

River Bedup Catchment Water Level Prediction Using Pre-developed ANN Model of Siniawan Catchment

Bustami, R.¹, Hong, C.² and Lim, D.³

Abstract—This study proposes the application of Artificial Neural Network (ANN) in the prediction of hourly water level under tidal influence for Sadong Basin. An ANN is undoubtedly a robust tool for forecasting various non-linear hydrologic processes, including the water level prediction. It is a flexible mathematical structure which is capable to generalize patterns in imprecise or noisy and ambiguous input and output data sets. In this study, ANN models were developed specifically to forecast the hourly water level for River Bedup Station. Distinctive networks were trained, validated and simulated using hourly data obtained from Department of Irrigation and Drainage, Sarawak in Kuching. The performances of ANN were evaluated based on the coefficient of efficiency, E^2 and the coefficient of correlation, R . The back propagation algorithm was adopted for this study. Models used in this study is trained, validated and simulated with scaled conjugate gradient algorithm (*trainscg*) with two hours of antecedent data, learning rate and the number of neurons in the hidden layer of 0.8 and 40 respectively. In this study, the models generated an accuracy of 100% for all training, validating and simulating stages. It has been found that the ANN has the potential to solve the problems of water level prediction.

Keywords: Artificial Neural Network, Water Level Prediction, Back propagation, River Bedup.

I. INTRODUCTION

Most civilization were developed along the rivers because the river provides access to and from the sea coast, irrigation for plantation, water and food supplies. However most of the benefits have always been counterbalanced by the dangers of flood as it is a common disaster faced ever since the older century. Floods can happen either in the rural area or in the urban area. Therefore in order to control or reduce flood, an accurate forecast of water level of river is extremely important.

Artificial Neural Networks (ANN) is model free estimators. This is because they are able to learn from examples and generalize the data given. Generally, they can be described as a human brain as they are able to predict or forecast the output precisely. Therefore, this makes them well suited to situations where problem complexity precludes the development of models based on the first principles [1].

In recent years, computer models of the principal hydrological and hydraulic components of river floods have been successfully developed. In a very small catchment, a good forecast can be made on the basis of the intensity of the rainfall. However, for larger catchments as the one considered in this study, the flow characteristics will strongly influence the ability to make a good prediction.

On the other hand, conventional models require a great amount of detailed data. Examples of the detailed data are such as topographical maps, river networks and characteristics, soil characteristics, rainfall and runoff data. However these data are often not available and therefore pose a great difficulty for model calibration.

In addition, a sufficiently long lead-time for forecasting the necessary flood evacuation measures is required. For dissemination of flood information and other flood evacuation measures, computational speed of the models used are of absolute importance. The excellent features of ANN are the reasons why they have become an attractive computational tool.

¹ Bustami, R. is a lecturer with the Department of Civil Engineering, Faculty of Engineering, Universiti Malaysia Sarawak. (e-mail: abrosmina@feng.unimas.my)

² Hong, C. is with Yunco Building Systems Sdn. Bhd., Kuching, Sarawak.

³ Lim, D. is with Ryobi Geotechnique Pte Ltd, Singapore.

It was found that they are able to recognize the relation between the input and output variables without explicit physical consideration. Furthermore, they work well even when the training sets contain noise and measurement errors. They are also able to adapt to solutions overtime to compensate for changing circumstances, and they possess other inherent information-processing characteristics and once trained, they are easy to be used.

Over the past decades, ANN have been applied to various hydrological engineering applications. There are ANN for water level prediction in a river under tidal influence [2], ANN for precipitation and water level prediction [3], routing in computer networks using ANN [4], prediction of precipitation by aneural network method [5], ANN approach for predicting transient water level in a multilayered groundwater system under variable state, simulation of flood flow in a river system using ANN [6], simulation of flood flow in a river system using ANN [7], stochastic modeling of Karasu River using ANN [8], flood forecasting on River Ouse using ANN [9],

II. OBJECTIVE

The objective of this study is to predict the hourly water level for Sg Bedup which is located at Sadong Basin using back propagation network developed in [2] for Siniawan Station located at Sg. Sarawak Basin.

