Artificial Neural Network Model for Rainfall-Runoff Relationship

P. Junsawang1,*, J. Asavanant1 and C. Lursinsap1

1Advanced Virtual and Intelligent Computing Center (AVIC).
Department of Mathematics, Faculty of Science, Chulalongkorn University,
Bangkok, 10330, Thailand

*Corresponding author: pj_prem@hotmail.com, jpremlc@gmail.com

Abstract

Conceptual models have been widely used and are considered to be the best choice for describing the runoff process in a watershed. In most cases, the solution accuracy is mainly based on the topographic and hydrologic information subject to certain requirements for model calibration. Thus, these types of model are inappropriate for watershed area with little hydrologic data. Artificial neural network (ANN) has become an alternative approach to model the runoff process in situations where explicit knowledge of the internal hydrologic processes is not available. ANN has a flexible mathematical structure which is capable of identifying complex nonlinear relationships between the sets of input and output data. In this paper, ANN is proposed as a tool to predict runoff hydrograph from only one hydrologic station for the Mae Tun River in Omkoi District, Chiang Mai Province which is located in the northern part of Thailand. The watershed area to be considered here is approximately 503 square kilometers. It has distinct hydrologic features with insufficient information on topography and rainfall-runoff data. In this study the problem is viewed as a time series prediction problem. A feed-forward artificial neural network is trained by using back-propagation algorithm. The training and testing data were collected during years 2000 to 2004. Since the original collected data contain only discharge and rainfall data, it is rather difficult to obtain accurate results. To improve the accuracy, the amount of water excess was calculated from the collected data and used as an additional input for training the network. The input to the network consists of discharge, rainfall measurement and discharge excess during the last 24 hours. The performance comparison is presented by statistic evaluation of percentage errors and correlations. From the simulations, it is found that random input pattern gives more accurate results than the well-order input pattern.

KEY WORDS: Artificial neural network, rainfall-runoff model, feed-forward multi-layer perceptron.

Introduction

Rainfall-runoff relationship is an essential component in the process of water resources evaluation and is considered as a central problem in hydrology. There have been extensive researches on the rainfall–runoff relationship with different models that can be classified into three main groups, namely distributed physically based models, lumped conceptual models and black box models.

Distributed physically based models describe the natural system using the basic fluid-flows derived from the energy and water budgets. The models consist of a set of linked partial differential equations together with various parameters, which, in principle, have direct physical significance and can be evaluated by independent measurements. One example of this type is the Sacramento model proposed by Hogue et al. (2006).

Conceptual models operate with different and mutually interrelated storages representing physical elements in a catchment area. Due to the lumped description, where all parameters and variables represent average values over the entire catchment areas, the description of the hydrological processes cannot be based directly on the equation derived for the individual soil column. Hence, the equations are semi-empirical, but with a physical basis. Therefore, the model parameters cannot usually be assessed from field data, but have to obtain through calibration. Examples of lumped
conceptual models are the Tank model suggested by Sugawara (1974) and the HEC-HMS developed in 2000 by the US Army Corps of Engineer (Pukdeboon et al., 2006).

Black box models are empirical, involving a mathematical equation that has been assessed. The models are not concerned with the physical processes in the catchment area, but from analyses of concurrent input and output time series. Examples of black box models are the unit hydrograph model (Yue et al., 2000) and the Neural Network model (Tan et al., 2006).

The first two types of rainfall-runoff models are important in understanding hydrological processes, but their applications may be a rough simplification of the real problem unless there are sufficient data for analyzing relevant parameters. On the other hand, the black box models use little information on physical description of the catchment area and hence provide an alternative way to represent this complex system. During the past few years, the use of artificial neural network (ANN) as a black box model has increased drastically among researchers due to its simplicity.

However, a severe drawback of this technique is that their parameters are not directly related to physical model conditions of the hydrological process. Nevertheless, ANN are widely accepted as a potentially useful way of modeling complex non-linear systems with a large amount of data. ANN is particularly useful in situation where the underlying physical process is not yet fully understood, and can be used as a substitute for the conventional and statistical models.

Recently, there have been many researches on rainfall-runoff relationship based on ANN in the line of hydrologic processes such as the use of ANN for predicting and forecasting water resource variables by Maier et al. (2000), and Smith et al. (1995) who used a feed-forward network to estimate the runoff peak value and the time to peak for spatially distributed rainfall. However their trainings were performed by using simulated data. Luk et al. (2001) studied the rainfall forecasting problem by using various ANNs and discussed the accuracies and discrepancies among these networks. Valenca et al. (2005) used a constructive neural network model (NSRBN) to forecast daily river flows for the Boa Esperanca Hydroelectric powerplant. In addition, the ANNs can also be applied in hydrologic processes by coupling with a simple linear regression model in order to construct a daily rainfall runoff models (Ruhurkar et al., 2004).

