River flow prediction: an artificial neural network approach

A. W. JAYAWARDENA & T. M. K. G. FERNANDO
Department of Civil Engineering, The University of Hong Kong, Pokfulam Road, Hong Kong
e-mail: hrecjaw@hkucc.hku.hk

Abstract A multi-layer perception neural network trained with the back-propagation algorithm is adopted to make daily river flow forecasts at the Nakhon Sawan gauging station on the Chao Phraya River in Thailand. Lead times from 1 day to 28 days have been used. The predicted flow rates when compared with observed ones agree within acceptable limits for both training and testing data for most of the considered lead times. The study also demonstrates that different network structures for the longer lead times can enhance the prediction.

Key words artificial neural networks; river flow prediction; cross-validation; Chao Phraya River

INTRODUCTION

River flow prediction is an important topic of research for hydrologists for a number of reasons. For example, it is useful and necessary for flood disaster mitigation. It helps in the planning, management and operation of water resources development projects. The topic is however a difficult one because of the complexity of the physical processes involved in the generation of river flows. An approach based on physics is still far from being realized and researchers have therefore focused attention on the use of “data driven” techniques in the recent years.

Data driven approaches do not lead to a better understanding of the process of the river flow generation, but they have their advantages. They are very practical for “data rich” but “theory weak” systems. Model formulations are often quite simple and nonlinearities in the system pose no problems. One disadvantage is that extrapolation of a model beyond the data range that has been used for calibration is not reliable if the system is nonlinear.

In this study, one such data driven technique, the artificial neural networks (ANN) approach is pursued with the objective of predicting daily river flows using lagged variables. Artificial neural networks are capable of performing nonlinear modelling without a priori knowledge about the relationship between input and output variables. They learn from given data and capture the functional relationships among the data even if the underlying relationships are unknown. After learning from sample data, they can be used to approximate a continuous function to any desired accuracy (Hornik et al., 1989; Hornik, 1991) using unseen input data. These distinguishing features make ANNs a general and flexible alternative modelling tool for hydrological time series prediction in recent years (Dawson & Wilby, 1998; Coulibaly et al., 2000; Jayawardena & Fernando, 2000; Jayawardena et al., 2000).
DATA SET

The data set used in this study consists of daily discharges at Nakhon Sawan gauging station on the Chao Phraya River in Thailand for April 1978–March 1994 (5844 samples). The Chao Phraya River basin is one of the intensively monitored basins in Asia under the auspices of the Global Energy and Water Cycle Experiment (GEWEX) and the GEWEX Asian Monsoon Experiment (GAME). It has a basin area of 110 569 km$^2$ at Nakhon Sawan (latitude 15.67°N; longitude 100.12°E) and can be considered as a macroscale basin. Several studies on the hydrology of this basin have been carried out in recent years (Oki et al., 1995; Raghunath et al., 2000). The data set was obtained from the Global Runoff Data Centre (GRDC # 2964100) in Germany.

METHODOLOGY

Artificial neural networks

Artificial neural networks are capable of learning the underlying behaviour of a system from available data sets. The learning process, also known as training, is carried out according to a learning rule. The synaptic weights are adjusted during the training so that the error surface converges to a minimum. Neural network architectures can basically be categorized into single-layer feedforward networks, multilayer feedforward networks and recurrent networks (Haykin, 1999). The five basic learning rules available in training neural networks include error correction learning, memory based learning, Hebbian learning, competitive learning and Boltzmann learning (Haykin, 1999).

A multilayer perception (MLP) type neural network trained with the back-propagation algorithm (Rumelhart et al., 1986), which belongs to the category of error correction learning rules, is the most widely used network paradigm in prediction (Zhang et al., 1998). This is particularly so in water resource modelling (Maier & Dandy, 2000). The MLP networks have gained popularity over other types of network paradigms mainly due to their inherent capability of arbitrary input–output mapping. In the present study too, a fully connected feedforward MLP network trained using the back-propagation algorithm with a momentum term has been adopted. Despite its popularity, the back-propagation algorithm suffers from limitations associated with its derivation. Many such limitations such as local minima, paralysis and over-fitting can be overcome by adopting necessary precautions during model formulation.

Data partition

The general approach of selecting a good training set from an available data series is to include all extreme events in the training set. The inclusion of biased samples in the training set is not recommended, as this will need a longer training time but not better testing results. A preliminary analysis of the data indicated that most significant events fall within the first quarter of the series. The latter part of the data series seems to contain a lot of biased samples. Therefore the first 1750 samples of the data series
were selected as the training set and the next 400 samples were taken as the validation set, despite the suggestion to use equal-sized training and validation sets (Reed & Marks, 1999). The choice of the length of the validation set is based on the recommendation to use 10–20% of the training samples as the validation set (Kasabov, 1996). The remaining samples were considered for testing the trained network.

