Development of Stage Discharge Rating Curve for River Kaduna, Nigeria Using Artificial Neural Network

John J. Musa, Martins Y. Otache, Elijah T. Musa, Otuaro, E. Akpoebidimiyen, Pius O. Dada, Michael Entonu

Abstract**—**The stage-discharge relationship is a critical component of quantitative water resources assessment, especially for surface water planning. This study thus attempts at evaluating the suitability of an Artificial Neural Network (ANN) approach in modelling the stage-discharge relationship of a particular river. To this end, daily stage and discharge data covering 8 years for River Kaduna, Nigeria were mobilised for this study along with ANN approach, the conventional method for the development of the stage-discharge relationship was employed in order to add objectivity to the discussion. Results obtained indicate that the stage-discharge model for River Kaduna using the conventional statistical approach is adequate; however, ANN statistical analysis showed that it outperformed the conventional method; especially the Mean Absolute Error (MAE) which was gotten to be 3417.43 for the validation process. Based on the results obtained, considering the short length of data used, useful generalisations cannot be feasible for long-term predictions. It is recommended that extensive data should be assembled taking note of hydro-climatic alterations.

Keywords**—**ANN, discharge, modelling, river, water resources.

I. Introduction

SUFFICIENT and reliable information about rate of flow (i.e. discharge) of a river is vital for several hydrologic applications like planning water resources and operation, water and sediment budget analysis, hydraulic and hydrologic modelling and design of storage and conveyance structures [1]. To obtain this information, measurements have to be taken continuously. However, the process of direct collection of the measurement is always expensive and tedious. For these reasons, a functional relationship between stage (height of water in the river per time which is more comfortable and less expensive to measure) and discharge is always established with the help of field measurements; this relationship is often referred to as rating curve. Unfortunately, the stage-discharge relationship is not always a simple, unique one [2]; this is because, in practical reality, rate of flow is not a function of stage alone, but is also dependent on other factors such as water surface slope, geometry of the channel, bed roughness and the unsteady nature of flow. In some cases, these factors may bring about a non-unique relationship which usually manifest as multiple loops in the observed stage-discharge measurement as such limits the practical use of rating curve for a particular river.

The rating curve has been developed in time past by various researchers using various methods like the conventional method which involves the establishment of the relationship between stage and discharge by statistical regression analysis this entails the use of a number of observations of stage (H) and discharge (Q) [3], [4]. This method, however, is tedious and not very reliable as it assumes discharge to be a function of the stage alone [5], [6]. Other methods that have been adopted by various researchers are data-driven models which have been adopted in water resources engineering in getting solutions to several other hydrologic problems. These methods of which were found to give more reliable results in modelling the stage-discharge relationship include genetic algorithm (GA) as used by Tripura et al., [7] alongside model tree (MT).

Considering the probable potential of the ANN and too that there has not been any reported work on modelling of stage-discharge relationship for river Kaduna by means of any soft computing framework, therefore, for this study the application of the ANN in this regard is examined with the sole purpose of evaluating its suitability or otherwise.

This study aims to develop a stage-discharge rating curve for river Kaduna in Kaduna state and establish the potential usage of ANN in stage – discharge forecasting or prediction under a condition of data austerity.

II. Materials and Methods

A. Hydrology of River Kaduna

The Kaduna River is a tributary of the Niger River which flows for 550 kilometres through Nigeria. It is located at an elevation of 55m above sea level, with coordinates of 8°45’0” N and 5°48’0” E. Because of the distinctive wet and dry seasons in Nigeria, the river experiences seeming seasonality effect; it is full during the peak raining period and extremely low around

J. J. Musa is with the Federal University of Technology, P. M. B. 65, Minna, Niger State Nigeria (phone: +234-803-668-2747; e-mail: johnmusa@futminna.edu.ng).
M. Y. Otache is with the Federal University of Technology, P. M. B. 65, Minna, Niger State Nigeria (phone: +234-803-268-0996; e-mail: martynso_pm@futminna.edu.ng).
E. T. Musa is with the Central Workshop Department of the University of Ilorin, Kwara State, Nigeria (phone: +234-803-289-8786; e-mail: elijahtmusa2013@gmail.com).
O. E. Akpoebidimiyen is with the Maritime University, Okerenkoko, Nigeria (phone: +234-803-553-6580; e-mail: e.otuaro@yahoo.com).
P. O. Dada is with the Federal University of Agriculture, Abeokuta, Ogun State Nigeria (phone: +234-803-268-0996; e-mail: dadapoo@funaab.edu.ng).
M. Entonu is with the Federal University of Technology, P. M. B. 65, Minna, Niger State Nigeria (phone: +234-803-647-6869; e-mail: michael.entonu@st.futminna.edu.ng).

