Rainfall-river discharge modelling for flood forecasting using Artificial Neural Network (ANN)

Arinze A. OBASI1), Kingsley N. OGBU2), Louis C. ORAKWE3), Isiguzo E. AHANEKU4)

1) Nnamdi Azikiwe University, Department of Agricultural and Bioresources Engineering, Ifite Road, 420110, Awka, Anambra State, Nigeria; e-mail: a.a.obasi@btinternet.com
2) Centre for Development Research, University of Bonn, Germany; e-mail: kn.ogbu@unizik.edu.ng
3) Nnamdi Azikiwe University, Awka, Anambra State, Nigeria; e-mail: lc.orakwe@unizik.edu.ng
4) Michael Okpara University of Agriculture, Umudike, Abia State, Nigeria; e-mail: ahaneku.isiguzo@mouau.edu.ng

Abstract

This study is aimed at evaluating the applicability of Artificial Neural Network (ANN) model technique for river discharge forecasting. Feed-forward multilayer perceptron neural network trained with back-propagation algorithm was employed for model development. Hydro-meteorological data for the Imo River watershed, that was collected from the Anambra-Imo River Basin Development Authority, Owerri – Imo State, South-East, Nigeria, was used to train, validate and test the model. Coefficients of determination results are 0.91, 0.91 and 0.93 for training, validation and testing periods respectively. River discharge forecasts were fitted against actual discharge data for one to five lead days. Model results gave $R^2$ values of 0.95, 0.95, 0.92, 0.96 and 0.94 for first, second, third, fourth, and fifth lead days of forecasts, respectively. It was generally observed that the $R^2$ values decreased with increase in lead days for the model. Generally, this technique proved to be effective in river discharge modelling for flood forecasting for shorter lead-day times, especially in areas with limited data sets.

Key words: artificial neural network (ANN), rainfall, flood forecasting, river discharge

INTRODUCTION

Population growth, socio-economic development and climate change have impacted heavily on the hydrologic cycle resulting to increased flood occurrences in some regions [Derdour et al. 2018]. According to these authors, reports from the United Nations Office for Disaster Risk Reduction (UNISDR) and Centre for Research on the Epidemiology of Disaster (CRED) show that between 1995–2015, 3062 flood disasters occurred globally out of which 47% was weather related. Today, the problems of rainfall and river discharge predictions at different temporal and spatial scales are of major concern to hydrologists. These hydrologic processes are non-linear, complex and vary both in time and space [Baby, Varjuk 2016]. Watershed management programs have therefore become pertinent through rainfall-river discharge studies in other to solve extreme hydrologic issues. In the past, over reliance on conventional methods such as deterministic and empirical models, to proffer solution to water resources problems have not been successful due to hydrological systems complexity, computational issues and dependence on large amount of data for model parameterization [Sarkar, Kumar 2012].

In southern Nigeria, the frequency of flooding has increased due to population growth and climate change and poses severe threats to human lives, environmental and water resources. Efforts by the Nigerian Federal Ministry
of Water Resources through the Nigerian Hydrological Services Agency (NHISA) to forecast flow processes of major rivers in Nigeria for early flood risk warning have yielded low success rate because of reliance on deterministic modelling approach. Paucity of hydro-meteorological gauge networks in Nigeria has majorly impacted on deterministic model outputs and affected scientific model-based policies [OBASI et al. 2017].

In recent times, rainfall-river discharge modelling lends itself well to ANN applications [HUSSAIN et al. 2017]. BABY and VARIJA [2016] reported that ANN has the capability of reproducing relationships between rainfall and river-discharge. Studies on this approach gave satisfactory results and have shown that ANN models can be effectively used for hydrologic predictions especially in data-scarce regions [BABY, VARIJA 2016; MACHADO et al. 2011; RAGHUWANSHI et al. 2006; SARKAR, KUMAR 2012; TOKAR, MARKUS 2000].

The aim of this study is to evaluate the use of ANN model for river discharge forecasting in Nigeria. Specifically, this study sought: to develop a rainfall-river discharge model for the Imo River watershed using ANN technique; forecast one to five lead days of streamflow using the ANN model; and evaluate the model performance and its potential for river-discharge modelling.