III. METHODOLOGY

For this study, the River Bedup Station of the Sadong Basin is selected. The is because the purpose of this study is to see the effect of recent model developed by [2] on Sinawan Station which has similar characteristics as River Bedup Station as both station are under tidal influence. The area of the whole Sadong Basin is about 3550 km², while the total length of the main rivers is 150 km.

In this study, there are a total of four rainfall stations and one station that measures both rainfall and water level data. The rainfall stations are Bunan Gega, River Busit, River Teb and Tebakang while River Bedup will be the station which provides both rainfall and water level data as it is located at the outlet of the basin. The tide level data is collected from the station at Muara Tebas. Figure 1 shows the location of the five stations.

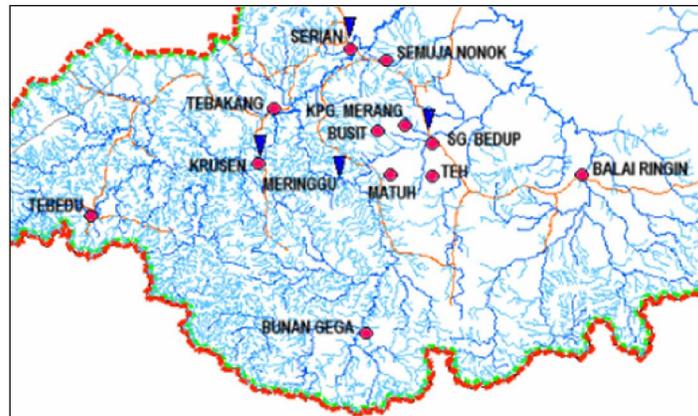


Figure 1 River Bedup Station, Sadong Basin

The selection of an appropriate input vector will allow ANN to be successfully mapped to the desired output vector. Besides that, a firm understanding of the hydrologic system under consideration is an important prerequisite for the successful application of ANN. For this study, the hourly rainfall and water level data from all five different rainfall stations and one river gauge station together with the tide level data were collected. The data collected period is from year 2005 to year 2007. The data used consists of two sets. One set is used for model calibration (training) and the other set is used for model verification (simulation). All the rainfall data were provided by the Department of Irrigation and Drainage (DID) of Kuching whereas the tide level data were provided by Marine Department Sarawak.

Measured hourly rainfall and water level for the selected gauging station from year 2005 to year 2007 were used to build the ANN models. Two different cases were investigated in this study. The first case consists of one year data for training and six month for simulation while the second case consists of two years data for training and one year data for simulation. The input data consisted of antecedent precipitation, antecedent water level, antecedent tide level and the rainfall for the current hour $Q(t-1)$, $Q(t-2)$... $Q(t-n)$, $P(t-1)$, $P(t-2)$... $P(t-n)$. The output was the water level at the current hour $Q(t)$. The matrix form for the input and output for the network are shown in Equation (1) and (2).

$$\text{Input} = \begin{bmatrix} P(t-2)_1 & P(t-2)_2 & P(t-2)_3 & \backslash & P(t-2)_n \\ P(t-1)_1 & P(t-1)_2 & P(t-1)_3 & \backslash & P(t-1)_n \\ P(t)_1 & P(t)_{21} & P(t)_3 & \backslash & P(t)_n \\ T(t-2)_1 & T(t-2)_2 & T(t-2)_3 & \backslash & T(t-2)_n \\ T(t-1)_1 & T(t-1)_2 & T(t-1)_3 & \backslash & T(t-1)_n \\ T(t)_1 & T(t)_2 & T(t)_3 & \backslash & T(t)_n \\ W(t-2)_1 & W(t-2)_2 & W(t-2)_3 & \backslash & W(t-2)_n \\ W(t-1)_1 & W(t-1)_2 & W(t-1)_3 & \backslash & W(t-1)_n \end{bmatrix} \quad \text{Equation (1)}$$

$$\text{Output} = [W(t)_1 \quad W(t)_2 \quad W(t)_3 \quad \backslash \quad W(t)_n] \quad \text{Equation (2)}$$

In this study, there were a few missing precipitation and water level data for the year 2006. Hence for the missing precipitation data, the normal ratio method was used to predict the missing precipitation data for 2006 data. As for the missing water level, the ANN model developed for one year training and six months simulation (Model A) is used to predict the missing water level which will be used in model B. The formula for normal ratio method is shown in equation (3).