February and March.

Fig. 1: Map of Nigeria showing the Kaduna River and its traverse

B. Data Collection

Daily stage and discharge data for River Kaduna were collected at Kaduna south for an eight-year period was used for this study. Both stage and discharge were collected over eight hours daily; that is, every eight hours. These are then arranged to obtain the mean daily stage and discharge, respectively.

C. Methods

In order to be disposed to objectively asses the soft computing framework, the conventional approach was employed along the ANN.

D. Conventional Method: Stage-Discharge Rating Curve

The conventional method used in this study is the parametric regression method which entails the development of the stage-discharge rating curve. To this end, it was assumed that the river is non-alluvial and thus exhibits permanent control; for the permanent control case, the relationship between stage and discharge is a single-valued relation; given as

$$Q = C\,(G - a)^{\beta} \tag{1}$$

where Q is the discharge, G the gauge height (stage) in meters, ‘a’ is a constant which represents the gauge height reading corresponding to zero discharge, β and C are rating curve constants. The values of C and β were obtained by the least square regression method. Taking logarithm, (1) becomes

$$Log\,Q = \beta Log(G - a) + Log\,c \tag{2}$$

or

$$Y = \beta X + b \tag{3}$$

Simplifying, the dependent variable $Y = Log\,Q$, whereas the independent variable $X = Log\,(G - a)$ and $b = Log\,C$. For the best fitting straight line of N observations of X and Y, by regressing $X = Log\,(G - a)$ on $Y = Log\,Q$.

$$\beta = \frac{N(\sum XY) - (\sum X)(\sum Y)}{N(\sum X^2)} \tag{4}$$

and

$$b = \frac{\sum Y - \beta(\sum X)}{N} \tag{5}$$

When ‘a’ is taken to be equal to zero, (1) transforms to

$$Q\;CG^{\beta} \tag{6}$$

Similarly, by implication,

$$Log\,Q = \beta Log\,G + Log\,c \tag{7}$$

Both situations, i.e., where ‘a’ assumes a zero and non-zero values were considered in the development of the rating curve.

E. Estimation of the Stage Corresponding to Zero Discharge

Two methods exist for the determination of the parameter ‘a’. The first method involved plotting Q vs G on an arithmetic graph paper and drew the best fit curve. By extrapolating the curve by eye judgment, 'a' was determined as the value of G corresponding to Q equal to zero. The value of 'a' obtained was used to plot a graph of log Q vs log (G-a) to verify whether the data plot is a straight line. On the other hand, Q Vs G was plotted to an arithmetic scale; three discharge values, Q_1, Q_2 and Q_3 were selected such that

$$\frac{Q_1}{Q_2} = \frac{Q_2}{Q_3} \tag{8}$$

The corresponding values of gauge reading, G_1, G_2 and G_3 were noted from the curve such that

$$\frac{(G_1 - a)}{(G_2 - a)} = \frac{(G_2 - a)}{(G_3 - a)} \tag{9}$$

i.e.

$$a = \frac{(G_1 G_3 - G_2^2)}{(G_1 + G)} - 2G_2 \tag{10}$$

F. Statistical Parameters

To effectively appraise the characteristics of the data set the following statistical parameters for the training and testing data set for the gauge station were obtained: The mean (μ), standard deviation (σ), coefficient of variation (cv), skewness (g), minimum value (X_{min}) and maximum value (X_{max}).

G.Soft Computing - ANN

The adopted soft computing framework for this study is the ANN. The use of the ANN is to optimise the values of the parameters c, β and a.

