressed in state space form as

$$\begin{aligned} x_1(k+1) &= \frac{x_1(k)x_2(k)\left[x_1(k)+2.5\right]}{1.0+x_1^2(k)+x_2^2(k)} + u(k) \\ x_2(k+1) &= x_1(k) \\ y(k) &= x_1(k) \end{aligned} \tag{33}$$

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 region $x_1 \in [-2\ \ 3]$, $x_2 \in [-2\ \ 3]$, $\mu \in [-0.5\ \ 0.5]$ into $40 \times 40 \times 40$ cells. By evaluating the difference equations in the given region and unravelling the cells, the cell diagram of the system projected onto the $y-u$ plane is shown in **Fig 2**. The fixed points of the system are identified as p-1 and p-3 cells which form the curve as shown in **Fig 2**. As the input u increases form -0.5 to -0.22, the system bifurcates and two more fixed points are created. As the input u increases further from -0.22 to 0.16, the system bifurcates again and the number of fixed points decreases to one. One advantage of cell mapping analysis is that both stable and unstable singular points of the system can be shown in the cell diagram. In **Fig 2**, the lower and upper sections correspond to the stable fixed points, while the middle section which connects the lower and upper sections represents the unstable fixed points. The singular points are computed in the whole cell state space instead of on the hyperplane $x_1 = x_2 = \ldots = x_{n_y}$ in all the simulations. The domains of attraction of the fixed point are not investigated in the present study. In all the simulations, the input signal to the network was a band limited uniformly distributed sequence in the range of [-2 2]. The test signal was a band limited square wave superimposed with a Gaussian signal of unit standard deviation. The input and output signals for the training and test sets are shown in **Figures 3** and **4** respectively.

4.1 Effects of Overparameterisation

The effects of overparameterisation on the dynamical properties of a trained radial basis function network for the system above can now be studied. The results are illustrated in **Figures 5** and **6**. Both figures show the same cell diagrams of a set of trained networks

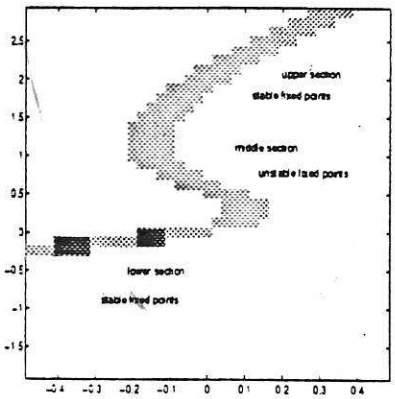

Figure 2: Cell mapping analysis of the difference equations (33) shows the P-1 cells (grey) and the P-3 cells (dark). Vertical axis - y, horizontal axis - u

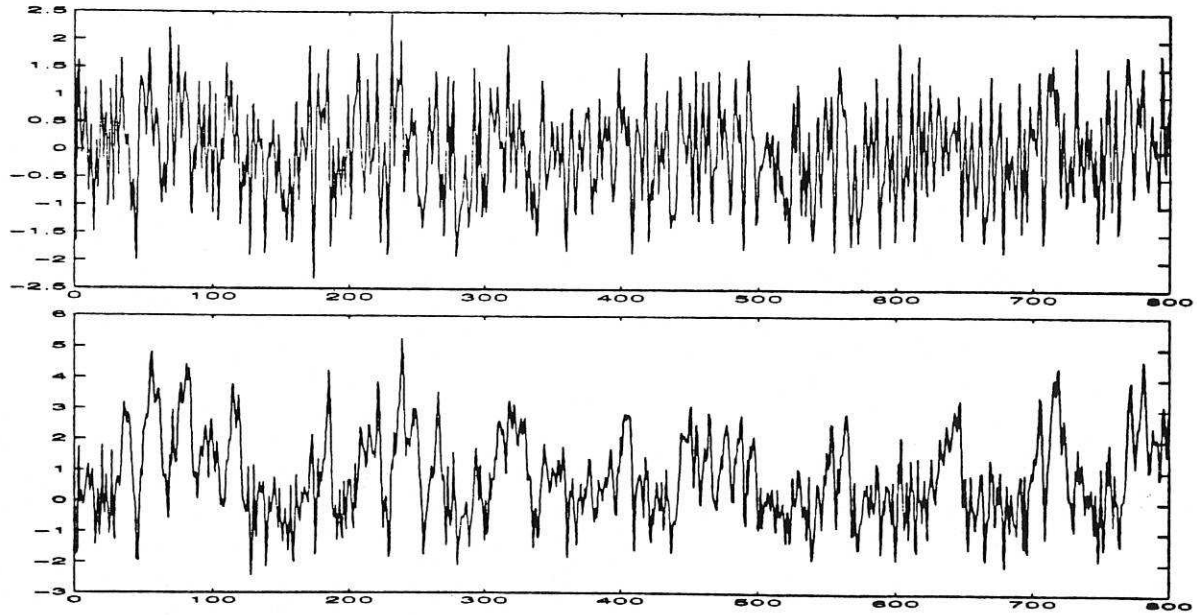

Figure 3: Input/output signals for network training. The input is a band limited uniformly distributed sequence. Top: input, bottom: output.

