MULTITEMPORAL SCALE HYDROGRAPH PREDICTION USING ARTIFICIAL NEURAL NETWORKS

Assefa M. Melesse and Xixi Wang

ABSTRACT: An artificial neural network (ANN) provides a mathematically flexible structure to identify complex nonlinear relationship between inputs and outputs. A multilayer perceptron ANN technique with an error back propagation algorithm was applied to a multitime-scale prediction of the stage of a hydrologically closed lake, Devils Lake (DL), and discharge of the Red River of the North at Grand Forks station (RR-GF) in North Dakota. The modeling exercise used 1 year (2002), 5 years (1998-2002), and 27 years (1975-2002) of data for the daily, weekly, and monthly predictions, respectively. The hydrometeorological data (precipitations $P(t)$, $P(t-1)$, $P(t-2)$, $P(t-3)$, antecedent runoff/lake stage $R(t-1)$, and air temperature $T(t)$) were partitioned for training and for testing to predict the current hydrograph at the selected DL and RR-GF stations. Performance of ANN was evaluated using three combinations of daily datasets (Input I = $P(t)$, $P(t-1)$, $P(t-2)$, $P(t-3)$, $R(t-1)$; Input II = Input-I less $P(t)$, $P(t-1)$, $P(t-2)$, $P(t-3)$; and Input III = Input-II less $T(t)$). Comparison of the model output using Input I data with the observed values showed average testing prediction efficiency ($E$) of 86 percent for DL basin and 46 percent for RR-GF basin, and higher efficiency for the daily than monthly simulations.

(KEY TERMS: artificial neural network; Devils Lake; Red River; flood forecasting; lake stage prediction; temporal scale.)


INTRODUCTION

Artificial Neural Networks

Neural networks, expert systems and fuzzy logic are the three branches of artificial intelligence. An artificial neural network (ANN) is a nonparametric attempt to model the human brain that uses machine learning based on the concept of self-adjustment of internal control parameters. Artificial neural networks are flexible mathematical structures that are capable of identifying complex nonlinear relationships between input and output datasets. The main differences between the different types of ANNs are arrangement of neurodes (network architecture) and the many ways to determine the weights and functions for inputs and neurodes (training) (Caudill and Butler, 1992). Training in ANN refers to pattern recognition using weight estimation and transfer function. The multilayer perceptron (MLP) neural network was designed to function well in nonlinear phenomena. A perceptron is a connected network that simulates an associative memory. The most basic perceptron is composed of an input layer and output layer of nodes, each of which is fully connected to the other. The multilayer in MLP refers to the use of the input, hidden, and output layers in the ANN analysis. A feed forward MLP network consists of an input layer, an output layer, and one or more hidden layers between them. Each layer has some number of neurons on it, input layers with input neurons and output layers with output neurons.
A typical ANN architecture is shown in Figure 1. The artificial neuron receives a set of inputs or signals \( x \), calculates a weighted average of them \( z \) using the summation function and weights \( w \) and then uses some activation function \( f \) to produce an output \( y \), where

\[
z = \sum_{i=1}^{n} x_i w_i
\]

(1)

The connections between the input layer and the middle or hidden layer contain weights, which are usually determined through training the system. The hidden layer sums the weighted inputs and uses the transfer function to create an output value. The transfer function (local memory) is a relationship between the internal activation level of the neuron (activation function) and the outputs. A commonly used typical transfer function is the sigmoid function, Equation (2), which varies from 0 to 1 for a range of inputs (Caudill and Butler, 1992). A function \( f(x) \) will be a sigmoid function if it is bounded and if its value always increases as \( x \) increases (Smith, 1993). A number of different functions have these characteristics and thus qualify as sigmoid functions. The sigmoid logistic nonlinear function is described in Equation (2) and illustrated in Figure 2.

\[
y = f(z) = \frac{1}{1 + e^{-x}}
\]

(2)

In time series prediction, supervised training is used in which the ANN is trained to minimize the difference between the network output and the target (observed). The most common training algorithm used in the ANN literature is called back propagation (BP). One of the basic requirements of the BP training is that the transfer function be continuous and differentiable. The sigmoid logistic nonlinear function explained above fulfills the requirement, making the implementation of BP easier.

