

## MODELING AND SIMULATION OF RIVER DISCHARGE USING ARTIFICIAL NEURAL NETWORKS

<sup>1</sup>Aiyelokun Oluwatobi, <sup>2</sup>Ogunsanwo Gbenga, <sup>2</sup>Adelere Joy and <sup>1</sup>Agbede Oluwole

<sup>1</sup>Department of Civil Engineering, University of Ibadan, Ibadan, Nigeria.

<sup>2</sup>Department of Computer Science and Information, Tai Solarin University of Education.

<sup>1</sup>E-mail for correspondence: aiyelokuntobi@gmail.com

(Received: 29<sup>th</sup> March, 2017; Accepted: 13<sup>th</sup> June, 2018)

### ABSTRACT

The study developed multiple artificial neural network models with the aim of establishing the most suitable non-linear discharge prediction model of Ibu River. A 12-year daily discharge of River Ibu gauged near Sagamu was obtained from the Ogun-Oshun River Basin Development Authority (OORBDA), Abeokuta Nigeria to model and simulate daily discharge. The back-propagation method was used in developing the artificial neural network model. The study revealed that only three artificial neural network (ANN) models out of fifteen developed, have overall results that are satisfactory for prediction, out of these, the model with the least error was used for validation. The results obtained with ANNs based on two hidden layers for 1-day ahead are better than those obtained by models with single layers. It was concluded that the general performance of ANN models depends solely on the data used. While it was recommended that additional basin characteristics such as slope, geology, morphology and surface roughness features should be included to obtain more robust river discharge models.

**Keywords:** Artificial Neural Network, Hydrologic record, Prediction

### INTRODUCTION

The predictions of discharge are expedient for water resources planning and management and flood management in a drainage basin. Majority of the river basins in developing countries, Nigeria inclusive have scarce records of hydrologic data because many of them are either partially gauged or not gauged at all. As a result, hydrologist and water resources engineers depend on modeling and simulation of historical empirical data available to make future predictions. Such predictions are useful in warning of the flood or drought conditions and in optimizing operation of hydrological systems at river basins. A large number of hydrological analyses require mapping and modelling of non-linear systems data.

Traditionally, such mapping is performed with the help of conceptual models or statistical tools such as regression and curve fitting. However, when the underlying physical laws are unknown or not precisely known, it is rather difficult to model the phenomenon adequately (Bisht *et al.*, 2010). Efforts have been made to establish a technique that does not require algorithms or rules development and thus reduces the complexity of the software; one of such technique is known as neurocomputing, and the networks laid out with

many parallel processing elements to do this neurocomputing are called artificial neural networks (Bisht *et al.*, 2010).

Stream discharge prediction model can play a key role in providing relevant information of possible impending floods in populated locations. The development of such models can reduce the damage in areas such as the Sagamu municipality by decreasing the economic and environmental impacts of floods. More importantly, discharge prediction models effectively lower the risk of harm and loss of life.

Artificial neural network (ANN) models can provide sufficiently accurate forecasts for one day ahead, while the lead time for flood warning can be extended, with the aim of making sure that subsequent flood emergency measures can be better planned and executed (Ramapulana, 2011). It is on this premise that this study seeks to reduce the problem of flooding in Sagamu by developing a rainfall-discharge flow prediction ANN model.

### DESCRIPTION OF STUDY AREA

River Ibu is a major tributary of the Ona river. It is located between 3.38333° N and 3.68333° N, and longitude 6.8000° E and 6.83333° E (Figure 1).

The drainage basin of Ibu occupies about 928 km<sup>2</sup>, which is about 14% of Ona River basin. Ibu takes its source at an altitude of about 180 meters above sea level; about 7.3 km west of Sadu in Odeda Local Government area of Ogun State, Nigeria, and flows for about 65.3 km before joining the Ona River.

River Ibu basin is influenced by the movement of the inter-tropical convergence zone (the ITCZ), a

quasi-stationary boundary zone which separates the sub-tropical continental air mass over the Sahara and the equatorial maritime air mass over the Atlantic Ocean. Geologically, sedimentary rocks of cretaceous and latter deposits are found in some parts of Ibu River basin which is mostly in the southern part. The remaining northern part consists of crystalline basement complex formation.



Figure 1: Map of Study Area

## MATERIALS AND METHOD

### Gathering of Hydrological Data

A 12-year daily (1984-1995) hydrologic record of Ibu gauged near Sagamu was obtained from the Ogun-Oshun River Basin Development Authority (OORBDA), Abeokuta Nigeria to generate an annual hydrograph for the study area. While 12-year daily rainfall (1984-1995) data for Abeokuta was collected from Nigeria Meteorological agency (NIMET) Abeokuta, Ogun state station. This rainfall station was chosen for the study since it was the only rainfall station with daily data close to the study area. Since only 12 years daily data was available for the partially gauged drainage basin of Ibu River, daily rainfall of this period was used for simulating discharge.

