Water Resources Systems—Water Availability and Global Change (Proceedings of symposium HS02a held during IUGG2003 at Sapporo, July 2003). IAHS Publ. no. 280. 2003.

Radial basis function network for prediction of hydrological time series

A. W. JAYAWARDENA
Department of Civil Engineering, The University of Hong Kong, Hong Kong, China
hrecjaw@hkucc.hku.hk

PENGCHENG XU
Academy of Mathematics and System Sciences, Chinese Academy of Sciences, Beijing, China

W. K. LI
Department of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong, China

Abstract In this study, a network using radial basis functions as the mapping function in the evolutionary equation for prediction of time series is presented. A radial basis function network requires the determination of the number of centres of the radial basis functions, their receptive field widths, and the linear weights of the network output layer. Methods to estimate the widths of the receptive fields, and the number of centres for the radial basis functions are introduced in the study. The latter is based on the concept of the Generalized Degrees of Freedom. The linear weights are determined by the least squares method. The predictions by the proposed method when compared with the actual values of four hydrometeorological data sets, are better than those by the traditional approach of fixing the number of centres.

Key words Chao Phraya River, chaos; generalized degrees of freedom; Mekong River; phase space; radial basis functions; S-index

INTRODUCTION

In the traditional approach to modelling hydrological time series, the series is considered to originate from a stochastic process, which, at least in theory, has an infinite number of degrees of freedom. In such cases, linear models of the Box-Jenkins type auto-regressive moving average (ARMA) have been used for the analysis and prediction of hydrological time series by many researchers over the years (Lawrance & Kottegoda, 1977; Box *et al.*, 1994; Jayawardena & Lai, 1989). Recently however, it has been realized that certain types of time series, which appear to be evolving from stochastic processes, can in fact be the outcome of fully deterministic processes. By treating the system that generates the time series as a deterministic one, it is possible to gain an understanding of the associated complicated dynamics, and to make more realistic short-term predictions. Such systems can exhibit stable properties, which are predictable with certainty at times, but may become "chaotic" under certain initial conditions. The study of chaotic systems has drawn the attention of many researchers in many disciplines in the recent past (Farmer & Sidorowich, 1987; Abarbanel *et al.*, 1990; Sugihara & May, 1990; Smith, 1992; Jayawardena & Lai, 1994; Sivakumar *et al.*, 1999).

In a deterministic system, predictions can generally be made using an evolutionary equation in which the future value is considered to be dependent upon present and past values. The prediction process therefore involves an accurate estimation of the mapping function, which transforms the present and past values to the future value. In a chaotic system, the predictive power is lost very quickly because of sensitivity to initial conditions.

The mapping function can be estimated using local models in which the function approximation at each time step is done from data sets of the local neighbourhood only in a piece-wise manner, or global models in which the function approximation is done for the whole domain. Local models include linear or polynomial function approximations in the local neighbourhoods whereas global models are generally of the polynomial type, although radial basis functions also have been used. In general, local models are more accurate locally than global models but they need to be trained step by step. The scope in this study is restricted to local models only. They include local linear models and local radial basis function models.

When radial basis functions are used in the local predicting model, a point set called radial basis points must be given. The main problem is how to obtain these points and how many points should be used for a better model. In this study, we introduce a method to select the radial basis points and to determine how many points would be needed for a better model using the concept of generalized degrees of freedom. The results obtained are encouraging.

RBF NETWORK

The first step in the analysis of a chaotic time series is to embed time series into a phase space. This requires two parameters: the time delay, τ, and the embedding dimension, d_e. Despite numerous suggestions for the time delay, there is no rigorous method of determining its optimal value. In this study, a value of unity was assumed for all data sets. There are several methods of estimating the embedding dimension (Grassberger & Procaccia, 1983; Abarbanel, 1995), but the False Nearest Neighbour method (Abarbanel, 1995) is used in this study.

