

WATER LEVEL PREDICTION BY ARTIFICIAL NEURAL NETWORK IN A FLASHY TRANSBOUNDARY RIVER OF BANGLADESH

BISWAS R.K.¹
JAYAWARDENA A.W.²

¹*Office of the Chief Planning
Bangladesh Water Development Board (BWDB)
Ministry of Water Resources
Motijheel C/A, Dhaka-1000, Bangladesh*

²*International Centre for Water Hazard and Risk Management
Public Works Research Institute (PWRI), Tsukuba, Japan*

Received: 17/09/2013
Accepted: 11/04/2014
Available online: 02/05/2014

**to whom all correspondence should be addressed:*
e-mail: biswasrobin@gmail.com

ABSTRACT

This paper presents the sensitivity analysis results of feed forward multilayer perceptron based Artificial Neural Network model for water level prediction in a data constraint transnational Surma River of Bangladesh. Catchment characteristics, hydro-geomorphological, meteorological and headwater information of the upper catchment area are not available to the authors. As such past daily total rainfall and water levels data available within the country are utilized in this study. Logistic sigmoid activation function with unit steepness parameter is exercised for non-linear transformations in both hidden and output layers. Synaptic weights are adjusted using modified delta rule through error back propagation algorithm. Batch mode of training is adopted for global error minimization. Finally, statistical indicators are used to evaluate the prediction performance of the neural network. The model is then applied to predict water levels with twenty four and forty eight hours lead time.

It is found that a single hidden layer with two hidden neurons are adequate to train the network. A higher number of hidden neurons is speeding up the training procedure, but with an unacceptable generalization for the application. The authors have successfully created a model that recognizes the intricate pattern of water levels, without having the spatially distributed geomorphic characteristics of the watershed and the time-series of the climatic factors.

Keywords: Flood Forecasting, Stochastic Hydrological Modelling, Artificial Intelligence, Feed Forward Neural Network, Surma River, Water Levels,

1. Introduction

The conversion of rainfall to river stage is influenced by a number of extremely complex, uncertain and nonlinear physical processes. Catchment's responses to any hydrological event are convoluted by various hydro-geomorphologic and climatic factors. These factors are distributed within a system in a heterogeneous manner both spatially and temporally (Anmala *et al.*, 2000). To understand the underlying physical processes, detailed information of the various hydro-geomorphologic and climatic factors along with catchment characteristics are required. Data collection except from being time consuming, is usually very costly. In the case of transnational rivers it also depends on the knotty political relationship among the shared basin countries.

Deterministic (process-based) and Stochastic are two distinctive approaches to model the complex and nonlinear relationship of rainfall and river stage. Process-based model describes the physical processes involved in the hydrologic cycle. This method is widely used to predict river flow. But according to Minns and Hall (1996) the quality of the parameter values obtained from doesn't always approach the values determined by the direct measurement, because of the high degree of empiricism in the representation of certain hydrological processes.

On the contrary, system analysis deals with the direct solution of technological problems, by using the constraints imposed from the available data sets without taking account of the physical processes of catchments (Amorocho and Hart, 1964). This technique was revitalized through the adoption of artificial intelligence techniques such as artificial neural networks (ANNs) and genetic algorithms. The artificial neural network (ANN) technique has been explored successfully to model the complex hydrologic behavior by numerous hydrologic researchers, practitioners and scientists e.g. Shamseldin (1997); Jayawardena *et al.* (1997); Dawson and Wilby (1998); Shamseldin and O'Connor (1999, 2001); Maier and Dandy (2000); ASCE (2000a, b); Madsen *et al.* (2003); Shamseldin *et al.* (2005); Melesse and Wang (2006); Shamseldin (2010); Siou *et al.* (2011) etc. of different countries, using a variety of training approaches and working principles. In most cases ANN technique successfully predicted the overall shape of the hydrograph with reasonable accuracy.

Multilayer Perceptron (MLP) coupled with the error Back Propagation (BP) algorithm and Radial Basis Function (RBF) networks were applied successfully by Jayawardena *et al.* (1997) to predict water levels at Tai Bin Chong gauging station (downstream part of the Liu Xie River, a tributary of the Pearl River) and Shang Qiao station (west of the Pearl River in southern China). Only one hidden layer was used in this study. But number of neurons were different for both stations and approaches. Similar results are also obtained by other researchers as well. For instance, Shamseldin (2010) also used MLP comprised of one hidden layer with two neurons for forecasting the Blue Nile river flows in Sudan. Based on the performance of the model it was concluded that neural network models have considerable potential and promise to be used as an alternative approach for river flow forecasting in developing countries. Maier and Dandy (2000) reviewed critically 43 papers majority of which were trained using the back propagation algorithm and dealing with the use of artificial neural network models for the prediction and forecasting of water resources variable. Based on the clinical analysis of the results and procedure of such models, it was suggested that undoubtedly ANN could be used as a tool for the prediction and forecasting of water resources variables.

Following such experiences, the authors intend to predict the water levels in the transnational Surma River, Bangladesh. Since monitoring procedures are not established, the underlying aim is to avoid the spatially distributed geomorphic characteristics of the watershed and the temporal inconsistent climatic factors to forecast water levels.

In Bangladesh, flood forecasting starts at the end of May and continues up to October. So, forecasting of lower water levels and pre-monsoon flash flood in transnational river are not covered by exiting flood forecasting system. In this regard, this approach could be useful for planning and managing water resources during dry season along with predicting flash flood in the data constraint transnational river.

2. Characteristics of the study area

The study area is located in the agriculturally and ecologically important northeast hydrological region of Bangladesh (Figure-1) where flash flood in pre-monsoon (March-May) season is very common and one of the major concerns for the economy. Total area of this region is about 8,700 km². The region has distinctive hydrological characteristics. The region receives copious river water from catchments on the slope of the Shillong Plateau across the border in India to the north and the Tripura hills in India to the southeast. Runoff from these catchments discharge very quickly into the depressed area (known as *haor*) of Bangladesh, much of which remain flooded for more than six months each year. Meghna River acts as the single outlet of this basin. As such drainage of this area is highly influenced by water levels of

Meghna River. The region is a low laying subsiding area, about 300 km away from the Bay of Bengal with bottom elevation close to the Mean Sea Level.