ANN is often characterised by an architecture that represents the pattern of connection between nodes, a method for determining the connection weights, and an activation function. The backward-feed forward multi-layer perception network was adopted for this study. It consists of three layers – the input, hidden and output layers; with the input layer consisting of six nodes, the hidden layer fifteen nodes and output layer of one node. The number of nodes for the hidden layer was determined by a trial and error method. The number was varied gradually until an acceptable mean squared error (MSE) was obtained. In the overall, a supervised training strategy was adopted; training was stopped when there is a seeming convergence in the sum squared weight (SSW), adequate number of parameters as well as sum squared error (SSE) in other to avoid overfitting of the network. The Bayesian Regularization training algorithm, tans signoid and purelin transfer function were employed, respectively, for the input and output layers with bias. The Bayesian regularisation algorithm was adopted ahead of the Leven Marquardt (lm) to avoid computational overhead and enhance generalisations.

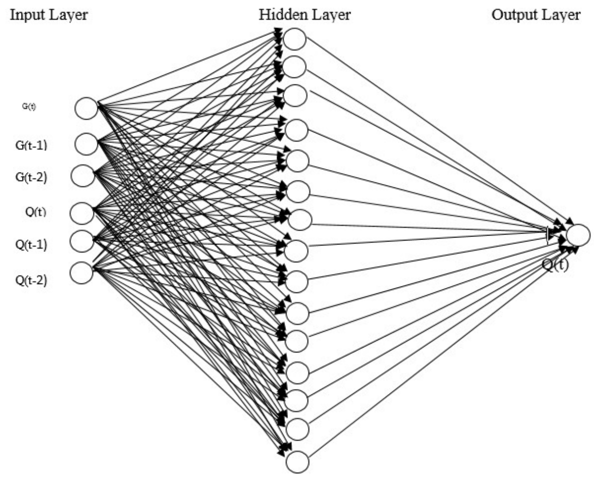

Fig. 2: The ANN Architecture

H.Data Base Management and Model Training

The eight years data were divided into two parts. The first part of six years was used for training and the other two years for testing purpose (split sampling). The standardised pre-processing option was adopted; this was informed by the desire to maintain the spectrum of the data. The standardisation was done accordingly.

Stage

$$G'(t) = \frac{G(t) - \mu}{\sigma} \tag{11}$$

Discharge:

$$G'(t) = \frac{Q(t) - \mu}{\sigma} \tag{12}$$

The determination of the input layer size was carried out by evaluating the correlation between the elements of the respective time series of the data sets; precisely at the lagged period where Q is the raw discharge data, G, raw stage data, μ and σ are the mean and standard deviation. Based on the correlation matrix, the following lagged series G(t), G(t-1), G(t-2) and Q(t), Q(t-1), Q(t-2) and Q(t-3) were noticed to have a seemingly high level of both intra and intercorrelations

I. Performance Statistics

The following statistical measures were used to evaluate the performance of the methods (i.e., conventional and ANN as applicable).

i. $$R_{max} = \frac{Forcasted\ maximum}{Observed\ maximum}\ x\ 100\% \tag{13}$$

$$R_{min} = \frac{Predicted\ maximum}{Observed\ maximum}\ x\ 100\% \tag{14}$$

where R depicts the correlation level.

ii. $$MAE = \frac{\sum_{j=1}^{N} |Y^{(o)} - Y^{(p)}|}{N} \tag{15}$$

where K_j is the rank of the residuals, and $Y_j^{(o)}$ and $Y_j^{(p)}$ refer to observed and predicted flows, respectively.

III. Results and Discussion

A. Conventional Method

The value for the parameter "a"; i.e. stage corresponding to zero discharge as estimated using the first method was 0.15. The relationship between stage and discharge for the station is as given in (16)

$$Q = 3.08(G - 0.15)^{3.37} \tag{16}$$

which can also be expressed as

$$Log\ Q = 3.37Log\ (G - 0.15) + 0.4884 \tag{17}$$

For the other case where 'a' is taken to be zero, i.e. a = 0; the model is expressed as

$$Q = 2.37G^{3.45} \tag{18}$$

Following the methodology for the conventional method (RC1), the stage corresponding to zero discharge 'a' was estimated to be 0.15 m. This result was used to estimate the other unknown parameters. The value of the Rating Curve constants β and c were gotten to be 3.45 and 3.37 and 3.08, respectively. For the second scenario (RC2), the values of β and c were gotten to be 3.45 and 2.37, respectively. These values were used in the Training and Validation processes