### Development of ANN Model

The neural network model based on multilayer

neural networks was used for the study because a single layer neural network can only be used for the classification of a special type of patterns which are linearly separable. The feed forward architecture (supervised network architecture) was employed and the model was trained with the back-propagation algorithm. The Neuralnet package installed in R programming language was used for designing, implementing, visualizing, validating and simulating neural networks. For easy coding of R, Rstudio Graphic User Interface (GUI) was used. The output discharge  $Q_t$  at time step  $t$  was mapped with rainfall  $R_t$  with specified lag time i.e.  $R_{t-1}$ ,  $R_{t-2}$ ,  $R_{t-3}$  etc. and the previous discharge that is  $Q_{t-1}$ ,  $Q_{t-2}$ ,  $Q_{t-3}$  etc as presented in table 1. Table 1 describes the developed ANN models with input variables, output variable and number of hidden layers.

Table 1: Model Development of Various ANN Rainfall-Stream flow Models

MODEL	No of Hidden Layer	Out put	In put variable
ANN – 1	One	Qt	Rt
ANN – 2	One	Qt	Rt, Rt-1
ANN – 3	One	Qt	Rt, Qt-1
ANN – 4	One	Qt	Rt, Rt-1, Qt-1
ANN – 5	One	Qt	Rt, Rt-1, Rt-2
ANN – 6	Two	Qt	Rt
ANN – 7	Two	Qt	Rt, Rt-1
ANN – 8	Two	Qt	Rt, Qt-1
ANN – 9	Two	Qt	Rt, Rt-1, Qt-1
ANN – 10	Two	Qt	Rt, Rt-1, Rt-2

The output layer used for the prediction has only one neuron which is the stream flow values. The stream flow values computed from the ANN model were used for the validation of the model. Numerous literatures have proven that a single hidden layer network is sufficient to uniformly approximate any continuous function. However, maximum of two hidden layers and various neurons were used. The number of hidden layers that yields the best result was adopted for the study.

**Neural Network Model Training**

The initial weight of the model was chosen at random by the Neuralnet software package of R programming. The learning process was maintained on an epoch-by-epoch basis until the synaptic weights stabilize and the average squared error over the entire training set converges to some minimum value. The training samples were randomized from one epoch to the next because it tends to make the search in weight space stochastic over the learning cycles, thus avoiding the possibility of limit cycles in the evolution of the synaptic weights vectors. For this study, the data was divided into two groups with a ratio of 75% to 25%. The 75% of the data was used for training of ANN, while the remaining 25% was used for the testing (validation).

**Performance Evaluation**

The model with the lowest Mean Square Error (MSE), Root Mean Square Error (RMSE) and highest positive R<sup>2</sup> – correlation coefficient was

chosen as the best. These computations were performed on the R Studio software for R programming. The RMSE which is the square root of MSE is represented by equation 1, while the correlation coefficient (R<sup>2</sup>) is represented by equation 2.

$$RMSE = \sqrt{\sum_{i=1}^n (y_i - x_i)^2 / n} \quad \text{eq. 1}$$

$$R^2 = \frac{\sum_{i=1}^n (x - \alpha)(y - \varepsilon)}{\sqrt{\sum_{i=1}^n (x - \alpha)^2 \sum_{i=1}^n (y - \varepsilon)^2}} \quad \text{eq. 2}$$

- Where, Y<sub>i</sub> = observed data
- X<sub>i</sub> = predicted data
- α = mean of predicted
- ε = mean of observed

**Validation and Visualization**

After series of training and testing of the neural network models, the model with best performance was validated using line chart and scatter plot.

**RESULT AND DISCUSSION**

The results of neural network training simulations that were developed are presented in this section. The goals of these experiments were to determine the data set composition and neural network architecture that would produce the best

predictive performance for stream flow.

### Discharge Prediction Models

The models were developed with rainfall of present day ( $R_t$ ), previous day's rainfall ( $R_{t-1}$ ) and previous day's stream flow as input ( $Q_{t-1}$ ) and present day's stream flow ( $Q$ ) as output for a major hydrological location on Ibu River at Sagamu. The Multilayered Perceptrons (MLPs) architecture, which is a feed forward-type ANN, was selected in

this study because it usually has good model performance, and also because it is the most frequently used configuration in hydrology. The ANN networks consisting of an input layer, one and two hidden layers composed of 1-20 nodes, and one output layer consisting of 1 node denoting the predicted discharge. Table 2 illustrates the ANN-MLP model structure in the present case for simulating daily stream flow data having different input variables.