The Surma and the Kushiara, the main distributaries of the Barak River, are border rivers between Bangladesh and India. The Barak River is divided into Surma (northern branch) and Kushiara (southern branch) at the Indo-Bangladesh border. Steep upstream basin topography, short concentration time, low elevation differences and sudden excessive rainfalls are the characteristics of the catchment area of Surma and Kushiara rivers. Meandering flood channels, abandoned river courses and fragileness of ecosystem are other important natural variables of this area that makes it difficult to implement flood control measures. This area is widely flooded and devastated by river flooding in every monsoon season and flash flooding during pre-monsoon season, which also has a negative effect to the local economic activity.

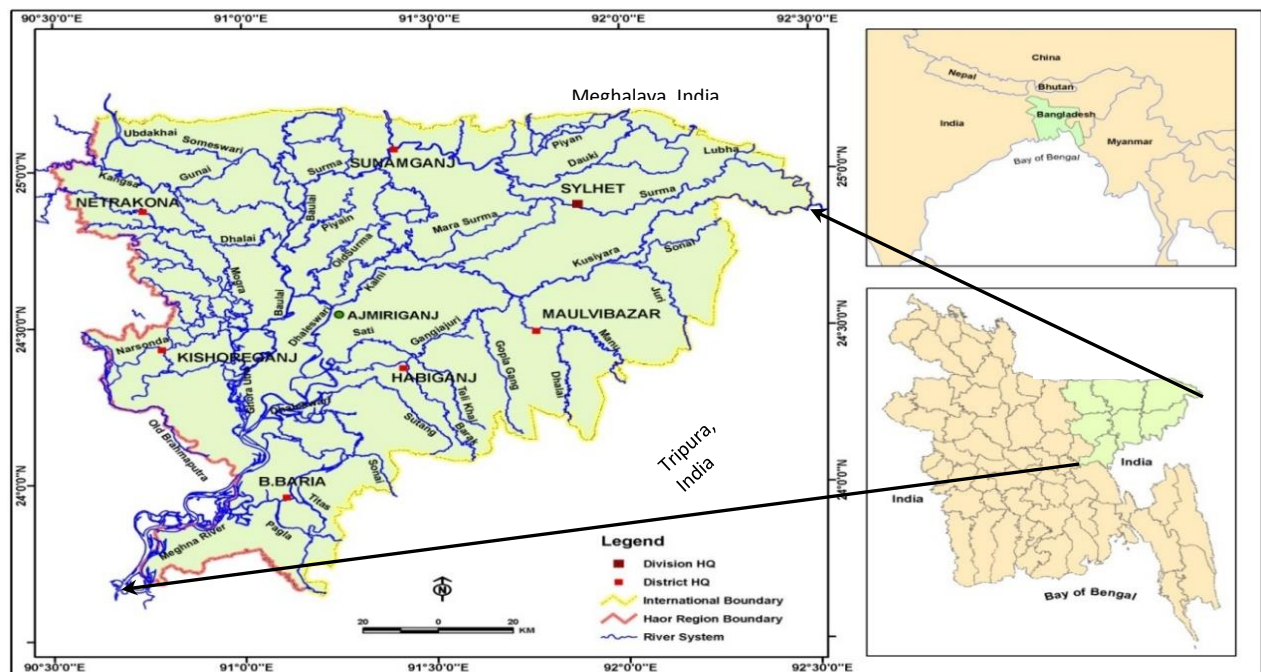


Figure 1. Northeast hydrological region of Bangladesh with major river system

The major parts of the catchments of the transnational rivers are in India and are divided into Meghalaya, Barak and Tripura River systems. The flow from these river systems occurs in the form of flash floods which usually rise to a peak within few hours. In the Bangladesh reaches of the rivers, the flood peaks are not diminished to any significant extent mainly because a huge volume of flash water enters in a short span of time. Since all the rivers tend to flood simultaneously, the hydrologic regimes of the main rivers of the region are complex and erratic (see Figure 2).

Flooding in *haor* is also largely affected by the rainfall on the adjacent upper catchments of Indian side and significant variation in rainfall in the upper catchments of different river system as well. The total average annual rainfall is higher in areas near the Meghalaya hilly region than that of the area close to the Tripura upland area with large average deviation. Most of the rainfall is concentrated in monsoon season. Annual rainfall ranges from 2200mm along the western boundary to 5800mm in its northeast corner and is as high as 12000mm in the headwaters of catchments extending to India. In Surma River, water level has two distinct peaks. One peak is in pre-monsoon season and another one during monsoon season.

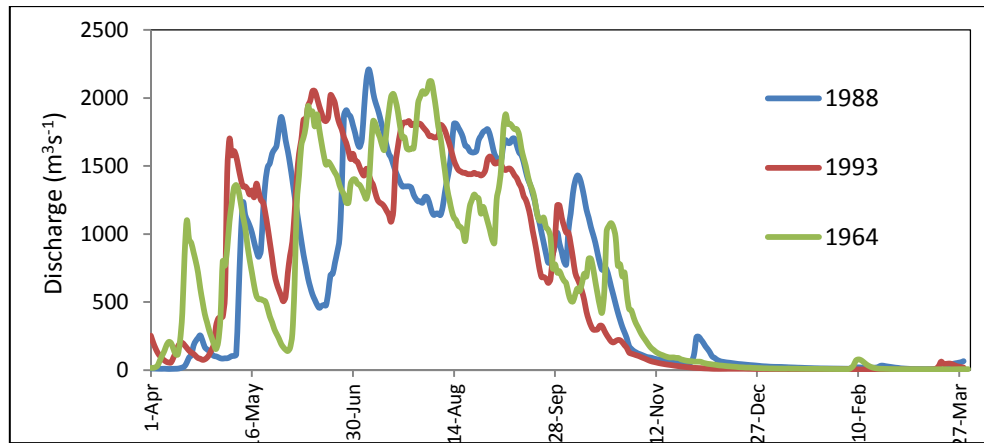


Figure 2. Discharge hydrograph of Surma River, Bangladesh at Sylhet Station

Table 1. Characteristics of rainfall and water levels at Sylhet Station, Bangladesh

