Research article

Event-Based Rainfall-Runoff Forecasting in Pampanga River Basin, Philippines using Artificial Neural Networks (ANN)

JCOB C. MALAGUIT*
University of the Philippines Los Baños, Laguna, Philippines
Email: jcmalaguit@up.edu.ph

MARRIANE F. MAKAHIYA
University of the Philippines Los Baños, Laguna, Philippines

MARK LEXTER D. DE LARA
University of the Philippines Los Baños, Laguna, Philippines

Received 15 December 2016 Accepted 5 May 2017 (*Corresponding Author)

Abstract This study developed a series of rainfall-runoff forecasting models that can be used in designing the flood warning system around Pampanga River Basin. The data regarding rainfall and water level of the river was obtained from the Hydrometeorological Division (HMD) of Philippine Atmospheric Geophysical and Astronomical Services Administration (PAGASA). Data were collected from 2014 and 2015 hourly water level reading and rainfall reading. Feedforward Backpropagation Model, a variant of the Artificial Neural Network (ANN), was used in the study along with Gradient Descent with Adaptive Learning Rate Algorithm as a learning technique for the network. MATLAB R2009b was used to train and design the networks. A total of 45 networks were trained. Results of the training gave reasonable predictions for most of the stations with a minimum accuracy of 96%. Inaccuracy of training in some stations were attributed to the inconsistency in data and other factors.

Keywords Artificial Neural Networks, river basin, event-based forecasting, Feedforward Backpropagation, rainfall-runoff model

INTRODUCTION

Flooding has been one of the most destructive phenomena in the Philippines. Flood causes loss of lives, damage to personal assets and real properties, and adverse economic, environment and agricultural impacts (Chen and Graham, 2005). In 2014, it was reported that 153 out of 324 natural disasters that occurred in the country were hydrological disasters (Guha-Sapir et al., 2015) and in the last decade, the country experienced frequent flooding due to continuous raining and typhoons brought about by its archipelagic characteristic.

In Central Luzon, Philippines, one of the regions that have been affected by recent disastrous events including Typhoon Nona (*Typhoon Melor*) and Lando (*Typhoon Koppu*), there situates the second largest river in Luzon region, the Pampanga River. It traverses the provinces of Pampanga, Nueva Ecija, Tarlac and Bulacan. Pampanga-Agno river basin serves as the catch basin for Tarlac, Nueva Ecija, Pampanga and Bulacan provinces. Severe flooding of this river can swallow the whole province of Pampanga and its neighboring provinces. Pampanga River Basin is continuously being monitored by Philippine Atmospheric Geophysical, Astronomical Services Administration (PAGASA) since it is one of the major land reservoir that affects the region.

The use of conventional deterministic models in this case requires extensive data and costs since it considers various physical factors that are not only tedious to collect but are also complex in nature.

Hence, the use of Artificial Neural Networks (ANN) attracted many researchers because of its simple yet powerful architecture to deal with complex nature of problems without sacrificing the results. The black-box nature of this model needs no requirement to fully understand the system in the aggregate (Brath and Toth, 2002). The model, given sufficient data, can be used as a reliable forecasting tool especially in forecasting the flood in Pampanga River Basin.

The use of ANN models as a rainfall-runoff model proves to produce qualitative forecasts (Lekkas et al., 2005); and the rainfall level and actual stream flow (water level) can be actually used as parameters for the development of the model (Elsafi, 2014). This study focuses on the application of ANN as a rainfall-runoff model in Pampanga River Basin with rainfall level and streamflow level (water level) as its parameters.

OBJECTIVE

The objective of this research is to forecast the flooding of the Pampanga River Basin. This will help develop a warning signal that will inform the residents living near the area ahead when to evacuate. In this way, a network will be trained to determine the model that will give the reasonable relationship between the water levels and rainfall in Pampanga River Basin specifically on how the rainfall level can forecast stream flow (water level) downstream.

