Artificial Neural Network for Precipitation and Water Level Predictions of Bedup River

Rosmina Bustami, Nabil Bessaih, Charles Bong, Suhaila Suhaili

***Abstract*—This study aims to improve water level prediction at Bedup River with estimations made to absent precipitation data, both using Artificial Neural Network (ANN). Studies to predict water level in the state of Sarawak, Malaysia have been actively carried out. However, among problem faced was absent precipitation readings, which inevitably affected water level precipitation accuracies. Backpropagation properties of ANN was used in the study to predict both missing precipitation and water level. ANN model developed in this study successfully estimates missing precipitation data of a recorder in Bedup River, Sarawak with 96.4% accuracy. The predicted values of precipitation were then used to forecast water level of the same gauging station and yielded accuracy value of 85.3%, compared to only 71.1% accuracy of water level prediction with no estimation made to its missing precipitation data. These results show that ANN is an effective tool in forecasting both missing precipitation and water level data, which are utmost essential to hydrologists around the globe.**

***Index Terms*—Artificial Neural Network, Backpropagation, Precipitation Prediction, Water Level Prediction**

I. Introduction

Sarawak is the largest state in Malaysia. Its coastal zone is generally flat and low lying. The area is very much influenced by tidal effect. Despite being 35 km from the sea, Kuching, capital city of Sarawak, experiences a tidal range of over 6m during spring tides, or locally known as King Tides. The annual rainfall for Kuching is very high, about 3800 mm, which makes the area prone to flooding. Over the last 40 years, there has been a number of significant hydrological events, all of which caused extensive flooding throughout Sarawak River. In January and February 1963, extremely heavy rainfall recorded at 2500 mm for two months brought in flood of over 7m depth in Sarawak River. The mishap claimed 4 lives, with 800 longhouses badly damaged and destroyed all over Sarawak. Flooding occurred again in January 1974, February 2003 and January 2004 resulting from prolonged rainfall and occurrence of spring tides.

Knowing that economic development in Sarawak takes place by the rivers, accurate forecasting of water level is therefore essential to warn public of potential rise in water level and call for necessary precautions.

Among the widely used methods for estimating missing precipitation are normal-ratio, arithmetic, inverse distance, isohyetal and thiessen polygon methods. With exception to normal-ratio and arithmetic methods, the rest of estimation methods require parameters such as distance and/or topographical conditions of the area.

Water level prediction via conventional method needs accurate estimation of runoff from a given rainfall event and an accurate hydraulic model for a given discharge. Runoff generation highly depends on catchment topography, river network, soil characteristics and antecedent moisture. On the other hand, hydraulic models are available only for a limited number of cross-sections. All these parameters are not all the time available, thus making estimation of water level very complex.

ANN was chosen for its ability to generalize results from unseen data and well-suited in modeling dynamic systems on a real-time basis. These properties of ANN are suitable to forecast water level and missing precipitation as their physical relationships are not well understood.

ANN has also been used in water resources engineering over the last decade. These include flood forecasting (R. Garcia Bartual 2002, Wright and Dastorani 2001), rainfall-runoff modeling (Tokar and Johnson 1999, Sobri Harun *et al.* 2002, Thurumalaiah and Deo 2000), streamflow prediction (Dolling and Varas 2001, Dastorani and Wright 2002, Wright *et al.* 2002), and water level prediction (Patrick and Collins 2002, Huang *et al.* 2003).

In this study, estimation of missing precipitation were made using Normal-ratio method and ANN. Results of the two simulations were then used as inputs along with available precipitation data to test the ANN ability in forecasting water level data. A third ANN water level predicting model was created with no estimations made to its missing precipitation data. This will enable us to see if there are any significant effects of estimating missing precipitation data on water level prediction.

Manuscript received March 13, 2006.

R. Bustami is with the Department of Civil Engineering, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia. (Phone: +60-82-583271; fax:+60-82-583409; e-mail: abrosmina@feng.unimas.my).

N. Bessaih was with Universiti Malaysia Sarawak, Malaysia. He is now with the Ministry of Finance, Algiers, Algeria (e-mail: nabil.bessaieh@mf.gov.dz).