Parameter	Average $\bar{X} = \frac{1}{N} \sum_{i=1}^N x_i$	Average deviation $\frac{1}{N} \sum_{i=1}^N x_i - \bar{X} $	Standard deviation $\sigma = \sqrt{C_v}$	Variance $C_v = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{X})^2$	Skewness $\frac{E(x_i - \bar{X})^3}{\sigma^3}$	Kurtosis $\frac{E(x_i - \bar{X})^4}{\sigma^4}$
Rainfall	12.07	16.38	27.52	757.32	4.26	27.06
Water levels	6.29	2.98	3.24	10.52	0.10	-1.58

The distribution pattern of rainfall and water level is shown in Figure 3. Both rainfall and water level are possessed a 'tail' pulled in the positive direction as skewness coefficients is positive (Table 1).

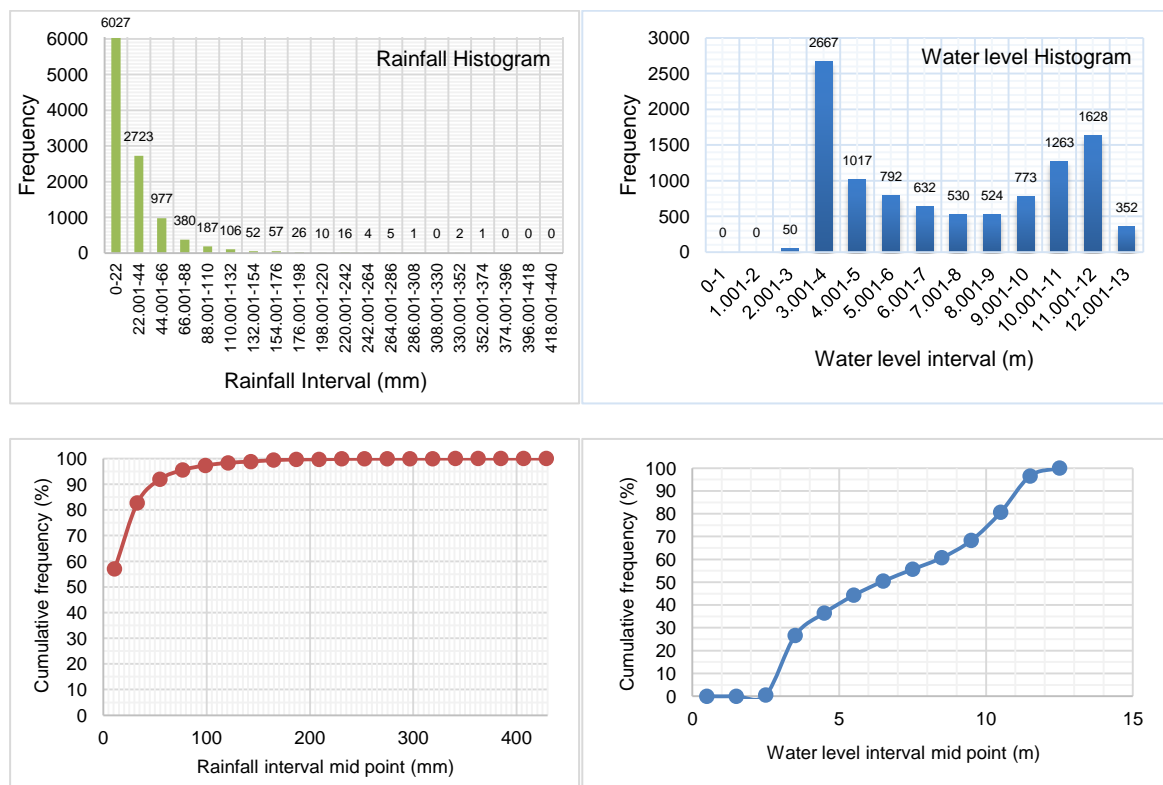


Figure 3. Rainfall and water level distribution pattern at Sylhet station, Bangladesh

But the degree of departure from symmetry is higher for rainfall than that of water level (Figure 3). The coefficient of kurtosis is positive for rainfall. This means rainfall in the study area has a higher peak with an asymmetric tail extending out towards right side than that of the normal distribution pattern. Nevertheless, water level in the Surma River has a lower peak with lighter tails than that of the normal distribution.

3. Multi-layer perceptron (MLP)

The general form of response function of two-layer perceptron network with single output neuron, I_n number of variables in an input pattern and "h" number of hidden neurons is expressed in Equation 1.

$$y = f \left(w_0^{(2)} + \sum_{\alpha=1}^h w_{\alpha}^{(2)} f \left(w_{\alpha 0}^{(1)} + \sum_{\beta=1}^{I_n} w_{\alpha\beta}^{(1)} x_{\beta} \right) \right) \quad (1)$$

where, x_{β} are the input variables, $f(\cdot)$ is logistic sigmoidal nonlinear activation function, $w_{\alpha\beta}^{(1)}$ and $w_{\alpha 0}^{(1)}$ are input weights and thresholds, $w_{\alpha}^{(2)}$ and $w_0^{(2)}$ are second layer weight and threshold. The architecture selection problem is involved to find appropriate value of I_n and h . The general form of the sigmoid activation function is defined below in Equation 2.

$$f(q) = \frac{1}{1 + e^{-Kq}} \quad (2)$$

where, q is the sum of incoming weighted signals plus associated threshold, K is a steepness parameter which can be assumed early. In this study, the value of K is equal to 1.0.