Model formulation

The use of lagged input variables leads to better predictions in a dynamic system. One way of finding out how many lagged variables to include is by way of the autocorrelation, which in this study revealed a high correlation even up to 50 days. However, the inclusion of such a large number of inputs from the same variable is not reliable and, furthermore, would give rise to a very large network. Therefore the simulations were first carried out by assuming that the predicted discharge depended only on the discharge values in the previous 7 days. Thus the input layer consists of seven nodes representing the daily discharge values at time levels \( t, t-1, \ldots, t-6 \). In a sensitivity analysis, it was found that these seven variables carried equal importance for 1-day lead-time predictions. The same input variables were therefore used for other lead-time predictions as well in the first instance as it has been pointed out (Coulibaly et al., 2000) that short-term predictions are not significantly affected by the selection of a particular network structure. However, it was seen that the results deteriorate rapidly with increasing lead time. Therefore different lagged input variables were used for longer lead times.

Number of nodes in the hidden layer

The number of nodes in the hidden layer affects the performance of the trained network and therefore should be optimally determined. In general, networks with fewer hidden nodes are preferable as they usually have better generalization capabilities and less over-fitting problems. The computational time required is also less with a smaller number of nodes. However, if the number of nodes is not large enough to capture the underlying behaviour of the data, then the performance of the network is impaired. In most cases, the selection of the optimal number of nodes in the hidden layer is a trial-and-error procedure with the help of some guidelines. A rule of thumb is that the number of samples in the training set should at least be greater than the number of synaptic weights (Rogers & Dowla, 1994; Tarassenko, 1998). This gives the upper limit of the number of nodes for the network. In this study, a trial-and-error procedure for hidden node selection was carried out by starting with two nodes and then gradually increasing the number until there is no further reduction in error, or up to the maximum number of nodes dictated by the above rule of thumb. This procedure led to four hidden nodes for 1-day lead time with seven inputs, which was adopted initially for all other lead times as well when the input layer had seven inputs. Subsequently, the effects of the numbers of nodes for networks with longer lead times but with the same inputs were examined. The numbers of hidden nodes for networks with other inputs and lead times longer than 14 days were selected on a case by case basis.
Learning rate and momentum term

The choice of the learning rate and the momentum term can have a significant effect on the training process. Too small learning rates will result in longer training times and too high values will lead to rapid learning but cause oscillations in the error surface. Addition of the momentum term to the back-propagation algorithm allows synaptic weights to be updated in the general direction of error reduction without getting trapped in a local minimum. The momentum term can take values between 0 and 1. Higher values would produce fluctuations in the error surface while smaller values would not have much effect on convergence speed. In this study, the optimal values for the learning rate and the momentum term, determined by trial and error, were 0.1 and 0.3, respectively.

Training and testing

The logistic function was selected as the activation function for both hidden and output layers. Hence the input variables as well as the target values were normalized linearly in the range of 0.1 to 0.9. The synaptic weights of the networks were initialized with normally distributed random numbers in the range of -1 to 1. The order of presenting the training samples to the network was also randomized from iteration to iteration. Such measures were taken to improve the performance of the back-propagation algorithm.

Training was carried out in a pattern mode; the weights were updated after presenting each sample in the training set to the network. After each iteration, the mean square errors (MSE) corresponding to the training set as well as the validation set were computed. The commonly used stopping criteria include a fixed number of iterations, the rate of change of MSE of the training set reaching a pre-determined value and cross-validation. When the ratio of the number of unbiased samples in the training set to the number of synaptic weights in the network is less than 30, the cross-validation criterion enhances the generalization of the trained network (Amari et al., 1997). Two stopping criteria, the cross-validation and the fixed number of iterations were adopted in this study. The maximum number of iterations was set at 1000.

When cross-validation is used as the stopping criterion, the common practice is to stop training when the MSE of the validation set just begins to increase. However, in real situations, the error reduction curve of the validation data is not a smooth one, and termination of training at the first local minimum may lead to incomplete training. In the present study, the program was developed in such a way that training can further be carried out for a pre-determined number of iterations after reaching a local minimum.

Model performance criteria

The performances of the models were evaluated using the root mean square error (RMSE), the coefficient of determination ($R^2$) and the coefficient of efficiency ($E$) (Nash & Sutcliffe, 1970) as error indicators. The coefficient of efficiency varies between $-\infty$ and +1, with higher positive values indicating better agreement. However,
$E$ cannot be interpreted easily as it has no lower bound (Legates & McCabe, 1999). Scatter plots, double mass curves and time series plots were also used for visual comparison of observed and predicted values.