The RBF (radial basis function) network is a feed-forward type of artificial neural network that consists of input, hidden and output layers of nodes. A general RBF network has one input node, L hidden nodes and one output node. Employing the widely used Gaussian function as the radial basis function for the network, the output of an RBF network can be described by the following equation:

$$x(t+\tau)=\sum_{i=1}^{M} w_i \exp\left(-\frac{\|\mathbf{Y}(t)-\mathbf{C}_i(t)\|^2}{2\sigma^2}\right) \tag{1}$$

Here the input, $\mathbf{Y}(t)$, and the centres, $\mathbf{C}_i(t)$, $1 \le i \le M$, both belong to the reconstructed phase space, and are respectively, denoted by:

$$\mathbf{Y}(t)=(x(t),x(t-\tau),x(t-2\tau),\cdots,x(t-(d_e-1)\tau)) \tag{2}$$

$$\mathbf{C}_i(t_i)=(c_i(t),\ c_i(t-\tau),\ \ldots,\ c_i(t-(d_e-1)\tau) \tag{3}$$

The output, $x(t+\tau)$, is the predicted signal of the input $\mathbf{Y}(t)$. Parameters w_i, $1 \le i \le M$ (M the number of centres), are the weights of the RBF network, σ the width of the RBF function, and $\|.\|$ the Euclidean norm. Several methods exist for the choice of centres. They include random selection of fixed centres (Broomhead & Lowe, 1998), self-organized selection of centres (Moody & Darken, 1989), supervised selection of centres (Poggio & Girosi, 1990), and the regularized interpolation exploiting the connection between and RBF network and Watson–Nadaraya regression kernel (Yee, 1998). In this study we use the following method to obtain the centres of the radial basis function network. Given a $\varepsilon > 0$, choose a point $\mathbf{C}_1(t)$ from the reconstructed phase space as a centre.

If there is a point $\mathbf{Y}(t)$ in the reconstructed phase space, such that $\min_{1\le j\le i}(\| \mathbf{Y}(t) - \mathbf{C}_j(t) \|) > \varepsilon$, then $\mathbf{C}_{i+1}(t) = \mathbf{Y}(t)$, where i indicates the number of centres.

The small parameter ε ensures that there are enough centres to use in the local radial basis function model. But too small value of ε makes two centres so close that they give similar functions in the local model. In this study it is assumed that $\varepsilon = 0.02 d_{\max}$, where $d_{\max} = \max(\| \mathbf{Y}_1(t) - \mathbf{Y}_2(t) \|)$. $\mathbf{Y}_1, \mathbf{Y}_2$ are two points in phase space.

TRAINING OF THE RBF NETWORK

Given a set of centres and a radial basis function, the training of a RBF network involves the determination of the network weights w_i, the width of the RBF, σ, and the number of the centres (a sub-set of the centre set). The parameters w_i can be determined by the least squares method for given values of σ and the number of centres. It can be done as follows: given a set of N nearest neighbours, $\mathbf{Y}^r(t)$; $r = 1,2,3,\cdots,N$ of $\mathbf{Y}(t)$ where:

$$\mathbf{Y}^r(t) = (x^r(t), x^r(t-\tau), \cdots, x^r(t-(d_e-1)\tau)) \tag{4}$$

which, at time level $t+\tau$, will evolve to $\mathbf{Y}^r(t+\tau)$, which will be in the neighbourhood of $\mathbf{Y}(t+\tau)$. The weights w_i are then determined by minimizing:

$$\sum_{r=1}^{N} \left| x^r(t+\tau) - \sum_{i=1}^{M} w_i \exp\left(-\frac{\| \mathbf{Y}^r(t) - \mathbf{C}_i(t) \|}{2\sigma^2}\right) \right|^2 \tag{5}$$

where $x^r(t+\tau)$ is the evolved point at time level $t+\tau$ of $x^r(t)$ at time level t. This is known.

It is difficult to optimize both the width and the number of centres. In general, the number of the centres is fixed and the width of the function is optimized by systematically testing several values. In this study, we use an iterative method to get the best value of the width and the weights of the RBF network for a given number of centres. The generalized degree of freedom (GDF) is used to evaluate the number of centres.