Average Square Error (ASE) has been determined by summing instantaneous square error over all training pattern and then normalizing with respect to the number of training pattern. This is described as follows in Equation 3.

$$ASE = \frac{1}{tp} \sum_{i=1}^{tp} E_i = \frac{1}{2 \times tp} \sum_{i=1}^{tp} (d_i - y_i^{(2)})^2 \quad (3)$$

The forward computation phase is terminated at the output layer by computing the error signal for each pattern. The back propagation of this error starts after the completion of error calculation for all input patterns.

The activity of the backward propagation starts with the backward movement of this error signals in layer by layer basis. The instantaneous value of error square and thus the average squared error is the function of all free parameters of the network such as synaptic weights and threshold of the network. The adjustments to the weights and thresholds have been made in accordance with the respective errors computed for each pattern. From the arithmetic average of these individual weight adjustments over the entire training pattern, actual modification to the weight vector is estimated. This correction is proportional to the derivative of the instantaneous value of error square with respect to weight. This is known as the delta rule. The general mathematical expression for computing this derivative is shown in Equation 4 below.

$$\frac{\partial E}{\partial w} = \frac{\partial E}{\partial e} \times \frac{\partial e}{\partial y} \times \frac{\partial y}{\partial q} \times \frac{\partial q}{\partial w} \quad (4)$$

where, E is the instantaneous value of error square, e is the error for an input output mapping, y is the output of a neuron (hidden or output), q is the weighted summation of incoming signals along with the bias and w is the weight.

The local gradient and correction to the weights according to delta rule is defined as Equation 5.

$$\delta = \frac{\partial E}{\partial e} \times \frac{\partial y}{\partial q} \Delta w = -\eta \frac{\partial E}{\partial w} \quad (5)$$

Where, δ is the local gradient, η is the learning rate parameter. The negative sign in weight change indicates gradient descent in weight space.

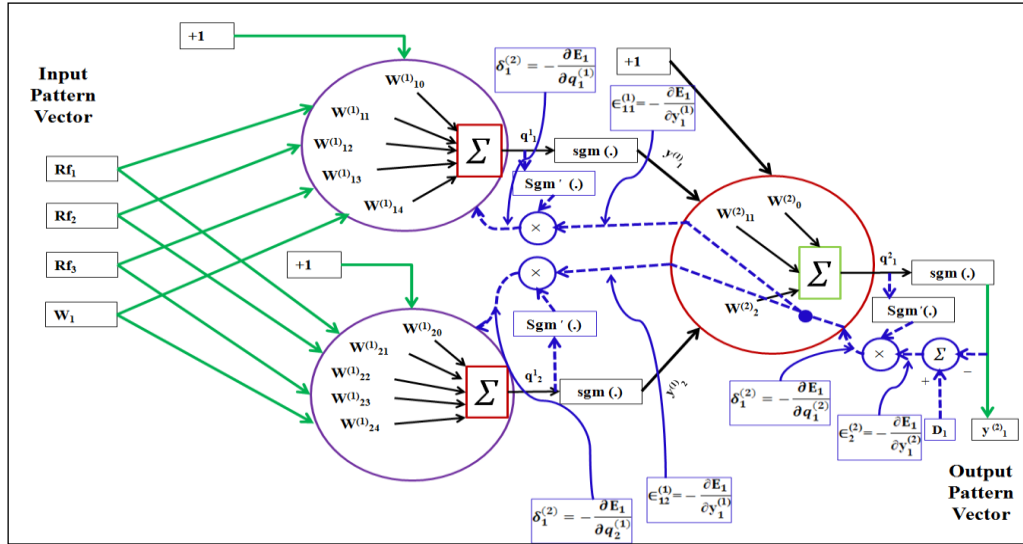


Figure 4. Forward and backward pass activities in MLP

Modified delta rule has been used to improve the slowness of learning which includes a momentum term. The modification has been made after accomplishing the computation of forward pass for all training patterns. The activities of back propagation algorithm have been stopped when the absolute rate of change in the averaged squared error (ASE) per epoch is very small or other way very close to zero. A FORTRAN programming code has been developed in order to complete the forward and backward pass calculation. The associated activities in forward and backward pass for an input pattern with four variables, two hidden neuron and one output neuron is shown in Figure 4.

The predicted results are then evaluated quantitatively by Mean absolute error (MAE), relative root mean square error (RRMSE) and coefficient of efficiency (EF) as expressed by Equation 6, 7, 8, 9.

$$MAE = \frac{\sum_{i=1}^N |O_i - P_i|}{N} \quad (6)$$

$$RRMSE = \sqrt{\frac{\sum_{i=1}^N (P_i - O_i)^2}{N}} \frac{1}{\bar{O}} \quad (7)$$

$$EF = 1 - \frac{\sum_{i=1}^N (P_i - O_i)^2}{\sum_{i=1}^N (O_i - \bar{O})^2} \quad (8)$$

$$R^2 = \frac{[\sum_{i=1}^n (P_i - \bar{P})(O_i - \bar{O})]^2}{\sum_{i=1}^n (P_i - \bar{P})^2 \sum_{i=1}^n (O_i - \bar{O})^2} \quad (9)$$

where, O_i is the observed value; P_i is the predicted value.

4. Model development

The activities in the forward pass start with the presentation of the input data to the network. These inputs are the stimulus signal of the network. Before presenting the data to the network, the input and

desired output data for training, validation and application are normalized using the relationship as delineated below in Equation 10:

$$NV = LB + \frac{(Value - Min)(UB - LB)}{(Max - Min)} \quad (10)$$

where “LB” and “UB” are expected minimum and maximum response of the network. Highest and lowest values of a data set have been defined as “Max” and “Min” and NV is the normalized value.

Daily water levels and rainfall data are collected manually by the Bangladesh Water Development Board (BWDB). Data from 18/8/1980 to 28/8/2008 at Sylhet and Sunamganj gauging stations of river Surma have been utilized for training, validating and application by splitting them into three equal parts. The first part (20/8/1980 to 22/12/1989) which contains the highest and lowest water levels has been used to train the network. Thus, the training range includes the flood event of 1987 and 1988. The second (23/12/1989 to 26/4/1999) part has been used to validate the network. The water levels of massive flood in 1998 are within the validation range. The rest of the data are used for application purpose, containing water levels of 2004, 2007 and 2008 flood events.