Table 2: Summary of ANN Model Results

Model	Topology	MSE	RMSE	Correlation
Ma1	1-1-1	36.5	6.04	0.3
Ma2	1-20-1	38.08	6.17	0.2
Ma3	1-20-10-1	36.68	6.05	0.3
Mb1	2-1-1	35.74	5.97	0.4
Mb2	2-20-1	34.43	5.87	0.4
Mb3	2-20-10-1	34.07	5.89	0.4
Mc1	2-1-1	3.26	1.80	0.96
Mc2	2-20-1	3.27	1.81	0.96
Mc3	2-20-10-1	3.20	1.79	0.96
Md1	3-1-1	79.95	8.94	0.26
Md2	3-20-1	87.54	9.36	0.22
Md3	3-20-10-1	87.02	9.32	0.21
Me1	3-1-1	33.55	5.79	0.4
Me2	3-20-1	32.96	5.74	0.3
Me3	3-20-10	32.45	5.70	0.3

As depicted in table 2, 15 ANN models were developed consisting of 5 categories of inputs (a,b,c,d,e). The first category of models had a single input depicting the rainfall of the present day ( $R_t$ ), the second category had two inputs depicting the present and previous day's rainfall ( $R_t, R_{t-1}$ ), the third category had two inputs depicting the present day's rainfall and previous day stream flow ( $R_t, Q_{t-1}$ ), the fourth category had three inputs depicting the present day's rainfall and previous day's rainfall and stream flow ( $R_t, R_{t-1}, Q_{t-1}$ ), while the fifth category of model had three inputs depicting the present day's and past two days' rainfall ( $R_t, R_{t-1}, R_{t-2}$ ). From general comparison it could be observed that the third category of the ANN models performed better having high correlation coefficient of 0.96. This implies that in order for the neural network to perform better in predicting stream flow, it is

required to provide stream condition of the previous day. These findings indicated that data set composition and neural network architecture had a critical influence on stream flow predictive performance.

### Performance Evaluation of ANN Models

Three performance measures were selected to evaluate the performance of ANN models, viz the root mean square error (RMSE), mean square error (MSE), and correlation coefficient (COR). As shown in figure 2, models Mc1, Mc2 and Mc3 had the lowest mean square error of 3.26, 3.27 and 3.20 respectively, while models Md1, Md2 and Md3 had the highest mean square errors. It can also be observed that the root mean square error follows the same pattern as the mean square error. This implies that ANN of "c" category has the highest performance in predicting stream flow.