First we introduce the method to estimate the best width σ and the weights of the radial basis function for the given number of centres. We rewrite equation (1) in matrix form as:

$$\mathbf{V} = \mathbf{WU} + \mathbf{E} \tag{6}$$

where

$$\mathbf{W} = (w_1, w_2, ..., w_M)$$

$$\mathbf{V} = (x^1(t+\tau), x^2(t+\tau), \cdots, x^N(t+\tau))$$

$$\mathbf{U} = \begin{bmatrix} \phi(\|\mathbf{Y}^1 - \mathbf{C}_1\|) & \phi(\|\mathbf{Y}^2 - \mathbf{C}_1\|) & \cdots & \phi(\|\mathbf{Y}^N - \mathbf{C}_1\|) \\ \phi(\|\mathbf{Y}^1 - \mathbf{C}_2\|) & \phi(\|\mathbf{Y}^2 - \mathbf{C}_2\|) & \cdots & \phi(\|\mathbf{Y}^N - \mathbf{C}_2\|) \\ \vdots & \vdots & \vdots & \vdots \\ \phi(\|\mathbf{Y}^1 - \mathbf{C}_M\|) & \phi(\|\mathbf{Y}^2 - \mathbf{C}_M\|) & \cdots & \phi(\|\mathbf{Y}^N - \mathbf{C}_M\|) \end{bmatrix}$$

Here the vector $\mathbf{E}$ is an error term, and ϕ is a function that contains the weights.

If an initial value of the width σ_n is assumed, then, the value $\mathbf{W}$ can be obtained by the least squares method as:

$$\mathbf{W}^n = \mathbf{VU}^{\mathrm{T}}(\mathbf{UU}^{\mathrm{T}})^{-1} \tag{7}$$

Then the next better estimate of the width value is:

$$\sigma_{n+1} = \sigma_n - \frac{A_n}{B_n} \tag{8}$$

where:

$$A_n = -\frac{1}{\sigma_n^2}\sum_{i=1}^{N}(x^r(t+\tau) - \sum_{j=1}^{M} w_j^n \phi(\|X^r - C_j\|))(\sum_{j=1}^{M} w_j^n \phi(\|X^r - C_j\|)\|X^r - C_j\|^2)$$

$$\mathrm{B_n} = -\frac{1}{\sigma_n^4}\sum_{i=1}^{N}(x^r(t+\tau) - \sum_{j=1}^{M} w_j^n \phi(\|X^r - C_j\|))(\sum_{j=1}^{M} w_j^n \phi(\|X^r - C_j\|)\|X^r - C_j\|^4)$$
$$+\frac{1}{\sigma_n^4}\sum_{i=1}^{N}(\sum_{j=1}^{M} w_j^n \phi(\|X^r - C_j\|)\|X^r - C_j\|^2)^2$$

In fact, the above procedure to estimate a better value of σ_{n+1} is just a standard Newton iteration procedure for the function (equation 5) with known w_i. By doing this iteration until two successive values of σ converge, better estimates of both the weights and width for the local RBF network can be obtained.

Next we show how to choose the best number of centres by using the generalized degrees of freedom method (GDF). Since the general linear method of the GDF cannot be used for this nonlinear modelling procedure, the Monte Carlo method introduced by Ye (1988) is employed here. According to Ye (1998), the unbiased estimate of the variance is given by:

$$v^2(M) = \frac{(\mathbf{V} - \hat{\boldsymbol{\mu}})(\mathbf{V} - \hat{\boldsymbol{\mu}})^T}{N - D} \tag{9}$$

where $\hat{\boldsymbol{\mu}}$ is the estimated value of the mean vector of $\mathbf{V}$ and D is the degrees of freedom. This provides a tool to evaluate the goodness of the model with the chosen number of centres.

PREDICTION BY RBF NETWORK

In order to make a prediction by the RBF model, we choose M centres nearest to $Y(t)$ from the centre set chosen before. In this study, $M = 11, 12, ...20$ were found to be suitable. The number of neighbours N needed to train the network was assigned a value of 30, which is not the only choice, but one that has to be greater than the number of centres. The weights, w_i's and the widths of the radial basis functions are obtained by least squares estimation as described before. By comparing the estimates of the variances v^2 for these modelling procedures, the best one that gives the minimum v^2 can then be selected. Lead time predictions can then be obtained for the desired number of time steps.