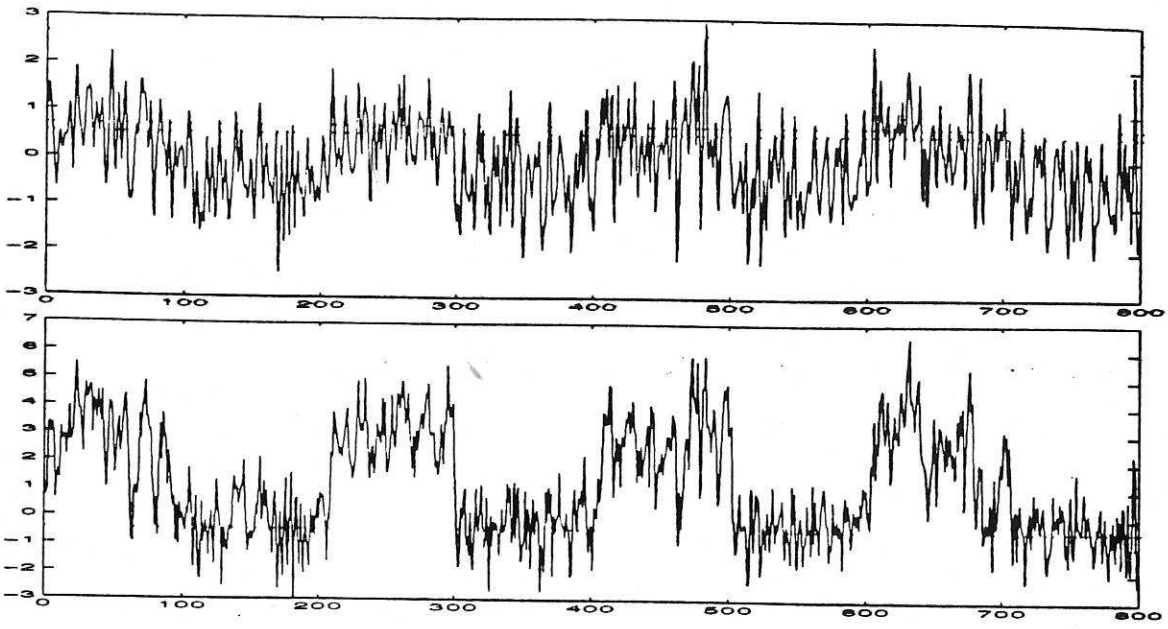

Figure 4: Input/output signals for network testing. The input consists of a band limited square wave superimposed by Gausian signal with unit standard deviation. Top: input, bottom: output.

and that of the original system. For easy of comparison, the cell diagram of the true system was displayed over that of the trained networks in **Fig 5**. To allow a better interpretation the results of **Fig 5** are replotted in **Fig 6** but now the order of the presentation has been reversed. In **Fig 5** the cell diagram of the true system was plotted over that of the network, but in **Fig 6** it is the other way around and the cell diagram of the network is plotted over that of the system. The inputs to the network were $y(k-1)$, $y(k-2)$, $u(k-1)$ which are exactly the same as the inputs to the true system. To accomodate the noise in the output, three linear noise terms were also selected as inputs to the network. The results clearly show that when the number of hidden layer nodes was 10, the network failed to capture the underlying system dynamics and the fixed points of the network were very different from those of the true system. In particular, the bifurcation point at 0.16 has been moved to 0.4 and the one at -0.22 has been shifted to -0.24. The branch of unstable fixed points is also far away from the original. As the number of hidden layer nodes was increased to 20, the bifurcation points of the network became the same as those of the original system but the fixed points in the upper and middle branches were slightly in error. Further increases in the number of hidden layer nodes resulted in a further shift in the bifurcation points and the fixed points in the upper and middle branches. When the number of hidden layer nodes was more than 40, the shape of the middle branch changed and the bifurcation points were shifted away from the original values. The bifurcation points of the networks are listed in **Table 1**.

In practice the correct lagged variables which form the input to the network will not be

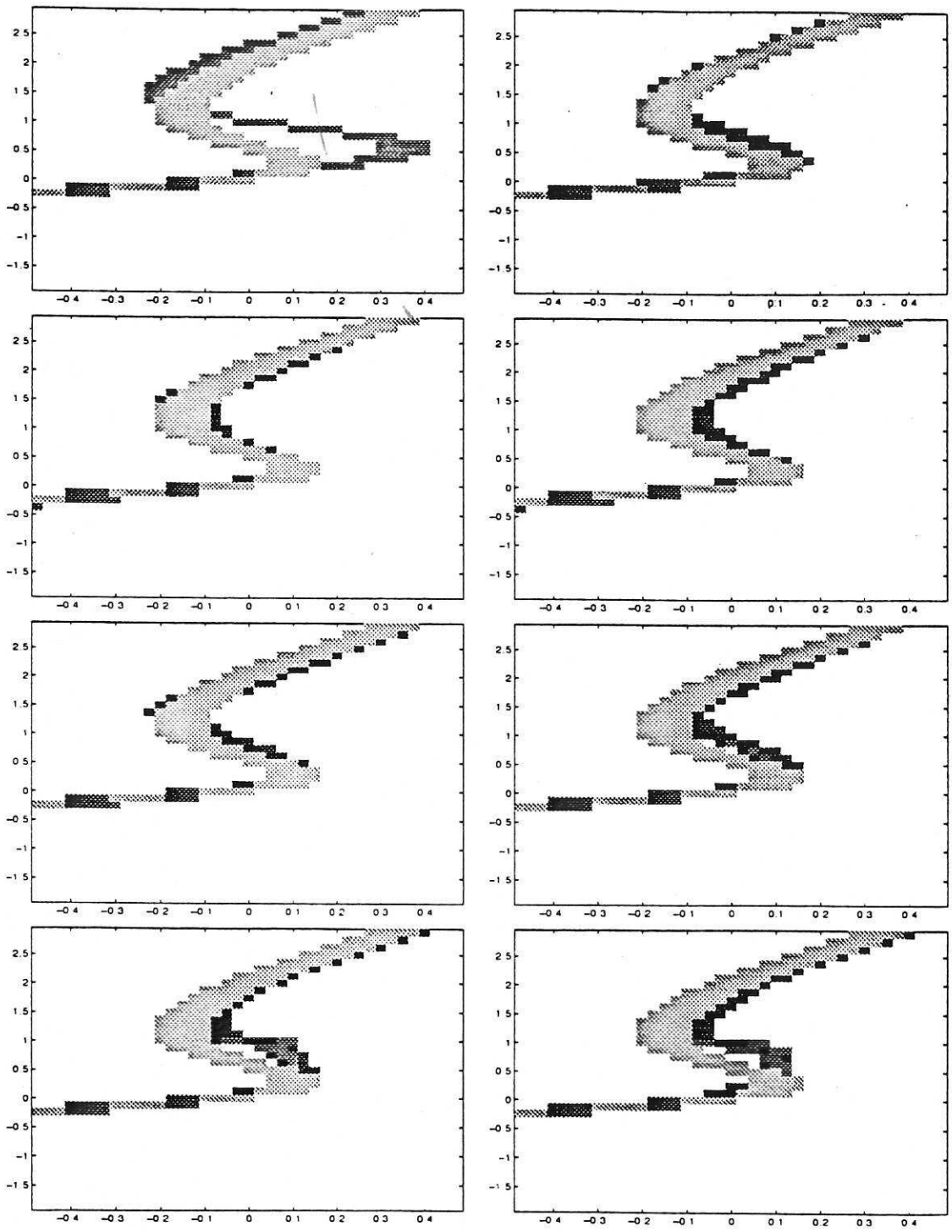

Figure 5: Cell mapping analysis of **RBF** networks. Top left: 10 centres, top right: 15 centres, left of second row: 20 centres, right of second row: 25 centres, left of third row: 30 centres, right of third row: 35 centres, bottom left: 40 centres, bottom right: 45 centres. In all cases the inputs to the networks were $y(t-1)$ $y(t-2)$ $u(t-1)$ and three linear noise terms were also selected as inputs to the networks. The signal to noise ratio was 40 db. The cell diagram of the true system was displayed over that of the network. Dark cells: trained network, Grey cells: true system.

