# Improving the Radial Basis Function Networks for Homogeneous Nonstationary Time Series Prediction

Eng-Siong CHNG †, Sheng CHEN ‡ and Bernard MULGREW †

† *Dept. of Electrical Eng., The University of Edinburgh, Edinburgh EH9 3JL, Scotland, U.K., Tel/Fax: +44 [31] 650 5655 / 650 6554. E-Mail: echng@ee.ed.ac.uk*

‡ *Dept. of Electrical and Electronic Eng., The University of Portsmouth, Anglesea Building, Portsmouth PO1 3DJ, U.K. E-Mail: schen@ee.port.ac.uk*

**Abstract.** The radial basis function (RBF) network has become a popular choice of neural network to be used for nonlinear time series prediction [1–3]. Although the results have been encouraging for modelling time invariant nonlinear systems, it is difficult to achieve the same level of success for tracking nonstationary signals [4]. In this article, we present a method of modifying the classical RBF networks, which improves the predictive accuracy for nonlinear and nonstationary data.

## 1 Introduction

Although the RBF network has achieved considerable success in the application to stationary nonlinear time series prediction, it is unable to achieve the same level of success for tracking nonstationary series. This is because the RBF network, like many other neural network models, does not characterise temporal variability well. Since real-world signals are often not only nonlinear but also nonstationary, it is desired to develop predictors which can handle signals that exhibit both such characteristics.

To improve the predictive performance for nonstationary data, we propose a gradient RBF (GRBF) network which is a modification of the classical RBF network. In the classical RBF network, the centers of the hidden nodes can be interpreted as prototype vectors which are used to sense the presence of the input pattern. That is, if a center matches the network input vector, the corresponding hidden node will fire strongly. While in the GRBF network, a hidden node's function is to sense the presence of a prototype vector's gradient. This significantly improves the predictive capability of the network in the situation where nonstationarity of the signal is due to the variations of mean and trend.

In using this GRBF network, we are exploiting the idea that, by performing a suitable difference operation on a nonstationary signal, the resulting signal becomes stationary. This idea is used in the auto-regressive integrated moving average (ARIMA) model [5] for linear prediction of nonstationary signals. By incorporating a similar mechanism into the RBF network, we can create a network model that is capable of dealing with nonlinear and nonstationary signals.

## 2 The Gradient Radial Basis Function Network

The GRBF network, like the RBF network, is a single-hidden-layer feedforward neural network [3]. It consists of a input layer with $M$ input elements, a hidden layer with $K$ hidden nodes and, in this study, an output layer with 1 node. There are however two main differences between the RBF network and the GRBF case.

Firstly, the input vector to the RBF network contains past samples of the time series $\{y_i\}$ while the input vector to the GRBF network is generated by differencing the raw data $\{y_i\}$. The order of differencing determines the order of the GRBF network. For example, if the input vector to the RBF network at time $i$ is given by

$$\mathbf{x}_i = [y_{i-1}, y_{i-2}, \cdots, y_{i-M}]^T \tag{1}$$

, then the input vector of the 1st-order GRBF network at time $i$ is

$$\begin{aligned} \mathbf{x}'_i &= \mathbf{x}_i - \mathbf{x}_{i-1} \\ &= [y_{i-1} - y_{i-2}, \cdots, y_{i-M} - y_{i-M-1}]^T \end{aligned} \tag{2}$$

The elements of $\mathbf{x}'_i$ show the rate of change in the time-series trajectory for the past $M$ samples.

Secondly, the function of the hidden node for the GRBF network is different from that of the RBF network. Figure 1 depicts the structure of the 1st-order GRBF network. Although the Gaussian function still serves as the nonlinear function which compares the similarity of the input vector to the hidden node's center, the response of the Gaussian function is now multiplied by an additional term $(y_{i-1} + \delta)$. The response of the $j$-th hidden node of a 1st-order GRBF network to the input vector $\mathbf{x}'_i$ is therefore given by

$$\phi'_{ij} = exp(-\alpha||\mathbf{x}'_i - \mathbf{c}'_j||) \times (y_{i-1} + \delta_j) \qquad (3)$$

where $\mathbf{c}'_j$ is the $M$-dimensional center vector of the $j$-th hidden node, $\alpha$ is a width parameter, and $\delta_j$ is a constant value associated with the center.