APPLICATION AND RESULTS

The hydrometeorological data sets used in this study include the daily discharges of the Chao Phraya River at Nakhon Sawan (15.67°N, 100.2°E, basin area 110 569 km^2, GRDC # 2964100) in Thailand for the period April 1978 to March 1994, the daily discharges of the Mekong River at Nong Khai (17.87°N, 102.72°E, basin area, 302 000 km^2, GRDC # 2969090) in Thailand, and of the same river at Pakse (15.12°N, 108.80°E, basin area, 545 000 km^2, GRDC # 2469260) in Lao for the period April 1980 to December 1991, and the monthly mean sea surface temperature (SST) anomaly over the region bounded approximately by 6°N–6°S and 180°W–90°W for the period January 1872 to December 1986, which has been defined as the S index by Wright (1984) and used to identify climatic anomalies attributed to El Niño and the Southern Oscillation. The first three data sets were obtained from the Global Runoff Data Centre (GRDC) in Germany and the last one from a table compiled by Wright (1989). A few missing records of the data sets were replaced by the long-term averages. All the data sets used were noise reduced by methods described in a separate study by the authors (Jayawardena & Gurung, 2000) before using in the modelling and prediction described in this study. Basic information of the data sets is given in Table 1.

For all the data sets used, the method in which the number of centres and the width of the RBF are optimized on the basis of the generalized degrees of freedom gives better predictions in the short term when compared with the method in which the number of centres and the width of RBF are fixed. A summary of the results is given in Table 2 and Table 3, which show the cumulative mean square errors for 10 steps as well as for 25 steps of lead-time prediction. For the 25-step lead-times, they are generally one order less than those using fixed N and linear model.

Table 1 Basic information of the data used in the study.

Data	Number of data	Embedding dimension	Time delay	Origin for prediction
Mekong (Nong Khai)	4292	3	1	4000
Mekong (Pakse)	4292	3	1	4000
Chao Phraya	5844	3	1	4000
SST	1380	3	1	1000

Table 2 Cumulative mean square errors for 10 lead time steps.

Data	RBFM with chosen centres	RBFM with given centres	Linear model
Mekong (Nong Khai)	1.52E-05	2.33E-05	3.29E-04
Mekong (Pakse)	6.08E-06	2.12E-05	2.70E-05
Chao Phraya	5.43E-05	7.84E-05	5.33E-03
SST	1.80E-03	4.71E-03	9.93E-03

Table 3 Cumulative mean square errors for 25 lead time steps.

Data	RBFM with chosen centres	RBFM with given centres	Linear model
Mekong (Nong Khai)	2.55E-03	1.16E-02	6.78E-03
Mekong (Pakse)	5.34E-05	3.68E-04	6.56E-04
Chao Phraya	4.68E-03	1.88E-02	5.41E-03
SST	4.45E-02	1.16E-01	3.38E-02

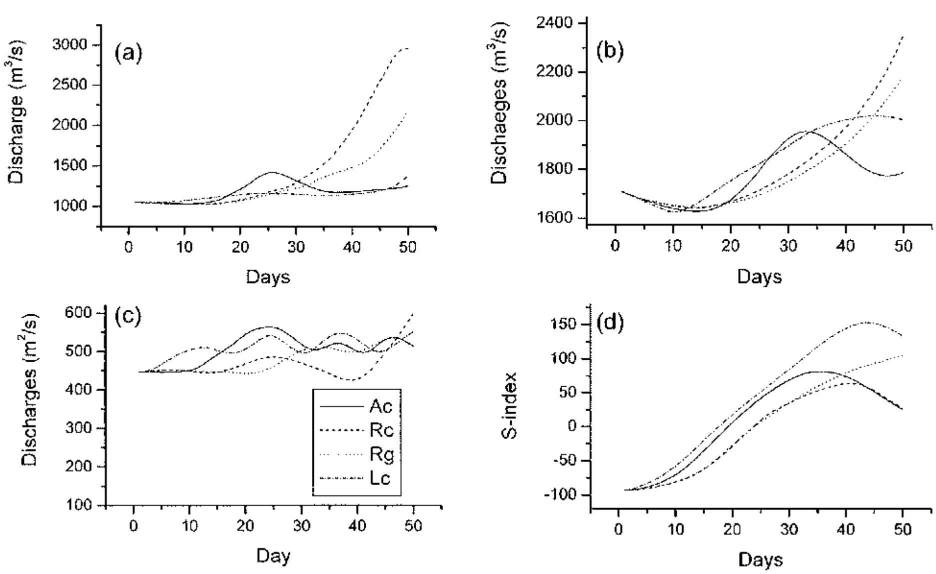

Fig. 1 Actual (observed) and predicted values for: (a) discharge in Mekong River at Nong Khai (origin for prediction t = 4000); (b) discharge in Mekong River at Pakse (origin for prediction t = 4000); (c) discharge in Chao Phyraya River at Nakhon Sawan (origin for prediction t = 4000); (d) sea surface temperature anomaly (S-index) (origin for prediction t = 1000) with Ac: actual or observed value; Rc: radial basis network model with chosen centres; Rg: radial basis network with given centres; Lc: local linear model with chosen neighbours.