TABLE I
Statistical Parameters for the Training and Testing Data Set

Data Set	Data Type	μ	σ	cv	G	X_{max}	X_{min}
Training	Stage (m)	7.82	11.41	1.46	23.42	318	1.87
	Flow (m^3/s)	6943.43	12295.5	1.77	7.76	231400	8
Testing	Stage (m)	6.06	2.75	0.45	0.19	9.61	0.36
	Flow (m^3/s)	397	6304.03	1.59	2.88	101415	1.74

TABLE II
Performance Evaluation for the Conventional Method

Model/Event		MAE (m^3/s)	R_{max}	R_{min}
RC1	Training	1857.56	67.15%	316.47%
	Testing	3564.84	37.50%	5.68%
RC2	Training	1864.267	67.40%	257.17%
	Testing	3409.00	39.00%	6.90%

Table 2 presents the statistical performances of the conventional method for both situations i.e. the case of 'a' estimated as the stage value corresponding to zero flow taken to be RC 1 and the other situation where 'a' is taken to be zero taken as RC2. RC1 conventional model resulted in a MAE of 3564.84 m^3/s for the testing period as against that of RC2 model, which yielded a MAE of 3409 indicating a better performance over the former. The R_{max} and R_{min} values of both also indicate superior performance of RC 2 over the RC 1 model.

B. Soft Computing Method

TABLE III
Correlation Matrix for Training data (ANN)

	Q(t)	Q(t - 1)	Q(t - 2)	Q(t-3)	Q(t)	G(t-1)	G(t-2)
G(t)	1						
G(t-1)	0.628	1					
G(t-2)	0.627	0.628	1				
G(t-3)	0.568	0.627	0.628	1			
G(t)	0.415	0.405	0.399	0.394	1		
G(t-1)	0.405	0.415	0.405	0.399	0.305	1	
G(t-2)	0.396	0.405	0.415	0.405	0.300	0.305	1

Correlation is significant at the 0.01 level; Elements are Pearson moment Correlation Vectors.