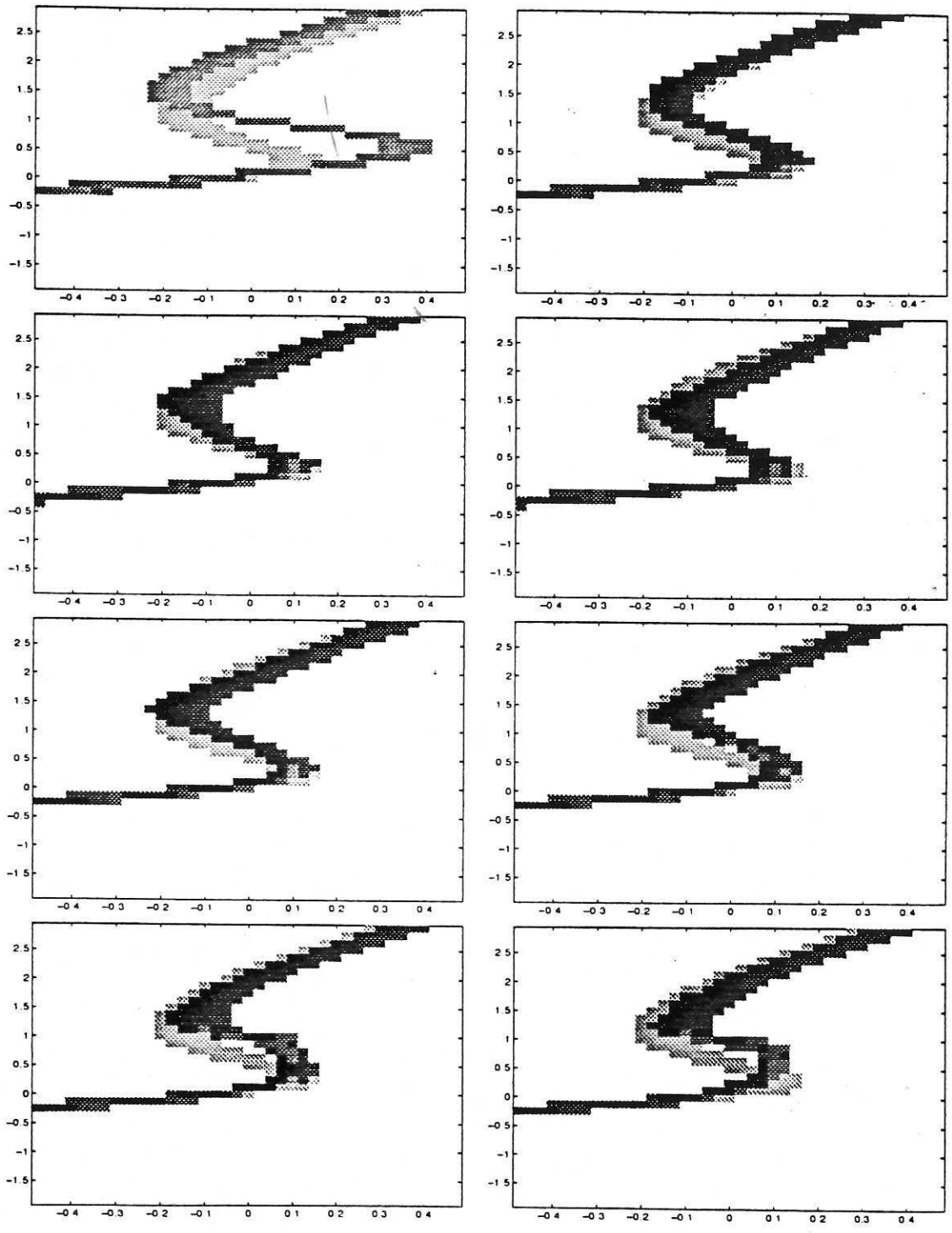

Figure 6: Cell mapping analysis of **RBF** networks. Top left: 10 centres, top right: 15 centres, left of second row: 20 centres, right of second row: 25 centres, left of third row: 30 centres, right of third row: 35 centres, bottom left: 40 centres, bottom right: 45 centres. In all cases the inputs to the networks were $y(t-1)$ $y(t-2)$ $u(t-1)$ and three linear noise terms were also selected as inputs to the networks. The signal to noise ratio was 40 db. Note that this is the same cell diagram as **Fig 5** except the order of plotting has now been reversed so now the cell diagram of the network is displayed over that of the true system. Dark cells: trained network, Grey cells: true system.

Table 1: **Bifurcation Points of RBF Networks**
(Network inputs were $y(k-1)$, $y(k-2)$, $u(k-1)$, the true bifurcation points are -0.22, 0.16)

No. of nodes	10	15	20	25	30	35	40	45
Bifurcation points	-.24, .4	-.19, .18	-.22, .16	-.19, .13	-.25, .16	-.19, .16	-.20, .16	-.19, .13

Table 2: **Bifurcation Points of RBF Networks**
(Network inputs were $y(k-1)$, $y(k-2), y(k-3)$ $u(k-1)$)
(The true bifurcation points are -0.22 and 0.16)

No. of nodes	10	15	20	25	30	35	40	45
Bifurcation points	-.07, .24	-.12, .18	-.12, .20	-.07, .22	-.09, .21	-.17, .27	-.17, .27	-.19, .21

known a priori. Qualitative validation can now be used to investigate the effects of incorrect input lag assignments. The results in **Figures 7** and **8** show the cell mapping analysis when the inputs to the network were incorrectly assigned as $y(k-1)$, $y(k-2)$, $y(k-3)$, $u(k-1)$. Again, both the figures show the same set of cell diagrams in reverse orders of overlap. It may be seen from these figures that more hidden layer nodes were required to capture the system dynamics when the number of input nodes was increased to include the spurious lag $y(t-3)$. The best result was obtained when the number of hidden layer nodes was 45, compared with 20 in the previous simulation. In general, the bifurcation points (see Table 2) and the fixed points in both the middle and upper branches were moved further away from their original positions compared to the case where the lagged inputs were correct. Note that the networks also have more parameters than in the previous simulations due to the increased dimension of the input vector. The increase in the number of input layer nodes actually made the network worse in the sense that the dynamic properties of the networks deviated even further from those of the true system. The results imply that the dynamic properties of the network are more sensitive to an increase in the number of input layer nodes than in the number of hidden layer nodes. Therefore, it is very important to select the correct input nodes in **RBF** network training.

