Forecasting the Congo River Discharge and Water-level using an Artificial Neural Network Approach

Salomon Salumu Zahera[1,2,3,5], Célestin Ngandu Balekelayi[1,2,4], Musandji Fuamba[3], Martin Kabantu Tshikeba [1,2], Simon Butsana[2], Vincent Lukanda[5] and Jean-Michel Hiver[6].

[1]Institut National du Bâtiment et des Travaux Publics (INBTP) BP: 4731 Kinshasa/Ngaliema RDC.
[2]Laboratoire de Modélisation et Analyse des Hydro-systèmes et des Constructions Hydrauliques (LMH-CH), BP: 11585 Kinshasa1 RDC
[3]École Polytechnique de Montréal, Département des génies civil, géologique et des mines.
[4]University of British Columbia, Kelowna, British Columbia, Canada
[5]Université de Kinshasa, Faculté de sciences
[6] Université Libre de Bruxelles

Abstract

Rating curves are used to estimate the discharge in rivers, especially in the Congo River. However, the rating curves used for the Congo River, established since 1959, are inadequate because of the changes recurrently occurring in the morphology of the river. The streamflow processes are complex and highly nonlinear and are a result of combination of different factors such as catchment, rainfall, geomorphologic and climatic characteristics. In fact, the sediment load, in the Congo River, is one of the highest load material in rivers in the world. Furthermore, the changes in climate and morphology, create quasi permanent unsteady flows in the river. On the other side, accurate forecasting of streamflow is essential for efficient operations and optimal allocation of water resources. Flow forecast for hours, days, months in advance is vital for hydropower agencies. This research aims at developing a flow forecast tool using Artificial Neural Network at the Inga dam. The input data are the water stage measured at the port of Kinshasa situated 350 km upstream the dam. The tool showed a better prediction of flow at Inga dam and will support the management team to have information days in advance.

***Keywords*:** Artificial Neural Network, Discharge, Congo River, Water-level, Flow forecast, Inga dam.

1. INTRODUCTION

Discharges and water levels are essential components of river hydrodynamics(Khan, Hasan, Panwar, & Chakrapani, 2016). Discharge measurement is time consuming, hazardous and costly. A cheaper alternative is the so-called rating curve that embodies a functional relationship between the water level (called stage when measured from a datum) and discharge with the help of field measurements. Once a reliable rating curve is available discharge can be estimated from the rating curve using the observed water level (stage)(Bhattacharya & Solomatine, 2005).

The relationship existing between the water-surface stage (i.e. the water level) and the simultaneous flow discharge in rivers and open channels is known as stage-discharge relation or rating curve, or also just rating (Braca, 2008). The rating curve is a very important tool in surface hydrology because the reliability of discharge data values is highly dependent on a satisfactory stage-discharge relationship at the gauging station(Braca, 2008). Although the preparation of rating curves seems to be an essentially empiric task, a wide theoretical background is needed to create a reliable tool to switch from measured water height to discharge.

The rating curve is established by concurrent measurements of stage and discharge (through velocity measurements, dilution methods, or other techniques) and the results are fitted graphically or statistically to yield the rating curves (Schmidt & Yen, 2001).

Reliable estimation of discharge in a river is the crucial component of efficient surface water management and planning. Once a relationship is established it can be used for forecasting discharge from future measurements of water level only. In the evolution of river science, a substantial amount of work has been done on predicting future discharges and water levels.