![Figure 1. A Typical Multilayer Perceptron ANN Architecture.](image1)

**BP Algorithm**

Back propagation, or backdrop, was provided and popularized by Rumelhart et al. (1986), and it is the most widely implemented of all neural network paradigms. BP is based on a multilayered feed forward topology with supervised learning involving mapping and learning. In the mapping mode, information flows forward through the network, from inputs to outputs. In the learning mode, the information flow alternates between forward and backward. A key element in the back propagation paradigm is the existence of a hidden layer of nodes. Back propagation uses a type of gradient descent method, following the slope of the error surface downward towards its minimum. If the network is linear in its output and has no hidden layers, the error surface will have a bowl type shape, and there will be only a global minimum. The error surface in networks with hidden layers may contain several local minima, and the steepest descent method may be trapped in one of them.

A high learning parameter in the BP algorithm corresponds to rapid learning, which may push the training toward a local minimum or cause oscillation. In turn, when applying small learning rates, the time to reach a global minimum will be considerably increased (Khan et al., 1993). The momentum factor can help in choosing the learning rate. The role of the momentum term is to smooth out the weight changes, which helps to protect network learning from oscillation (Anderson et al., 1993).
ANNs Application

ANNs have been used for a wide range of different learning from data application problems and input-output correlations of nonlinear processes. Various researchers have used ANNs for hydrologic studies, including for time series predictions of runoff (Hsu et al., 1995; Shamseldin, 1997; Zealand et al., 1999; Imrie et al., 2000; Tingsanchali and Gautam, 2000), water table management (Yang et al., 1998), reservoir operation optimization (Solomatine and Torres, 1996), water quality management system (ChingGung and ChihSheng, 1998), estimating water quality parameters (Zhang and Stanley, 1997), sediment prediction (Abrahart and White, 2001; Nagy et al., 2002; Yitian and Gu, 2003; Cigizoglu, 2004) and nonpoint source contamination (Brion and Lingireddy, 1999). In recent years, ANNs have been increasingly used for prediction and forecasting of variables involved in hydroclimatologic processes. Application of ANN to the rainfall-runoff process uses precipitation (input) as stimulus and river discharge or lake stage (output) as response. Evaluating the ANN technique for different temporal scales would be important to assess its prediction ability. The technique is one of the most useful tools available in modeling nonlinear relationships when given continuous data. The limitations are that it remains unexplained how ANN accomplishes its results, and it is data intensive. It has been demonstrated that ANN is superior to most deterministic and physically based approaches where nonlinearity is an important factor.

Since the flow at any point in time is effectively composed of contributions from different subareas whose time of travel to the outlet covers a range of values, both the concurrent and antecedent precipitation should be considered as stimuli (Hall and Minns, 1993; Hsu et al., 1995; Dawson and Wilby, 1998; Zealand et al., 1999; Tingsanchali and Gautam, 2000). The importance of antecedent precipitation and flow will be greater for watersheds with longer travel time to the outlet and snowmelt dominated runoff. The number of antecedent rainfall ordinates required is broadly related to the lag time of the drainage area. In addition, the use of previous output values in the input pattern is also important and used in recurrent network architectures.

In this study, both the current and antecedent inputs and prior output values were used in the prediction of lake stage and river discharge on daily, weekly, and monthly bases.

The overall objective of the study is to employ multilayer perceptron ANN model with error BP algorithm for multitemporal scale prediction of lake level or stage at DL and discharge at RR-GF. The specific objectives of the study are to assess sets of input hydrologic and meteorological variables for improved lake stage and river discharge predictions; to compare the performance of the model for three time steps (daily, weekly, and monthly); and to evaluate the performance of the ANN model in reproducing the observed values.

Description of Study Areas and Data Sets

The modeling exercise used data from two subbasins in the Red River of the North basin, the DL subbasin and the RR-GF subbasin (Figure 3).

Devils Lake Subbasin

The DL subbasin is a 9,870 km² watershed in the Red River of the North basin in northeastern North Dakota (Figure 3). About 8,600 km² of the total 9,870 km² is tributary to DL. It contributes to the Red River of the North when the level of DL is greater than 444.7 m above mean sea level (amsl) (Wiche and Vecchia, 1996).

