

## River flow modelling for flood prediction using artificial neural network in ungauged Perkerra catchment, Baringo County, Kenya

Sally Jeruto Chebii<sup>a,\*</sup>, Micah Mukungu Mukolwe<sup>b</sup> and Basil Iro Ong'or<sup>c</sup>

<sup>a</sup> MSc Finalist, Civil and Structural Engineering Department, Masinde Muliro University of Science and Technology, Kakamega, KENYA

<sup>b</sup> Dr., Chairman, Department of Civil and Structural Engineering, Masinde Muliro University of Science and Technology, Kakamega, KENYA

<sup>c</sup> Professor, Civil and Structural Engineering Department, Masinde Muliro University of Science and Technology, Kakamega, KENYA

\*Corresponding author. E-mail: sally92.sc@gmail.com

### ABSTRACT

Artificial neural network (ANN) modelling has been applied successfully in hydrology to predict future flows based on previous rainfall-runoff values. For a long time, flooding has been experienced in the surrounding areas of the Rift Valley lakes including Lake Baringo, fed by the River Perkerra, due to the rising water levels because of the above-normal rainfall season, resulting in massive socioeconomic losses. The study aims at predicting the occurrence of floods in River Perkerra using an ANN model with the input data being 417 consistent pairs of daily rainfall and discharge, and simulated runoff as the output. The model was trained, tested and validated producing a best fit regression with  $R^2$  of 0.951 for training, 0.938 for validation, 0.953 for testing giving an average of 0.949 indicating a close relationship between the input and output values. The overall best validation performance, RMSE, was 0.9204 m<sup>3</sup>/s indicating high efficiency of the FFNN model developed to predict floods. Flows greater than 14 m<sup>3</sup>/s, Q1, were the extreme flood events closely associated with socioeconomic losses. This prediction of Q1 value is crucial in the formulation and implementation of measures and policies by the County Government that will mitigate adverse impacts of predicted floods in the catchment.

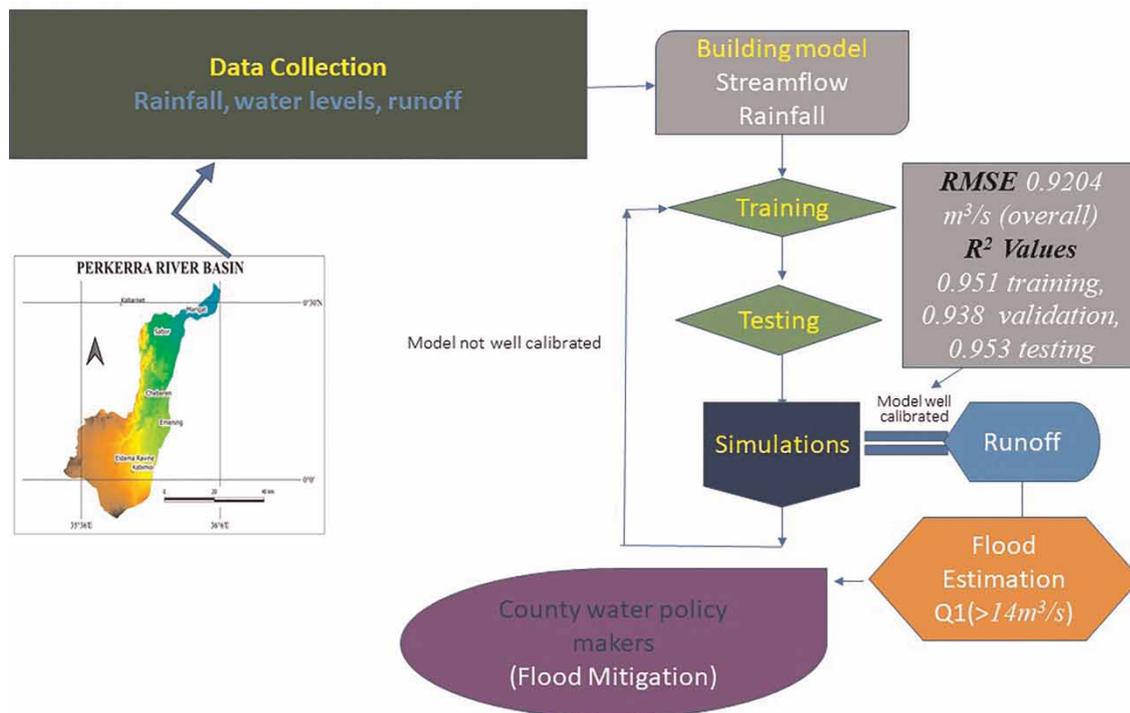
**Key words:** artificial neural network, feed forward neural network, flooding, prediction, rainfall-runoff