METHODOLOGY

Study Site

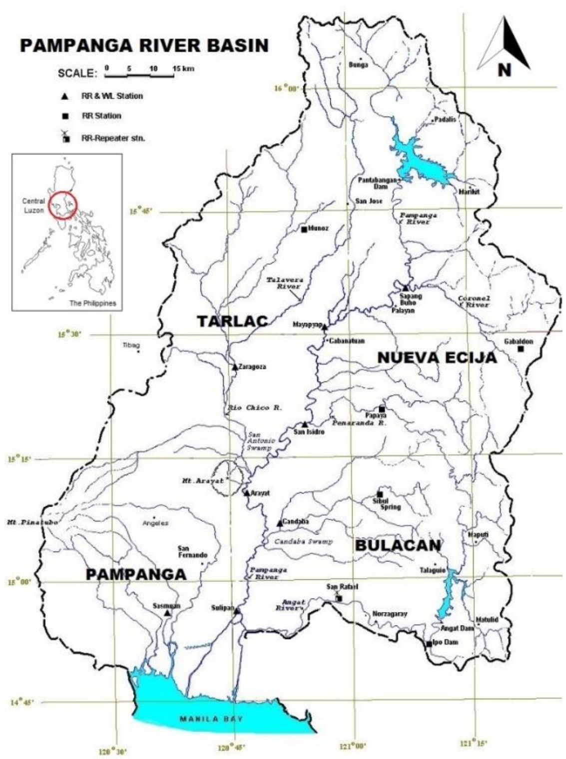

Fig. 1 Hydrological map of Pampanga River Basin, Philippines

The Pampanga River Basin (Fig. 1), the 4th largest basin in the Philippines, has an aggregate area of 10,434 km^2. About ninety five percent (95%) of the basin transcends the bounds of four provinces, namely Nueva Ecija, Tarlac, Pampanga and Bulacan. The remaining five percent is part of provinces including Aurora, Zambales, Rizal, Quezon, Pangasinan, Bataan and Nueva Vizcaya.

Data Preparation

Rainfall and water level data of the Pampanga River Basin for the years 2014 and 2015 were gathered from the Hydrometeorological Division (HMD) in Philippine Atmospheric Geophysical and Astronomical Services Administration (PAGASA). An hourly data from January 2014 to October 2015 of the water level and rainfall reading was provided by the office. Since this study focuses on event-based modelling, cumulative 24-hour rainfall level of the wet season of Philippines including June 2014 – Dec 2014 and from June 2015 – October 2015 were considered.

In the data given, there were seventeen rainfall reading stations, but only thirteen stations were considered since the data from other stations have problems in it including incomplete or erroneous data. The rainfall data from these stations will serve as input data for the model. Nevertheless, for the water level reading stations, there are ten of them but only nine stations were considered. These nine stations will represent the nine sets of models which will serve as basis for the clustering of the data set. Data from these nine stations will represent the output/target data for the model. The input data is, then, normalized using the normalization formula intrinsic in the use of the MATLAB Neural Network Toolbox given in Eq. (1).

$$p' = 2\frac{p - \min(p)}{\max(p) - \min(p)} - 1 \qquad (1)$$

In Eq. (1), p represents the original data before normalization, p' represents the normalized data and; min(p) and max(p) represents the minimum and the maximum data in the set, respectively. The data was split into two computational components: training set and validation set which will be used as calibration for the model developed. Using MATLAB Neural Network Toolbox, the input data can be randomly classified into seventy percent (70%) for the training set and the thirty percent (30%) for the validation set.