Devils Lake has a continental climate characterized by relatively warm, short summers and long, cold winters. Precipitation averages about 432 mm annually, some 75 percent of which falls between April and September. The one-fourth of the annual precipitation that is received during the colder months contributes the most to spring runoff and subsequent recharge of DL. The maximum recorded temperature is 44°C, and the minimum is -43°C. The growing season is short, averaging 131 days (Wiche and Vecchia, 1996).

Historically the level of the lake has fluctuated (Figure 4). Since the subbasin is a closed drainage, the only outlet of water is evapotranspiration. A high volume of precipitation and less evaporation caused flooding of more than 20,000 ha of land in 1993 (Williams-Setter et al., 1996). The surface area of the DL increased by 117 percent between 1993 and 2003 (Melesse and Nangia, 2004). Since 1993, more than $450 million in damages have occurred as a result of flooding throughout the DL basin (Williams-Setter et al., 1996).

Red River of the North Subbasin at Grand Forks

The Red River of the North (Figure 3) has a meandering length of 880 km along the channel and 456 km in a straight line. It flows north from the confluence of the Bois de Sioux and Otter Tail Rivers in southern North Dakota and Minnesota, USA, to Lake
Winnipeg, Manitoba, Canada, draining about 100,480 km$^2$. The Red River flows through the very flat Red River Valley, where natural topographic variations are subtle (Macek-Rowland, 2001).

The climate of the Red River of the North basin is continental, ranging from dry subhumid in the western part of the basin to subhumid in the eastern. Winter precipitation as snow comprises about 15 percent of the total precipitation. Average annual precipitation ranges from about 457 mm at the northwest corner of the basin to about 686 mm toward the southeast corner. About three-fourths of the annual precipitation falls from April through September. Average annual temperatures vary between 2.8 and 6.1 $^\circ$C. Average monthly temperatures range from -18.3 $^\circ$C in January to 21.7 $^\circ$C in July.

In the Red River Valley of North Dakota and Minnesota, average annual ground surface temperature has increased 2.5 $^\circ$C, and average annual air temperature has increased 2 $^\circ$C during the past century (Gosnold et al., 1997). Precipitation dropped from 1890 until the mid-1930s and has been rising since 1950 (Groisman and Easterling, 1994). Historically, the Red River Valley floods frequently and peaked to its historic high in April 1997, when 1,590 m$^3$/sec for April flow was measured at the Grand Forks gaging station. The 1997 flood caused $3.6$ billion losses within the Grand Forks (North Dakota)/East Grand Forks (Minnesota) metropolitan area.

The Red River of the North receives most of its flow from its eastern tributaries, largely as result of regional patterns in precipitation. Annual runoff varies greatly, and most runoff occurs in spring and early summer from rains falling on soils saturated by melting winter snow.
The study considers part of the basin contributing to the Grand Forks gaging station (Figure 3). The subbasin has a contributing drainage area of 67,300 km² with inflows from the Red River and Red Lake River watersheds from North Dakota and Minnesota.

Historical records of the lake at DL and discharge at RR-GF (Figure 4) show greater variability in the flood response of the two study areas. It is shown that since 1993, the recorded level and discharges were higher than the all time average, causing flood damages to the cities of Devils Lake and Grand Forks, respectively. Predicting floods in these areas requires accurate techniques in understanding the hydroclimatological processes and identification of tools for flood prediction at various time scales.

Datasets

The study used hydrometeorological data from the National Climatic Data Center (NCDC) and U.S. Geological Survey (USGS). Daily, weekly, and monthly precipitation and air temperature for both of subbasins were obtained from NCDC (2004). The daily, weekly, and monthly lake stage and river discharges were obtained from USGS (2004). Figure 5 shows a network of weather stations in the study area.

![Figure 5. Network of Weather Stations in the Study Areas.](image)

Given the sizes of the watersheds and the flat topography, antecedent precipitation and flow/lake stage are important to the study. As snowmelt is one of the major factors in the hydrographs of the watersheds, inclusion of air temperature can improve hydrographic predictions. A separate study in the region has shown that precipitation and air temperature together with the antecedent discharge or lake stage play important roles in forecasting the hydrographs (Melesse, 2004).