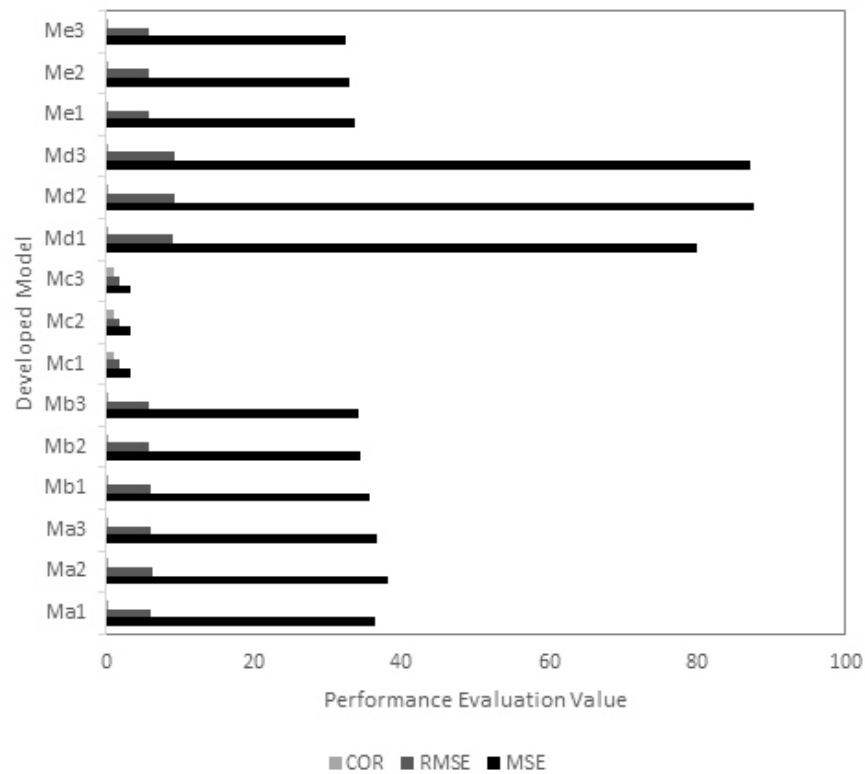


Figure 2: Plot showing Performance Evaluation Value against Developed Models

Only three ANN models (Mc1, Mc2 and Mc3) produced overall results that were satisfactory for forecasting as shown in table 2. The best models performed far efficiently than others because of the employment of previous day's discharge used in predicting the present stream discharge along with present day rainfall. Many researchers such as Ozgür (2005) and Solaimani (2009) have used past stream flow in predicting future stream flow in their studies. The models were also evaluated using the correlation coefficient ( $r$ ). The correlation coefficient is used in statistics to provide information in terms of the strength of a linear relationship between the simulated or forecasts and the observed values. In this study the correlation coefficient of observed stream flow data and simulated results were calculated for ANN model as shown in table 2 and figure 2.

When the correlation coefficient value is between 1.0 and 0.7, it indicates a strong positive correlation; 0.3 and 0.1 indicates weak correlation, and less than 0.1 indicates very little correlation. The correlation coefficient obtained was positive, ranging from 0.2 to 0.96 which implies that models Mc1, Mc2 and Mc3 are acceptable for discharge prediction. In addition to this assertion, Mc3 with the lowest MSE and RMSE outperformed Mc2 and Mc3. Figure 3 shows the plot of the best model with two hidden layers of node 20 and 10. Simulation results for testing set show that the neural network with two hidden layers could outperform the network with one hidden layer. Therefore, the multi-layer neural network with two hidden layers (Mc3) was selected to simulate and predict the stream flow time-series.



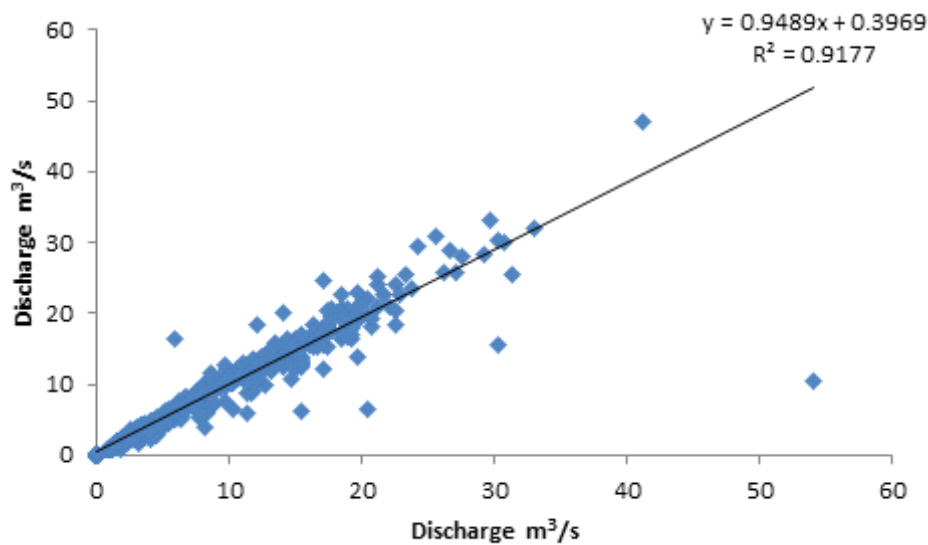


Figure 5: Observed and Computed (ANN model) Flows Scatter Plot

### DISCUSSION OF FINDINGS

Results show that the neural network provides a high accuracy prediction of 1-day ahead discharge of Ibu River. One of the advantages of the presented model compared to the ordinary numerical models is that it is not dependent on the initial and boundary conditions. However, the reliability of results is dependent on how much data is available for training the network.

Multi-layer neural networks can have more than 1 hidden layer, although theoretical studies have shown that a single hidden layer is sufficient for ANNs to approximate any complex nonlinear function (Cybenko, 1989; Hornik *et al.*, 1989). Indeed, many experimental results seem to confirm that one hidden layer may be enough for most forecasting problems (Zhang *et al.*, 1998; Coulibaly *et al.*, 1999). In contrast, this study has shown that models with two hidden layers are as well adequate for prediction.