Model Development and Implementation

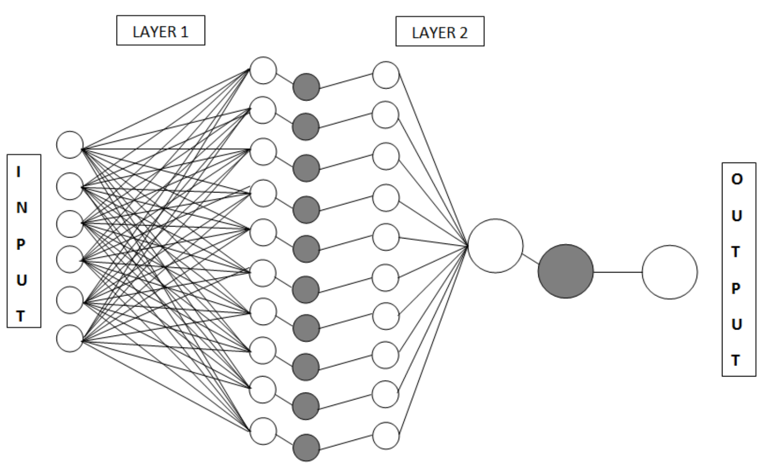

Fig. 2 Network visualization of the architecture of ANN

ANN's are designed to mimic the behavior of a neuron inside a human brain while it sends information (Sarkar and Kumar, 2012; Gershenson, 2003; Zhang and Govindaru, 2000; Chakraborty et al., 1992). It requires inputs (like neuron synapses) which are multiplied by numerical weights (strength of respective signals), and mathematical functions that determine the activation of the neuron. In the end, a function computes for the total output of an artificial neuron. In this study, a *Feedforward Backpropagation* ANN Model was used with fixed network structure consisting of two hidden layers and ten nodes each as shown in Fig. 2.

A hyperbolic tangent sigmoid (*tansig*) function is used in the problem for all the transfer (or activation) functions in the model. The *tansig* function is given by Eq. (2).

$$\operatorname{tansig}(n) = \frac{2}{1+e^{-2n}} - 1 \tag{2}$$

Before the training process starts, several parameters were set. These values will be used in the process and is important to achieve the best possible model using the Mean Square Error (MSE) criterion. Parameters (Table 1) are set accordingly before training the model.

Table 1 Values of the model parameters before the model training

Parameters	Values assigned
Maximum number of epochs to train	1000.00
Maximum validation failures	1000.00
Initial learning rate	0.01
Learning rate increase factor	1.05
Learning rate decrease factor	0.70
Maximum performance increase	1.04
Momentum	0.90
Minimum performance gradient	10^{-10}
Epochs between displays	25.00

Values shown were the pre-set of MATLAB Neural Network Toolbox.

RESULTS AND DISCUSSION

Network Model Assessment

Forty-five (45) models using the set architecture of ten neurons in the first layer and a single node in the last layer for the output were developed after the training process. For each area cluster, there are five models that were developed – each forecasting from three hours water level advance up to seven hours water level advance. In general, most of the models showed good fit in its regression analysis in terms of their forecasting capability and validity. Using the Mean Square Error (MSE) criterion, it measured the average squares of errors or deviation from the predicted water level data to the actual recorded water level data and the Pearson Correlation (R) that defines the strength of linear fit for the ANN trained.

The other four to seven-hour models follow the same behavior as the initial three-hour model for each area cluster. Overall the average square of deviation of all these models ranges from 0.0007 to 0.7624 which is relatively low compared to the actual predicted data. Most of the models showed significantly strong linear relationship (R between values 0.8 to 1) for the predicted data except for the set of Sasmuan area cluster with weak linear relationship (R between values 0.4 to 0.6).

Table 2 MSE and R of the three-hour model for each of the nine (9) area clusters

Clustered Station	Mean Square Error (MSE)	Pearson Correlation (R)
Arayat	0.5832	0.9916
Candaba	0.0060	0.9991
Mexico	0.0019	0.9922
Penaranda	0.0090	0.9962
San Isidro	0.0295	0.9979
Sapang Buho	0.0473	0.9926
Sasmuan	0.0246	0.5448
Sulipan	0.2000	0.9889
Zaragoza	0.0135	0.9948

Data Plots

The actual plot of area model with the highest value of R (three-hour model in Candaba station) shows a clear visual captivity of behavior of the actual recorded water level versus the predicted data in Candaba station of water level as shown in Fig. 3.

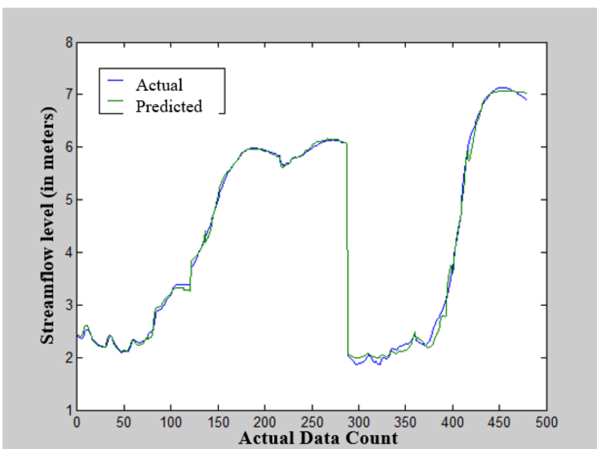

Fig. 3 Plot of target data vs predicted data for three-hour ahead forecasting in Candaba station