The rating curve can be constructed with the help of polynomial regression or auto-correlation-based statistical method. A number of models have been proposed to predict water discharges and water levels in a river (Abrahart & See, 2000; Atiya, El-Shoura, Shaheen, & El-Sherif, 1999). Birgand, Lellouche, and Appelboom (2013) calculated rating curves from random sampling of reference flow and stage data. Based on the Jones formula, Petersen-Øverleir (2006) proposed a methodology utilizing nonlinear regression as a solution for situations in which the stage-discharge relationship is affected by hysteresis due to unsteady flow; Clemmens, and Wahlin (2006) evaluated the accuracy of various methods for finding stage-discharge relationships. Guven, and Aytek (2009) also reported that Liao, and Knight (2007) proposed three analytic stage-discharge formulas for prismatic open channels that are suitable for manual calculation. As mentioned by Guven, and Aytek (2009), the body of literature contains many applications of other soft computing techniques in stage-discharge modeling. Jain, and Chalisgaonkar (2000) established a stage-discharge relationship based on three layer feed forward ANNs, Sudheer, and Jain (2003) explored the effectiveness of a radial basis function (RBF), and Bhattacharya, and Solomatine (2005) observed that ANNs and M5 model trees predicted the stage-discharge relationship much more accurately than the traditional rating curves. Deka, and Chandramouli (2003) compared the performance of an ANN model, a modularized ANN model, a conventional curve-fitting approach, and a neuro-fuzzy model for deriving the rating curve using a case study.

Rating curve used to estimate the discharge in the Congo River at Kinshasa station, was established since 1959. It became inadequate for forecasting water level and discharge because of the changes recurrently occurring in the river morphology.

In this research water flow at Inga dam.is forecasted from observed water levels at the Port of Kinshasa, situated 350 km upstream, using Artificial Neural Network. Data availability and the objective of protecting Inga dam from flooding through appropriate decision making are the reason why this port is selected.

