RADIAL BASIS FUNCTION (RBF) FOR NON-LINEAR DYNAMIC SYSTEM IDENTIFICATION

ROBIAH AHMAD[1] & HISHAMUDDIN JAMALUDDIN[2]

Abstract.   One of the key problem in system identification is finding a suitable model structure. In this paper, radial basis function (RBF) network using various basis functions are trained to represent discrete-time nonlinear dynamic systems and the results are compared. The orthogonal least square algorithm is employed to select parsimonious RBF models. To demonstrate the identification procedure, two examples of modelling nonlinear system were included.

Keywords: radial basis function, system identification, non-linear system modelling, orthogonal least square algorithm

Abstrak: Masalah utama dalam pengenalpastian sistem ialah memilih struktur model yang sesuai. Dalam artikel ini, rangkaian fungsi asas jejarian menggunakan pelbagai fungsi asas dilatih untuk mewakili sistem dinamik tak linear masa diskrit dan keputusannya dibandingkan. Algoritma kuasa dua terkecil ortogon digunakan untuk memilih model rangkaian asas jejarian termudah. Untuk menerangkan tatacara pengenalpastian, dua contoh pemodelan sistem tak linear dibincangkan.

Kata kunci: fungsi asas jejarian, pengenalpastian sistem, pemodelan sistem tak linear, algoritma kuasa dua terkecil ortogon

1.0   INTRODUCTION

In system identification, it is important to detect the model structure of a system or to determine terms to be include in the final model. In linear AutoRegressive Moving Average model with eXogenous inputs (ARMAX), the structure of the model can be determined by summing up the input and output lags and the noise terms in the model. The structure of the model can also be determined by increasing the number of lags until the model adequately fits the true system.

For nonlinear systems, the number of terms to be include in the final model will increase. In NARMAX (Nonlinear AutoRegressive Moving Average with eXogenous inputs) model, as the number of lags (input and output) and degree of nonlinearity increases, the number of terms increases [1]. Therefore, it is important to use algorithm that can detect the structure of the model to ensure that the model to be identified is best fitted.

[1&2]   Department of Applied Mechanics, Faculty of Mechanical Engineering, Universiti Teknologi Malaysia. e-mail: robiah@fkm.utm.my, hishamj@fkm.utm.my

Researchers used different methods such as group method of data handling GMDH [2], Volterra and Weiner Series Model [3] and also linear-in-parameter nonlinear models [4]. Different structure selection methods for system identification were summarized [5].

For structure identification of linear-in-parameter nonlinear systems, the system is modeled from its input and output components and the residuals are computed. The objective is to select the components from all the possible structure. The present study investigates the use of radial basis function (RBF) network with different basis functions to model nonlinear system. In searching for all the possible structure, the parameters for those structures have to be calculated. The problem remain is how to select an appropriate set of RBF centres.

However, by considering RBF as a special two-layered network of a linear-in-the-unknown-parameters and therefore, a linear least square procedures can be used to train the network. An orthogonal least squares (OLS) learning algorithm for RBF network is adopted [6]. In OLS, the significance of each of the model term contributed to the overall model is calculated based on the error reduction ratios (ERR). The larger the value of ERR, the more significant the term will be in the final model.

This paper investigates the use of radial basis function network in modeling discrete-time nonlinear system. The adequacy of the fitted model is determined using model validity tests and finally, the identification results of two nonlinear processes are presented to illustrate the method. The use of different basis functions for RBF network models have not been widely reported except for gaussian and thin-plate-spline function [1]. In this study the adequacy of RBF network with five different basis functions namely linear, cubic, thin-plate-spline and multiquadratics and inverse multiquadratics are compared.

2.0   PROBLEM DEFINITION

For discrete-time linear system, the representation is largely based on linear difference equation model

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

where $y(t)$ and $u(t)$ are the output and input of the system and $e(t)$ is noise, while n_y, n_u and n_e are the maximum output, input and noise lags respectively. This model is known as ARMAX model.

In early research, a nonlinear system can be represented by the Volterra series

$$y(t) = \sum_{i=0}^{\infty} h_{xo/ki}(u(1), \ldots, u(k)) \tag{2}$$

where $h_{x0/ki}$ is a homogeneous degree i polynomial in $u(1),...,u(k)$. The functional series expansion of Volterra map past inputs into the present output and at the end an excessive parameter set will be needed to describe even a simple nonlinear model and this will lead to excessive computation. The analysis of this model was reviewed [7].

Alternatively [4] represented a non-linear system based on Nonlinear AutoRegressive Moving Average models with eXogeneous inputs or NARMAX models

$$y(t) = f(y(t-1), \dots y(t-n_y), u(t-1), \dots, u(t-n_u), e(t)) \qquad (3)$$

where $y(t)$ and $u(t)$ are the output and input of the system, $f(.)$ is the nonlinear function and $e(t)$ is noise. If this NARMAX model with first order system expanded to second order nonlinearity, the system equation would be represented as

$$y(t) = f[(y(t-1), u(t-1)]$$
$$= a_1 y(t-1) + a_2 u(t-1) + a_3 y(t-1) u(t-1) + a_4 y^2(t-1) + a_5 u^2(t-1) \qquad (4)$$

As the input and output lags and the nonlinearity increased, the number of parameter to be estimated would also be increased. In designing the model for nonlinear system, the parsimonious principle [8] is critical because a nonlinear model involves an excessive number of parameters. A lot of work needs to be done by searching over a wide range of possible solution in terms of number of lags and nonlinearity of the system to ensure that the best model fits with the smallest number of parameters. Therefore, in system identification it is important to have an algorithm that could detect the accurate model structure of the model.