This paper is concerned with the use of ANN for modeling the rainfall-runoff relationship in small watershed area with little hydrologic data. This is a major problem in many underdeveloped rural areas in Thailand. Here we select the watershed area of Mae Tun River in Omkoi District, Chiangmai Province located in the northern part of Thailand to test the ANN model. This watershed covers the area of approximately 503 square kilometers. It has distinct hydrologic features with insufficient information on topography and runoff data. A feed-forward ANN is trained by using the back-propagation algorithm. Two input feature selections from little hydrologic data for the network are proposed. Comparison between two distinct input pattern groups for training and testing the ANN models are discussed.

Artificial Neural Networks

Haykin (1994) defined the ANN as a massively parallel distributed processor that has a natural propensity for storing the experimental knowledge and making it available for use. It resemble the brain in two aspects namely knowledge required by the network through a learning process and inter-neural connection strengths known as synaptic weights used to store the knowledge. The ANN can be categorized into different groups such as pattern association, mapping and clustering. In pattern mapping, the input-output relationship is captured by means of a suitable “learning” strategy so the black-box type modeling of the rainfall-runoff relation can be classified into category of pattern mapping. Two type of networks, the feed-forward multi-layer perceptron (MLP) network (Rumelhart et al., 1986) and the Counter propagation network (CPN) network (Hecht-Nielsen, 1987), are usually used for pattern mapping problems. Of the two types of network, MLP is selected for application in this paper.

MLP is consisted of a number of computational elements described as neurons. The neurons are organized in three layer: the input layer contains the input unit that receive information from the outside world, one or more hidden layer of computation neurons and output layer of computation neurons. Each neuron is fully connected to neurons in the next layer (Figure 1). All input to a neuron
in a particular layer is from the proceeding layer and these unidirectional strengths are known as weights. A gradient descent procedure known as generalized error-back propagation or back-propagation is usually employed for training the MLP network. Thus MLP network is known as a back-propagation network (BPN). The back-propagation consists of two passes: a forward pass and backward pass. In the forward pass, an input pattern is applied to the input layer and its effect is propagated layer by layer through network. The activity at a neuron is computed as the weighted sum of the outputs of the neurons of the previous layer. The output of the neuron is computed from a nonlinear activation function. The most commonly used in this type of network is the logistic sigmoid function. This activation function is continuously differentiable, symmetric and bounded between 0 and 1. The mathematical expression of the logistic sigmoid function is given by

\[ f(n) = \frac{1}{1 + e^{-n}} \]  

[1]

In the backward pass, the sum of squared deviation of the output from the target value at neurons of output layer defines the error signal propagated back to the previous layers. The parameters are adjusted to minimize errors in the computations.

Mae Tuen River

Mae Tun river is the small branch of Ping river in Omkoi District located in the lower part of Chieng Mai Province. The natural topography of Mae Tun watershed is mountainous covered mostly by forests. Due to its geographic location and soil property, there have been persistent problems in flooding and poor agricultural harvesting in the summer period.

Omkoi catchment area is located in the upper part of Mae Tun river covering the area of approximately 503 square kilometers. Geographically, this area is bounded by the latitude of 17°47' N to 18°2’ N and the longitude of 98°12’ E to 98°30’ E. Ten-year record of hourly and three hourly data of rainfall and runoff discharges can be found at the P64 station, situated in the lower part of Mae Tun watershed, which is the only nearby station in the watershed. These hydrologic data for rainfall and runoff were collected by the Meteorological Department and the Royal Irrigation Department of Thailand, respectively. Runoff data were recorded only during the day-time period.

Methodology

The training and testing data set collected during years 2000 to 2004 were used. Since the original recorded data contain only discharge and rainfall data, it is generally difficult to achieve highly accurate simulated results based on these data. An attempt to improve the accuracy is to use

data on the discharge excess and the sum of rainfall during the last 24 hours from the prediction time as additional inputs to the network model. The expression can be written in mathematical form as follows

\[
Q(t) = f(SR, DQ, R(t_i - 3), R(t_i - 2), R(t_i - 2), Q(t - 3t_s), Q(t - 2t_s), Q(t - t_s), Dq) \quad [2]
\]

\begin{align*}
&t - \text{Time of prediction.} \\
&t_i - \text{Time to incorporate rainfall (in this case, } t_i = t - 4). \\
&t_s - \text{Time period (3 hours).} \\
&R - \text{Rainfall intensity (mm/hr).} \\
&Q - \text{Discharge (CMS).} \\
&SR - \text{Summation of rainfall value from } t - 8t_s \text{ to } t - 3t_s \text{ (mm/hr).} \\
&DQ - \text{Discharge excess between } Q(t - 8t_s) \text{ and } Q(t - 3t_s) \text{ (CMS).} \\
&Dq - \text{Discharge excess between } Q(t - 3t_s) \text{ and } Q(t - t_s) \text{ (CMS).}
\end{align*}

In this study, we consider the total input of 2,306 patterns for training and testing. Two types of input patterns are used with MLP neural network. For the first type, Type I, the total input patterns is provided in consecutive order both for training and testing procedures. The training patterns consist of the data during the year 2000-2002 (1442 patterns) while the testing patterns are the data collected during the year 2003-2004 (864 patterns). For the second type of input patterns, Type II, the data for training and testing patterns are randomly selected to 1383 and 923 input patterns respectively.