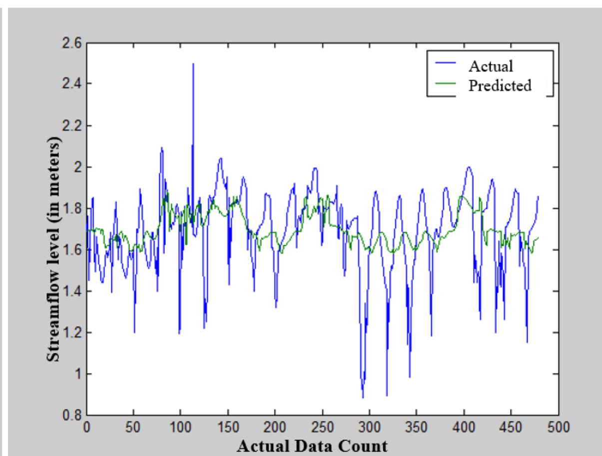

Fig. 4 Plot of target data vs predicted data for five-hour ahead forecasting in Sasmuan station

On the contrary, the model with the lowest value of R (five-hour model in Sasmuan station) showed weak captivity of behavior of the actual water level data (Fig.4). The model cannot linearly capture the trend of the actual data and the actual data shows periodic up and down of plot. This may be attributed to the nonlinear relationship of the rainfall and water level in the area and to the other physical phenomena or factors not considered in this study.

CONCLUSION

The application of Artificial Neural Networks (ANN) to the hydrology in Philippine setting has been explored by creating forty-five (45) forecasting models under nine (9) area clusters in Pampanga river basin. The training of the model resulted in good-fitted models except for the case of the Sasmuan station. Inaccuracy of the model in Sasmuan station may be attributed to the other factors not considered in this study and hence can be further investigated and studied so that the model can be recalibrated. Furthermore, the use of ANN is not governed by theoretical rules nor any rule-of-thumb suggestions that will result in a global optimal model. This side of the ANN can be further explored in its application to hydrology. In addition to this, ANN is a data-dependent method and reasonable amount of data must be used in order to build a much reliable network model.

ACKNOWLEDGEMENTS

The data provided by the Hydrometeorological Division (HMD) of PAGASA thru Engr. Dominic Joshua Lopez is hereby gratefully acknowledged.

REFERENCES

Brath A. and Toth, E. 2002. Flood forecasting using artificial neural networks in black-box and conceptual rainfall-runoff modelling. Proceeding of the 1st Biennial meeting of the International Environmental and Software Society, 166-171.

Chakraborty, K., Mehrota, K., Mohan, C.K. and Ranka, S. 1992. Neural networks and their applications. Review of Scientific Instruments, 65, 1803-1832.

Chen, Y. and Graham, G. 2005. Flood forecasting in real time. Waterworld. [Online]. Available: http://www.waterworld.com/articles/wwi/print/volume-21/issue-4/features/flood-forecasting-in-real-time.html.

Elsafi, S. 2014. Artificial neural networks (ANN's) for flood forecasting at Dongola Station in the River Nile, Sudan. Alexandria Engineering Journal, 53, 655-662.

Gerhenson, C. 2003. Artificial neural networks for beginners (August 2003) [Online]. Available: http://arxiv.org/pdf/cs/0308031

Guha-Sapir, D., Hoyois, P. and Below, R. 2015. Annual disaster statistical review 2014, The numbers and trends. Brussels: CRED [Online]. Available: http://ww.cred.be/sites/default/files/ADSR_2014

Lekkas, D., Onof, C., Lee, M. and Baltas, E. 2004. Application of artificial neural networks for flood forecasting. Global Nest: The International, 6 (3), 205-211.

Sarkar, A. and Kumar, R. 2012. Artificial neural networks for event based rainfall-runoff modeling. Journal of Water Resource and Protection, 4, 891-897.

Zhang B. and Govindaru S. 2000. Prediction of watershed runoff using bayesian concepts and modular neural networks. Water Resources Research, 36 (3), 753-762.