$$\frac{P_x}{N_x} = \frac{1}{n} \left(\frac{P_1}{N_1} + \frac{P_2}{N_2} + \dots + \frac{P_n}{N_n} \right) \quad \text{Equation (3)}$$

Where P_x = Missing precipitation fro station X
 N_x = Normal long-term precipitation at station X
 P_1, P_2, P_n = Precipitation at neighboring station for the concurrent period
 N_1, N_2, N_n = Normal long-term precipitation for neighboring station

The Multi-layer Perceptron is adopted in this study. 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 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. Backpropagation was created by generalizing the Widrow-Hoff learning rule to multiple-layer networks and non-linear differentiable transfer functions [10]. The transfer functions selected are *tan-sigmoid* for hidden layer and *pure-linear* for the output layer.

The networks are built by utilizing one to two years of water level and tide level data and were simulated using the back propagation algorithm. The back propagation algorithm used in this study is the scaled conjugate gradient (*trainscg*). In the scaled conjugate gradient, a search is performed along conjugate directions, which produces generally faster convergence than the steepest descent directions used by the basic back propagation. The scaled conjugate gradient was designed to avoid the time consuming line search.

During the simulation, the target output is compared with network output, and the difference or error is minimized by adjusting the weights and biases. The weights and biases are adjusted by moving them along the negative gradient or error function during each iteration until the convergence is completed.

Validation is to be done in order to check and compare the simulated result with the real time data. The validation period will be 2 months for both of the model namely January 2006 and February 2006 for Model A whereas January 2007 and February 2007 for Model B. Table 1 shows the training, simulation and validation period for both models.

Table 1 Training, Simulation and Validation Period for Model A and Model B

	Model A	Model B
Training	1 year	2 years
Validation	2 months	2 months
Simulation	6 months	1 year

Hence as a conclusion, the model is simulated using the scale conjugate (*trainscg*) with learning rates of 0.8, number of neurons in hidden layer of 40 and antecedent hours of 2 hours. The model performance is measured by using the coefficient of correlation, R and The Nash-Sutcliffe coefficient, E^2 . Value of coefficient of correlation, R, is given by the Equation (4).

$$R = \frac{\sum_{i=1}^{i=n} [Q_m(t_i) - \overline{Q_m(t_i)}][Q_s(t_i) - \overline{Q_s(t_i)}]}{\sqrt{\sum_{i=1}^{i=n} [Q_m(t_i) - \overline{Q_m(t_i)}]^2 \sum_{i=1}^{i=n} [Q_s(t_i) - \overline{Q_s(t_i)}]^2}} \quad \text{(Equation 4)}$$

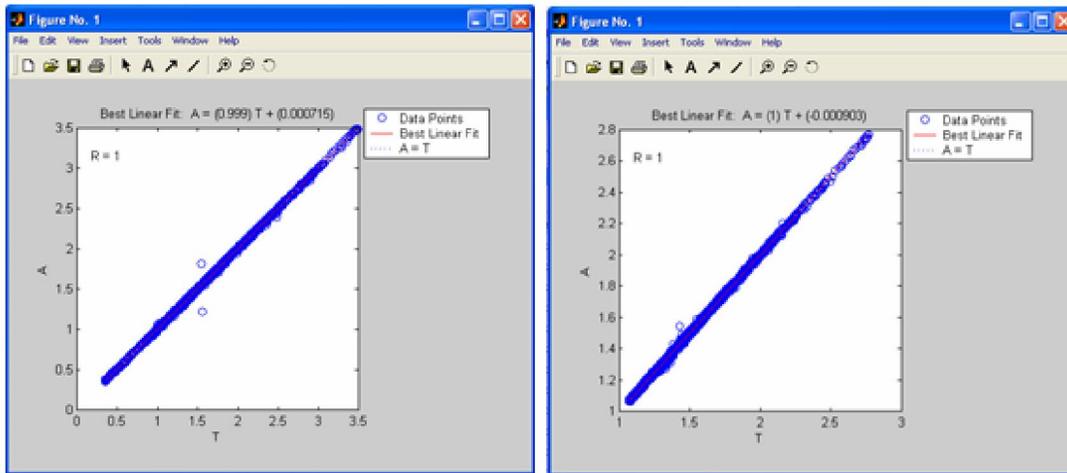
where subscript m and s represent the measured and simulated water level respectively while $\overline{Q_m}$ and $\overline{Q_s}$ are average of the measured and simulated water level for the investigated station.