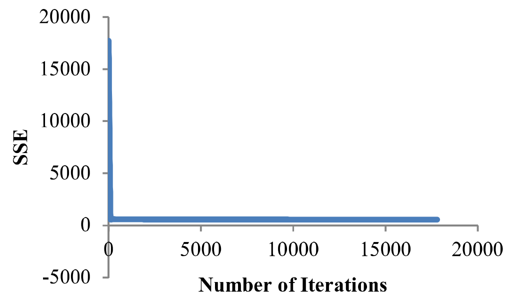

Fig. 3 Number of Iterations and SSE

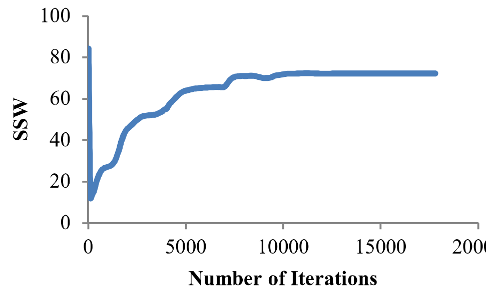

Fig. 4 Number iterations and SSW

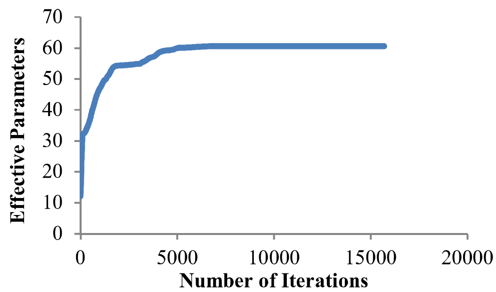

Fig. 5 Number Iteration and the Effective number of Parameter

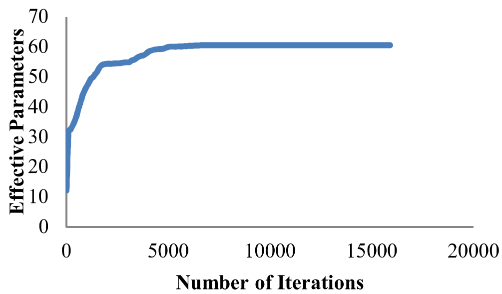

Fig. 6 Number of Iterations and Effective number of Parameters

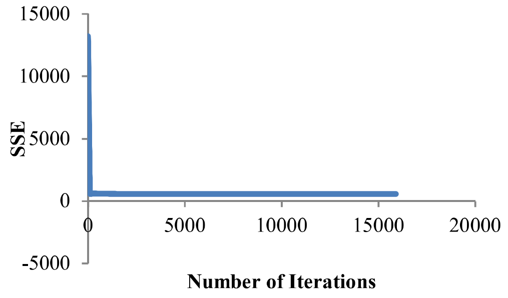

Fig. 7 Number of Iterations moreover, SSE

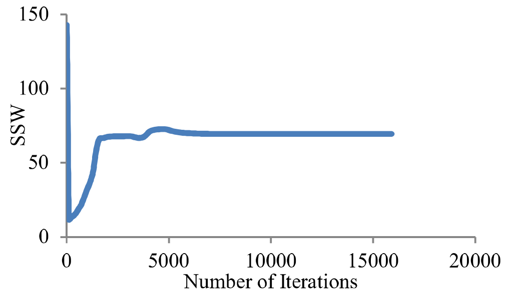

Fig. 8 Number Iteration and SSW

Figures 3 to 8 show the relationship between the Number of Iterations and the SSW, SSE and the Effective Number of Parameters both for the Training and Validation periods. For each of the plots, a change is noticed from the beginning of the Iterations until a point where the line becomes horizontal, indicating a point of convergence which shows that the ANN has finished leaving and any other further Training will amount only to a waste of time. The Training and Validation process was stopped where no change was noticeable in the Parameters and indicated in the plots above which is similar to the findings of [8]. MSE for this point for Training Validation process was gotten to be 0.2587 and 0.2603 respectively.

TABLE IV
PERFORMANCE EVALUATION FOR ANN MODEL

		MAE (m^3/s)	R_{max}	R_{min}
ANN	Training	1168.8	49.52%	11%
	Testing	3417.34	74.18%	14.57%

Table 4 shows the performance evaluation of the ANN model. With a MAE value of 3417.34 m^3/s for the testing period, the model shows slightly better performance over the two conventional model cases examined above. The R_{max} and the R_{min} values of the model which are 74.18% and 14.57% respectively for the testing period show in comparison with the conventional models that the ANN model is more reliable in replicating the characteristics of the flow regime in terms of low and peak flow indexes. Also, Figures 9 and 10 show that the ANN model was able to mimic the trend of observed flows at the Kaduna river, doing better at low flows.