The input and target values of each feature (rainfall, runoff or discharge and runoff difference) are normalized into appropriate scale by using the maximum and minimum values such that both are contained in the unit interval [0,1]. Following Sajikumar et al. (1999), the general scale down procedure is given by

\[
X_n = F_{\text{min}} + \left( \frac{X_u - \text{fact min}}{\text{fact max} - \text{fact min}} \right) (F_{\text{max}} - F_{\text{min}}) \quad [3]
\]

Here \(X_u\) and \(X_n\) represent the value to be scaled down and its normalized value respectively. \(F_{\text{min}}\) and \(F_{\text{max}}\) are the minimum and maximum of normalized value and fact max and fact min are the maximum and minimum observed values of each feature of inputs (rainfall intensity, runoff discharge, and discharge excess). In this paper, the Levenberg-Marquardt training algorithm (Singh et al., 2006) is used in the training procedure.

### Results and Discussions

Traditional statistical criteria is adopted here to help select the desired optimal network model. The selection procedure is based on the following statistics: correlation (COR), coefficient of determination (\(R^2\)), mean absolute relative error (MARE), and root average square relative error (RASRE). These are defined by
\[ \text{COR} = \frac{\sum_{i=1}^{n} (Q_i - \overline{Q})(\hat{Q}_i - \overline{\hat{Q}})}{\sqrt{\sum_{i=1}^{n} (Q_i - \overline{Q})^2 \sum_{i=1}^{n} (\hat{Q}_i - \overline{\hat{Q}})^2}} \]  

\[ R^2 = 1 - \frac{\sum_{i=1}^{n} (Q_i - \hat{Q}_i)^2}{\sum_{i=1}^{n} (Q_i - \overline{Q})^2} \]  

\[ \text{MARE} = \frac{\sum_{i=1}^{n} \left| \frac{Q_i - \hat{Q}_i}{Q_i} \right|}{n} \times 100 \]  

\[ \text{RASRE} = \sqrt{\frac{\sum_{i=1}^{n} (Q_i - \hat{Q}_i)^2}{n}} \]

\( n \) is the number of data,
\( Q \) is the observed value,
\( \hat{Q} \) is the predicted value,
\( \overline{Q} \) is the average of observed value, and
\( \overline{\hat{Q}} \) is the average of predicted value.

Table 1: Statistics of the two types of input pattern for testing phase.

<table>
<thead>
<tr>
<th>Input Type</th>
<th>COR</th>
<th>RASRE</th>
<th>( R^2 )</th>
<th>MARE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type I</td>
<td>0.9316</td>
<td>0.15546</td>
<td>0.84724</td>
<td>6.8</td>
</tr>
<tr>
<td>Type II</td>
<td>0.96746</td>
<td>0.13129</td>
<td>0.93592</td>
<td>6.79</td>
</tr>
</tbody>
</table>

By the definitions of Types I and II data, the input patterns are less scattered for Type I than Type II. It is natural to expect better results from the randomly distribution of data. As shown in Table 1, all the basic statistics have uniformly reasonable values. A 0.97 correlation coefficient is found for Type II input pattern which is better than a 0.93 correlation coefficient of Type I. This is in agreement with the determination coefficient \( R^2 \). The errors (MARE and RASRE) also illustrate the better accuracy of ANN model by using Type II input than Type I. These statistics remain consistent after the experimentation on more specific (smaller) set of data with more rainfall intensity.
Due to a large amount of input pattern, only 80 patterns are selected to illustrate in comparison between the observed and predicted values for Types I and II data as shown in Figures 2 and 3. Though both types of the input pattern seem to have good agreement with the observations, there are a few extreme events that ANN suffered large errors resulting from drought or flood. Despite of the deteriorating of model accuracy in such extreme cases, Type II still gives better runoff estimation than Type I in terms of the maximum relative error.

Conclusion

The research presented in this paper focused on the estimation of runoff discharge using the artificial neural network approach. This can be generalized to determine the flood hydrograph in the studied area. Here the Mae Tun watershed in Omkoi district, Chiangmai Province is selected in this case study to demonstrate the application of the proposed technique. The distinct feature of this selected area is due to its geographical location and the limited rainfall-runoff data recorded from one single hydrologic station. A feed-forward ANN is trained on the available historic data using the backward propagation algorithm. The input parameters used herein are antecedent rainfall index, runoff discharge, and discharge excess. As output, the network is trained to generate information on the discharge in the next time level and then used to develop a complete runoff hydrograph. Here we
compare two types of input patterns, Type I and Type II, regarding the ordered sequence of input patterns as Type I and the random input patterns as Type II, and evaluate the statistics of percentage errors and correlations. From the simulations, it is found that Type II input pattern gives better accurate results than Type I.

The ANN approach used in this study does not take into account the physical processes nor information about the catchment area involved in runoff generation and hence makes it very attractive as a substitute for conceptual watershed modeling. However, the ANN does have some difficulties managing the extreme events. The improvement for such cases can be done by incorporating the discharge excess prior to the prediction time. Research on this extreme event is still in progress.

References


Sugawara, M., 1974. Tank model with snow component. national research center for disaster prevention, Japan.