Nash-Sutcliffe coefficient, E^2 is expressed as:

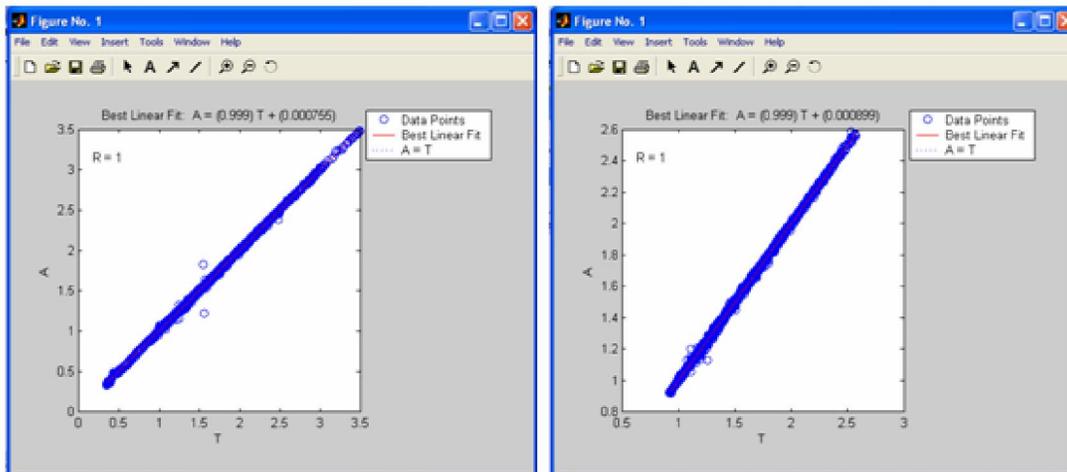
$$E^2 = 1 - \frac{\sum_{i=1}^{i=n} [Q_m(t_i) - Q_s(t_i)]^2}{\sum_{i=1}^{i=n} [Q_m(t_i) - \overline{Q_m}]^2} \quad \text{(Equation 5)}$$

in which the subscripts m and s imply the measured and the simulated water level, respectively, while $\overline{Q_m}$ is the average of the measured water level for the station concerned.

IV. RESULTS



(a) (b)
Figure 2 (a) Training and (b) Simulation Results for Model A



(a) (b)
Figure 3 (a) Training and (b) Simulation Results for Model B

From Figure 2 and 3, for all models, the accuracy of training and simulation is 100%, as depicted in their given R and E^2 values. Hence the graph is a linear passing through point 0 as shown in the figure above. Both model results can be summarized as in table 2.

Table 2 Summary of the Results

Model	Training Algorithm	Learning Rates	Neurons	Training		Simulation		Validation	
				R	E^2	R	E^2	R	E^2
A	Trainscg	0.8	40	0.9997	0.9995	0.9996	0.9993	0.9997	0.9994
B	Trainscg	0.8	40	0.9997	0.9995	0.9998	0.9994	0.9997	0.9995

When plotted, trained, validated and simulated water level for Model A and Model B mimicked previously recorded real-time water level data in River Bedup. Figure 4 and figure 5 compares real-time water level data recorded in River Bedup for testing and simulation stages for Model A and B, respectively. The simulated data is approximately the same as the real time data when compared. Besides that, the R and E^2 for the validation is also approximately 1.00. Hence both model can be use to predict the water level in the future.