**METHODOLOGY**

In a neural network architecture used in this study, the output node represents the lake levels (DL) and discharges (RR-GF) to be predicted. The hidden nodes, the internal part of the system, enable ANN to learn the nonlinear relationship between the output and input. The parameters on which the forecast value depends with some function represent the input nodes (precipitation, P, air temperature, T, and lake levels or river discharges, R). The input data for ANN's modeling consist of current precipitation, three antecedent precipitations, air temperature, and one antecedent runoff. In a separate analysis (Melesse, 2004), these parameters have been shown to have high correlation with lake level and river discharge for the watersheds considered in this study. In order to evaluate the effect of the different hydrometeorological variables on the model output, three input datasets were used and evaluated. These were Input I, \( P(t), P(t-1), P(t-2), P(t-3), T(t), \) and \( R(t-1); \) Input II, I-1 less \( P(t), P(t-1), P(t-2), P(t-3); \) and Input III, I-2 less \( T(t). \) Figure 6 shows the ANN architecture and inputs used in the study. For all of the time scale simulations using Input-I, six inputs as shown in Figure 6 were used, where \( P(t), P(t-1), P(t-2), \) and \( P(t-3), \) refer to current precipitation (mm) and precipitation at time steps \( t-1, t-2 \) and \( t-3, \) respectively, and \( T \) and \( R \) refer to air temperature (°C) and lake stage (meters above mean sea level)/river discharge (m³/s), respectively. In a separate analysis, model performance was evaluated for three different input datasets.

The above input data were processed and normalized between 0.01 and 0.99 using Equation (3) in order to input them to the neural network utilizing logistic sigmoid transfer function, which takes the value between 0 and 1.

\[
X_s = \frac{0.99(X_i - X_{min})}{(X_{max} - X_{min})} + 0.01
\]

where \( X_s \) is the scaled input value, \( X_i \) is actual unscaled observed input value, and \( X_{min} \) and \( X_{max} \) refer to the minimum and maximum values of the data, respectively.

In this study, 1-year daily (2002), 5-year weekly (1998-2002) and 27-year monthly (1975-2002) input-output data (precipitation, air temperature, lake
level/river discharge) were divided into training and verification periods. Table 1 shows period of record and partitioning of the data used in the study. Previous hydrometeorological analysis of the two study areas shows longer lag time due to the larger watershed and flat topography. In both study areas, discharge peaks in spring from snowmelt and rainfall falling on wet soils (Melesse and Nangia, 2003; Melesse, 2004).

![Figure 6. An ANN Architecture and Input Data Structure Used in the Study.](image)

For the daily simulation, the first nine months (January-September, 2002) of data were used for training purposes. For the weekly modeling, the first three years (1998-2000) of data were used for training, and the remaining two years (104 weeks) of data (2001-2002) were used for testing purposes. The monthly data from 1975 through 2002 were arranged in such a way that the 1975 through 1995 (20 years) data were used for training, and the seven-year data from 1996 through 2002 were used for verification purposes.

Due to the nature of the region’s predominantly snowmelt driven rainfall-runoff process, flat topography, and large size of the watersheds, it was shown that air temperature, antecedent precipitation, and prior runoff values play an important role in modeling future storm runoff. Air temperature is an important parameter for the snow melt process in the spring flow. Because of the flat topography in the Red River Valley and large size of the watersheds considered in this study, the time of concentration is greater, and antecedent flow and precipitation are important input variables in the modeling.

The training time is directly dependent on the size of the network: the larger the network, the longer the training time. The training process depends on the number of input units, number of hidden units, and convergence criteria. The study used ANN architecture with three layers: input layer, hidden layer, and output layer. A small number of hidden nodes may not be able to train the network, and a very large number of hidden nodes poses difficulty to the training and may weaken the effective learning strength of the networks. Therefore, the determination of the hidden layers and nodes is made by the trial-and-error process, depending on the condition of the problem. In this study, for the sake of simplicity the number of hidden nodes was equal to the number of inputs. This is the recommended arrangement by the NeuNet Pro software (NeuNet, 2005), which was used in this study.