The term $(y_{i-1} + \delta_j)$ can be interpreted as a local single-step prediction of $y_i$ by the $j$-th hidden node. From (3), if the input vector is similar to the $j$-th center, the value of the Gaussian function will be close to 1.0 and the predictor $(y_{i-1} + \delta_j)$ becomes fully active. As in the case of the RBF network, the output layer is a linear combiner with weights $h_j$, $1 \leq j \leq K$. Similar to the selection of RBF centers, $\mathbf{c}'_j$ and $\delta_j$, $1 \leq j \leq K$, can be selected during training from the training data set $\{\mathbf{x}'_k\}_{k=1}^{N}$, where $N$ is the number of training data. For each training input vector $\mathbf{x}'_k$, define $d_k = y_k - y_{k-1}$. If $\mathbf{x}'_k$ is chosen as the $j$-th center $\mathbf{c}'_j$, the values of $\delta_j$ is set to $d_k$. This ensures that the $j$-th hidden node is a perfect predictor of $y_k$.

The rationale behind the GRBF model become obvious when the network performs predictive operation. Each hidden node compares the network input vector $\mathbf{x}'_i$ with its center $\mathbf{c}'_j$. The Gaussian response of each hidden node indicates the degree of matching between $\mathbf{x}'_i$ and $\mathbf{c}'_j$. The hidden nodes thus sense the gradient of the time series rather than the series itself as in the case of the RBF model. The term $(y_{i-1}+\delta_j)$ also has a clear geometric meaning; if the $j$-th center $\mathbf{c}'_j$ matches the gradient $\mathbf{x}'_i$ of the series, $(y_{i-1} + \delta_j)$ is likely to be a very good prediction of $y_i$. Although the complexity of a GRBF hidden node is greater than that of a RBF hidden node, the GRBF has better generalisation property, particularly in predicting nonstationary time series. This often results in a smaller GRBF network. Therefore, the overall complexity of the GRBF network may not necessarily be greater than that of the RBF network in practical applications.

# 3 Simulation Results

We present some simulation results of time series prediction using the RBF and GRBF predictors. Initial full models were created by using all the available data in the training set as RBF and/or GRBF centers. Some linear terms were also included into the full models. Subset models were then selected from these large full models using the OLS [2] scheme, and used to evaluate single-step and multi-step prediction performance.

## 3.1 Results for Stationary Series

The Mackey-Glass (figure 2) chaotic time-series was used to evaluate model predictive performance. Data samples of point 100-600 were used as the training set and samples 601 to 1100 were used as the validation set. The values of $M$ was chosen to be 6, and the width of Gaussian function was set to $\alpha = 1.0$. The following types of models were used:

i) **L-model** - The linear model of order 50.

ii) **L0-model** - A combination of the linear model and the classical RBF model.

iii) **L01-model** - A combination of the linear model, the classical RBF and 1st-order GRBF models.

The results of single-step performance for the predictors in training phase are shown in figure 3, where the vertical axis indicates the normalised mean square error (NMSE) in dB. As expected, as the size of each selected subset model increases, the accuracy of the model continued to improve. However, the rate of improvement was not the same for each model. The predictors with GRBF expansion, i.e. **L01-model**, achieved better error reduction with a smaller model size. This GRBF subset model also performed better on the validation set compared with the linear and classical RBF models, as can be seen in figure 4.

## 3.2 Results for Nonstationary Series

To examine how the predictors behave for nonstationary series, we used a modified Mackey-Glass time-series (figure 5). This new series was formed by adding sinusoid with amplitude 0.3 and a period of 3000 samples to the Mackey-Glass time series used in the previous example. As the training data were formed from samples 100 to 600 and the validation data consisted of samples from 601-1100, the predictors were trained without being exposed to the change in the level and trend of the test data. The results for the single-step prediction in the validation phase (figure 6) suggest that the GRBF network can perform better than the classical RBF network in a nonstationary environment.

## 4 Conclusions

We have presented a GRBF network for nonlinear and nonstationary time series prediction. The hidden layer of this GRBF network is designed to respond to the gradient of time-series rather than the trajectory itself. This can usually improve predictive accuracy, particularly for homogeneous nonstationary time series as are demonstrated in the simulation results. Although the discussion was based on time series prediction, this GRBF network can be applied to other signal processing applications.

## References


[1] T.POGGIO and F.GIROSI, "Networks for approximation and learning", *Proc. IEEE*, vol. 78, pp. 1481–1497, 1990.

[2] S.CHEN, C.F.COWAN, and P.M.GRANT, "Orthogonal least squares learning algorithm for radial basis function networks", *IEEE Trans. Neural Networks*, vol. 2, pp. 302–309, 1991.

[3] D.S. BROOMHEAD and D.LOWE, "Multivariable functional interpolation and adaptive networks", *Complex Systems*, vol. 2, pp. 321–355, 1988.