5. Parameter selection by sensitivity analysis

5.1 Combination of Input-Rainfall and Water-Levels

A number of rainfall and water-level combinations have been used for both stations to select the appropriate rainfall and water-level arrangement that predicts the water levels with twenty four hours lead-time. MLP with two hidden neurons and one output neuron, momentum coefficient and learning rate parameter equal to 0.50 and 0.10 respectively and normalization range 0.10-0.90 is used for this sensitivity analysis. Performance of different combinations is evaluated quantitatively by statistical indicators. Based on the results, it has been observed that rainfall with one, two and three day lag-time in addition to the water levels with one day lag-time is able to predict the water levels with twenty four hours lead-time more accurately. The combination is selected for further sensitivity analysis and application. Exclusion of rainfall or inclusion of one or more water levels of higher lag-time is not bringing any improvement, on the contrary it diminishes the performance of prediction. Some statistical indicators are performing relatively shoddier.

5.2 Influence of learning rate parameter and momentum coefficient

The converging patterns using momentum coefficients equal to 0.40, 0.50 and 0.60 of total relative root mean square error in each epoch for different learning rate parameter for Sylhet and Sunamganj station are studied.

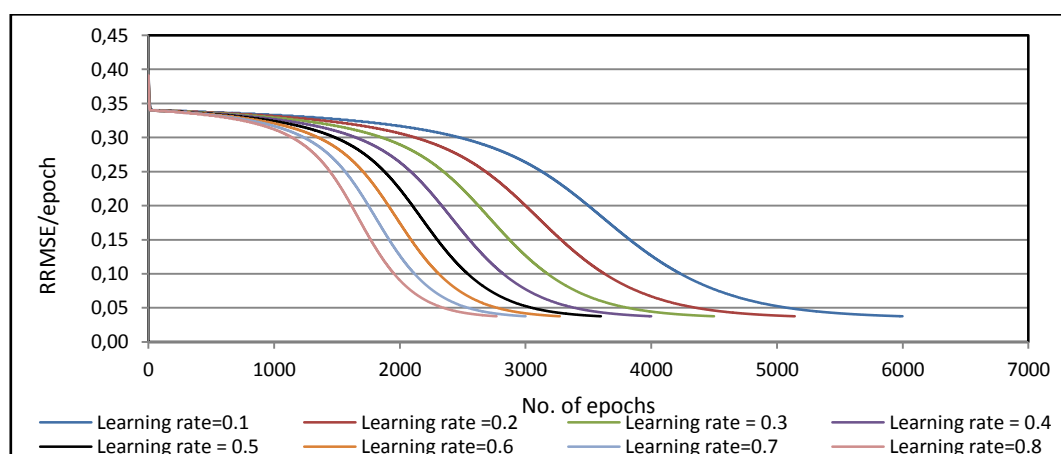


Figure 5. Convergence pattern during training for different learning rate at Sunamganj station with Momentum coefficient=0.50

Figure 5 shows the converging pattern for momentum coefficient 0.50 and for different learning rate parameter for Sunamganj station. Similar pattern has also been observed for Sylhet station as well.

For lower momentum coefficient and learning rate parameter, the required numbers of epochs to attain the convergence point are higher than that of the higher value of these parameters. The observations are similar to the experimental findings described by Haykin (1998). This signifies that smaller learning rate parameter is causing small change to the synaptic weights in the network from one to the other iteration. Larger momentum coefficients are able to improve this slowness. But there exists the potential danger of oscillation in the generating error surface. A constant learning rate with a momentum coefficient equal to 0.10 and 0.50 respectively gives best generalization. Based on the definition of optimal learning and momentum constant provided by Haykin (1998) and good convergence by Boyan and Moore (1995), 0.1 and 0.5 are selected as the optimal value of learning rate parameter and the momentum coefficient. Although speedy convergence is attained for higher values of learning rate and momentum constants, because of the potential danger of oscillation, these values are avoided.

5.3 Sensitiveness of number of hidden neurons

The convergence patterns during training for different number of hidden layer neuron at Sylhet and Sunamganj gauging stations are observed by using one, two and three day lagged rainfall and one day lagged water levels as the input pattern. The momentum coefficient and learning rate are 0.50 and 0.10 correspondingly. These parameters are kept same for both layers. In both stations with the increase of numbers of hidden nodes, the initial relative root mean square errors are increasing. The initial rate of declination of the RRMSE is very swift. With the increase of number of iteration, the rate of declination or in other words, the slope of the relative root mean square error curve is becoming very gentle. The initial and final RRMSE at each simulation together with the number of required epoch is needed to reach the stopping criteria are given in Table 2.

Table 2. Initial and final RRMSE for different hidden neurons

No. of hidden neurons	Initial RRMSE		Final RRMSE		No. of epochs	
	Sylhet	Sunamganj	Sylhet	Sunamganj	Sylhet	Sunamganj
2	0.509	0.391	0.047	0.037	9845	5174
4	0.781	0.619	0.051	0.043	3429	3436
5	0.807	0.641	0.053	0.043	3121	3465
6	0.854	0.683	0.053	0.043	3059	3514

At both stations, the required numbers of epochs are the highest when hidden neurons equals to two but final RRMSE is the lowest. This signifies that increment of number of hidden nodes is improving the required time to train a network. At the same time, the performance is deteriorating. Better generalization is one of the conditions of optimality of better trained networks which can perform the recognition from the experience. The relation between measured and predicted water levels for different number of neurons at hidden layer in case of training, validation and application for both stations are investigated. It is observed that the strength of linear relationship between measured and predicted water levels, decreases in the simulation for higher number of hidden neurons. Rapid declination of the slope of the error gradient can be one of the reasons for this poor generalization. Contrastingly, the performance using two hidden neurons is reasonably close to the measured water levels. Better generalization is also obtained using two hidden neurons. On this basis, it is concluded that two hidden neurons are adequate for the nonlinear mapping of input to the output.