The ANN structure contains a number of layers, namely the input, hidden and output layers as shown in Figure 1, respectively, which are interconnected by a network of neurons, called synaptic weights \((w)\) and addition of bias \((b)\) to boost the network optimization [Kisi et al. 2012].

Research has shown that the arrangement of neurons in an artificial neural network follows an indefinite arrangement but highly interconnected and structured into three layers [BABY, VARIJA 2016]. However, the artificial neural network architecture as shown in Figure 1 simply defines a certain way artificial neurons are structured to perform a given task. Some classes of artificial neural network architecture are the feed-forward neural network architecture (FFNNA), radial basis-function neural network (RBFNN), generalized regression neural network (GRNN) and the recurrent neural network architecture (RNN). The feed-forward neural network architecture (FFNNA) was employed in this study, using rainfall and antecedent discharge values that had earlier occurred at times \(t_{n-1}\) to forecast the river discharge at time \(t_n\) [RAJURKAR et al. 2004], and is found to perform best for one time-step forecasting, when applied to data for which the measuring time interval is less than or equal to 24 hours [VAROONCHOTIKUL 2003]. The choice for feedforward neural network architecture was justified by the study of VAROONCHOTIKUL [2003] for scarcity of data and measurement of daily river discharge, respectively, which are peculiar to the study area. The \(X_{t,4}\) are the network input vectors feeding into the neurons in the input layer, the output from the input layer becomes input to the hidden layer which are interconnected with synaptic weight \(w_{1,4}\); the process terminates at the output layer.

**MATERIALS AND STUDY METHODS**

**STUDY AREA AND HYDRO-METEOROLOGICAL DATA**

The study area is the Imo River located in Imo River watershed as shown in Figure 2 covering an area of 1450 km². Figure 2 is the location map showing Imo River watershed in the map of Nigeria/Africa. The Imo River is one of the major rivers in South-East Nigeria and its basin is very fertile for agricultural activities. The Imo River watershed lies between latitudes 05°33’N and 06°07’N and longitudes 007°08’E and 007°61’E. The gauging station located at Umuopara village is maintained by the Anambra–Imo River Basin Development Authority (AIRBDA). The basin experiences rainy season from April to October and dry season from November to March with rainfall peaks in April and September. The topography of the study area is generally plain in the South but consists of gentle to high undulating ridges in the North and ranges from about 350–790 m above sea level. The basin experiences an annual rainfall of between 2000 and 2400 mm and mean annual temperature of above 20°C [AMANGABARA 2015].

Daily rainfall and river discharge data were obtained from Anambra–Imo River Basin Development Authority for the period 2014–2017 and used for this study. This study primarily focused on the development and use of an artificial neural network (ANN) model that was trained, validated and tested based on hydro-meteorological data obtained for the study.

![Fig. 1. Structure of an Artificial Neural Network model](image-url)
DEVELOPMENT OF ARTIFICIAL NEURAL NETWORK (ANN) MODEL

In this study, neural network software, Alyuda Neuro-Intelligence (ver. 2.2) [ALYUDA 2018] was used for the development of the ANN model. NeuroIntelligence is a neural network software application designed to assist modelling experts in solving real-world problems such as analysing and pre-processing datasets, find the best neural network architecture, train, test and optimize neural networks, and apply the designed neural network to new data. The hydro-meteorological data collected were divided into three sets: the training dataset, the validation set and testing set. The training set consisted of 80% of the total data serially selected, the validation set consisted of 13% of the total data while the testing dataset consisted of 7% of the total historical data. Consequently, a one to five lead day river discharge forecasts were also modelled and model performance evaluated to determine its suitability for rainfall-river discharge modelling.