4.2 Effects of Output Noise

The effects of output noise on the dynamic properties of **RBF** networks can be investigated using an analogous approach. Noise is often neglected by most authors so it is important to study if this is justified.

In the simulations, the input signal was taken to be the same as in the previous section, but the output noise was increased from 40 db to 35 db. The inputs to the networks were

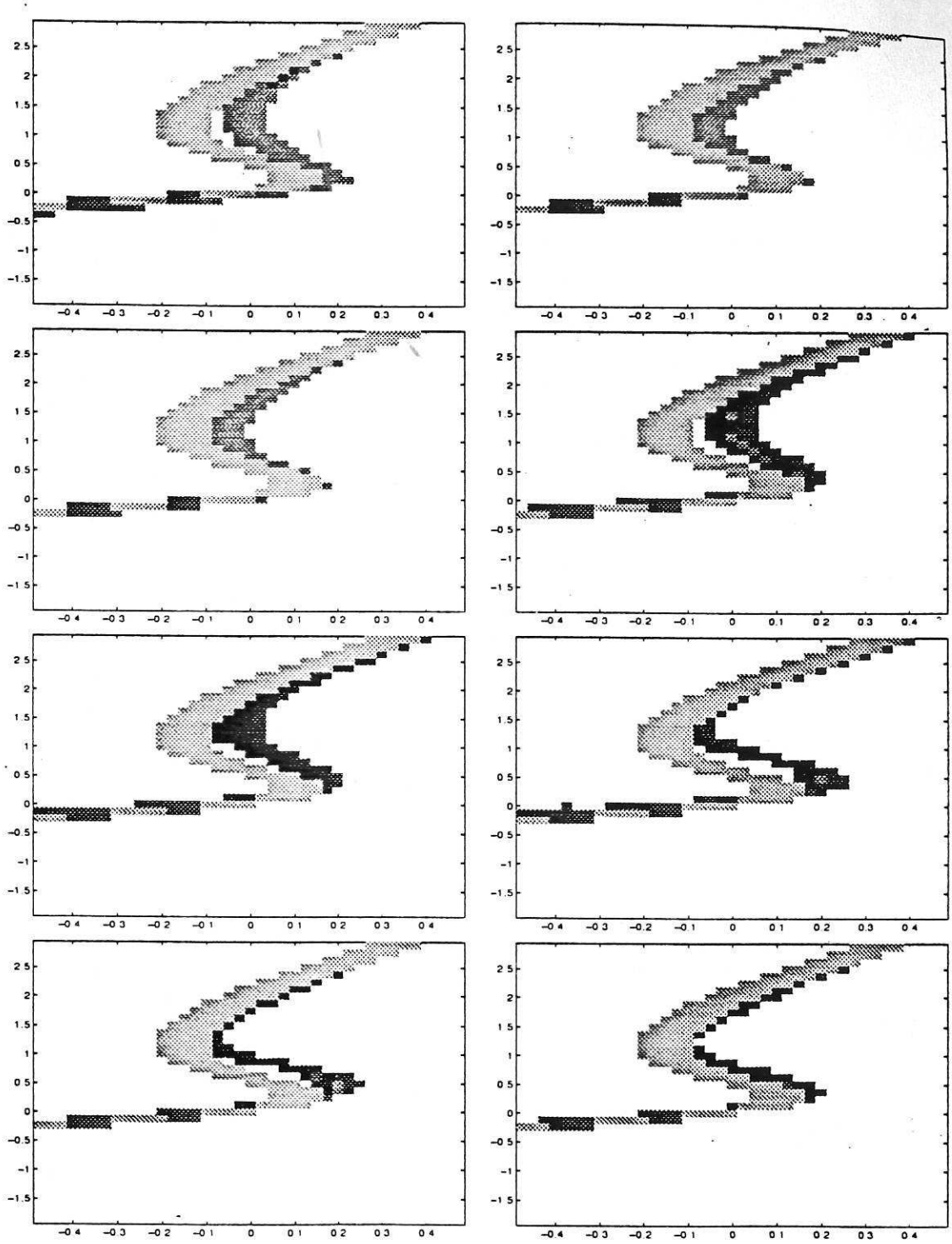

Figure 7: Cell mapping analysis of **RBF** networks. Top left: 10 centres, top right: 15 centres, left of second row: 20 centres, right of second row: 25 centres, left of third row: 30 centres, right of third row: 35 centres, bottom left: 40 centres, bottom right: 45 centres. In all cases the inputs to the networks were incorrectly assigned as $y(t-1)$ $y(t-2)$ $y(t-3)$ $u(t-1)$ and three linear noise terms were also selected as inputs to the networks. The signal to noise ratio was 40 db. The cell diagram of the true system was displayed over that of the network. Dark cells: trained network, Grey cells: true system.