The comparisons of predictions by the three methods are shown in Fig. 1(a)–(d) for the Chao Phraya data, Mekong data at Nong Khai, Mekong data at Pakse and SST data, respectively. These clearly indicate that predictions are possible in the short term, but the predictive power is not long lasting. Another point of important aspect in this kind of prediction is the effect of the predicting origin. Several origins have been tried for the data sets, and for most cases, the results are still valid.

CONCLUSION

In this study, a method, based on the generalized degrees of freedom method, for optimizing the number of centres and the width of the RBF in a RBF network is proposed. The results of the applications are encouraging, and it can be concluded that the GDF criterion certainly leads to a better RBF network. The main finding in this study is that the number of centres and the width of RBF needed for modelling in the phase space can be best determined by the GDF method.

Acknowledgements This work is partially supported by the Hong Kong Research Grants Council Grant no. 7003/97E. Their financial support is gratefully appreciated.

REFERENCES

Abarbanel, H. D. I. (1995) *Analysis of Observed Chaotic Data* (2nd edn). Springer, New York, USA.

Abarbanel, H. D. I., Brown, R. & Kadtke, J. B. (1990) Prediction in chaotic nonlinear systems: methods for time series with broad band Fourier spectra. *Phys. Rev.* **A41**(4), 1782–1807.

Box, G. E. P., Jenkins, G. M. & Reinsel, G. C. (1994) *Time Series Analysis: Forecasting and Control* (3rd edn). Prentice Hall, Englewood Cliffs, New Jersey, USA.

Broomhead, D. S.& Lowe, D. (1998) Multivariable functional interpolation and adaptive networks. *Complex Systems* **2**, 321–355.

Farmer, J. D. & Sidorowich, J. J. (1987) Predicting chaotic time series. *Phys. Rev. Lett.* **59**, 845–848.

Grassberger, P. & Procaccia, I. (1983) Measuring the strangeness of strange attractors. *Physica D* **9**, 189–208.

Jayawardena, A. W. & Gurung, A. B. (2000) Noise reduction and prediction of hydrometeorological time series: dynamical systems approach *vs* stochastic approach. *J. Hydrol.* **228**(3–4), 242–264.

Jayawardena, A. W. & Lai, F. Z. (1989) Time series analysis of water quality data in Pearl River, China. *J. Environ. Engng, ASCE* **115**(3), 590–607.

Jayawardena, A. W. & Lai, F. Z. (1994) Analysis and prediction of chaos in rainfall and streamflow time series. *J. Hydrol.* **153**(1–4), 23–52.

Lawrance, A. J. & Kottegoda, N. T. (1977) Stochastic modeling of river flow time series. *J. Roy. Statist. Soc.* **140**, 1–4.

Moody, J. & Darken, C. (1989) Fast learning in networks of locally-tuned processing unites. *Neural Computation* **1**, 267–273.

Poggio, T. & Girosi, F. (1990) Networks for approximation and learning. *Proc. IEEE* **78**, 1418–1497.

Sivakumar, B., Liong, S. Y., Liaw, C. Y. & Phoon, K. K. (1999) Singapore rainfall behavior: chaotic? *J. Hydrol. Engng, ASCE* **4**(1), 38–48.

Smith, L. A. (1992) Identification and prediction of low dimensional dynamics. *Physica D* **58**, 50–76.

Sugihara, J. & May, R. M. (1990) Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature* **344**(19), 734–741.

Wright, P. B. (1984) Relationship between the indices of the southern oscillation. *Mon. Weath. Rev.* **112**, 1913–1919.

Wright, P. B. (1989) Homogenized long-period southern oscillation indices. *Int. J. Climatol.* **9**, 33–54.

Ye, Jianming (1998) On measuring and correcting the effects of data mining and model selection. *J. Am. Statist. Assoc.* **93**(441), 120–131.

Yee, P. V. (1998) Regularized radial basis function networks: theory and applications to probability estimation, classification, and time series prediction. PhD Thesis, McMaster University, Hamilton, Ontario, Canada.