### HIGHLIGHTS

- Application of ANN in ungauged catchment to predict floods, which have been a menace for years.
- It will aid the county policy and decision makers in water resources planning and management purposes.
- Creating flooding awareness.
- Aids in the design and maintenance of irrigation infrastructures/hydraulic structures, hence reducing the negative impacts of floods.
- Calls for application of other hydrological models.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY-NC-ND 4.0), which permits copying and redistribution for non-commercial purposes with no derivatives, provided the original work is properly cited (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## GRAPHICAL ABSTRACT



## INTRODUCTION

Flooding is a recurring natural hazard with potentially devastating consequences. Among other natural disasters, flood impact is one of the most significant, threatening human life, agriculture, property and socioeconomic systems in the world resulting in approximately a third of socio-economic losses due to floods (WMO 2011).

Floods occur when a river's flow exceeds its channel flow capacity due to increased surface runoff or breaking of big water retention structures (Roy & Husain 2014). Flooding continues to be a regular threat in various parts of Kenya. The surrounding areas of the Rift Valley lakes including Lake Baringo, fed by the River Perkerra, have experienced rising water levels because of the above-normal 2019 October-December rainfall season (Kenya Meteorological Department 2020) causing economic losses, destruction of infrastructure, disruption of transport and social services, displacement and even deaths of people and livestock.

Although extreme climate events such as flooding can be managed, in the future, damage to infrastructure may be rampant, which in turn increases economic losses. Hence, it is essential to develop a method to deal with associated massive socio-economic losses (Farhadi & Najafzadeh 2021). It is therefore a priority for the Government to plan for a sustainable flood risk management strategy that aids in flood prevention, protection and preparedness to avert the negative impacts of flooding. This is achieved through reliable and accurate prediction of floods.

Flood prediction is essential in water resource planning and the design of flood protection structures in a catchment that experiences periodic flooding (Michael & Patience 2018). Mitigation of damage caused by floods, hazard assessment and extreme event management can be achieved by implementation of flood prediction models that use historical hydrological data as well as catchment characteristics to predict short-term and long-term flood occurrences.

Robust and accurate prediction provides useful information for analysis of a catchment's characteristics, formulation of water policies and strategies with higher efficiency, provision of information for catchment protection, soil and water conservation, design of hydroelectric projects, real-time operation of water resources tasks and design of hydraulic structures to minimize damage impacts on the environment of the constantly changing climatic events (Brown 2009; Anusree & Varghese 2016).

Today, there are a wide range of machine learning (ML) techniques applied in flood modelling, which include empirical black box, event-driven, stochastic, lumped and distributed, continuous, deterministic and hybrid

models (Mosavi *et al.* 2018a, 2018b). There have been continuous advancements of these techniques over the past two decades. They are mainly data-specific and involve various simplified assumptions (Lohani Goel & Bhatia 2014). They are reported to perform better than conventional methods in flood prediction. A study by Ortiz-García *et al.* (2014) indicates the efficiency of ML techniques in modelling complex hydrological systems including floods by applying specific techniques.

Machine learning defines a subfield of artificial intelligence (AI) that deals with the designing and developing of algorithms necessary to enhance the performance of computers (machines) over time, based on data (Mahdavinejad *et al.* 2018). techniques detects efficiently general flood patterns and variations through learning of hydro-meteorological parameters and measured climatic indices of a catchment. The most common flood frequency analysis methods includes multiple linear regression (MLR) (Adamowski *et al.* 2012), autoregressive moving average (ARMA) and autoregressive integrated moving average (ARIMA) (Valipour *et al.* 2012) among other early statistical models.