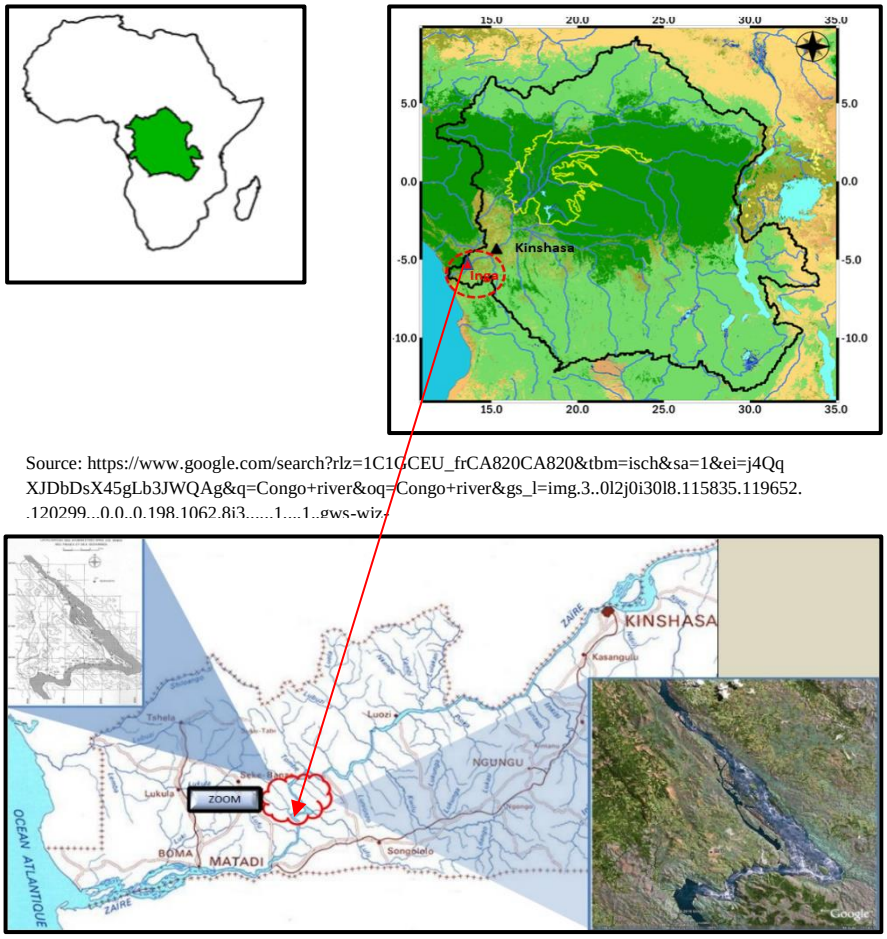
2. STUDY SITE

The Inga dams' site is located in western Democratic Republic of the Congo, 150 km upstream of the mouth of the Congo River, and 225 km southwest of Kinshasa on the Congo River (Fig. 1). The Congo River is the world's second largest in terms of flow (42,000 m^3/s), after the Amazon, and the second longest river in Africa (4,700 km), after the Nile River. It empties into the equatorial Atlantic Ocean creating what is famously known as the Congo Plume. The plume is a high-productivity area arising from the rich nutrient flow from the river and is detected as far as 800 km offshore. The plume accounts for 40-80% of total carbon productivity and is one of the largest carbon sinks in the world.

The river is unique in that it has large rapids and waterfalls very close to the mouth while most rivers have these features upstream. The dam site is on the largest waterfall in the world by volume, the Inga Falls. Inga Falls is a series of falls and rapids that drop in elevation via small rapids. The main falls are 4 km wide, dropping to about 21.37 metres near a bend and forming hundreds of channels and rivulets and many small islands.

3. AVAILABLE DATA

The data set used in this study was obtained from the waterway agency of the Democratic Republic of Congo. The time series of daily stage and discharge data was taken from the port station of Kinshasa from 1903 to 2010 (Fig. 2).



Source: https://www.google.com/search?rlz=1C1GCEU_frCA820CA820&tbm=isch&sa=1&ei=j4QqXJDbDsX45gLb3JWQAg&q=Congo+river&oq=Congo+river&gs_l=img.3..0l2j0i30l8.115835.119652.120299...0.0..0.198.1062.8j3.....1....1..gws-wiz-

Source:(Tonino-J., 2015)

Figure 1. Study site location

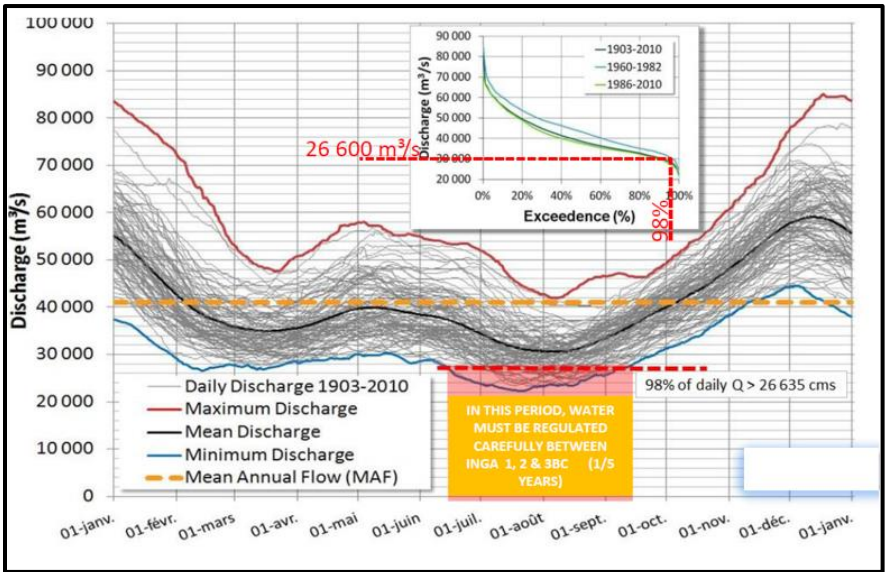


Figure 2. Daily discharge of the Congo River at Inga 1903-2010 (Société Nationale d'Électricité, 2012).

4. METHODOLOGY

From past and present observations of water level at the port of Kinshasa, the future water flow at Inga dam is predicted for the next day. To reduce the uncertainty and to account for the complexity of hydrological processes occurring between Inga and Kinshasa, only four previous days and the actual observation have been incorporated in the model. In addition, the two points define a blackbox with the input water levels at Kinshasa and output flows at Inga. ANN selection is justified.