In recent years, the applications of (ANN) techniques in hydrological modeling have received increasing attention. The ANN has the capability to identify complex nonlinear relationships between input and output data sets without the necessity of understanding the nature of the phenomena and without making any underlying assumptions regarding linearity or normality (Abudu *et al.*, 2011). ANN also has the capability to handle noisy data. Multilayered perceptrons (MLPs) are the simplest and most commonly used

neural network architectures. MLPs can be trained using many different learning algorithms. This study corroborates the findings of Tokar and Markus (2000) that used MLP for Fraser River in Colorado, US for flow forecast of one month ahead. The study of long-term discharge prediction using ANN and other data-driven models were also presented by Cheng *et al.* (2008), Lin *et al.* (2006), Wang *et al.* (2009), and Wu *et al.* (2009). Data-driven models, such as artificial neural networks (ANNs), have often been used instead of physically based models because of their simplicity and promising capability to simulate input-output relations (Wu *et al.* 2005).

### CONCLUSION AND RECOMMENDATIONS

The potential of ANN models for simulating the hydrologic behavior of stream flow has been presented in this study. The results obtained with ANNs based on two hidden layers for one-day ahead are better than those obtained by models' single layers. The study therefore concludes that the general performance of ANN models depends solely on the data used in modeling. Based on the findings of the study, it is recommended that similar work should be extended to other river basins in Nigeria where there is limited discharge; future similar studies should incorporate additional basin characteristics such as slope, geology, morphology and surface roughness features should be included to obtain more robust stream flow prediction; the

application of data mining techniques should be encouraged in Nigeria in solving numerous environmental problems that are related to stream flows such as flooding. Finally, majority of the river gauging stations in Nigeria are not operational leading to scarce data, hence majority of the gauge stations should be refurbished for consistent data collection.

## REFERENCES

- Abudu, S. Cui, C. King, J. P. Abudukadeer, K. (2010). Comparison of performance of statistical models in forecasting monthly streamflow of Kizil River, China. *Water Science and Engineering*. 3(3): 269-281.
- Bisht, D.C. Raju, M. M. Joshi M. C. (2010). ANN Based River Stage – Discharge Modelling For Godavari River, India. *Computer Modelling and New Technologies*. 14(3): 48–62.
- Cheng, C.T. Wang, W.C. Xu, D.M. and Chau, K.W. (2008). Optimizing hydropower reservoir operation using hybrid genetic algorithm and chaos. *Water Resources Management*. 7: 895-909.
- Coulibaly, P. Ancil, F. and Bobe'e, B. (1999). Pre'vision hydrologique par re'seaux de neurones arti\_ciels: e'tat de l'art" *Canadian Journal of Civil Engineering*. 26: 293-304.
- Cybenko, G. (1989). Approximation by superposition of a sigmoidal function, *Mathematics of Control, Signals, and Systems*. 2: 303-314.
- Hornik, K., Stinchcombe, M. and White, H., (1989). Multilayer Feedforward Networks Are Universal Approximators. *Neural Networks*. 2: 359-366.
- Lin, J.Y. Cheng, C.T. and Chau, K.W. (2006). Using Support Vector Machine for Long Term Discharge Prediction. *Hydrological Sciences Journal*. 43: 47-66.
- Ozgür, K. (2005). Daily River Flow Forecasting Using Artificial Neural Networks and Auto-Regressive Models. *Turkish Journal of Engineering and Environmental Sciences*. 29: 9-20.
- Ramapulana, N. (2011). Artificial Neural Network Modelling of Flood Prediction and Early Warning. Master thesis presented in University of Free State.
- Solaimani, K. (2009). A Study of Rainfall Forecasting Models Based on Artificial Neural Network. *Asian Journal of Applied Sciences*. 2(6): 486-498.
- Tokar, A. S. and Markus, M. (2000). Precipitation runoff modeling using artificial neural networks and conceptual models. *Journal of Hydrologic Engineering*. 5(2): 156-161.
- Wang, Y.M. Kerh, T. and Traore, S. (2009) Neural network approach for modeling river suspended sediment concentration due to tropical storms. *Journal of Global Nest*. 11: 457-466.
- Wu, C.L. Chau, K.W. and Li, Y.S. (2009). Methods to improve neural network performance in daily flows prediction. *Journal of Hydrology*, 372(1-4): 80-93.
- Wu, J. S. Han, J. Annambhotla, S. and Bryant, S. (2005) Artificial neural networks for forecasting watershed runoff and stream flows. *Journal of Hydrologic Engineering*, 10(3): 216–222.
- Zhang, G., Patuwo, B.E. and Hu, M.Y. (1998). Forecasting with Artificial Neural Networks: The State of the Art. *International Journal of Forecasting*. 14: 35-62.