A three-layer, feed-forward multilayer perceptron of artificial neural network model was adopted for this study. This is in line with HSU et al. [1995] who opined that only a three-layer feed forward multilayer perceptron ANN is the most suitable in a real-world situation to model relationships that may be unknown or having poorly defined complexity/form. In feed-forward multi-layer perceptron (FFMLP), there exists relationship between input and output layers, the hidden layer. Information is transmitted through the connections called synaptic weight between the neurons in a layer-by-layer basis. All the synaptic weights in the neural network were randomized between ±0.5 as the learning rate and the network momentum were fixed between 0.4 and 0.5. Prior to the execution of flow of information from the input layer through the output layer, a function called “Normalization” was also created in the neuron class to normalize the inputs that were received in the input layer to values between 0 and 1. Normalization of the data set was highly essential to enable the network outputs to remain within the range of the network output function and also for all data to receive equal treatment during training as well as to enhance the efficiency of the network training algorithm. The training data were normalized according to ABRAHART et al. [2016], using Equation (1).

\[ N_k = \frac{R_k - \text{min}_k}{\text{max}_k - \text{min}_k} \]  (1)

Where: \( R_k \) is the real value applied to neuron \( k \); \( N_k \) is the normalization value calculated for neuron \( k \); \( \text{max}_k \) is the maximum value; \( \text{min}_k \) is the minimum value.

The processes on how transfer functions perform its task and how each neuron computes a linear combination of the inputs vector from the connections feeding into them, follow the pattern set forth by AWU et al. [2017]. The two-input values (rainfall and antecedent discharge) were coupled following the works of RAJURKAR et al. [2004], and in-turn, multiplied by the synaptic weight \( w \) from the accompanying neurons \( x \). The net combined input vectors were transformed using sigmoidal transfer function, \( f(x) \), as implemented by DAWSON and WILBY [2001], giving in Equation (2):

\[ f(x) = \tan h(x) = \frac{2}{1 + e^{-2x}} - 1 \]  (2)

Where \( x \) is the weighted sum of the inputs to the unit.
Rainfall-river discharge modelling for flood forecasting using Artificial Neural Network (ANN)

The input values from preceding neurons \(x\) are multiplied by the synaptic weights \(w\) that accompany their connections. The results are summed \((net = \sum(x \cdot w))\) and an additional value bias \((b)\) is commonly added to this value to give neuron output value \((Y = f(net))\). Thus, for \(n\) input units, the unit sum \(\nu\) or the activation function is calculated, according to ABRAHART et al. [2005], using Equation (3):

\[
\nu_{x} = \sum_{j=1}^{n} w_{ij} x_{j} + b
\]  

(3)

The output obtained serves as an input to next neurons in the next layer. The output signal then becomes the response of the neural network to the given input stimulus.

The supervised algorithm employed in this study for model training using the conventional gradient descent optimization techniques was also allowed to iterate until the error falls between a given threshold or converged to an error point of 0.001. At this point the backpropagation training algorithm was considered to have learned the full function of the input-output mapping and then, training terminated.

Figures 3 and 4 showed the descriptive diagram of the feedforward neural network and feedbackward propagation respectively. It also showed the step by step computational procedures of the feedforward and feedbackward propagation of the neural network during training. The network input vectors remained unchanged during the network training otherwise the neural network training would be retrained. This is to ensure that the training process is not interrupted for good model network performance with best result.

Where, \(y_{1,n}\) are the input neurons, \(v_{j}\) is the linear combination of the input neurons, \(\phi(v_{j})\) is the transformed input neuron, \(y_{j}\) is the output of the input neurons which serves as the input to the hidden layer, \(v_{k}\) is the linear summation of the hidden neurons, \(\phi(v_{k})\) is the transformed hidden neurons, \(y_{k}\) is the output of the hidden neurons which also serves as the input to the output neurons, \(d_{k}\) is the desired output, \(e_{k}\) is the total error factor, \(\phi(v_{k,mdl})\) is the prime transformed output neurons back to the network, \(\delta_{l,mdl}\) is the local gradient decent of the hidden layer, \(w_{j,k}\) are the synaptic weights connecting the output layer to the hidden layer, and \(\delta_{l}\) is the local gradient decent of the hidden neuron. Consequently, as the neural network model was trained, validated and tested satisfactorily, one to five lead day river discharge forecasts were predicted based on the neural network model developed. Coefficient of determination \((R^2)\) as given by Equation 4 [LEGATES, MCCABE 1999] was used as the statistical measurement to evaluate the developed neural network model performance:

\[
R^2 = 1 - \frac{\sum_{i=1}^{n}(Q_{i}-\bar{Q})^2}{\sum_{i=1}^{n}(Q_{i}-\bar{Q})^2 + \sum_{i=1}^{n}(Q_{i}-\bar{Q})^2}
\]  

(4)

Where: \(Q_{i}\) are the \(n\) modelled flows; \(\bar{Q}_{i}\) are the \(n\) observed flows; \(\bar{Q}\) is the mean of the observed flows and \(\bar{Q}\) is the mean of the modelled flows.