5.4 Influence of data normalization

A number of simulations have been conducted in this study in order to find the influence of data normalization. No normalization has also been applied, while rainfall and water-level combinations have been normalized between the ranges of 0.10-0.90, 0.15-0.85 and 0.20-0.80. The desired water levels are always normalized.

For smaller number of hidden neurons, oscillations are observed at both stations when approaching the stopping criteria when the simulation doesn't normalize the input data. Increasing the number of hidden neurons, results to an improvement in the oscillation pattern, but still the network attains stopping criteria with higher relative root mean square error, after several iterations. In order to improve this condition, the input rainfall and water levels and the desired water levels are normalized within the range of 0.10 to 0.90 (see Figure 6). Normalization results to an improvement in the oscillation and early stopping problems. But the predicted lower water levels are higher those measured. In order to improve this shortcoming, in the third case, the input data are normalized to the range of 0.15 to 0.85 which enhance the strength of the collinearity between the measured and predicted water levels. In fourth case, the input data has been normalized in the range 0.2 to 0.80 with better agreement between measured and predicted water levels, compared to previous three normalization cases. Sigmoid function is used as the non-linear transfer function both in hidden layer and output layer, which has the bounded output range between zero and one. The actual observed output of the network is outside of the bounded range of the sigmoid transfer function. So data normalization is prerequisite and it is one of the important aspects in designing the topology of the network. When the rescaled values are close to the limiting boundary of the activation function, the network is producing better results. The performance of error convergence gives the similar pattern as described by Van Ooyen and Nienhuis (1992).

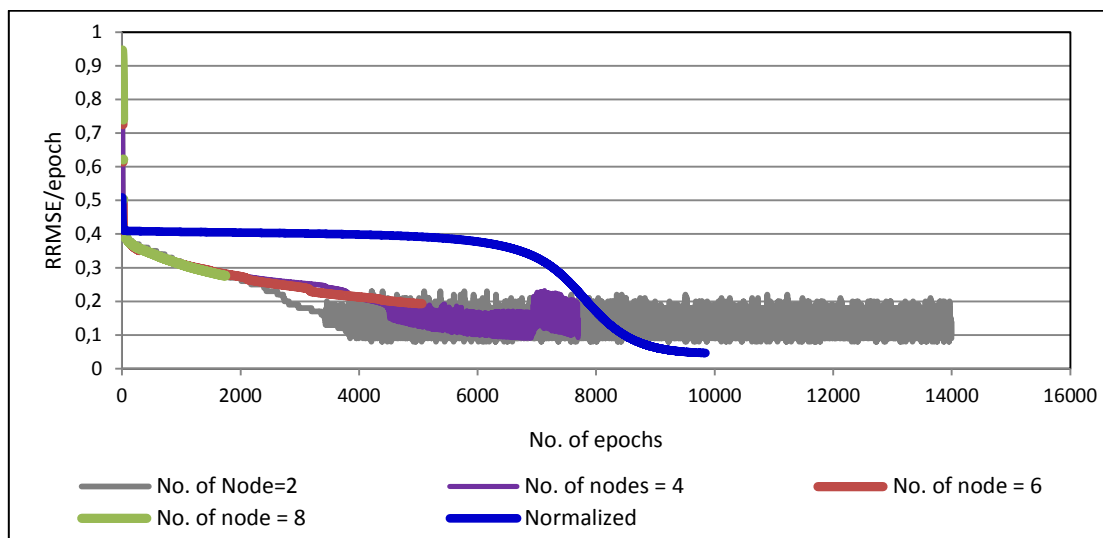


Figure 6. Converging pattern of non-normalized and normalized data for different no. of hidden neurons at Sylhet station

6. Outline of selected MLP topology

Designing the architecture of the network consists of the selections of the number of input variables in a pattern, number of hidden layers and number of hidden neurons in each layer. The numbers of output neurons are predefined. In this study, single output neuron is predetermined. Based on the results of the sensitivity analysis one, two and three day lagged rainfall, in conjunction with one day lagged water levels, are selected as the variables in an input pattern. The number of hidden layer is one and it is

composed of two hidden neurons. Water level is the single output, expected from the network. Learning rate parameter, momentum coefficient and synaptic weight of the network are the other free parameters of the network.

7. Application for forecasting

The selected data combination (rainfall with one, two and three along with water levels with one day lag-time), learning rate parameter (0.10), momentum coefficient (0.50) and data normalization range (0.20-0.80) and model topology (two hidden neurons and one output neurons) are applied to predict the water levels with 24 and 48 hrs lead time for Sylhet and Sunamganj station.

Comparing the results of the application phase of simulation with the measured water levels in a time series plot (Figure 7 & 8), indicates that continuous prediction of river water levels with ANN is possible. The actual water levels are very close to the neural network predictions, at both stations. The data region where the pattern is defined clearly is predicted fairly. The lower water levels are over predicted. The scatter diagrams show an almost linear relation of the prediction with the observations. The Mean Absolute Error (MAE) is a measure of the closeness of the predicted water level to the measured water level. MAE value is close to the optimum value (0.0) which indicates that the measured water levels are well-modelled. On the other hand, mean absolute error for earlier, lead-time predictions. This occurs because the deviation of predicted values from the measured values is increasing. Relative Root Mean Square Error (RRMSE) evaluates the concentration around the mean of the observed values and it is close to zero. This means that the predicted water levels are not dispersed around the mean observed value. Nash-Sutcliffe coefficient (EF) is used to assess the predictive power of a model. Its value is very close to 1.0 which means the model is capable to capture the overall pattern of the water level hydrograph with reasonable accuracy. The performance of statistical indices are shown in Table 3.