Due to its flexibility, radial basis function networks proved to be successful in dealing with nonlinear system [9]. The structure of RBF is nonlinear but have a linear-in-
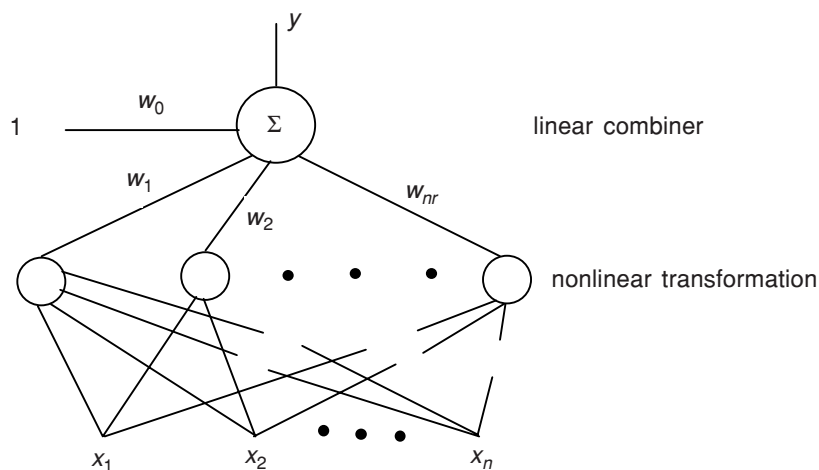

Figure 1   A radial basis function architecture

the-parameter formulation. The output layer is a linear combination of Euclidean distance and the nonlinear function and its architecture is shown in Figure 1.

In searching for the best structure for the model, it is important to perform structure determination for all the possible structure for the chosen system. Least square method is done based on certain performance criteria. In this study, the linear-in-the-parameter structure of RBF expansion can be utilized for model selection and orthogonal least square (OLS) algorithm will select centres so that adequate and parsimonious RBF networks can be obtained. Each component of the model is checked based on error reduction ratio (ERR) and simultaneously the parameter θ_i are calculated.

3.0 RADIAL BASIS FUNCTION AND LINEAR REGRESSION MODEL

The alternative neural network architecture besides multilayer perceptron (MLP) is radial basis function. It is a technique used for interpolating in multidimensional space and the theoretical properties of this method have been carefully investigated [10,11]. RBF is a two layer processing structure of neural network where the first hidden layer consists of an array of nodes where the nodes will calculate the Euclidean distance between the input signal x and the center of the neuron c_i. Then the activation function known as basis function ϕ is applied. The centers c_i are usually chosen to be a subset of the data or distributed uniformly in the input domain. The outputs of the hidden layer are combined linearly by the neuron of the second layer based on the equation

$$f(x) = \sum_{i=1}^{n_H} \theta_i \, \phi \left(\left\| x - c_i \right\| \right) \tag{5}$$

where $x = [x_1 ... x_m]^{\mathrm{T}}$ is the input vector with m inputs, $\phi(.)$ is the basis function, $\| . \|$ is the Euclidean norm, θ_i are weights, $c_i = [c_{1,i} ... x_{m,i}]^{\mathrm{T}}$ are the RBF centres and n_H is the number of hidden nodes. Some choices of the functions considered are the linear function

$\phi(v)=v$

the cubic function

$\phi(v)=v^3$

the thin plate spline function

$\phi(v)=v^2 \times \log(v)$

the multiquadratic function

$\phi(v)=(v^2 + \beta^2)^{1/2}$

and inverse multiquadratic function

$\phi(v)=1/(v^2 + \beta^2)^{1/2}$

where β is a real constant and in all cases v is the scaled radius ($\|x-c_i\|)/d_i$, and d_i is scaling factor.

RBF network is a special case of linear regression model and can be represented as

$$y(t) = \sum_{i=1}^{N} \theta_i(t)\phi_i + e(t), 1 \leq t \leq N \tag{6}$$

where $y(t)$ is the output corresponds to network input of $x(t)$ with N equal number of data, θ_i are the estimated unknown parameters and $\phi_i(t)=\phi(\|\mathbf{x}(t)-\mathbf{c}_i\|)$ using every $x(t)$ as a centre, that is, $c_i=x(i)$ for $1 \leq i \leq N$ and $e(t)$ is the error between $y(t)$ and $f(x(t))$. By defining

$$\mathbf{y} = [y(1) \ldots y(N)]^{\mathrm{T}}$$

$$\mathbf{\phi_i} = [\phi_i(1) \ldots \phi_i(N)]^{\mathrm{T}}$$

$$\mathbf{\theta} = [\theta_1 \ldots \theta_N]^{\mathrm{T}}$$

Therefore, equation (6) can be regarded as a linear regression model in the form of

$$\mathbf{y} = \mathbf{\phi}\mathbf{\theta} + \mathbf{e} \tag{7}$$

3.1 Orthogonal Least Square Algorithms

The OLS method transforms the regressors ϕ_i into a set of orthogonal basis vector. Using a classical Gram-Schmidt orthogonal procedure, the regressor is decomposed into

$$\mathbf{\phi} = \mathbf{WA} \tag{8}$$

where $\mathbf{A} \in \Re^{MxM}$ is an upper triangular matrix with diagonal elements and $\mathbf{W} \in \Re^{NxN}$ is a matrix with orthogonal columns w_i such that

$$\mathbf{W}^{\mathbf{T}}\mathbf{W} = \mathbf{H} \tag{9}$$

and $\mathbf{H}$ is diagonal with elements h_i such that

$$h_i = w_i^T w_i = \sum_{i=1}^{N} w_i(t)w_i(t), \quad 1 \leq i \leq M \tag{10}$$

By substituting (7) into (6) gives and auxiliary equation of

$$\mathbf{y} = \mathbf{WA}\mathbf{\theta} + \mathbf{e} = \mathbf{Wg} + \mathbf{e} \tag{11}$$

where $\mathbf{g} = \mathbf{A}\mathbf{\theta}$. Since w_i and w_j are orthogonal for $i \neq j$, the sum square of y can be calculated

$$y^T y = \sum_{i=1}^{M} g_i^2 w_i^T w_i \tag{12}$$

dividing both sides with N indicates that $g_i^2 w_i^T w_i/N$ is the increment to the output variance by w_i and can be defined as

$$[err]_i = \frac{g_i^2 u_i^T u_i}{y^T y} \tag{13}$$

which provides an indication of which term to be included in the model. Then the parameter estimate $\boldsymbol{\theta}$ can be computed from

$$\mathbf{A}\boldsymbol{\theta} = \mathbf{g} \tag{14}$$

3.2   Implementing the Algorithm

The basic idea of OLS algorithm is to transform equation (7) into equivalent orthogonal equation as in equation (11). The algorithm can be calculated by applying Gram-Schmidt procedures as described by [12,13]. The procedure for the regressor selection is summarised as follows:

At the first step, all the $p_i(t)$, i=1,....,M are considered as possible candidates for $w_1(t)$. For $1 \le i \le M$, compute

$$w_1^{(i)}(t) = p_i(t), \quad g_1^{(i)} = \frac{\sum_{t=1}^{N} w_1^{(i)}(t)\,y(t)}{\sum_{t=1}^{N} (w_1^{(i)}(t))^2}, \quad [err]_1^{(i)} = \frac{(g_1^{(i)})^2 \sum_{t=1}^{N} (w_1^{(i)}(t))^2}{\sum_{t=1}^{N} y^2(t)}$$

Find

$$[err]_1^{(i)} = \max\{[err]_1^{(i)}\}, \text{where } 1 \le i \le M$$

and select

$$w_1 = w_1^{(ik)} = p_{i1}$$

At kth step, for $k \ge 2$, $1 \le i \le M$, $i \ne j$, compute

$$w_k^{(i)} = p_i - \sum_{j=1}^{k-1} \alpha_{jk}^{(i)}\, w_j, \quad 1 \le i \le k, \quad \alpha_{jk}^{(i)} = \frac{\sum_{t=1}^{N} p_i(t) w_j(t)}{\sum_{t=1}^{N} w_j^2(t)}, \quad g_k^{(i)} = \frac{\sum_{t=1}^{N} w_k^{(i)}(t)\,y(t)}{\sum_{t=1}^{N} (w_k^{(i)}(t))^2}$$

$$[err]_k^{(i)} = \frac{(g_k^{(i)})^2 \sum_{t=1}^{N} (w_k^{(i)}(t))^2}{\sum_{t=1}^{N} y^2(t)}$$

Find

$$[err]_k^{(ik)} = \max\{[err]_k^{(i)}\},$$

and select

$$w_k = w_k^{(ik)} = p_{ik} - \sum_{j=1}^{k-1} \alpha_{jk} w_j$$

The procedure is terminated at Mth step when

$$1 - \sum_{i=1}^{n_r} ERR_i < p_t$$

The parameter estimators are calculated from the equation below as

$$\theta_M = g_M, \text{ and } \theta_i = g_i - \sum_{j=i+1}^{M} \alpha_{ij}\theta_j$$

where $i = M-1, M-2, ...,1$.

4.0   MODEL VALIDITY TESTS

Model validity tests are developed to determine the adequacy of the fitted model. The unbiased model should be uncorrelated to the other variables including inputs and outputs. The tests are based on the following correlation functions [14]:

$$\phi_{\varepsilon\varepsilon}(\tau) = 0, \quad \tau \neq 0$$

$$\phi_{u\varepsilon}(\tau) = 0, \quad \text{for all } \tau$$

$$\phi_{\varepsilon\varepsilon u}(\tau) = 0, \quad \tau \geq 0 \tag{15}$$

$$\phi_{\varepsilon\varepsilon uu}(\tau) = 0, \quad \text{for all } \tau$$

$$\phi_{\varepsilon uu}(\tau) = 0, \quad \text{for all } \tau$$

Generally, if the correlation functions are within 95% confidence interval, $\pm 1.96\sqrt{N}$, the model is regarded as adequate. The predictive accuracy of the identification model can also be used as validity test and computed by defining the normalized root mean square of the residuals as an error index defined by

$$\text{Error index} = \left[\frac{\sum (\hat{y}(t) - y(t))^2}{\sum y^2(t)} \right]^{\frac{1}{2}} \tag{16}$$

The variance of the predicted models is also calculated and it is a measure of sum of squared distance between the residuals divided by the total number of data and is defined as

$$\text{Variance} = \frac{\sum (\hat{y}(t) - y(t))^2}{N} \tag{17}$$

5.0 SIMULATION

Computer simulation was carried out to evaluate the models found using OLS learning algorithm to RBF network models. For multiquadratic and inverse multiquadratic function, the value of k is 0.01. Two examples were used in this application to illustrate the method and results were compared between different basis functions. The data is well known and frequently used as an example for testing identification algorithms.

5.1 Model 1: Gas Furnace

An example of gas furnace data [15] was used to illustrate OLS algorithm. The input $u(t)$ is gas flow rate into of the furnace and the output $y(t)$ is CO_2 concentration into outlet gas. RBF networks were given as $x(t)=[y(t-1)\ y(t-2)\ y(t-3)\ u(t-1)\ u(t-2)\ u(t-3)]^{\text{T}}$ and there were 296 pairs of input-output data. The OLS algorithms identified RBF models with 10 centres. The first 200 data points were used as the estimation set and the remaining 96 were used as the testing set.
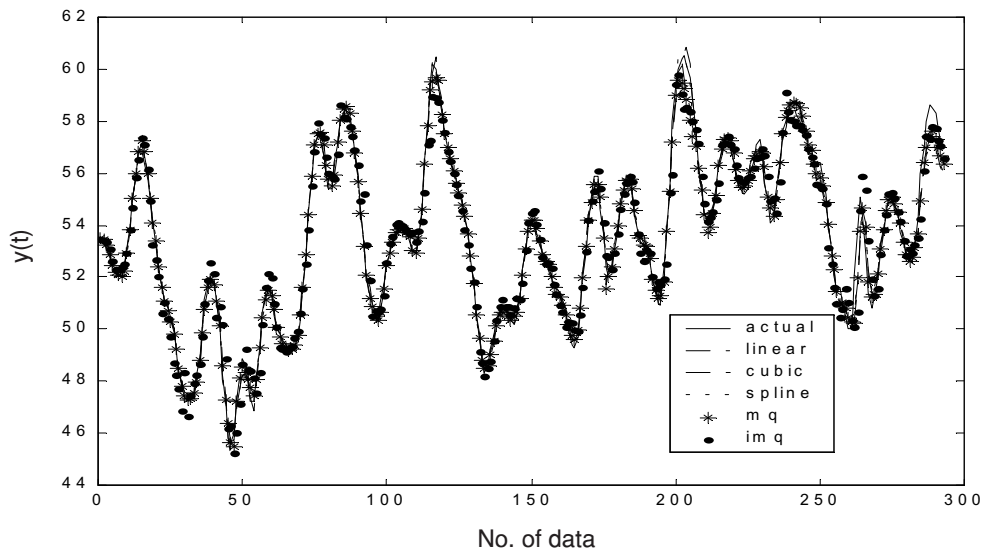