The most frequently used AI techniques to predict floods with greater accuracy in comparison to traditional statistical models (Xu & Li 2002) include K-nearest-neighbour (K-NN) methods (Kramer 2013), artificial neural networks (ANNs) (Lafdani *et al.* 2013), genetic programming (GP) (Liong Khu & Babovic 2007), adaptive neurofuzzy inference system (ANFIS) (Ren *et al.* (2013) and support vector machines (SVMs) (Araghinejad 2013). They are effective for both short-term and long-term flood predictions (Mosavi *et al.* 2018a, 2018b).

ML methods are as good as their training, hence relying on the quality of the input data for learning of the target task. Analysis of the data characterising the system under study using computational intelligence methods and data-driven modelling is crucial in solving problems related to river basin management. They perform well when the data is robust, covering varieties of the task, with minimum gaps and of sufficient quantity. However, it is practically impossible to recommend one particular type of data-driven model to solve river basin management problems since the data for water-related applications are characterized as noisy and of poor quality (Solomatine & Ostfeld 2008).

Some robust AI techniques such as GP have been widely used in gauged catchments with reliable data to predict floods giving a reliable mathematical expression for recognizing behavioural patterns of complex systems. The input variables with low level of contribution to the performance of the model are automatically reduced.

Additionally, GP in flood forecasting requires reliable input parameters and provides functional relationships which can be analyzed and interpreted. This is unlike physically based and conceptual models, which require many parameters for development and calibration (Liong Khu & Babovic 2007). This method is suitable for gauged catchments with reliable datasets. For the perkerra catchment, with unreliable data, this method will not be suitable.

Incomplete historic rainfall data recorded using rain gauged are available. Due to several gaps in the hydrological data sets from the River Perkerra catchment, the study area is referred to as ungauged. Lack of input data remains a challenge in modelling of river flow in ungauged catchments using hydrologic models, which relies on the quality and availability of input data and other hydrologic parameters, and the model structure (Meresa *et al.* 2016; Hadush 2018). Previous studies in the ungauged Perkerra river catchment on run-off prediction are not available. However, several countries have experienced rising interest for the past decade in understanding the hydrological processes in runoff estimation in ungauged catchments by hydrologists, water resources planners and policy makers (Blöschl 2005; Laaha & Blöschl 2007; He *et al.* 2011).

The most popular learning algorithms among all the ML techniques are ANNs. They gained popularity in modelling flood prediction from 1990s (Wu & Chau 2010). They can mimic flow observations in hydrology, without any mathematical descriptions of the physical processes being studied, hence, increasing the applicability of black box-based-AI models in ungauged catchments in comparison with formulation-based AI models such as GP.

The fact that ANNs provide a quick and flexible approach for data integration and development give them an advantage over other networks (Elsafi 2014). They approximate measurable functions up to a random degree of accuracy, hence they are appropriate for this research (Parasuraman 2007). A single 'good' result is normally selected as the ultimate result without much explanation of the initial weight parameter (Kim & Seo 2015).

Computational modeling is not necessary when using ANN technique because ANN imitates and memorizes whichever complicated association between outputs and inputs by learning and training using past data, and also perform the computational analysis by relating to the inputs. ANN techniques have relatively low computational demands as well as integrate with other approaches easily.

There is continual research in the field of prediction of river flow to increase the accuracy. ANN models may be used to increase the accuracy of the prediction. A comparison of ANN model with a traditional statistical model such as auto-regression suggests that ANN provides better accuracy in prediction (Tsakiri Marsellos & Zurbenko 2014).

Some of the successful applications of ANN with a satisfying accuracy of output despite absence of accurate information about the physical mechanisms underlying the river flow in a catchment are predictions of river flow, water quality parameters (precipitation, suspended sediment, stream flow etc.), rainfall-runoff process, forecasting of rainfall-runoff, modelling of event-based rainfall-runoff, as well as for characterization of soil pollution (Nagesh Kumar *et al.* 2004; Riad *et al.* 2004; Diamantopoulou *et al.* 2005; Sarkar & Kumar 2012; Singh *et al.* 2014; Aichouri *et al.* 2015; Kostić *et al.* 2016).