4.1 Weka Interface

Weka (Waikato Environment for Knowledge Analysis) is a popular suite of machine learning software written in Java, developed at the University of Waikato, New Zealand (Arora, 2012). The Weka suite contains a collection of visualization tools and algorithms for data analysis and predictive modeling, together with graphical user interfaces for easy access to this functionality.
The original non-Java version of Weka was TCL/TK frontend software used to model algorithms implemented in other programming languages, plus data preprocessing utilities in C, and a Make file-based system for running machine learning experiments. This Java-based version (Weka 3) is used in many different application areas, in particular for educational purposes and research. There are various advantages of Weka:

- It is freely available under the GNU General Public License
- It is portable, since it is fully implemented in the Java programming language and thus runs on almost any architecture
- It is a huge collection of data preprocessing and modeling techniques
- It is easy to use due to its graphical user interface

Weka supports several standard data mining tasks, more specifically, data preprocessing, clustering, classification, regression, visualization, and feature selection. All techniques of Weka's software are predicated on the assumption that the data is available as a single flat file or relation, where each data point is described by a fixed number of attributes.

4.2 Classification Function Multilayer Perceptron

Multilayer Perceptron classifier is based upon backpropagation algorithm to classify instances. The network is created by an MLP algorithm. The network can also be monitored and modified during training time. The nodes in this network are all sigmoid (except for when the class is numeric in which case the output nodes become unthresholded linear units).

The backpropagation neural network is essentially a network of simple processing elements working together to produce a complex output. The backpropagation algorithm performs learning on a multilayer feed-forward neural network. It iteratively learns a set of weights for prediction of the class label of tuples. A multilayer feed-forward neural network consists of an input layer, one or more hidden layers, and an output layer. An example of a multilayer feed-forward network is shown in Fig.3(Khan et al., 2016).

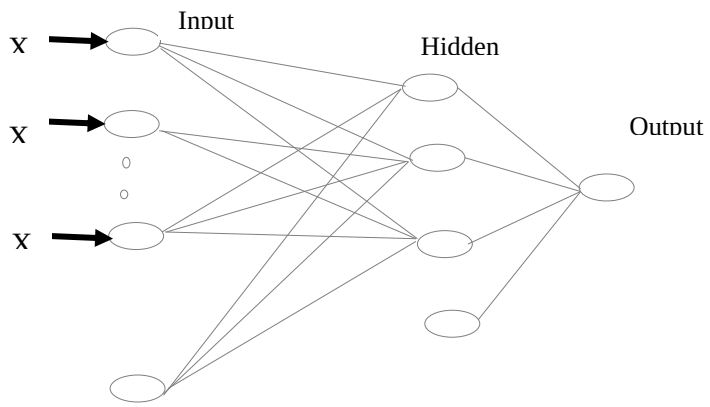


Figure 3. The structure of the artificial neural networks

Each layer is made up of units. The inputs to the network correspond to the attributes measured for each training tuple. The inputs are fed simultaneously into the units making up the input layer. These inputs pass through the input layer and are then weighted and fed simultaneously to a second layer of "neuronlike" units, known as a hidden layer. The outputs of the hidden layer units can be input to another hidden layer, and so on. The number of hidden layers is arbitrary, although in practice, usually only one is used (Witten, Frank, Hall, & Pal, 2016). At the core, backpropagation is simply an efficient and exact method for calculating all the derivatives of a single target quantity (such as pattern classification error) with respect to a large set of input quantities (such as the parameters or weights in a classification rule) (Werbos, 1990). To improve the classification accuracy we should reduce the training time of neural network and reduce the number of input units of the network (Lu, Setiono, & Liu, 1996).

5. RESULTS AND DISCUSSION

Two models have been developed and tested againt a testing set data that represent 25 % of initial data. Table 1 gives the results evaluation parameters. Model 1 uses only water levels at the port of Kinshasa as input parameters. From Fig. 4, it is seen that the ANN captured the pattern of water flow at the Inga dam but could not reproduce high flows. This can be explained in the fact that there are several incoming water flows between the two points (e.g. the Inkisi creek could at some point discharge high flows in the river that have great change in the flow at Inga and is not predictable from Kinshasa). The model is well calibrated since there is not much changes between the Root Mean Squared Error (RMSE) of the training and the testing runs. The same observation occurs for the Mean Absolute Error (MAE) and the Relative Error (RE).