C. Bong is with the Department of Civil Engineering, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia. (e-mail: bhjcharles@feng.unimas.my).

S. Suhaili was with Universiti Malaysia Sarawak, Malaysia. She is now working in a private firm in Kuching, Sarawak. (e-mail: 8135@upgmail.unimas.my)

II. Study Area

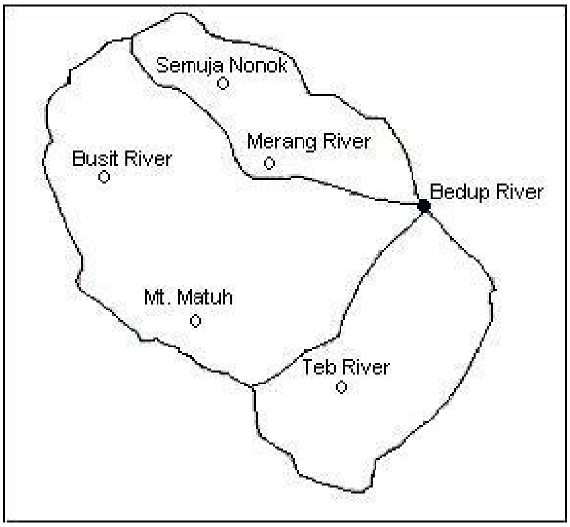

Fig. 1 Bedup River Catchment

The study area selected for this study is Bedup River Catchment, shown in Fig. 1. This catchment is part of Sadong Basin and is situated approximately 80 km away from Kuching. The surface area of the catchment is 48 km^2. The catchment comprises few villages namely Kampung Sungai Berok, Kampung Merjan and Kampung Longgo. The catchment elevation varies from 8m to 686m above mean sea level. The vegetation cover consists mainly of shrubs, low plant and forest. Bedup River basin has a dendritic type of channel system. The maximum stream length for the basin is approximately 10 km, which is measured from the most remote point of the stream to the basin outlet.

There are five rainfall stations and one water level station distributed over the catchment. Rain gauge stations are located at Merang River, Teb River, Busit River, Semuja Nonok and Mount Matuh. A water level station is located at the outlet of the basin, Bedup River Station. Data used for this study were from the years 2000 to 2004, obtained from Department of Irrigation and Drainage, Kota Samarahan, Sarawak.

III. Methodology

Methodology of this study is divided into three parts. The first part explain methods used for both precipitation and water level predicting ANN, the second part discuss methods used to develop ANN model estimating missing precipitations, where as the third part discuss methodology carried out in water level prediction.

A. Precipitation and Water Level Prediction

For both precipitation and water level predicting ANN, a two-layer Multilayer Perceptron (MLP) backpropagation network was chosen, as shown in Fig. 2. In this network, the input data are fed to input nodes and then they will pass to the hidden nodes after multiplying by a weight. A hidden layer node

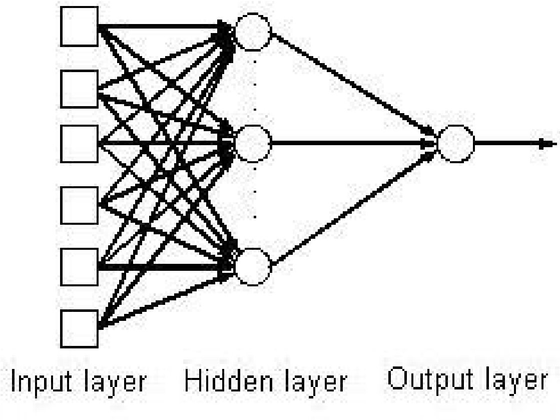

Fig. 2 MLP with one hidden layer

adds up the weighted input received from the input nodes, associates it with the bias and then passes the result on through a nonlinear transfer function. The output node does the same operation as that of a hidden layer. This type of network is preferred as backpropagation learning is a popular algorithm to adjust the interconnection weights during training, based upon the generalized delta rule proposed by [4]. Furthermore, it has shown outstanding forecasting performance in hydrological simulations.