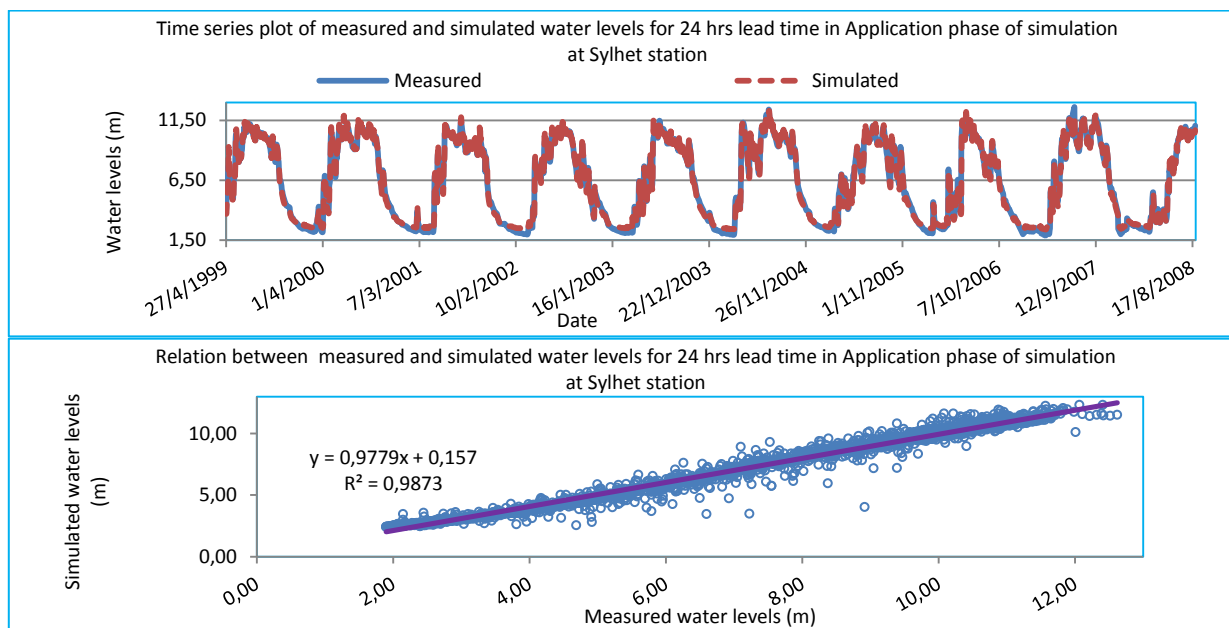


Figure 7. Comparison of measured and predicted water levels in application phase of simulation at Sylhet station for 24 hrs lead time

The stopping criteria adopted in this study, improved the performance of the model as compared to Thirumalaiah & Deo (1998). The value of R^2 is close to 1.0 which signifies a very good linear regression correlation between the measured and predicted water levels. The linear regression correlation between the measured and predicted water levels is also better for 1-day lead time prediction.

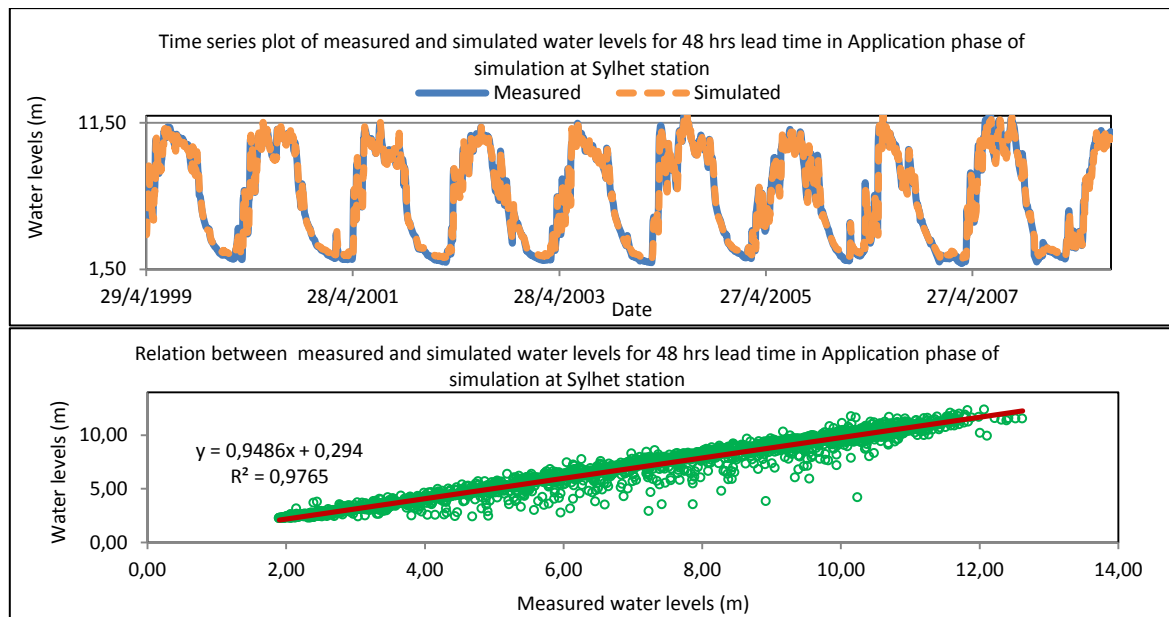


Figure 8. Comparison of measured and predicted water levels in application phase of simulation at Sylhet station for 48 hrs lead time

Table 3. Performance of statistical indices for 24 and 48 hrs lead-time

Lead time (hrs)	Phase of Simulation	Mean Absolute Error (MAE) (m)		Relative Root Mean Square Error (RRMSE)		Coefficient of Efficiency (EF)		R ²	
		Sylhet	Sunamganj	Sylhet	Sunamganj	Sylhet	Sunamganj	Sylhet	Sunamganj
24	Training	0.27	0.20	0.06	0.05	0.99	0.99	0.99	0.99
	Validation	0.29	0.19	0.06	0.05	0.99	0.99	0.99	0.99
	Application	0.26	0.17	0.06	0.04	0.99	0.99	0.99	0.99
48	Training	0.35	0.24	0.08	0.06	0.98	0.98	0.98	0.98
	Validation	0.38	0.22	0.09	0.07	0.97	0.98	0.97	0.98
	Application	0.31	0.23	0.08	0.06	0.98	0.98	0.98	0.98

8. Conclusions

Based on the performance of the statistical indicators, it can be concluded that MLP based ANN model successfully predicts river water levels in data scarce transnational flashy Surma River without using the geomorphic characteristics of the watershed and climatic factors. For lower momentum coefficient and learning rate parameter, required number of epochs to reach the stopping criteria are higher. Larger momentum coefficients improve this delay. But oscillations are observed during error surface generation. In this study, a constant learning rate (0.10) and a momentum coefficient (0.50) produces best generalization.