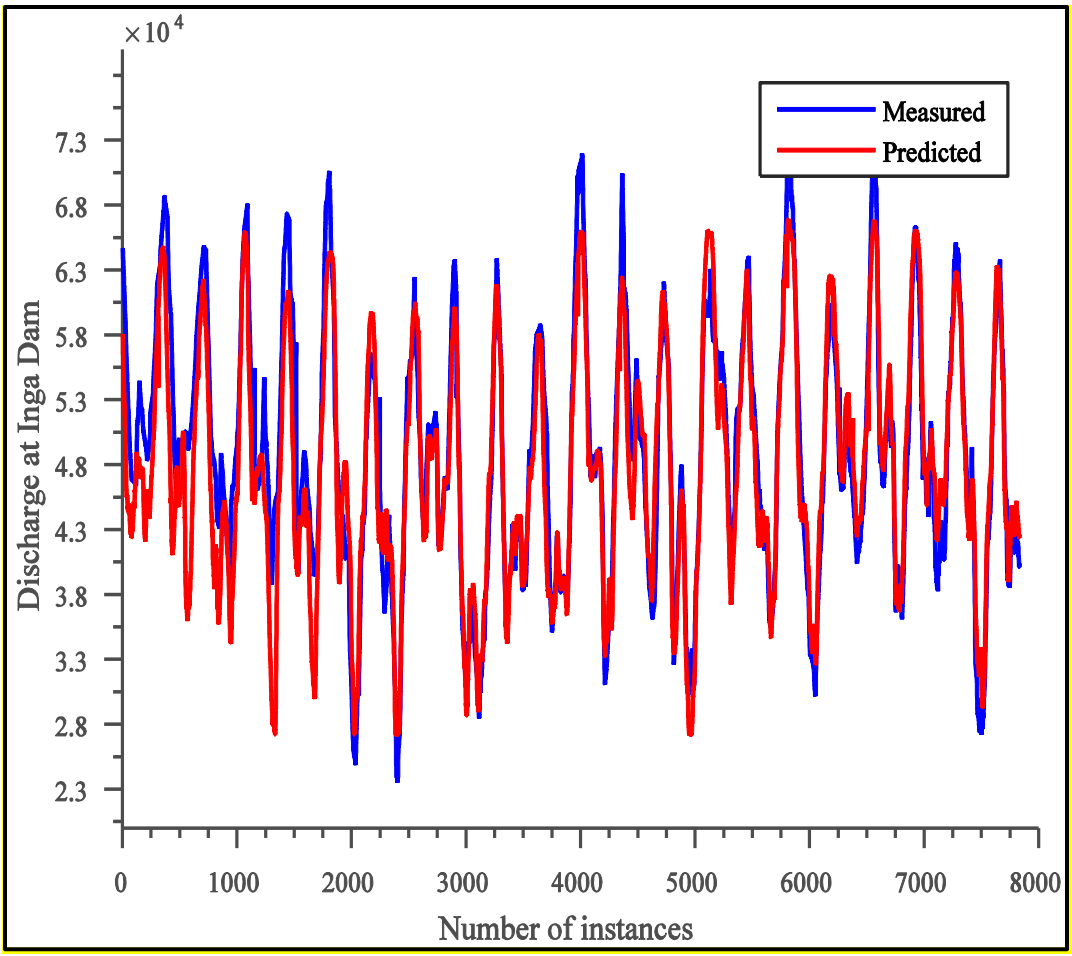


Figure 4: Model 1 training data

Table 1: Models results comparison

	Model 1		Model 2	
	Training	**Testing**	**Training**	**Testing**
Correlation Coefficient	0.93	0.92	0.99	0.98
MAE	2592.24	2770.38	195.48	197.33
RMSE	3480.38	3510.19	374.47	388.92
RE	33.27	35.79	2.51	2.53