Performances of all networks are measured by coefficient of correlation, *R*, given by (1).

$$R = \frac{\sum_{i=1}^{i=n}[Q_m - \overline{Q_m})][Q_s - \overline{Q_s}])}{\sqrt{\sum_{i=1}^{i=n}[Q_m - \overline{Q_m}]^2 \sum [Q_s - \overline{Q_s}]^2}} \tag{1}$$

where:

Q_m : Actual value of precipitation/ water level

Q_s : Simulated value of precipitation/ water level

$\overline{Q_m}$: Average of actual value of precipitation/ water level

$\overline{Q_s}$: Average of simulated value of precipitation/ water level

Different transfer functions for hidden and output layers were used to find the best ANN structure for this study. Transfer function used in hidden layer of the backpropagation network is tangent-sigmoid while pure linear transfer function is used in output layer.

ANN in this study was trained and simulated using MATLAB 6.5 developed by The Math Works Inc, Natick, Massachusetts.

B. Missing Precipitation Prediction

For the study, two techniques are used to predict missing precipitation data recorded at the gauging stations in Bedup River catchment. They are normal ratio method and ANN. Both methods are discussed further in the following sub-chapters.

B.1 Normal Ratio Method

Missing precipitation values at a site can be estimated from concurrent observations that are located as close to and evenly spaced from the missing data station as possible, known as index station [16]. The normal ratio method is:

$$\frac{P_x}{N_x} = \frac{1}{n}\left(\frac{P_1}{N_1} + \frac{P_2}{N_2} + \frac{P_3}{N_3} + ... + \frac{P_n}{N_n}\right) \quad (2)$$

where:

P_x : Missing precipitation for station X

P_1, P_2, P_3, P_n : Precipitation at neighboring station for the concurrent period

N_x : Normal long-term precipitation at station X

N_1, N_2, N_3, N_n: Normal long-term precipitation for neighboring station

B.2 Artificial Neural Network (ANN)

ANN developed for prediction of precipitation is trained with different learning algorithms, learning rates, and number of neurons in its hidden layer. The aim is to create a network which gives an optimum result.

The network was simulated using 3 different backpropagation learning algorithms. They are Resilient Backpropagation (*trainrp*), Fletcher-Reeves Conjugate Gradient (*traincgf*) and Scale Conjugate Gradient (*trainscg*).

The Resilient Backpropagation (*trainrp)* eliminates the effect of gradient with small magnitude. As magnitudes of the derivative have no effect on the weight update, only the sign of the derivative is used to determine the direction of the weight update. *Trainrp* is generally much faster than standard steepest descent algorithms, and require only a modest increase in memory requirements which suits network with sigmoidal transfer function.

Fletcher-Reeves Conjugate Gradient (*traincgf*) generally converges in fewer iteration than *trainrp*, although there is more computation required in each iteration. The conjugate gradient algorithms are usually much faster than variable learning rate backpropagation, and are sometimes faster than *trainrp*. *Traincgf* also require only a little more storage than simpler algorithms, thus they are often a good choice for networks with a large number of weights.

The third algorithm, Scale Conjugate Gradient (*trainscg*) was designed to avoid the time-consuming line search. This differs from other conjugate gradient algorithm which requires a line search at each iteration. The *trainscg* routine may require more iteration to converge, but the number of computations in each iteration is significantly reduced because no line search is performed. *Trainscg* require modest storage.

Learning rates were applied to the networks during simulations. The values ranged from 0.2 to 0.8.

There is only one hidden layer constructed for the network. Due to the complexity of input elements, number of neurons in the hidden layer must be sufficient enough to withhold the mass of inputs. Number of neurons in the hidden layer was investigated by trial and error method. The values investigated are 20, 40, 60, 80, and 100.