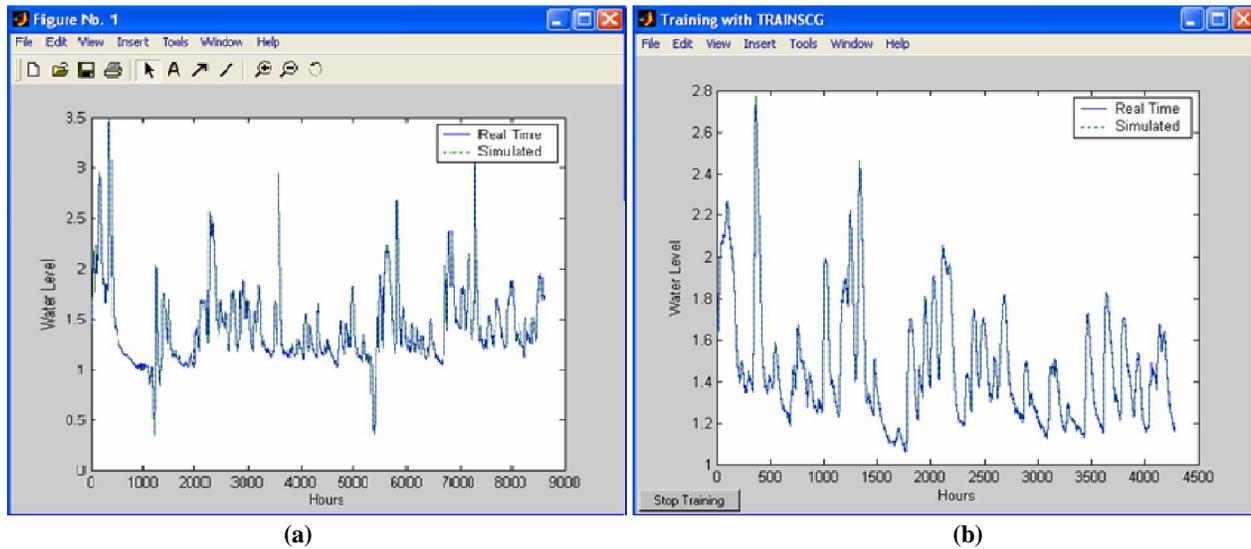


Figure 4 Comparison between real-time and simulated water level data for (a) training and (b) simulation for Model A

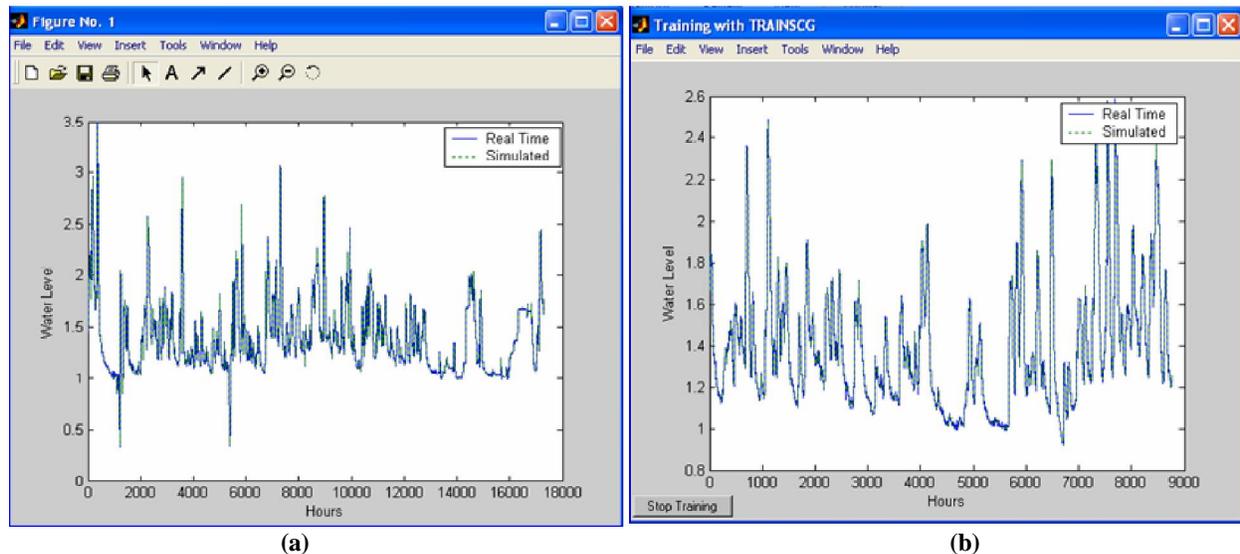


Figure 5 Comparison between real-time and simulated water level data for (a) training and (b) simulation for Model B

V. DISCUSSION

Two different ANN models were built in this study, which represent two different set of training and simulation data. Set 1 contains the models with 1 year of hourly data for training and 6 months hourly data for simulation while Set 2 contains models with 2 years of hourly data for training and 1 year of hourly data for simulation. For both set of network, the models were built with 2 antecedent hours network.