From the literature review, a comparative study between autoregressive models (ARMA and ARIMA) and ANN on the ability to simulate runoff in a catchment was done. ANN models gave more accurate improvements than autoregressive models, hence the reason ANN is chosen for this study for flood prediction over the statistical techniques (Shamseldin & O'Connor 2001; Brath *et al.* 2002). Asati & Rathore (2012) and Nasri *et al.* (2008) developed a non-linear relationship between rainfall and runoff using ANN, MLR and an autoregressive model with results indicating that ANNs consistently gave superior predictions.

Dibike (2002) states that the use of SVM in hydraulics and hydrology offers an attractive technique of data modeling; however, it has a slow response as well as high computational demands. Handling a large number of mediums sustenance vectors (that consist of a series taking a greater length of time as compared to those in an ANN approach) is a challenge. Unlike SVM, the ANN approach is likely to result in best performance when predicting activities that are within the scope of the training information as there is no technique check of the final model.

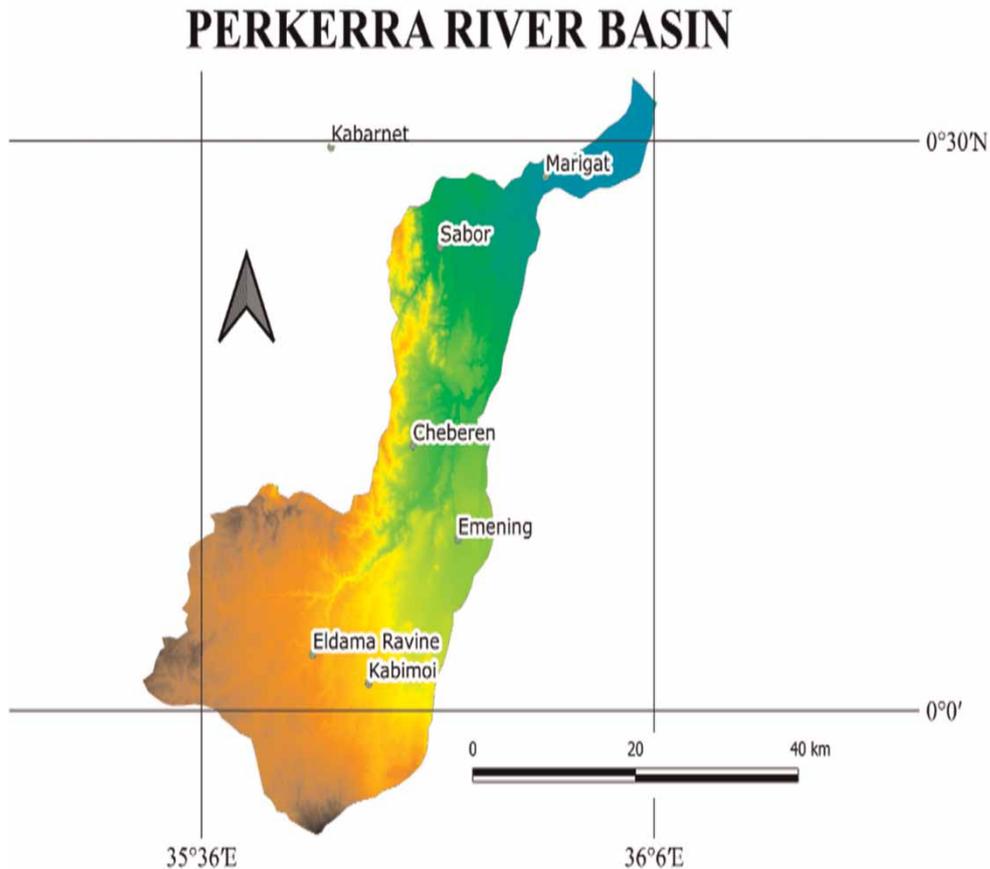
The K-NN technique is not attempting to isolate an output/input mapping function and therefore has no capability to extrapolate an unknown input vector into the future. Contrasting this, additional non-linear modelling, including ANNs, attempts the identification mapping function to output from the input and have the ability to extrapolate.