Daily precipitation data were chosen for training and testing. Networks were trained with data from year 2000 to 2004. The output of the built the networks are the missing precipitation of station X of the day, $P_x(t)$, with inputs being precipitation of the other 5 neighboring station of the same day. Equation for the model developed is as shown in (3).

$$P_x(t) = f[P_1(t), P_2(t), P_3(t), P_4(t), P_5(t)] \quad (3)$$

where:

P_x : Precipitation at station X (missing)

$P_1, P_2, P_3, P_4, P_5,$: Precipitation at 5 neighboring stations

C. Water Level Prediction

The ANN model used to predict water level was applied from the model recommended in. [9]. The input node consists of antecedent water level, antecedent precipitation and precipitation for the current day. Expected output for the network is water level for the current day. Equation (4) represents operation to predict water level with 4 days of antecedent data, as recommended in [9].

$$W(t) = f[P(t-4), P(t-3), P(t-2), P(t-1), P(t), W(t-4), \cdot W(t-3), W(t-2), W(t-1)] \quad (4)$$

where:

t : Time (days)

P : Precipitation

W : Water level

Predicted missing precipitation values earlier are used to predict water level of River Bedup. A total of 3 different sets of input data were developed, namely Set A, B and C.

For set A, missing precipitation data are predicted using Normal ratio method. Data for set B utilized ANN to predict its missing precipitation, where as in set C, predictions were not made to missing data.

Parameters of the recommended network from [9] are tabulated in Table I.

Table I Recommended ANN model for daily water level prediction from [9]

No. of nodes in hidden layer	20
Antecedent time	4 days
Learning Algorithm	*Trainscg* (Scale Conjugate)
Learning Rate	0.8

IV. RESULTS

Results of simulations for missing precipitation and water level predictions are discussed in the following sub-chapters:

A. Missing Precipitation Prediction

Among the different learning algorithms used in the study, *trainrp* proved to be the best algorithm in simulating missing precipitation, shown by its high value of *R* during simulation process. Results of ANN models developed with all learning algorithms analyzed are tabulated in Table II.

Table II Comparison of different learning algorithm used

Learning Algorithm	R (Training)	R (Testing)
Trainrp	1	0.964
Traincgf	1	0.905
Trainscg	1	0.949

Varying value of learning rates introduced to the network gave no significant effect on the simulation process. All learning rate values tested are able to converge in both training and testing phases. However, simulation with larger learning rate values slowed down the convergence process and is not preferred. Therefore, the smallest learning rate investigated on the network, 0.2, was chosen as the optimum learning rate.

Simulation for precipitation estimation using ANN was done with 20 to 100 neurons in the hidden layer. Comparison on number of neurons fed to the network with *R* values is shown in Table III. Simulation results show that ANN fed with 60 neurons in its hidden layer gave the highest correlation value.

Table III Influence of number of neurons in hidden layer

Number of Neurons	R (Training)	R (Testing)
20	0.998	0.905
40	0.999	0.894
60	1	0.964
80	1	0.897
100	1	0.907

B. *Water Level Prediction*

Results of water level simulation using ANN model recommended in [9] for all 3 sets created in this study is tabulated in Table IV.

Table IV Performance of ANN in water level estimation of Sets A, B and C

Set	R (Training)	R (Testing)
A	1	0.830
B	1	0.853
C	1	0.711

From Table IV, Set B with missing precipitation predicted using ANN, gave the highest correlation in predicting water level for Bedup River for the study period, which is 0.853. Set A, with missing precipitation data estimated using normal ratio method has a correlation of 0.83, while ANN network fed with no estimation made to the precipitation data (Set C) gave the lowest correlation of 0.711.

V. DISCUSSION

A. *Missing Precipitation Prediction*

It was found that backpropagation ANN developed in this study performed very well in simulating missing precipitation. Fig. 3 compares simulated precipitation with actual precipitation for the optimum network, yielding an accuracy of 96.4%. Parameters of the optimum network for missing precipitation simulations are as shown in Table V.