The weights leading into the output node and weights leading to the hidden nodes are adjusted backward so that the correction in error is propagated backward in an iterative process until the system converges after a predefined acceptable limit. In this study the limits were visual fit, correlation coefficients, and residuals.

**Model Comparison**

The performance of the ANN model was evaluated using statistical comparisons of the predicted and observed outputs. The comparison uses the two most common error measures (RMSE and MAPE) and model efficiency.

The root mean squares of errors (RMSE) in m (DL) or m³/s (RR-GF) can be calculated using Equation (4),

<table>
<thead>
<tr>
<th>Time Scale</th>
<th>Period</th>
<th>Training</th>
<th>Testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daily</td>
<td>January 1, 2002 to September 30, 2002</td>
<td>First 270 days</td>
<td>90 days</td>
</tr>
<tr>
<td>Weekly</td>
<td>January 1, 1998 to December 31, 2002</td>
<td>First 156 weeks</td>
<td>104 weeks</td>
</tr>
<tr>
<td>Monthly</td>
<td>January 1, 1975 to December 31, 2002</td>
<td>First 240 months</td>
<td>84 months</td>
</tr>
</tbody>
</table>
where \( y_{o,i} \) is the observed hydrograph value of observation \( i \) in (DL) or \( m^3/s \) (RR-GF); \( y_{p,i} \) is the predicted hydrograph value of observation \( i \) in (DL) or \( m^3/s \) (RR-GF); and \( n \) is the number of observations. The RMSE is very sensitive to outliers (extreme forecast errors), and it is not scale free. To avoid the problem of scaling, many use the scale free error measurement, mean absolute percent error (MAPE) in percent, as defined as

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_{o,i} - y_{p,i})^2}
\]  

(4)

\[
MAPE = \frac{\sum_{i=1}^{n} 100 \frac{|y_{o,i} - y_{p,i}|}{y_{o,i}}}{n}
\]  

(5)

MAPE calculates the forecast error as a percentage of the actual value. The drawback of the MAPE measure is that it puts a heavier penalty on the forecasts that exceed the actual value than on those that fall below the actual value (Melesse and Nangia, 2005).

Another widely used goodness-of-fit measure based on the error variance is the model efficiency (\( E \)) (Nash and Sutcliffe, 1970), defined as

\[
E = \left[ 1 - \frac{\sigma^2_{\varepsilon}}{\sigma^2_o} \right]
\]  

(6)

The error variance \( \sigma^2_{\varepsilon} \), is defined as

\[
\sigma^2_{\varepsilon} = \frac{1}{n-1} \sum_{i=1}^{n} (y_{o,i} - y_{p,i})^2
\]  

(7)

The variance of the observed lake stage/discharge \( \sigma^2_o \), defined as

\[
\sigma^2_o = \frac{1}{n-1} \sum_{i=1}^{n} (\bar{y}_o - \bar{y}_o)^2
\]  

(8)

\[
\bar{y}_o = \text{average observed lake stage/discharge in m (DL) or m}^3/\text{s (RR-GF)}.
\]

The model efficiency, \( E \), is similar to the statistical coefficient of determination. It has the value of 1 for perfect fit when \( \sigma^2_{\varepsilon} = 0 \); it has the value of 0 when \( \sigma^2_{\varepsilon} = \sigma^2_o \), which is equivalent to saying that the hydrological model is not better than a one-parameter “no knowledge” model that gives a prediction of the mean of the observations for all time steps. A negative efficiency indicates that the variance of the errors is larger than the variance of the observed flow. Negative efficiency is an indication of poor performance and higher deviations of the predicted values from the observed values.