RESULTS AND DISCUSSION

The initial results with the same network structure for all lead-time forecasts are promising for lead times up to 7 days. The error indicators computed for training and testing sets are summarized in Table 1. They are computed for the whole training and testing sets. The illustrations (Figs 1–3) are for the period from day 2601 to day 2800, which represents one of several peaks in the testing set. The scatter plots for the 1-day and 7-day lead times illustrate the higher accuracy with shorter lead time (Fig. 1). Although the mass curves for the same data show a marginal variation with increasing lead time (Fig. 2), the time series plots (Fig. 3) clearly demonstrate the increase in error with increasing lead time.

Table 1 Error indicators for training and testing data: network structure 7-4-1.

<table>
<thead>
<tr>
<th>Lead time (days)</th>
<th>RMSE (m$^3$ s$^-1$) (Training)</th>
<th>RMSE (m$^3$ s$^-1$) (Testing)</th>
<th>$R^2$ (Training)</th>
<th>$R^2$ (Testing)</th>
<th>$E$ (Training)</th>
<th>$E$ (Testing)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>108</td>
<td>113</td>
<td>0.990</td>
<td>0.972</td>
<td>0.972</td>
<td>0.875</td>
</tr>
<tr>
<td>2</td>
<td>126</td>
<td>127</td>
<td>0.981</td>
<td>0.945</td>
<td>0.962</td>
<td>0.842</td>
</tr>
<tr>
<td>7</td>
<td>213</td>
<td>196</td>
<td>0.901</td>
<td>0.727</td>
<td>0.891</td>
<td>0.623</td>
</tr>
<tr>
<td>14</td>
<td>346</td>
<td>252</td>
<td>0.723</td>
<td>0.541</td>
<td>0.711</td>
<td>0.375</td>
</tr>
<tr>
<td>21</td>
<td>425</td>
<td>290</td>
<td>0.575</td>
<td>0.405</td>
<td>0.564</td>
<td>0.174</td>
</tr>
<tr>
<td>28</td>
<td>469</td>
<td>322</td>
<td>0.477</td>
<td>0.313</td>
<td>0.468</td>
<td>-0.023</td>
</tr>
</tbody>
</table>

$$E = 1 - \frac{\sum (O_i - P_i)^2}{\sum (O_i - \bar{O})^2}$$

where $N$ is the length of the data set and $O$ and $P$ are the observed and predicted values respectively. Overbar denotes the mean for the considered data set.

Network structure 7-4-1 refers to seven nodes in the input layer, four nodes in the hidden layer and one node in the output layer.

The ratio between the number of samples in the training set to the number of synaptic weights in the network was about 55 for these simulations. As a result, the stopping criterion should be based on the maximum number of iterations (Amari et al., 1997). However, training was terminated using the cross-validation criterion, except for forecasts up to 7-day lead times. In an attempt to obtain the optimal networks for higher lead times, it was assumed that the predicted discharge at lead times greater than or equal to 7 days depended on the discharge values at times $t$, $t - 1$, ..., $t - 14$ and $t$, $t - 1$, ..., $t - 21$. The accuracy of the results for the 7-day lead time does not vary significantly by varying the inputs and the hidden nodes. However the computed values seem to be sensitive for the network structure for the remaining lead times considered. The optimal network structures obtained for different lead times and the corresponding error indicators are given in Table 2. The results have improved from
Fig. 1 Scatter plot for part of the testing data.

Fig. 2 Double mass curve for part of the testing data.

Fig. 3 Observed and predicted discharge values for part of the testing data.
Table 2 Error indicators for higher lead times with optimal network structure.

<table>
<thead>
<tr>
<th>Lead time (days)</th>
<th>Network structure</th>
<th>R² (Training)</th>
<th>R² (Testing)</th>
<th>E (Training)</th>
<th>E (Testing)</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>14-30-1</td>
<td>0.763</td>
<td>0.534</td>
<td>0.759</td>
<td>0.438</td>
</tr>
<tr>
<td>21</td>
<td>21-15-1</td>
<td>0.619</td>
<td>0.403</td>
<td>0.613</td>
<td>0.246</td>
</tr>
<tr>
<td>28</td>
<td>21-30-1</td>
<td>0.523</td>
<td>0.315</td>
<td>0.517</td>
<td>0.043</td>
</tr>
</tbody>
</table>

the initial study. However, it is difficult to obtain a good model for lead times greater than two weeks. This study clearly demonstrates the potential of ANN in making reasonably reliable daily discharge predictions for short lead times using lagged discharges as inputs.

REFERENCES