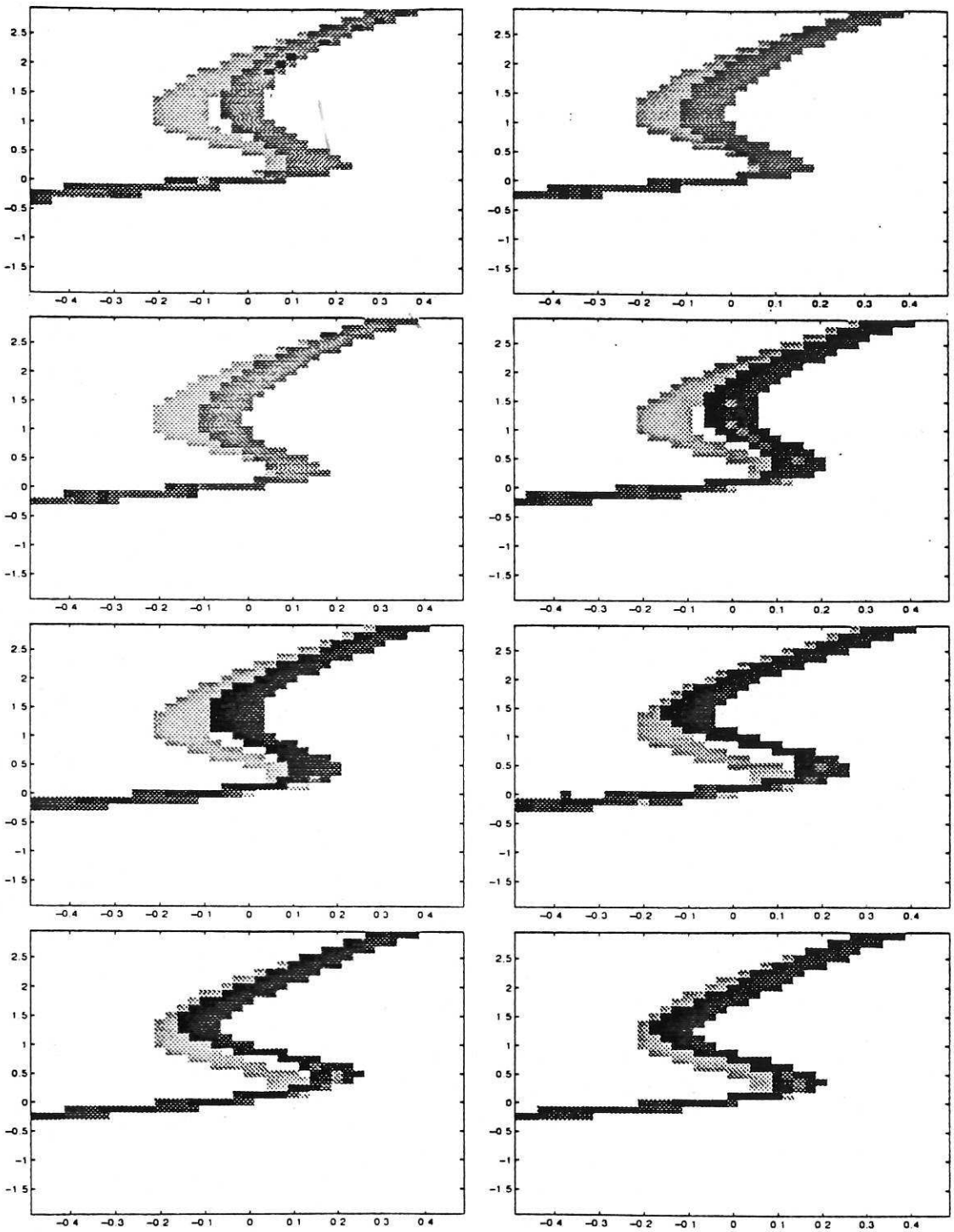

Figure 8: Cell mapping analysis of **RBF** networks. Top left: 10 centres, top right: 15 centres, left of second row: 20 centres, right of second row: 25 centres, left of third row: 30 centres, right of third row: 35 centres, bottom left: 40 centres, bottom right: 45 centres. In all cases the inputs to the networks were incorrectly assigned as $y(t-1)$ $y(t-2)$ $y(t-3)$ $u(t-1)$ and three linear noise terms were also selected as inputs to the networks. The signal to noise ratio was 40 db. Note that this is the same cell diagram as **Fig 7** except the order of plotting has now been reversed so now the cell diagram of the network is displayed over that of the true system. Dark cells: trained network, Grey cells: true system.

Table 3: **Bifurcation Points of RBF Networks**
(Network inputs were $y(k-1)$, $y(k-2)$, $u(k-1)$, Signal to noise ratio was 35 db)
(The true bifurcation points are -0.22, 0.16)

No. of nodes	10	15	20	25	30	35	40	45
Bifurcation points	-.14, .27	-.17, .18	-.14, .16	-.16, .16	-.13, .11	-.12, .14	-.13, .16	-.12, .18

$y(k-1)$, $y(k-2)$, $u(k-1)$ and three linear noise terms were also selected by using the **OLS** algorithm. The results illustrated in **Figures 9** and **10** and Table 3 show that how the bifurcation points and the fixed points in the middle and the upper branches have been moved even further away from their true positions compared with those in **Figures 5** and **6** except when the number of centres was 10 or 25. As the noise level increases, the network exhibits a higher possibility of overfitting to the data and creating spurious fixed points. When the number of centres was increased to 30, the shape of the middle branch was distorted and the resulting network had very different dynamic behaviours from the true system. It can also be seen that the dynamic properties of the network were much more sensitive to changes in the number of hidden layer nodes compared to the cases with lower noise levels. To cope with the increased noise level, more hidden layer nodes were required for the network to capture the system dynamics. This in turn increased the possibility of the network overfitting to the data.

When the signal to noise ratio was further reduced to 20 db, the cell mapping diagrams of the networks for the same input sequence and the same structures are shown in **Fig 11**. The bifurcation points were now far away from their true positions and the middle branches were moved further towards the right. There may be two factors which caused these failures of the networks to capture the system dynamics. First, the training algorithm used in this study selects the data points as the **RBF** centres. Therefore as the noise level was increased, the centres were placed further away from the real data points. Second, since only linear noise terms were used in the networks, as the noise level was increased, the network weights are likely to become increasingly biased. The network performance may be improved if a clustering algorithm is combined with the **OLS** algorithm to select the **RBF** centres. The clustering algorithm could be used to average the effect of large noise levels to a certain extent and to reduce the computational cost of the **OLS** algorithm. To overcome the bias in the network weights, nonlinear noise terms would need to be used. In addition, the training data may be pre-processed using a smoothing algorithm to reduce the noise level in the data prior to centre selection and training.