## RESULTS AND DISCUSSION

### Input Data Scenarios

The MLP-BP algorithm applied to a multitemporal-scale flood prediction problem of lake and river watersheds is reported in this paper. The effect of the different input datasets on model performance is also evaluated. Based on the three input data structures considered, Input I with daily simulation showed better results, with most of the performance indices used indicating good agreement with the observed values, than the Input II and Input III datasets (Table 2a, 2b, and 2c). It is shown that the inclusion of precipitation and antecedent discharge/lake stage improved the modeling result (Table 2b). Dropping the precipitation (Input II) increased the RMSE and MAPE and reduced the \( R^2 \) (Table 2b). On the other hand, the use of precipitation \((P_{i(t)}, P_{i(t-1)}, P_{i(t-2)} \) and \( P_{i(t-3)} \)) alone resulted in a poor performance, indicating that the combined effect of the inputs selected is more important than individual parameters. The ANN modeling seems to perform poorly in predicting larger observed values (Figures 7b, 8b, and 9b) for RR-GF. For Input I data structure, the model efficiency, \( E \), for the testing phase ranged from 84 to 88 percent for DL and 38 to 60 percent for RR-GF, respectively. The lower efficiency and fit for the RR-GF was due to the model’s inability to capture the extreme discharges in all the time steps. The larger discharges especially in 1997 and 2001 can be attributed to the combination of various hydroclimatological conditions that could not be represented as input in the ANN model structure.

### Temporal Comparison

The ANN with Input I datasets predicted values for the three time scales used in the study are shown in Figures 7 through 9. In all cases, the predicted values correspond well with observed values. Table 2 shows some statistical performance indicators for training and testing phases of modeling. In all cases, the training dataset showed a better result than the testing. In general, both the RMSE (m or m\(^3/\text{sec}\)) and MAPE were lower, and \( E \) was higher for daily simulation of DL. The monthly simulations for RR-GF showed lower performance than the daily and weekly simulation results.
As shown in the figures, predicted results were in better agreement for the DL subbasin than for RR-GF. The slow rise of the lake level in DL was predicted better than a rise at the RR-GF. Since the antecedent lake level and river flow were used as input in the model, the use of predicted versus observed correlation will be less informative than the change in predicted relationship versus change in observed relationship.

The ability of ANN to explain the change in hydrographs was clearly presented (Figure 10). The correlation coefficients, $R^2$ (Table 2 and Figure 10), were higher for DL than for RR-GF. The unexplained (error) variance was higher for RR-GF than DL in all simulations. Lower $E$ values indicate higher error variance ($\sigma^2_e$), as shown by Equation (6).

## CONCLUSION AND RECOMMENDATION

The study presents a comparative evaluation of the performance of an artificial neural network to the problem of flood prediction for three time scales of two subbasins of the Red River of the North basin. A
multilayer neural network model with error back propagation algorithm for flood prediction at DL and RR-GF was developed. The modeling exercise seems to work well and to predict lake stage and river discharge for the three time scales considered. The results showed better promise for DL than for RR-GF. Comparison of the different input sets of data indicated that the inclusion of antecedent precipitation and lake stage/river discharge (Input I) improved the prediction, with good correspondence of the observed and predicted values. Comparison of the model output using Input I data with the observed values showed average testing prediction efficiency ($E$) of 86 percent for DL and 46 percent for RR-GF subbasins, with higher efficiency for daily than for monthly simulations. For both DL and RR-GF stations, the daily time step was predicted better than the weekly and monthly scales.

![Figure 7. Daily Observed and ANN Hydrographs for the Testing Phase Using Input I Dataset: (a) DL and (b) RR-GF.](image_url)

The main problems with ANN seem to be its lack of explanation capabilities and of a proper building methodology to define the network architecture. At present, the ANN modeling process is basically empirical. There are no established modeling methodologies or tests for network selection. On the other hand, in contrast with expert systems, it is impossible to know how the network has arrived at a particular result. Moreover, since there are no statistical considerations involved in the ANN modeling, it can only supply point predictions. An alternative approach to alleviate this will be the use of the neuro fuzzy technique, in which prior knowledge and data-driven information are used to maximize the forecasting capability of the neural networks. Such an approach requires external information that can be used in the analysis.

ACKNOWLEDGMENTS

The authors acknowledge George Seielstad, Vijay Nangia, and David Baumgartner and other members of the Upper Midwest Aerospace Consortium for their comments. The research was partially funded by NASA grant NAG 13-02047.
Figure 9. Average Monthly Observed and ANN Hydrographs for the Testing Phase Using Input I Dataset: (a) DL and (b) RR-GF.

LITERATURE CITED


Figure 10. Change in Observed Versus Change in Predicted Values Using Input I Dataset for the Testing Phase.