Figure 2 System output (solid data) superimposed on one-step ahead prediction of RBF models with different functions

Table 1   Variance and error index for different RBF functions

RBF Functions	Variance		Error Index	
	Estimation	**Training**	**Estimation**	**Training**
Linear	0.051	0.334	0.0043	0.0111
Cubic	0.038	0.346	0.0038	0.0113
Spline	0.027	0.270	0.0032	0.0099
Multiquadratics	0.048	0.358	0.0042	0.0115
Inverse multiquadratics	0.326	0.770	0.0110	0.0170

Figure 2 shows the actual process output in comparison with the identified models with different RBF functions and it shows that the identified models performed a good fitting to the actual process. It also shows that there is slight difference between different RBF functions that were used. The results are shown in Table 1 where the corresponding variance and error index for different functions are provided to give the insight of the accuracy of the identification models. RBF with thin-plate-spline has the lowest variance and error index if compared with the others.

The correlation tests are shown in Figures 3, 4, 5, 6 and 7 respectively. It is observed almost all correlation tests were within the 95% confidence bands. It can be concluded that the tests reveal that the models are adequate.
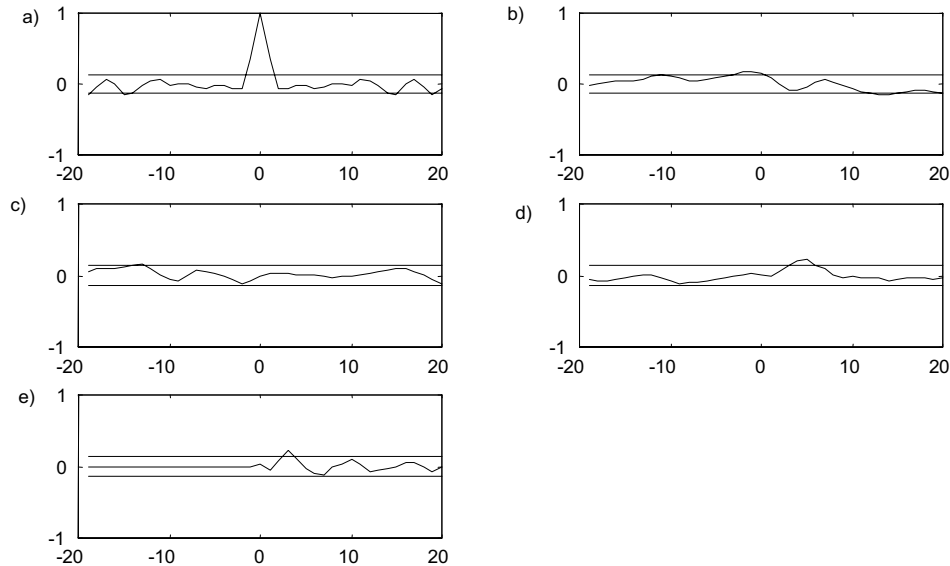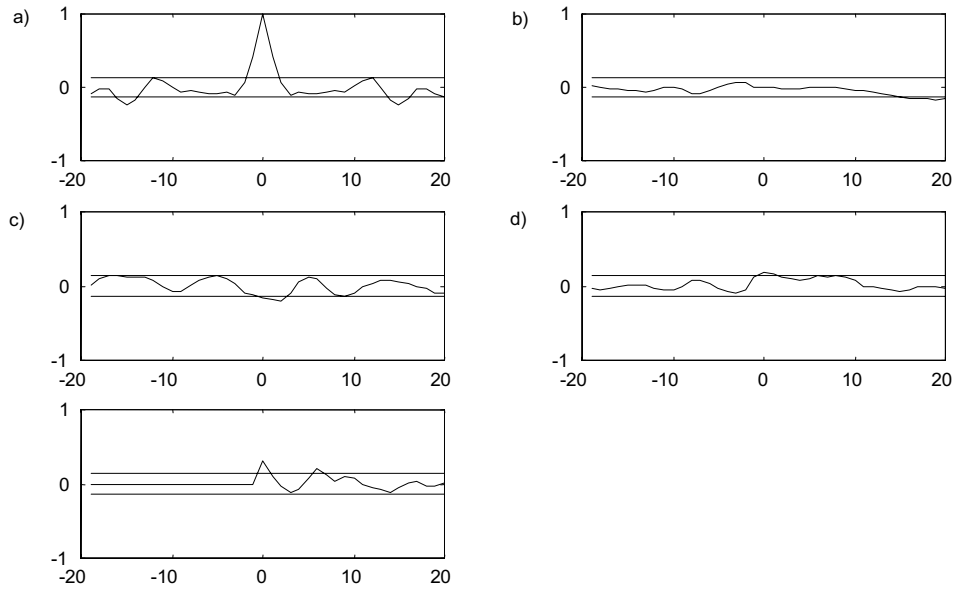

Figure 3   Correlation tests for linear function a) $\phi_{\varepsilon\varepsilon}(\tau)$, b) $\phi_{\varepsilon u}(\tau)$, c) $\phi_{uu\varepsilon}(\tau)$, d) $\phi_{u^2\varepsilon^2}(\tau)$, e) $\phi_{\varepsilon\varepsilon u}(\tau)$