4.3 Comparison with Network Predictions

The mean square prediction errors of the networks are plotted in **Fig 12**. The network predictions were computed by substituting the delayed outputs $y(k-1), \ldots, y(k-n_y)$ with

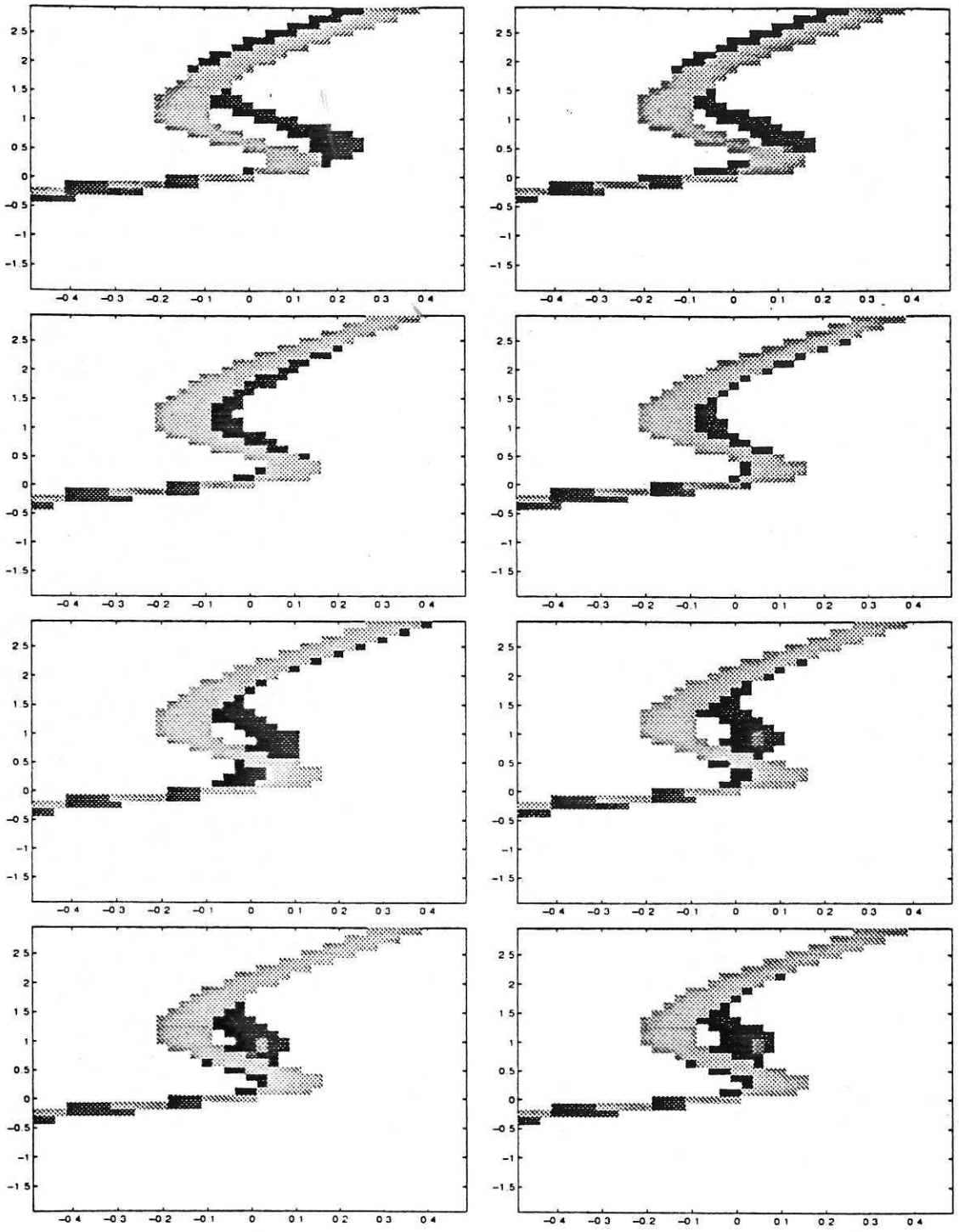

Figure 9: Cell mapping analysis of **RBF** networks. Top left: 10 centres, top right: 15 centres, left of second row: 20 centres, right of second row: 25 centres, left of third row: 30 centres, right of third row: 35 centres, bottom left: 40 centres, bottom right: 45 centres. In all cases the input to the networks were $y(t-1)$ $y(t-2)$ $u(t-1)$ and three linear noise terms were also selected as inputs to the networks. The signal to noise ratio was 35 db. The cell diagram of the true system was displayed over that of the network. Dark cells: trained network, Grey cells: true system.