Table V Optimum parameters for missing precipitation prediction

No. of nodes in hidden layer	60
Learning Algorithm	*Trainrp* (Resilient Backpropagation)
Learning Rate	0.2

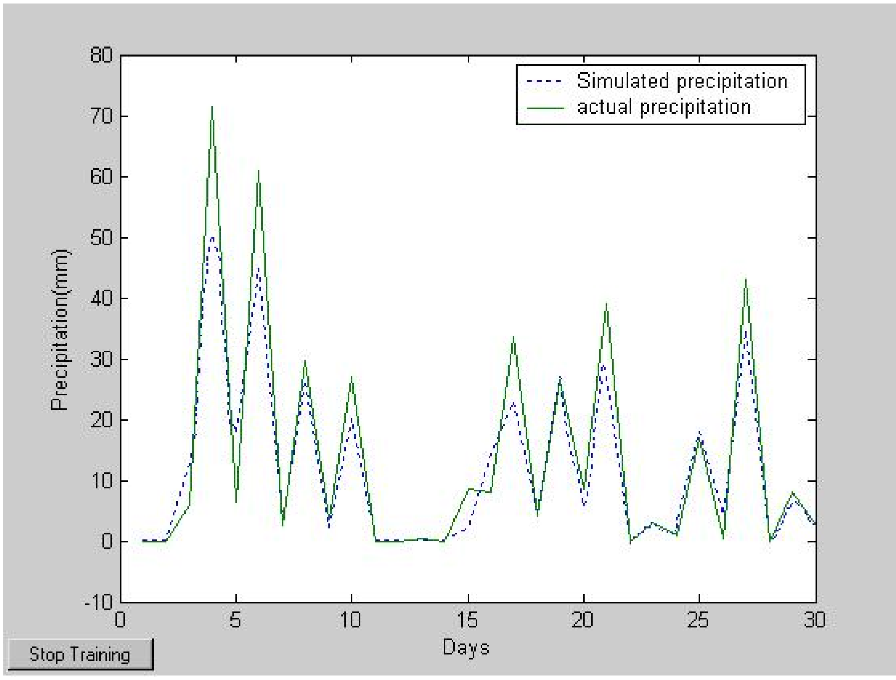

Fig. 3 Comparison between simulated and actual precipitation

B. Water Level Prediction

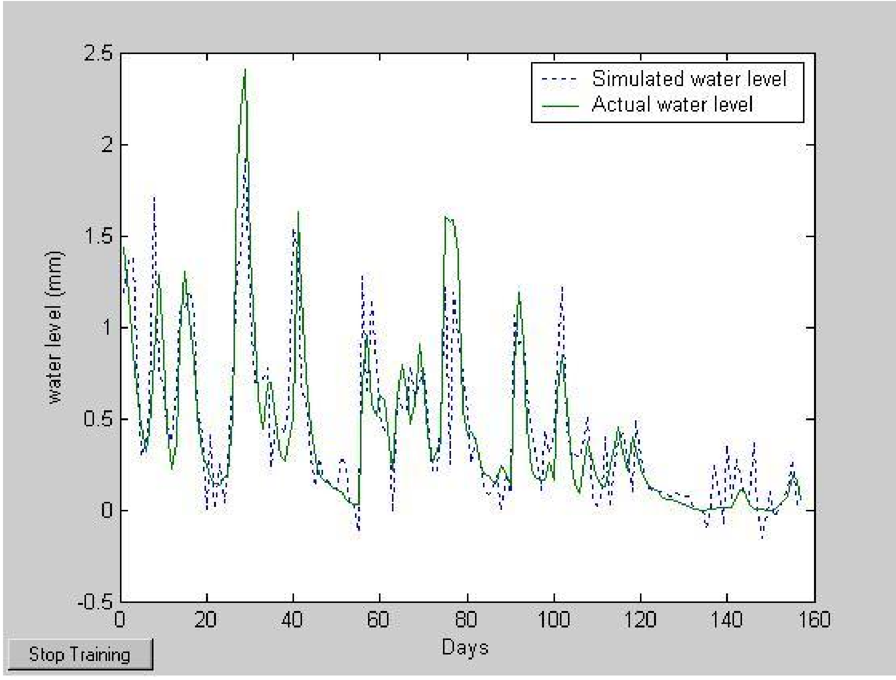

Fig. 4 Comparison between simulated and actual water level for Set B

Set B model built in this study gives the highest correlation value in predicting water level at River Bedup. Modifications were made to missing precipitation data in Bedup River by substituting them with simulated data from the first part of this study.