ROBIAH AHMAD & HISHAMUDDIN JAMALUDDIN

Figure 4   Correlation tests for cubic function a)$\phi_{\varepsilon\varepsilon}(\tau)$, b)$\phi_{\varepsilon u}(\tau)$, c)$\phi_{uu\varepsilon}(\tau)$, d)$\phi_{u^2\varepsilon^2}(\tau)$, e)$\phi_{\varepsilon\varepsilon u}(\tau)$
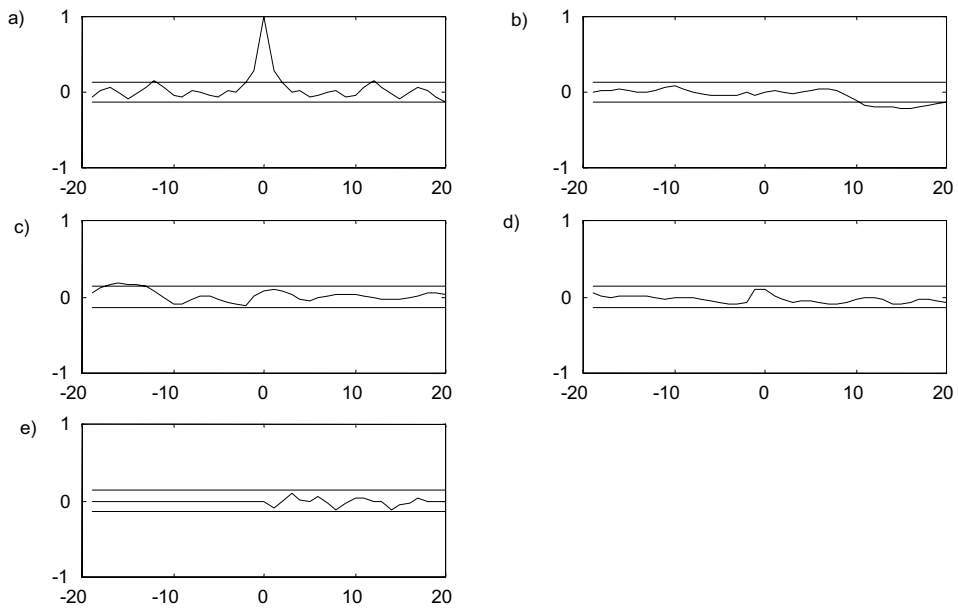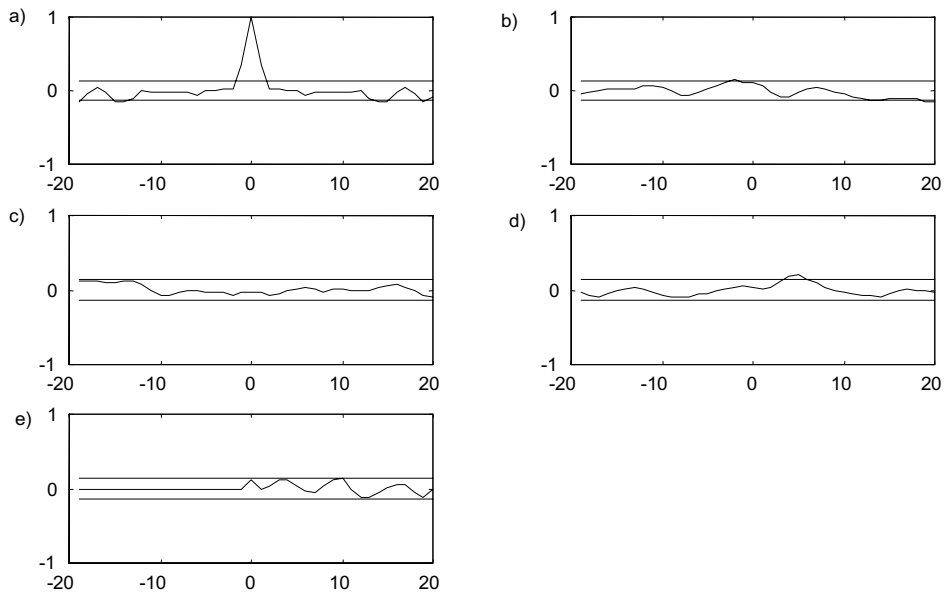

Figure 5   Correlation tests for thin-plate-spline function a)$\phi_{\varepsilon\varepsilon}(\tau)$, b)$\phi_{\varepsilon u}(\tau)$, c)$\phi_{uu\varepsilon}(\tau)$, d)$\phi_{u^2\varepsilon^2}(\tau)$, e)$\phi_{\varepsilon\varepsilon u}(\tau)$
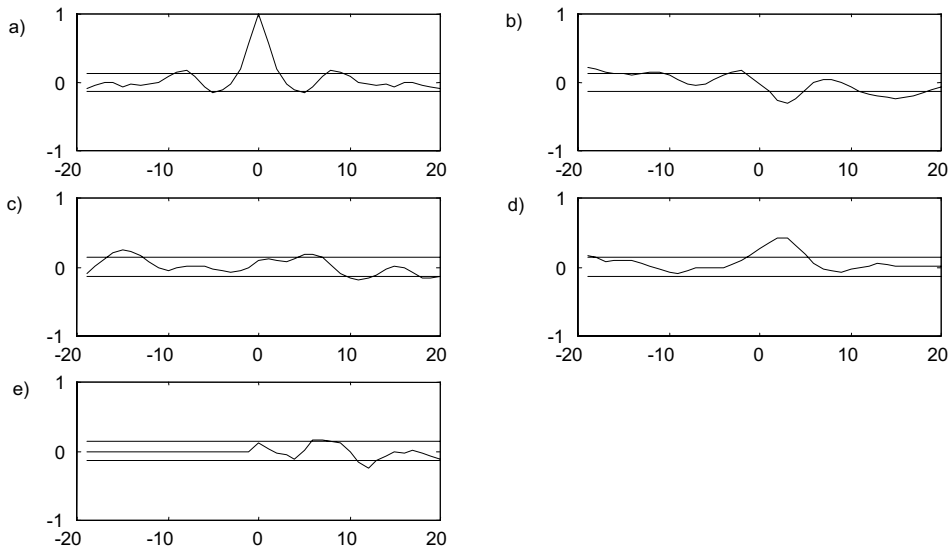
Figure 6 Correlation tests for multiquadratic function a)$\phi_{\varepsilon\varepsilon}(\tau)$, b)$\phi_{\varepsilon u}(\tau)$, c)$\phi_{uu\varepsilon}(\tau)$, d)$\phi_{u^2\varepsilon^2}(\tau)$, e)$\phi_{\varepsilon\varepsilon u}(\tau)$

Figure 7 Correlation tests for inverse multiquadratic function a)$\phi_{\varepsilon\varepsilon}(\tau)$, b)$\phi_{\varepsilon u}(\tau)$, c)$\phi_{uu\varepsilon}(\tau)$, d)$\phi_{u^2\varepsilon^2}(\tau)$, e)$\phi_{\varepsilon\varepsilon u}(\tau)$

5.2 Model 2: Dynamic Nonlinear System

The second system to be identified is governed by the difference equation

$$y(t + 1) = [0.3y(t) + 0.6y(t - 1) + 0.6\sin(\pi u(t)) + 0.4\sin(3\pi u(t))]/5.5 + e(t)$$

where the input $u(t)$ is chosen to be

$$u(t) = \sin(2\pi t / 250)$$

Five hundred input-output data pairs were generated. The input of the RBF network were given as $x(t) = [y(t-1)\ y(t-2)\ u(t-1)\ u(t2)]^{T}$ and the number of data is 500. The OLS algorithms identified RBF model with 10 centres. 250 data points were used as the estimation set and a further 250 data points were used as the testing set.