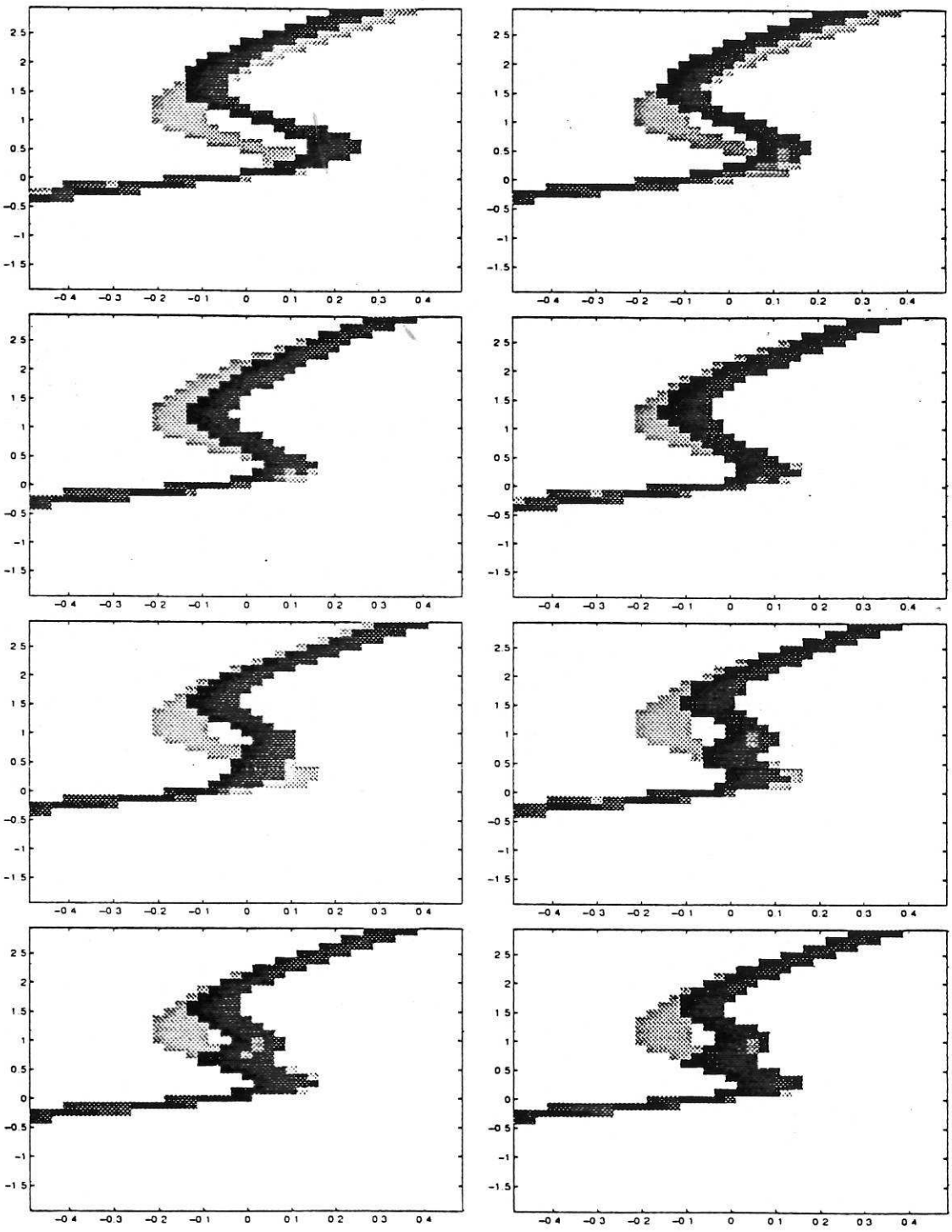

Figure 10: Cell mapping analysis of **RBF** networks. Top left: 10 centres, top right: 15 centres, left of second row: 20 centres, right of second row: 25 centres, left of third row: 30 centres, right of third row: 35 centres, bottom left: 40 centres, bottom right: 45 centres. In all cases the input to the networks were $y(t-1)$ $y(t-2)$ $u(t-1)$ and three linear noise terms were also selected as inputs to the networks. The signal to noise ratio was 35 db. Note that this is the same cell diagram as **Fig 9** except the order of plotting has now been reversed so now the cell diagram of the network is displayed over that of the true system. Dark cells: trained network. Grey cells: true system.

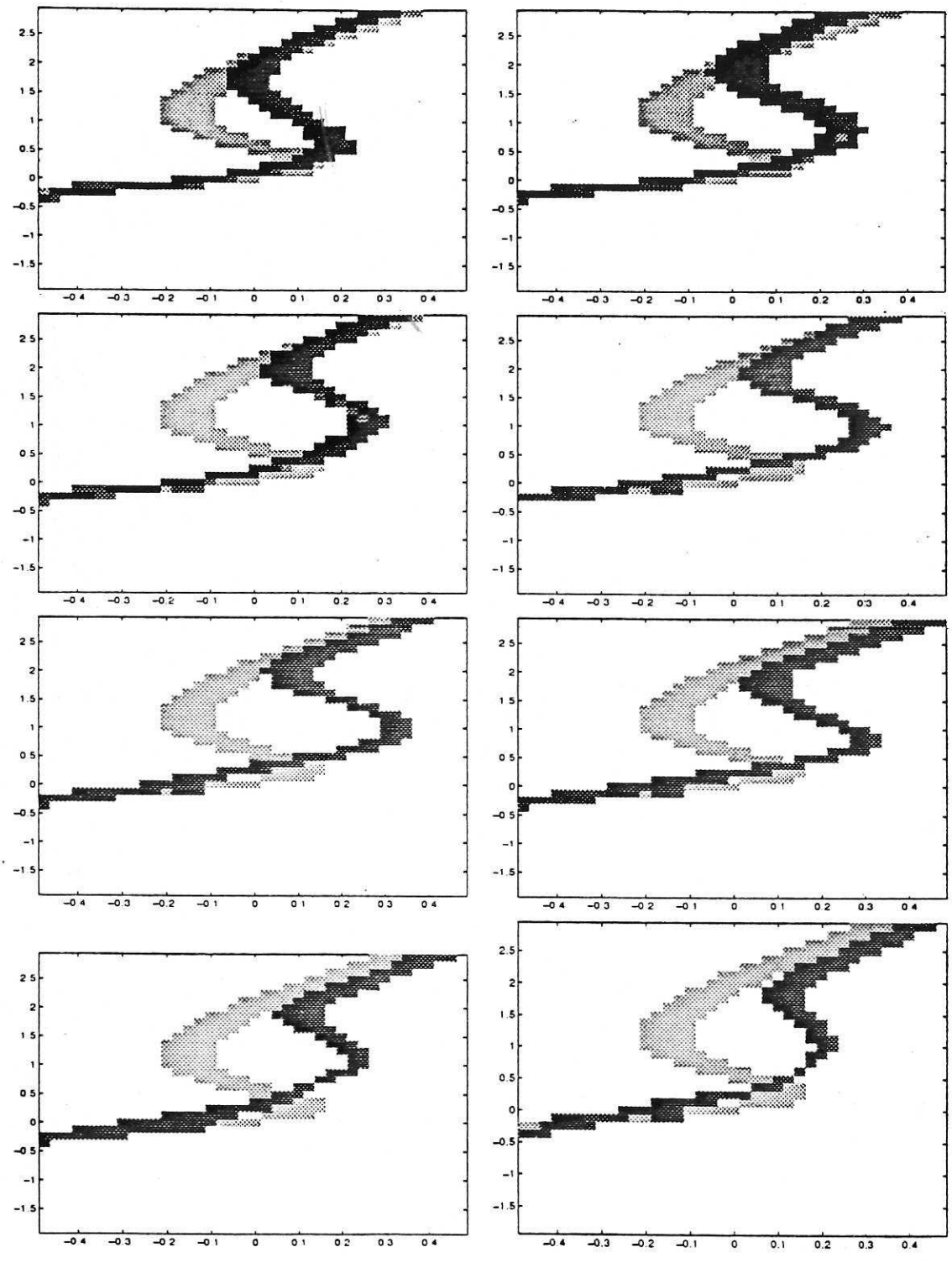

Figure 11: Cell mapping analysis of **RBF** networks. Top left: 10 centres, top right: 15 centres, left of second row: 20 centres, right of second row: 25 centres, left of third row: 30 centres, right of third row: 35 centres, bottom left: 40 centres, bottom right: 45 centres. In all cases the inputs to the networks were $y(t-1)$ $y(t-2)$ $u(t-1)$ and three linear noise terms were also selected as inputs to the networks. The signal to noise ratio was 20 db. The cell diagram of the network was displayed over that of the true system. Dark cells: trained network, Grey cells: true system.

past predictions of the network $y_p(k-1), \ldots, y_p(k-n_y)$ in equation (20) to give.