Ren *et al.* (2013) experimented on a monthly runoff prediction with ANFIS and wavelet examination, in which a wavelet analysis technique was employed and determined that it could be used for prediction but further development is needed.

In the current research, the performance of ANNs is examined for prediction of floods of the River Perkerra, Baringo County, Kenya, where rain gauges were used. Spatial flood prediction for identification of flood location was not part of the scope. Intelligent sensing systems for data collection are not available for this catchment, hence this study covers only riverine flooding and not flash floods. The prediction efficiency levels of ANNs are evaluated by various statistical benchmarks. Using a ANN approach to estimate runoff in this catchment is very innovative and original in terms of methodology, study area and framework approach. Also, ANNs can be more easily accepted by decision-makers due to their reliance on simple linear models (Solomatine & Ostfeld 2008). Hence, it is necessary to predict floods in the catchment to mitigate the adverse impacts of floods in the county.

## OVERVIEW OF CASE STUDY

This study was conducted in the River Perkerra, located in Baringo County, Kenya. The Perkerra River watershed was delineated and the map presented in Figure 1. It covers an approximate area of 1,314.93 sq km, draining Baringo central, Baringo south and Eldama Ravine sub-counties, covering at longest a length of 95 km. It serves as the main water source for Perkerra Irrigation Scheme, which was started in 1952 and aims at boosting productivity and increasing food production of the semi-arid areas in Kenya (Keitany 2016). The river originates from the Tugen hills and drains to Lake Baringo. The annual rainfall around the Tugen hills ranges between 1,100 and 2,700 mm and around Lake Baringo between 500 and 750 mm. Temperatures are high in low altitude areas and vice versa. The top part of the Tugen hills (upper catchment) has very steep slopes and a high altitude of 2,400 m above the sea level while the lower catchment is a semi-arid area with a low altitude of 980 m above sea level, characterized by scattered shrubs and little undergrowth. The average gradient of the basin is 8.9%. The river has faced a great sediment load resulting from increased rate of soil erosion and siltation. This has caused a shallowing effect from a depth of 8 m in 1972 to 2.5 m in 2008 (Akivaga *et al.* 2010). The land use is



**Figure 1** | River Perkerra catchment.

mainly forest with little or no undergrowth, cultivated lands and few settlement points in rural and urban centres such as Kabarnet, Marigat and Eldama Ravine. For the past two decades, anthropogenic activities such as deforestation, cultivation of lands and settlements within the catchment have resulted in frequent floods.

In the Tugen hills and in areas where farmers practise soil conservation, the soils are highly productive with a higher content of organic matter. The soils at lower elevation areas around the Perkerra Irrigation Scheme and Lake Baringo are not so leached as the soils on the uplands. They are calcareous and saline in some areas since the rainfall is scarce and temperatures are perennially high in low altitude areas. The area is prone to severe sheet and gully erosions and characterized by mostly unproductive land since in some areas. An entire surface layer of soil has been removed resulting in a badland with a major contribution from frequent floods in the area (Sketchley *et al.* 1978).

Recently, incidences of flooding were reported in various parts of the country. Water levels rose in Lake Victoria, Lake Baringo, Lake Bogoria and other Rift Valley lakes. The water levels are also likely to remain high, causing flooding in the surrounding areas, whereby families in Baringo south were left homeless after their houses were submerged in water. Some infrastructure such as hotels worth millions are submerged because the catchments feeding the lakes have continued to receive above-normal rainfall from the 2019 October-November-December rainfall season to date (Kenya Meteorological Department 2020).

To curb the occurrence of numerous floods causing massive socio-economic effects, there is a need to predict the occurrence of floods from Perkerra River to protect lives and livelihoods.