In this study, normal ratio method was chosen to predict the missing precipitation for some particular month in year 2006 instead of using Matlab software. This is due to two rainfall stations having missing precipitation at the same time. Besides that, Model A was used to predict the missing water level for Model B as the simulation and validation for Model A gives 1.00 for both R and E².

From the results, it can be concluded that both model performed very well with ANN model developed in [2] as both models give the R value of 1.00.

VI. CONCLUSION

The objective of this study which is to predict the hourly water level for River Bedup Station using the optimum model developed by Bessaih et al. (2004) in [2] using ANN had been successfully achieved. The models were developed using the rainfall data or precipitation, historical water level and the tide level data. These models are expected to give a good prediction on the current hour water level.

In this study, two sets of model were used for training and simulation. Model A comprises of one year training data with six months of simulation data whereas Model B consists of two years of training data with one year of simulation data. As for the validation period, both models will be validated for two months namely January 2006 and February 2006 for Model A whereas January 2007 and February 2007 for Model B

In this study, both of the models were training using *trainscg* with the numbers of neurons in the hidden layer of 40, learning rates of 0.8 and antecedent hours of 2 hours. Both models give almost the same R and E² value which is approximately 1.00 respectively during training, simulation and validation.

However due to different length of data, Model B gives a slightly better result compare to Model A as it provide more data for training even though there were a number of missing precipitation and water level from July till December. All the missing precipitation was overcome by using normal ratio method whereas all the missing water level data was obtained by simulating result from Model A.

As conclusion, both these model is good not only in Sinawan Station but also in River Bedup Station. Hence it is proven that both these model can be use for other basin and catchments with similar characteristics.

ACKNOWLEDGEMENT

Authors are thankful to Universiti Malaysia Sarawak for the support and to Department of Irrigation and Drainage Sarawak for the rainfall and water level data supplied to the research team.

REFERENCES

- [1] Haykin, S (1994). *Neural Network Compressive Foundation*. New Jersey: Prentice Hall.
- [2] Bessaih, N., Bustami, R.A., and Maliana, S. (2004) Water Level Estimation for Sarawak River. *Proceedings of 1st International Conference on Managing Rivers in the 21st Century*, Penang, Malaysia
- [3] Bustami, R., Bessaih, N., Bong, C., & Suhaili S. (2007) Artificial Neural Network for Precipitation and Water Level Predictions of Bedup River. *IAENG International Journal of Computer Science*. Vol. 34 (2), pp. 228-233.
- [4] Pierre. S., Said. H. and Probst, W.G. (2000) Routing In Computer Networks Using Artificial Neural Networks. *Artificial Intelligence in Engineering*. Vol. 14 (4), pp. 295-305.
- [5] N. Maeda, S. Kobayashi, K. Izumi, S. Kouno and M. Amenomori (2001) Prediction of Precipitation by Aneural network method. *Journal of Natural Disaster Science*. Vol. 23 (1), pp 23-33.
- [6] Coppola E., Szidarovszky F., Poulton M. and Charles E. (2003) Artificial Neural Network Approach for Predicting Transient Water Levels in a Multilayered Groundwater System under Variable State, Pumping and Climate Conditions. *Journal of Hydrological Engineering*. Vol. 8 (6), pp 348-360.
- [7] Shrestha R.R, Theobald S. and Nestmann F. (2005) Simulation of Flood Flow in a River System Using Artificial Neural Network. *Hydrology and Earth System Sciences*. Vol. 9 (4), pp 313-321.
- [8] Can. I., Yerdelen. C. and Kahya. E. (2007) Stochastic modeling of Karasu River (Turkey) using the methods of Artificial Neural Networks. *Proceedings of Hydrology Days 2007*, Colorado State University, CO, USA.
- [9] See, L., Corne, S., Dougherty, M. and Openshaw, S. (1997) *Some initial Experiments with Neural Network Models of Flood Forecasting on the River Ouse*. 2nd Annual Conference of Geo Computation 97 & SIRC 97.
- [10] Demuth H. and Beale M. (2001). *Neural Network toolbox user's guide version 4*. Toronto: Prentice-Hall.