RESULTS AND DISCUSSION

Artificial neural network (ANN) model for rainfall-river discharge of the Imo River was developed, trained, validated and tested for the study area. One to five lead day river discharge forecasts were performed and the performance of the developed model was evaluated for its fitness to simulate river discharge. Figure 5 showed the relationship of the hydro-meteorological data (rainfall and river discharge) used in this study. Each of the rainfall and river discharge stands as a clear signature of hydrological process and catchment activities in the study area. The \(R^2\) of the model performance results are shown in Table 1 – for model training, validation and testing, and Table 2 – for one-to-five lead day forecasts.

From Tables 1 and 2, results show that the ANN model performed satisfactorily with coefficient of determination \((R^2)\) values ranging from 91% to 96%. The model gave \(R^2\) values of 0.91, 0.91 and 0.93 for training, validation and testing respectively as shown in Table 1. From Table 2, it was observed that the coefficient of determination decreased with increase in lead days up to the third day, after which the model gave a high \(R^2\) value of 0.96 on fourth day but subsequently decreased on fifth day with \(R^2\) value of 0.94. Study result generally conforms to studies of AWU et al. [2017], RAJURKAR et al. [2004], HUSSAIN et al. [2017], JOSHI and PATEL [2011], MOTAMEDNIA et al. [2015] SARKAR and KUMAR [2012] and JAFAR et al. (2010). The \(R^2\) values have been the statistical method of evaluation used in this study showed variance only in two decimal places, both for training, validation and testing set, respectively. Although, the coefficient of determination \((R^2)\) values for the model’s validation were very close to +1 which indicates good positive correlation.
Table 1. Model evaluation statistics for the artificial neural network model

<table>
<thead>
<tr>
<th>Description</th>
<th>$R^2$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>91</td>
</tr>
<tr>
<td>Validation</td>
<td>91</td>
</tr>
<tr>
<td>Testing</td>
<td>93</td>
</tr>
</tbody>
</table>

Source: own study.

Table 2. The coefficient of determination ($R^2$) values of one to five lead day forecasts for the artificial neural network model

<table>
<thead>
<tr>
<th>Lead day</th>
<th>$R^2$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>95.3</td>
</tr>
<tr>
<td>2nd</td>
<td>95.1</td>
</tr>
<tr>
<td>3rd</td>
<td>92.0</td>
</tr>
<tr>
<td>4th</td>
<td>96.0</td>
</tr>
<tr>
<td>5th</td>
<td>94.0</td>
</tr>
</tbody>
</table>

Source: own study.

Figures 6 and 7 show hydrographs of daily observed river discharge against simulated river discharge during ANN model training, validation and testing respectively for the Imo River.

Generally, both observed and simulated flows showed the same trend during model training, validating and testing. However, the trained ANN model from Figure 6 simulated the discharge very well but with little or insignificant differences. It was observed that the trained ANN model seems to be excited at the beginning not to have picked the first day discharge and overestimated the river discharge when it started giving output. Also, the descriptive graph of the trained ANN model revealed that the model could not estimate perfectly whenever there is a rise of the river discharge, and this agrees with the findings of AWU et al. [2017]. Likewise, the validation ANN model showed that the model under-predicted the actual flow at 75 m$^3$·s$^{-1}$ and below, but however, continually simulated the observed river discharge pattern. The descriptive test ANN model shown in Figure 8 revealed that the model slightly delayed in estimating the river discharge at the beginning but was able to accurately capture peak flows and likewise showed a perfect and continual simulation of the observed discharge pattern. The ability to point out this grey area of importance was as a result of presenting the ANN modelled result in a descriptive graph which could be difficult using tabulated statistical data. Hence, the modelled result showed that ANN perfectly simulated the observed river discharge, thus, agreeing to the work of BABY and VARIJAK [2016] and can be perfectly used for hydrologic modelling.
Rainfall-river discharge modelling for flood forecasting using Artificial Neural Network (ANN)