The second model (Model 2) in Table 1 is built from the previous model with the inclusion of previous observed water flows at the Inga dam. The observed water flows are transmitted to Kinshasa each morning. Thus, the decision maker can use this model to forecast the coming water flows at the dam and orient team on field. Testing model 2 (Fig. 5) shows perfect match between the prediction and observations. The correaltion coefficient makes a good jump up to 0.98 for the testing data. Most important, the RMSE error dropped by 90 %. The reduction of RMSE and all the relative errors shows that the blackbox should at least have insigth of .

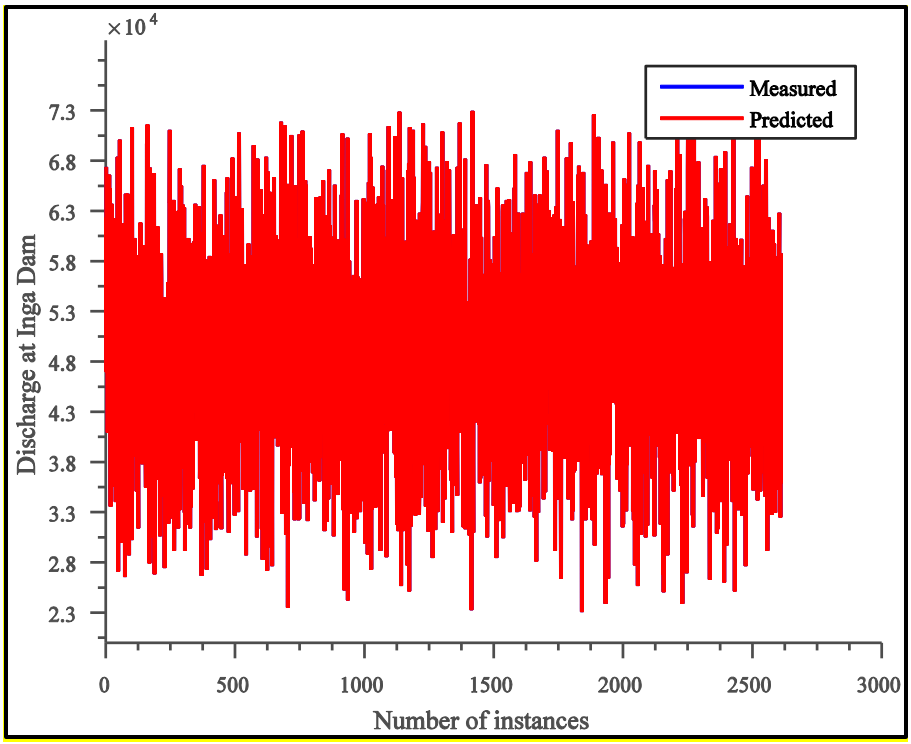


Figure 5: Model 2 testing data

The inclusion of previous water flows shows an enhancement of the model 1 that is based only on the water levels at the port of Kinshasa. From this study, it can be observed that the contribution of other creeks (e.g. Inkisi creek) between the port of Kinshasa and Inga dam is not negligible. However, for an early decision making, the use of water level at the port of Kinshasa gives an orientation on the attitude to adopt in the definition of operating policies especially in rainy seasons (9 months of the year in the region).

This study provides a tool to support decision making from water levels observed at the port of Kinshasa. It shows that the lack of incoming water flows between measurement stations (Port of Kinshasa and Inga dam) can be overcome through the application of well-developed models. ANN is perfect for this time of application and is more flexible to include more data as they become available.

6. CONCLUSIONS AND RECOMMENDATIONS

In this paper, the potential use of ANN to predict the discharge in a river from stage recorded data is demonstrated. The discharge at time t at 350 km downstream the gauge station is first calculated using the recorded water stage for the 5 previous days. The increased number of variable is aims at the reduction of error in the predictions. The high RMSE obtained from this model will lead to overestimation of the discharge with high consequences on the operations of the dam. The reduction of the error is obtained by introducing in the prediction the discharge of the two previous days. ANN is a powerful tool and lead to accurate prediction if the model is well construct. Future work will investigate the optimal number of variable (H and Q) to include in the model for discharge prediction with a given level of error.