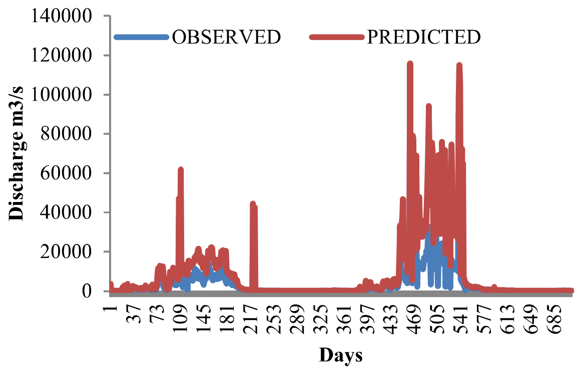

Fig. 9 Model validation hydrograph

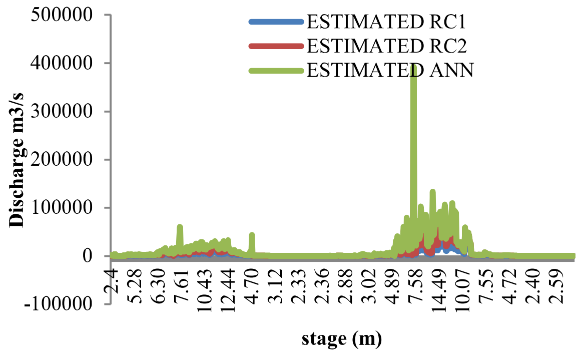

Fig. 10 Computed discharge for different modelling schemes

ANN model outperformed the conventional model in terms of the performance criteria used and also as observed on Figures 9 and 10 regarding its ability to better mimic the trend of the observed flow at the river using the data length available. With a MAE value of 3417.34 for the Testing period when compared with the values gotten from other researchers [9]-[11] who used the ANN to forecast upper and lower uncertainty bends in of rivers flood discharge, suspended sediment load prediction on river discharge information and stage-discharge relationship modelling using data mining models respectively which showed slightly too high values. This may be due to the length of data available for their study, which rather too short and also climatic variations which are on its own capable of affecting the outcome of the study.

IV. CONCLUSION

The results obtained in this study have been able to demonstrate the capacity of the ANN in developing a Stage-Discharge Rating Curve for the given River within the prevailing circumstances, i.e., given the length of data available for this study. The ANN shows superior performance over the conventional method as observed in the result MAE as compared with that of the conventional method and from R_{max} value obtained for the validation, it also shows that ANN performs better primarily in predicting high lows for the given river situation. The MAE value obtained when in comparison with the research links carried out on other rivers and the climatic situation shows air compliance when the data length and variability in climate are consideration.

REFERENCES

[1] S. J. Cho, P. R. Wilcock, P. Belmont, K. B. Gran, and B. F. Hobbs. "Simulation model for collaborative decision making on sediment source reduction in an intensively managed watershed". *Water Resources Research*, 2019, *55*(2), 1544-1564.

[2] A. R. Vatankhah, and M. Mohammadi. "Stage–discharge equation for simple flumes with semi-cylinder contractions". *SN Applied Sciences*, 2020, *2*(3), 1-13.

[3] H. Iraji, M. Mohammadi, B. Shakouri, and S. G. Meshram. "Predicting reservoir volume reduction using artificial neural network". *Arabian Journal of Geosciences*, 2020, *13*(17), 1-13.

[4] Z. Cheng, K. Lee, D. Kim, M. Muste, P. Vidmar, and J. Hulme. "Experimental evidence on the performance of rating curves for continuous discharge estimation in complex flow situations". *Journal of Hydrology*, 2019, *568*, 959-971.

[5] N. Taguchi. "Description and explanation of pragmatic development: Quantitative, qualitative, and mixed methods research". *System*, 2018, *75*, 23-32.

[6] S. Barbetta, T. Moramarco, and M. Perumal. "A Muskingum-based methodology for river discharge estimation and rating curve development under significant lateral inflow conditions". *Journal of Hydrology*, 2017, *554*, 216-232.

[7] J. Tripura, P. Roy, and A. K. Barbhuiya, "Simultaneous streamflow forecasting based on hybridized neuro-fuzzy method for a river system". *Neural Computing and Applications*, 2020, 1-13.

[8] E. Hoffer, I. Hubara, and D. Soudry. "Train longer, generalize better: closing the generalization gap in large batch training of neural networks". *Advances in neural information processing systems*, 2017, *30*, 1731-1741.

[9] J. Leandro, A. Gander, M. N. A. Beg, P. Bhola, I. Konnerth, W. Willems, R. Carvalho, and M. Disse. "Forecasting upper and lower uncertainty bands of river flood discharges with high predictive skill". *Journal of Hydrology*, 2019, *576*, 749-763.

[10] S. Q. Salih, A. Sharafati, K. Khosravi, H. Faris, O. Kisi, H. Tao, M. Ali, and Z. M. Yaseen. "River suspended sediment load prediction based on river discharge information: application of newly developed data mining models". *Hydrological Sciences Journal*, 2020, *65*(4), 624-637.

[11] A. Al-Aboodi, H. Ibrahim, and W. S. Al-Rekabi. "Stage-discharge relationship modeling using data mining techniques in an arid region". *International Journal of Applied Engineering Research*, 2018, *13*(1), 326-336.