Figure 8 shows the actual system in comparison with the identified models with different RBF functions. It shows that all the identified models performed a good fitting to the actual process except for model with inverse multiquadratic basis function. The results in Table 2 show the corresponding variance and error index for
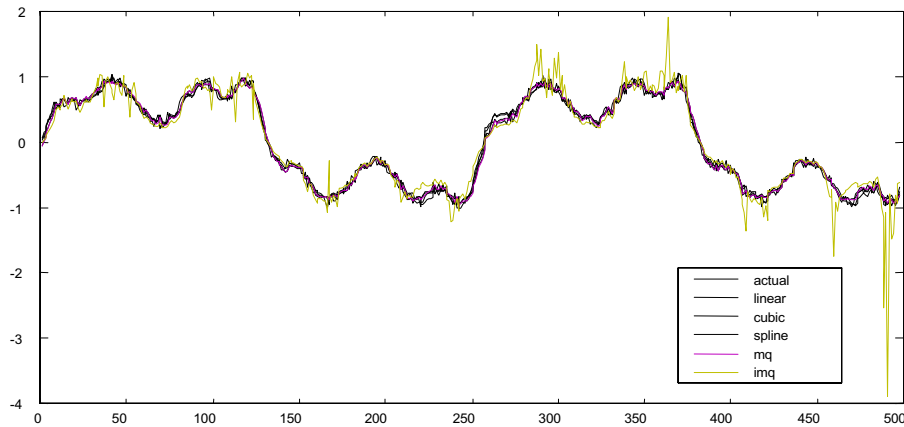

Figure 8   System output (solid data) superimposed on one-step prediction of RBF models with different functions

Table 2   Variance and error index for different RBF functions

RBF Functions	Variance		Error Index	
	Estimation	**Training**	**Estimation**	**Training**
Linear	0.0054	0.0060	0.109	0.115
Cubic	0.0059	0.0072	0.113	0.127
Spline	0.0058	0.0071	0.113	0.125
Multiquadratics	0.0054	0.0061	0.110	0.116
Inverse multiquadratics	0.0174	0.0870	0.196	0.216

different functions. As shown, RBF with inverse multiquadratic provides the largest variance and error index if compared with the others.

The correlation tests are shown in Figures 9,10,11,12 and 13 respectively. The correlation tests for cubic function indicates that $\phi_{\varepsilon\varepsilon}(\tau)$ is outside the band, showing that the process model is adequate but is biased. Generally, it is observed that most portions of the functions lie inside the 95% confidence and it can be concluded that the tests reveal that the models are adequate.

Results from model 1 indicate that all the basis function give good predictive accuracies and difficult to distinguish. For model 2, all the basis function except the cubic function gave comparable predictive accuracies. All results clearly showed that RBF network produced good prediction over the estimation set but the prediction over the data from test set gives adequate predictions since the radial basis function is considered as performing a curve-fitting operation over the estimation set.

The algorithm was also tested for larger number of centers but the data is not shown. With larger number of centres, the complexity of the network will be increased. A smaller number of error index will be produced and the one step ahead prediction also decreased. However, the results produced quite similar prediction. For the purpose of comparison and illustration, in this application the value of k is set to be 0.1 and the number of centres is set to be 10. The predictive accuracy of the estimation set will tend to increase but the results are not significantly distinguishable.
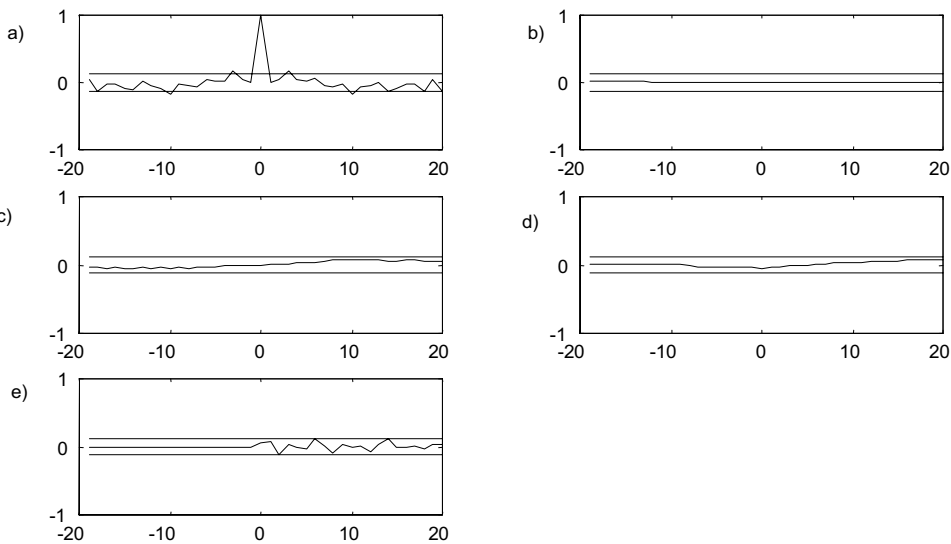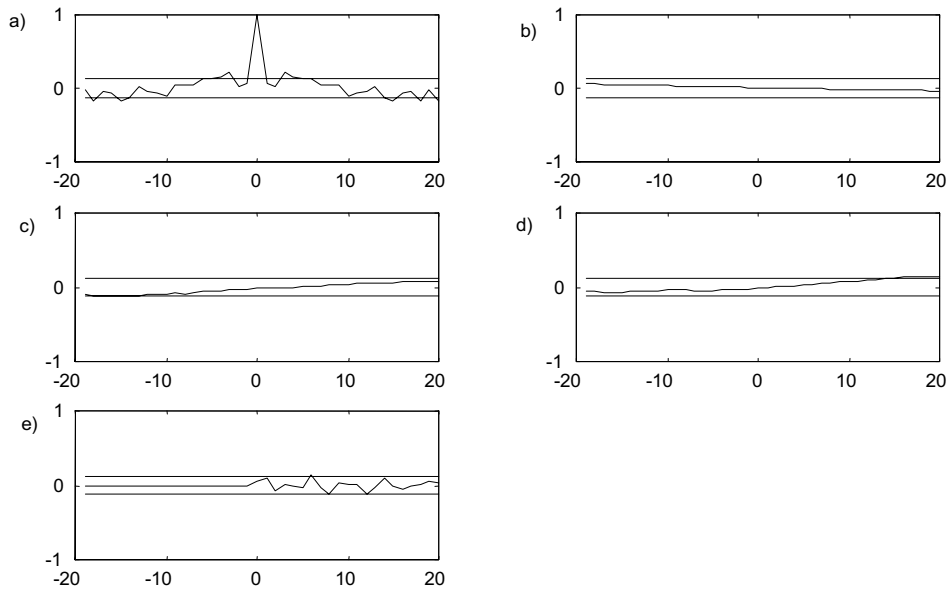