7. REFERENCES

1. Abrahart, R. J., & See, L. (2000). *Comparing neural network and autoregressive moving average techniques for the provision of continuous river flow forecasts in two contrasting catchments. Hydrological processes, 14(11-12), 2157-2172.*
2. Arora, R. (2012). *Comparative analysis of classification algorithms on different datasets using WEKA. International Journal of Computer Applications, 54(13).*
3. Atiya, A. F., El-Shoura, S. M., Shaheen, S. I., & El-Sherif, M. S. (1999). *A comparison between neural-network forecasting techniques-case study: river flow forecasting. IEEE Transactions on neural networks, 10(2), 402-409.*
4. Bhattacharya, B., & Solomatine, D. P. (2005). *Neural networks and M5 model trees in modelling water level–discharge relationship. Neurocomputing, 63 381-396.*
5. Birgand, F., Lellouche, G., & Appelboom, T. (2013). *Measuring flow in non-ideal conditions for short-term projects: Uncertainties associated with the use of stage-discharge rating curves. Journal of hydrology, 503 186-195.*
6. Braca, G. (2008). *Stage-discharge relationships in open channels: Practices and problems: Univ. degli Studi di Trento, Dipartimento di Ingegneria Civile e Ambientale.*
7. Clemmens, A. J., & Wahlin, B. T. (2006). *Accuracy of annual volume from current-meter-based stage discharges. Journal of hydrologic engineering, 11(5), 489-501.*
8. Deka, P., & Chandramouli, V. (2003). *A fuzzy neural network model for deriving the river stage—discharge relationship. Hydrological sciences journal, 48(2), 197-209.*
9. Guven, A., & Aytek, A. (2009). *New approach for stage–discharge relationship: gene-expression programming. Journal of Hydrologic Engineering, 14(8), 812-820.*
10. Jain, S., & Chalisgaonkar, D. (2000). *Setting up stage-discharge relations using ANN. Journal of Hydrologic Engineering, 5(4), 428-433.*
11. Khan, M., Hasan, F., Panwar, S., & Chakrapani, G. (2016). *Neural network model for discharge and water-level prediction for Ramganga River catchment of Ganga Basin, India. Hydrological Sciences Journal, 61(11), 2084-2095.*
12. Liao, H., & Knight, D. W. (2007). *Analytic stage-discharge formulas for flow in straight prismatic channels. Journal of Hydraulic Engineering, 133(10), 1111-1122.*
13. Lu, H., Setiono, R., & Liu, H. (1996). *Effective data mining using neural networks. IEEE transactions on knowledge and data engineering, 8(6), 957-961.*
14. Petersen-Øverleir, A. (2006). *Modelling stage—discharge relationships affected by hysteresis using the Jones formula and nonlinear regression. Hydrological sciences journal, 51(3), 365-388.*
15. Schmidt, A., & Yen, B. (2001). *Stage-discharge relationship in open channels. Paper presented at the Proc. 3rd Intl. Symp. on Envr. Hydr., Tempe, AZ.*
16. Société Nationale d'Électricité, S. (2012). *Étude de développement du site hydroélectrique d'Inga et des interconnexions associées Kinshasa:*
17. Sudheer, K., & Jain, S. (2003). *Radial basis function neural network for modeling rating curves. Journal of Hydrologic Engineering, 8(3), 161-164.*
18. Tonino-J., N. (2015). *[Étude de développement du site hydroélectrique d'Inga et des interconnexions électriques associées]*.
19. Werbos, P. J. (1990). *Backpropagation through time: what it does and how to do it. Proceedings of the IEEE, 78(10), 1550-1560.*
20. Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data Mining: Practical machine learning tools and techniques*: Morgan Kaufmann.