$$y_p(k) = F(y_p(k-1), \ldots, y_p(k-n_y), u(k-1), \ldots, u(k-n_u)) \tag{34}$$

The variance of the noise in the training data was 1.9940e-2 and that in the test data was 0.3472. Our experience indicates that the network prediction in equation (34) is a better measure of network performance than the one step ahead prediction used by most authors (Zheng and Billings, 1996). However, the network prediction is computed with a particular input sequence and therefore it is a 'local' measure of the dynamic properties of the network. Good prediction performance is necessary for a network to capture the dynamic properties of the underlying system but it is not sufficient. This can be seen from **Fig 12.** The network with 30 centres and 3 input nodes exhibited the best prediction over the training set but the shift in the bifurcation points compared to the true system was worse than for a network with 20 centres and 3 input nodes which produced the best prediction over the test data. When the noise level was increased from 40 db to 35 db, the network with 40 centres and 3 input nodes achieved the best prediction over the training set but the shape of the middle branch was distorted, while the network with 25 centres had the best prediction over the test set and the cell diagram was closer to the original. When the inputs to the network was incorrectly assigned, the prediction over the training set decreases as the number of centres increases, however the network with 40 centres exhibited the best prediction over the test set. As expected the mean squared prediction errors over the training set are less sensitive to changes in the number of hidden layer nodes than over the test set.

In practice the trained network will be an approximation to the true system. Therefore, it is not completely certain that networks which show smaller shifts in bifurcation and fixed points will achieve better predictions on any sequence of data if the dynamic properties of the networks are different from the original system. It is equally uncertain whether a network which exhibits better predictions on a particular data set will show smaller shifts in the bifurcation and fixed points. But prediction performance must be regarded as a 'local' measure while the cell diagram may be regarded as a 'global' measure of the dynamic properties of the network. These should be used as complementary tools. For applications where prediction is the main objective, it might be sufficient to select the network according to predictive performance. But for more general applications where the objective is to use the trained network as a model of the system which can be simulated for different inputs and used in analysis and controller design, it may be necessary to select the network based on the dynamic properties of the system. Methods of achieving this when the true system model is known have been discussed above. These can be used to investigate what is important in network training and as a basis to develop qualitative validation methods which can be used in the realistic situation where the underlying system model is unknown.

5 Conclusions

In this paper, a cell mapping analysis was introduced to qualitatively validate **RBF** networks. It was shown that the cell mapping method can be applied to the global analysis of a diverse range of nonlinear systems including neural networks. The important characteristics of the original nonlinear system and the network, including the equilibrium points,

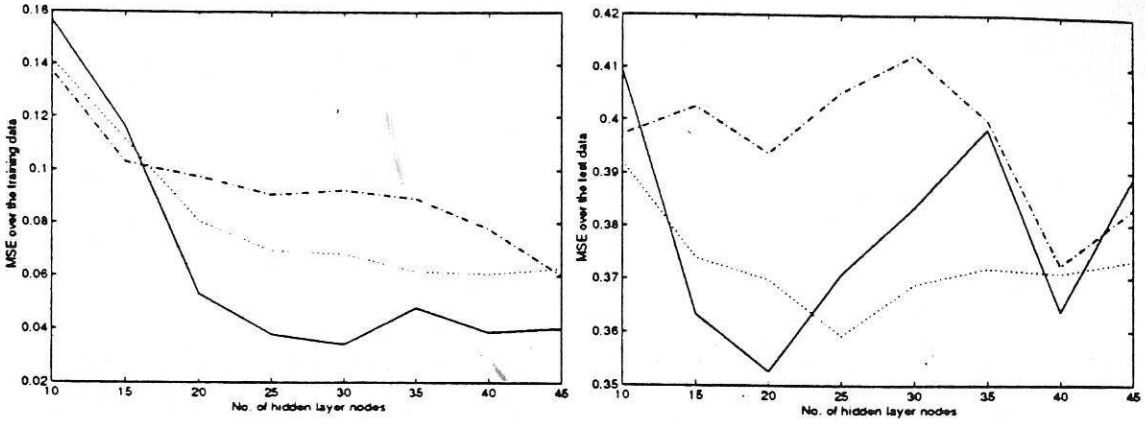

Figure 12: Mean squared iterative prediction errors of the trained networks based on equation (34). Solid line: the input to the networks were $y(t-1)$ $y(t-2)$ $u(t-1)$, and the signal to noise ratio was 40 db. Broken line: the input to the networks were $y(t-1)$ $y(t-2)$ $y(t-3)$ $u(t-1)$, and the signal to noise ratio was 40 db. Dotted line: the input to the networks were $y(t-1)$ $y(t-2)$ $u(t-1)$, and the signal to noise ratio was 35 db

the fixed points, the limit cycles, the domains of attraction and the bifurcation points can be graphically displayed in a cell state space. The algorithm can be used to pin point deficiencies in trained networks by comparing the resulting cell map diagrams with those of the original system. Used in this way the method provides a useful tool for investigating the effects of overparameterisation, output noise, input selection and many other aspects of network training algorithms on the dynamic properties of the network.

Underparameterisation or overparameterisation during training will result in a network with dynamic properties which differ from those of the underlying system, such as a shift in the bifurcation and fixed points, or even spurious singular points. The simulation results also show that the network can be sensitive to a change in the assignment of input nodes. It is therefore important to select the correct input nodes in network training. Noise on the training data will also severely impair network performance and it is strongly recommended that noise models are used in network training or that the noise is accomodated in some other way.

Network predictions and most of the statistical validation methods may be regarded as 'local'. These 'local' methods and the 'global' validation methods of which the cell mapping analysis is an example should be used as complementary tools in network training to increase the possibility of identifying a representative model.

Acknowledgements

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

References

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. (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.

Broomhead, D. S. and Lowe, D. (1988). Multivariable functional interpolation and adaptive networks. *Complex Systems*, 2:321 – 355.

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. (1987). Model selection and validation methods for nonlinear systems. *Int. J. Control*, 45(1):311 – 341.

Moody, J. and Darken, C. (1989). Fast learning in networks of locally-tunned processing units. *Neural Computation*, 1:281 – 294.

Narendra, K. S. and Parthasarathy, K. (1990). Identification and control of dynamical systems using neural networks. *IEEE Trans. on Neural Networks*, 1(1):4–27.

Poggio, T. and Girosi, F. (1990). Network for approximation and learning. *Proceddings of IEEE*, 78(9):1481 – 1497.

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

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

List of Figures