The findings show that ANN models predicts water level better with its missing precipitation simulated using ANN, compared to simulation of water level with missing precipitation data calculated using normal ratio formula, or simulation with missing precipitation data. Fig. 4 distinguishes simulated from observed water level for Set B with accuracy of 85.3%.

VI. Conclusion

Results from this study evidently prove that ANN models developed are reliable to estimate missing precipitation and produce higher accuracy of water level prediction at Bedup River in Kota Samarahan, Sarawak. This study has initiated a new development in water resources engineering in Sarawak, in particular, in areas where precipitation data are absent.

References

[1] A.R. Patrick, W.G. Collins, P.E. Tissot, A. Drikitis, J. Stearns, P.R. Michaud, and D.T. Cox, "Use of the NCEP MesoEta data in a water level predicting artificial neural network", in the *Proceedings of 19th Conf. on weather Analysis and Forecasting/15th Conf. on Numerical Weather Prediction, 2002.*

[2] A.S. Tokar, and P.A. Johnson, "Rainfall-runoff modeling using Artificial neural networks", *Journal of Hydrologic Engineering*, 4(3), 1999, pp. 223-239.

[3] ASCE Task Committee on Application of Artificial neural networks in Hydrology, "Artificial neural networks in hydrology (I & II)", *Journal of Hydrologic Engineering*, 5(2), April 2000, pp. 115-136.

[4] D.E. Rummelhart, G.E. Hinton, and R.J. Williams, "Learning internal representations by error backpropagation", *Parallel Distributed Processing*, vol. 1, Cambridge, Mass: MIT Press, 1986.

[5] H. Demuth, and M. Beale, *Neural Network Toolbox For Use with MATLAB.* Natick, MA: The Mathworks Inc., 2001.

[6] H. Sobri, M.K. Amir Hashim, and A.N. Nor Irwan, "Rainfall-runoff modeling using artificial neural network", *Proceedings of 2nd World Engineering Congress*. Kuching, 2002.

[7] K. Thirumalaiah, and M.C. Deo, "Hydrological forecasting using artificial neural network", *Journal of Hydrologic Engineering*, 5(2), 2000, pp. 180-189.

[8] M.T. Dastorani, and N.G. Wright, "Artificial neural network based real-time river flow prediction", in the *Proceedings of Hydroinformatics 2002*, 2002.

[9] N. Bessaih, R. Bustami, and M. Saad, "Water level estimation for Sarawak River", *Proceeding of Rivers' 04*, USM, Penang, September 2004.

[10] N.G. Wright, and M.T. Dastorani, "Effects of river basin classification on Artificial neural networks based ungauged catchment flood prediction", *Proceedings of the International Symposium on Environmental Hydraulics*, Phoenix, 2001.

[11] N.G. Wright, and M.T. Dastorani, "Effects of river basin classification on artificial neural networks based ungauged catchment flood prediction", in the *Proceedings of the International Symposium on Environmental Hydraulics*. Phoenix, 2001.

[12] N.G. Wright, M.T. Dastorani, P. Goodwin, and C.W. Slaughter, "A combination of artificial neural networks and hydrodynamic models for river flow prediction", in the *Proceedings of Hydroinformatics 2002*, 2002.

[13] O.R. Dolling, and E.A. Varas, "Artificial neural networks for streamflow prediction", *Journal of Hydraulic Research*. 40(5), 2002, pp. 547-554.

[14] R. Garcia-Bartual, "Short term river flood forecasting with neural networks", *Proceedings of iEMSs 2002*, Switzerland, 2002.

[15] R.H. Mc.Cuen, *Hydrologic Analysis and Design 2nd Ed.* Upper Saddle River, N.J: Prentice Hall, 1997.

[16] R.S. Gupta, *Hydrology and Hydraulic Systems*. New Jersey: Prentice Hall, 1989.

[17] S. Haykin, *Neural Networks. A Comprehensive Foundation.* Upper Saddle River, N.J.: Prentice Hall, 1994.

[18] W. Huang, C. Murray, N. Kraus, and J. Rosati, "Development of a regional artificial neural network prediction for coastal water level prediction". *Ocean Engineering* vol.30, 2003, pp. 2275-2295.