[4] E.LEVIN, "Hidden control neural architecture modelling of nonlinear time varying systems and its applications", *IEEE Trans. Neural Networks*, vol. 4, pp. 109–116, 1993.

[5] G.E.P.BOX and G.M.JENKINS, *Time Series Analysis : Forecasting and Control*, Holden-Day, 1976.


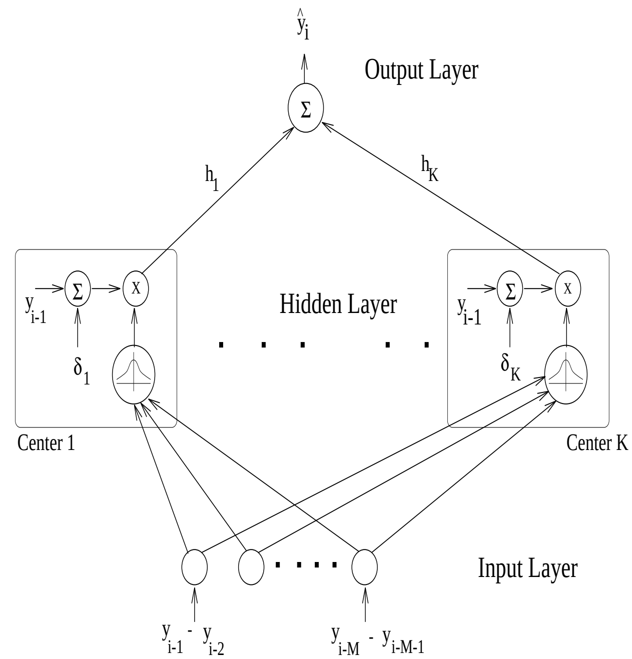


Figure 1: Topology of 1st-order GRBF network

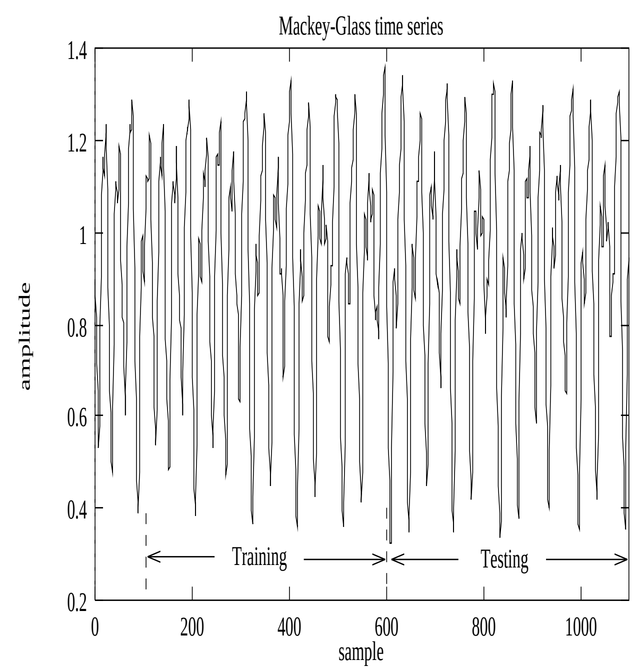


Figure 2: Mackey-Glass time series

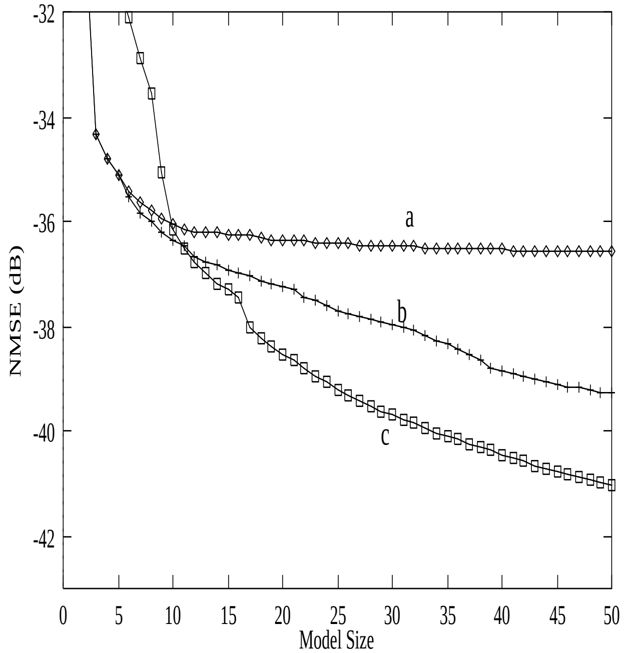


Figure 3: Performance of predictors in training phase for Mackey-Glass series
a) Linear model, b) Linear & RBF model, c) Linear, RBF & 1st order GRBF model