In the case of non-normalized data with smaller number of hidden neurons, oscillations are also observed. Improvement over oscillation pattern is achieved by increasing the number of hidden neurons. But the stopping criteria are attained with poor generalization. On the contrary, normalized data generate better outcomes. This is due to the sigmoid activation function which was used in both

hidden and output layers as the nonlinear transfer function. When the rescaled values are not close to the limiting boundary of the activation function, the network is capable to comprehend the measured water level patterns more accurately. This implies that data normalization is very useful in case of multilayer perceptron feed forward neural network where sigmoid function has been used as the nonlinear transfer function. Normalized range 0.20-0.80 produces better generalization during application in this study.

With the increase of number of hidden neurons, the required number of iterations to attain the stopping criteria are lower. But this gives poor generalization in the application phase of simulation because of the rapid change in the gradient of the error surface. It has been found that the performance with two hidden neurons is better than that of higher number of hidden neurons. Predicted water levels are close to the measured water levels as well. Higher water levels of the river are found to be trained more fairly than that of the lower water levels. Pre-monsoon water level is also predicted quite reasonably by the model.

REFERENCES

- Amorocho J. and Hart W.E. (1964), A critique of current methods in hydrologic systems investigation, *Trans. AGU*, **45**, 307-321.
- Anmala J., Zhang B. and Govindaraju S. (2000), Comparison of ANNs and empirical approaches for predicting watershed runoff, *Journal of Water Resources Planning and Management*, **126**(3).
- ASCE (2000a), Task Committee on Application of Artificial Neural Networks in Hydrology, Artificial Neural Networks in Hydrology. II: Hydrologic Application, *Journal of Hydrologic Engineering*, **5**, pp 124-136.
- ASCE (2000b), Task Committee on Application of Artificial Neural Networks in Hydrology, Artificial Neural Networks in Hydrology. I: Preliminary Concepts, *Journal of Hydrologic Engineering*, **5**, pp 115-123.
- Boyan J.A. and Moore A.W. (1995), Generalization in reinforcement learning: safely approximating the value function. In: Tesauro G., Touretzky D.S. and Leen T.K. (eds), *Advances in neural information processing system 7*, MIT press, Cambridge MA.
- Dawson C.W. and Wilby R. (1998), An artificial neural network approach to rainfall-runoff modelling, *Journal of Hydrological Science*, **43**(1).
- Haykin S. (1998), *Neural Networks: A Comprehensive Foundation* (2nd Edition). Prentice Hall, New Jersey.
- Jayawardena A.W., Fernando D.A.K. and Zhou M.C. (1997), Comparison of Multilayer Perceptron and Radial Basis Function networks as tools for flood forecasting. Destructive Water: Water-Caused Natural Disasters, their Abatement and Control (Proceedings of the Conference held at Anaheim, California, June 1996). IAHS Publ. no. 239.
- Madsen H., Rosbjerg D., Jesper D. and Hansen F.S. (2003), Data assimilation in the MIKE 11 Flood Forecasting system using Kalman filtering. Water Resources Systems- Hydrological Risk, Management and Development (Proceedings of symposium IIS02b held during IUGG2003 at Sapporo. July 2003). IAHS Publ. no. 281.
- Maier H.R. and Dandy G.C. (2000), Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications, *Environmental Modelling & Software*, **15**, pp 101-124.
- Melesse A.M. and Wang X. (2006), Multitemporal Scale Hydrograph Prediction Using Artificial Neural Networks, *Journal of the American Water Resources Association*, **42**(6), 1647-1657.
- Minns A.W. and Hall M.J. (1996), Artificial neural networks as rainfall-runoff models, *Journal of Hydrological Sciences*, **41**(3), 399-417.
- Shamseldin A.Y. and O'Connor K.M. (1999), A real-time combination method for the outputs of different rainfall-runoff models, *Hydrological Sciences Journal*, **44**, pp 895-912.
- Shamseldin A.Y. and O'Connor K.M. (2001), A non-linear neural network technique for updating river flow forecasts, *Hydrology and Earth System Sciences*, **5**, pp 577-597.
- Shamseldin A.Y., Abrahart R.J. and See L.M. (2005), Neural network river discharge forecasters: an empirical investigation of hidden unit processing functions based on two different catchments, *Neural Networks*, **4**, 2655 – 2660.

- Shamseldin A.Y. (1997), Application of a neural network techniques to rainfall-runoff modelling, *Journal of Hydrology*, **199**, pp 272-294.
- Shamseldin A.Y. (2010), Artificial neural network model for river flow forecasting in a developing country, *Journal of Hydroinformatics*, **12**(1), pp 22-35.
- Siou L.K.A., Johannet A., Borrell V. and Pistre S. (2011), Complexity selection of a neural network model for karst flood forecasting: The case of the Lez Basin (southern France), *Journal of Hydrology*, **403**(3-4), pp 367-380.
- Thirumalaiah K. and Deo M.C. (1998), River stage forecasting using artificial neural networks, *Journal of Hydrologic Engineering*, **3**, No.1.
- Van Ooyen A. and Nienhuis B. (1992), Improving the convergence of the back-propagation algorithm, *Neural Networks*, **5**, pp 465-471.