Figure 9   Correlation tests for linear function a) $\phi_{\varepsilon\varepsilon}(\tau)$, b) $\phi_{\varepsilon u}(\tau)$, c) $\phi_{uu\varepsilon}(\tau)$, d) $\phi_{u^2\varepsilon^2}(\tau)$, e) $\phi_{\varepsilon\varepsilon u}(\tau)$

ROBIAH AHMAD & HISHAMUDDIN JAMALUDDIN

Figure 10   Correlation tests for cubic function a)$\phi_{\varepsilon\varepsilon}(\tau)$, b)$\phi_{\varepsilon u}(\tau)$, c)$\phi_{uu\varepsilon}(\tau)$, d)$\phi_{u^2{'}\varepsilon^2}(\tau)$, e)$\phi_{\varepsilon\varepsilon u}(\tau)$
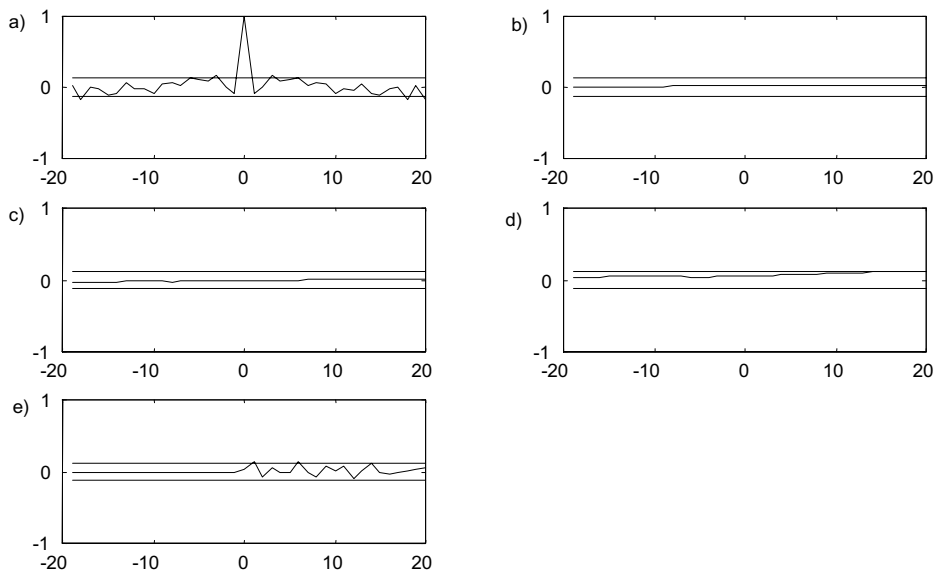
Figure 11   Correlation tests for thin-plate-spline function a)$\phi_{\varepsilon\varepsilon}(\tau)$, b)$\phi_{\varepsilon u}(\tau)$, c)$\phi_{uu\varepsilon}(\tau)$, d)$\phi_{u^2{'}\varepsilon^2}(\tau)$,
e)$\phi_{\varepsilon\varepsilon u}(\tau)$