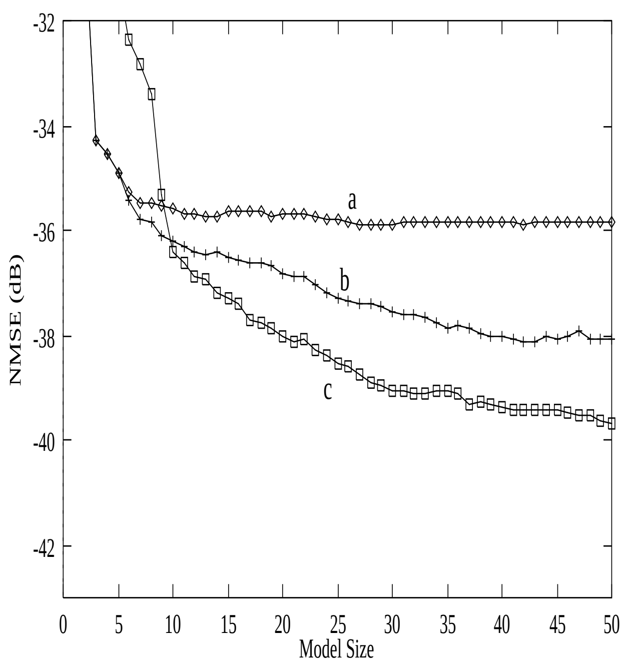


Figure 4: Performance of predictors in testing phase for Mackey-Glass series
a) Linear model, b) Linear & RBF model, c) Linear, RBF & 1st order GRBF model

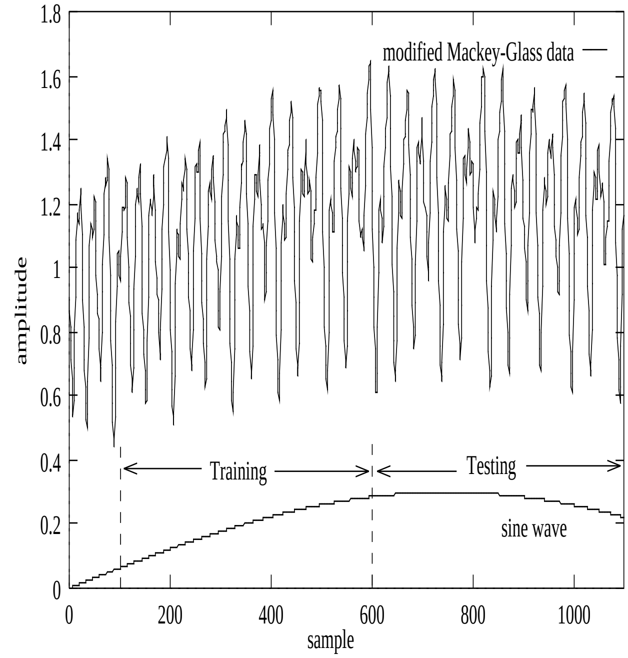


Figure 5: Modified Mackey-Glass time series

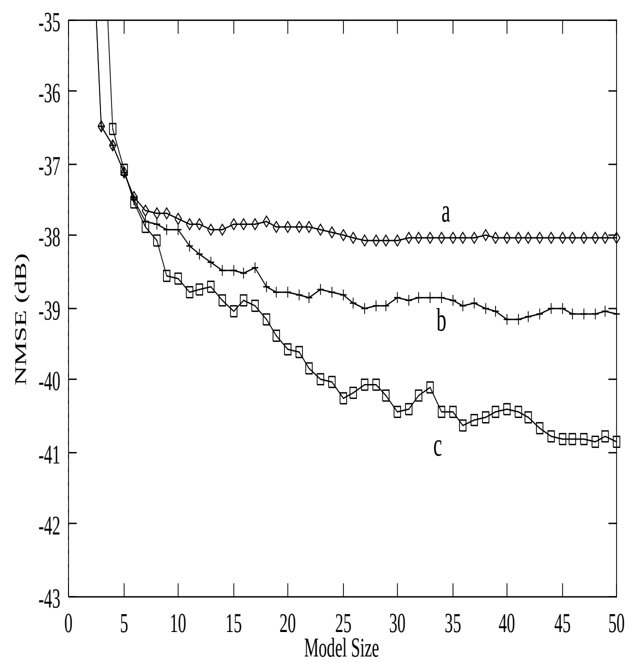


Figure 6: Performance of predictors in testing phase for modified Mackey-Glass series
a) Linear model, b) Linear & RBF model,
c) Linear, RBF & 1st order GRBF model.