After testing, the ANN model was adopted as the truly developed ANN model for the Imo River, having performed well with coefficient of correlation close to +1. The developed ANN model was further subjected to one to five lead day forecasts as shown in Figure 8, which appears to follow the same pattern for the first day to the fifth day, with exception of the tip, being the exploded area, which is vividly different for the first lead day through the fifth lead day as shown in Figure 9a–e.

The graphical description (Fig 8, further differentiated through Fig. 9a–e) showed that the model gave a good prediction in the first to third lead days but slightly over-predicted discharge during the fourth lead day forecast and under-predicted discharge during the fifth lead day forecast, respectively.

The exploded areas of Figure 8, have been presented as Figure 9a–e, for first, second, third, fourth and fifth lead day forecasts, respectively. The generalized Figure 8, appears to look as the same for the first to fifth lead days, until its tip was exploded, to show difference in trend of the forecast. From these exploded plots, the forecast results of the model developed, changed progressively in values, from first lead-day forecast to the fifth lead-day forecast. It is observed from the results on Table 2 and Figure 9a–e, that the forecast results decreased as the lead-days increased, with the accuracy of the ANN model decreasing as the lead-days increased. Hence, ANN is very accurate for a short-time flood forecasting, in hours or few days.

The plots of the exploded area, as shown in Figure 9a–e depict the plots of the exploded areas of the one-to-five lead day forecasts, shown in Figure 8, which is a harmonized diagram of the plots for the first, second, third, fourth and fifth lead day forecasts, each of which resemble one another in outlook, with the exception of their trends depicted at their exploded regions.

Generally, model results showed satisfactory performance for short term forecasts and underscored the utility of ANN technique for rainfall–river discharge modelling at watershed scale. Good modelled and forecasted results achieved in this study showed that the feed-forward multi-layer neural network architecture remains the best neural network architecture for rainfall–river discharge modelling where there is scarcity of data; this conformed to the work of Varoonchotikul [2003].

Fig. 7. Observed vs simulated hydrograph during artificial neural network models: a) validation, b) testing; source: own study

![Fig. 7. Observed vs simulated hydrograph during artificial neural network models: a) validation, b) testing; source: own study](image)

Fig. 8. Generalized diagram for first to fifth lead day forecasts for the Imo River; source: own study

![Fig. 8. Generalized diagram for first to fifth lead day forecasts for the Imo River; source: own study](image)
CONCLUSIONS

This study presents the application of ANN technique for rainfall-river discharge of Imo River. The feed-forward multilayer perceptron neural network and supervised back-propagation training algorithm were employed for model development. The model was trained, validated and tested using hydro-meteorological records for the study area with a view of forecasting one to five lead day discharge. The one to five lead day forecasts were made and the results showed satisfactory performance of the ANN and its suitability for rainfall-river discharge modelling for the study area. The developed model perfectly simulated the river discharge with good coefficient of multiple determination close to +1, thus, proving that ANN is a fast and reliable technique for rainfall-river discharge modelling, and thus can be adopted by the Federal Ministry of Water Resources through the Nigeria Hydrological Services Agency (NHISA) to forecast flow processes of major rivers in Nigeria for early flood risk warning. This study further showed that this technique could be applied in data-scarce watersheds, where it is not practical to apply process-based hydrologic models for hydrologic modelling. Again, the Federal Ministry of Water Resources through its River Basin Development Authorities will find this work useful for flood and erosion control, and watershed management to effectively provide and maintain irrigation infrastructures across Nigeria.

REFERENCES


---

Fig. 9. Exploded area of tip of Figure 8 – showing consecutive lead days forecast for the Imo River: a) first day, b) second day, c) third day, fourth day, e) fifth day; source: own study
Rainfall-river discharge modelling for flood forecasting using Artificial Neural Network (ANN)