Figure 12   Correlation tests for multiquadratic function a)$\phi_{\varepsilon\varepsilon}(\tau)$, b)$\phi_{\varepsilon u}(\tau)$, c)$\phi_{uu\varepsilon}(\tau)$, d)$\phi_{u^2\varepsilon^2}(\tau)$, e)$\phi_{\varepsilon\varepsilon u}(\tau)$
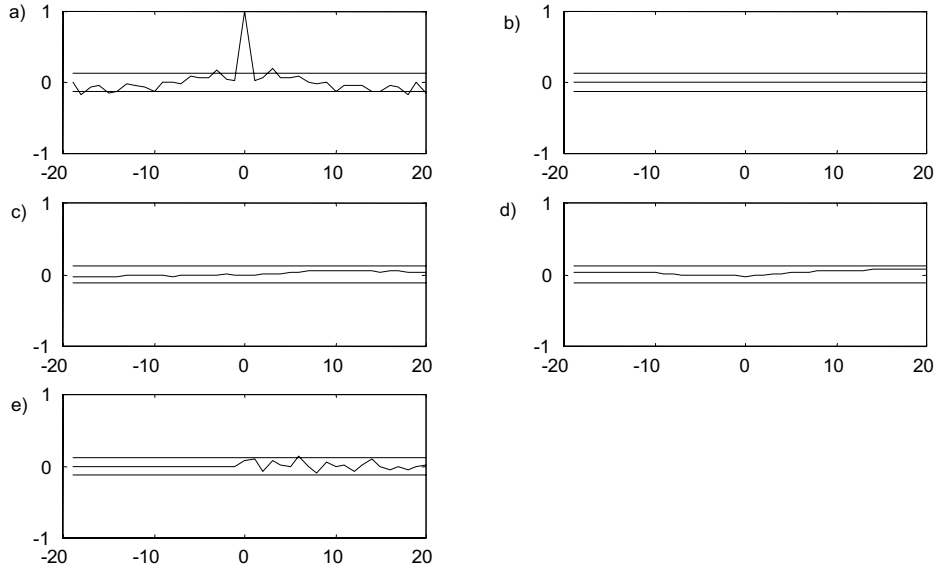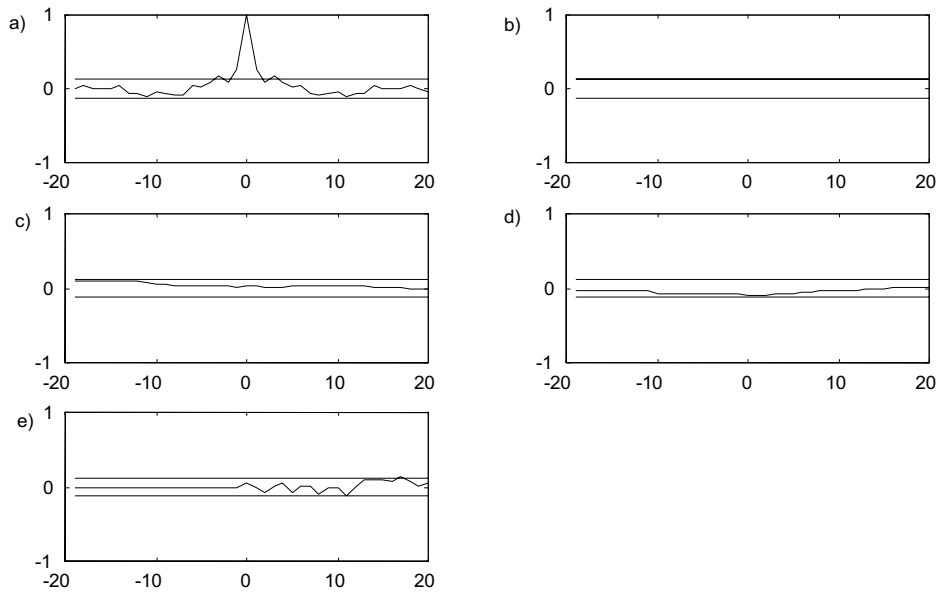

Figure 13   Correlation tests for inverse multiquadratic function a)$\phi_{\varepsilon\varepsilon}(\tau)$, b)$\phi_{\varepsilon u}(\tau)$, c)$\phi_{uu\varepsilon}(\tau)$, d)$\phi_{u^2\varepsilon^2}(\tau)$, e)$\phi_{\varepsilon\varepsilon u}(\tau)$

6.0 CONCLUSION

In this paper, the RBF network model was used for the purpose of identifying the nonlinear dynamic systems. Orthogonal least square algorithm was used to identify adequate network structures for RBF network models. Five different basis functions were considered: linear, cubic, thin-plate-spline, multiquadratic and inverse multiquadratic. The use of *ERR*, the by-product of OLS algorithm, provides appropriate set of RBF centers and estimates of the corresponding parameters can be determined in an efficient manner. The fundamental properties of the OLS were highlighted, illustrating the advantages of this approach to identifying RBF network models. The predictive accuracies of various basis functions were also investigated.

The properties of different basis functions for RBF networks have been investigated. For both examples, the basis functions give similar predictive accuracies and the results are comparable. However, for both systems, inverse multiquadratic function was shown to produce relatively poor accuracy as compared to other basis functions. The results also indicate that the performance of the algorithm depends very much on the different systems that were used.

REFERENCES

[1]   Chen, S., and S. A. Billings. 1989. Representations of non-linear systems: the NARMAX model. *International Journal of Control*. **49**(3): 1013 – 1032.

[2]   Desrochers, A., and S. Mohseni. 1984. On determining the structure of a non-linear system. *International Journal of Control*. **40**(5): 923 – 938.

[3]   Schetzen, M. 1980. *The Volterra and Wiener theories for nonlinear systems*. New York: John Wiley and Son Inc.

[4]   Leontaritis, I. J., and Billings. 1985. Input-output parametric models for non-linear systems, Part I:deterministic non-linear systems. *International Journal of Control*. **41**(2): 303 – 328.

[5]   Haber, R., and H. Unbehauen. 1990. Structure identification of nonlinear dynamics systems- a survey on input/output approaches. *Automatica*. **26**(4): 651 – 677.

[6]   Chen, S., C. F. N. Cowan, and P. M. Grant. 1991. Orthogonal least squares learning algorithm for radial basis function networks. *IEEE Transactions on Neural Networks*. **2**(2): 302 – 209.

[7]   Billings, S. A. 1980. Identification of nonlinear systems: a survey. *IEE Proceedings*. **127**(6): 272 – 285.

[8]   Soderstrom, T., and P. Stoica. 1989. System Identification, United Kingdom: Prentice Hall International Ltd.

[9]   Chen, S., and S. A. Billings. 1992. Neural networks for nonlinear dynamic systems modelling and identification. *International Journal of Control*. **56**(2): 319 – 346.

[10] Powell, M. J. D. 1985. Radial basis function for multivariable interpolation: a review, *IMA Conference on Algorithms for the Approximation of Functions and Data*, RMCS, Shrivenham, pp. 143 – 167.

[11] Powell, M. J. D. 1987. Radial basis function approximations to polynomials, *Proceedings of 12th Biennial Numerical Analysis Conference*, Dundee, pp. 223 – 241.

[12] Chen, S., S. A. Billings, and W. Luo. 1989. Orthogonal least square methods and their application to nonlinear system identification. *International Journal of Control*. **50**(5): 1873 – 1896.

[13] Korenberg, M., S. A. Billings, Y. P. Liu, and P. J. McIlroy. 1988. Orthogonal parameter estimation algorithm for non-linear stochastic systems. *International Journal of Control*. **48**(1): 193 – 210.

[14] Billings, S. A., and W. S. F. Voon. 1986. Correlation based model validity tests for non-linear models. *International Journal of Control*. **44**(1): 235 – 244.

[15] Box, G. E. P., G. M. Jenkins, and G. C. Reinsel. 1994. *Time series Analysis Forecasting and Control*. New Jersey: Prentice-Hall Inc.