## METHODOLOGY OF THE RESEARCH

### Choice of model parameters

Maier *et al.* (2010) stated that there are various catchment characteristics that have an impact in prediction of floods, for example, rainfall, streamflow, groundwater level, rainfall stage, water level, river flood, soil moisture,

rainfall-runoff, precipitation, daily flows, river inflow, peak flow and extreme flow. A study by Lafdani *et al.* (2013) indicates that among the listed key influencing flood resource variables, rainfall and the spatial study of the hydrologic cycle had the most remarkable role in runoff and flood modelling. On this note, traditionally in short-term flood prediction, quantitative rainfall prediction is used (Collier 2007). High-resolution precipitation forecasting is essential, other flood resource variables were considered in the (Grecu & Krajewski 2000). However, for accurate long-term flood prediction, rainfall prediction was reported to be inadequate. In addition, rainfall and soil moisture estimates are crucial in streamflow prediction in a catchment (Seo & Breidenbach 2002). Thus, the methodology of this literature review aims to include rainfall and riverflow parameters in rainfall-runoff modelling for floods prediction.

### Collection of model input data

The model input parameters include the events and occasions of rainfall and observed runoff data from April 2005 to August 2007, giving a large set of data of up to 417 pairs of rainfall and runoff values. The runoff generated stood as the output parameter. The daily rainfall data of the period between April 2005 and August 2007 was used for this study because they were highly consistent with minimum gaps and give a better representation of the current state of the catchment.

The rainfall data was obtained from the meteorological department in Kabarnet town located within the Perkerra catchment boundaries. The observed stream flow data was obtained from the Water Resources Management Authority (WRMA), Nakuru regional office. Marigat monitoring site (upstream gauging site) demonstrates the river flow at the recording station especially near the irrigation site and it is the only existing site for gauging the River Perkerra.

### Data partitioning

In neural networks, the data are partitioned into training/learning sets, test sets and validation sets. The training/learning set is used to determine the adjusted weights and biases of a network. A good training data set generally contains all possible minimum and maximum values in the training set from the available data series. It also contains the bulk of the data. The test set consists of a representative data set used for calibration and to prevent overtraining of the networks. For development of ANN model, the observed data were used 80% for training, 10% for testing and 10% for validation.

### Estimation of missing rainfall data

For this study, the arithmetic averaging (AA) method (Equation 1) was used to estimate the missing precipitation since the computational results of the study done by Sattari *et al.* (2017) demonstrated that among classical statistical methods, the AA method performed best.

$$P_x = \frac{1}{n} \sum_{i=1}^n P_i \quad (1)$$

where  $n$  is the number of nearby stations,  $P_i$  is precipitation at the  $i^{\text{th}}$  station and  $P_x$  is missing precipitation.

### Implementation of ANN

The ANN comprises many processing elements with simple neuron-like and numerous connections weighted between the elements, and it can have one layer, double layers or multiple layers depending on the number of layers. It can be categorized as feed-forward, recurrent and self-organizing based on the direction of flow of information and processing (Govindaraju 2000).

A wide range of network architectures are suitable for use in addressing different problems. Different types of feed-forward networks (FFNNs) include the single layer FFNN multi-layer FFNN and radial-basis function networks (Augusto & Nugent 2006).

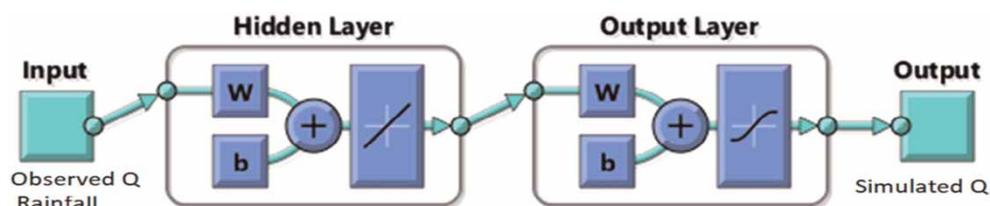
Among these different architectures, the multilayer FFNNs, which have an input layer, many hidden layers and an output layer, have been widely utilized. However, the FFNN architecture with a single hidden layer was used in this study because studies have determined that having more than one hidden layer yields no significant improvement in performance of a network with a single hidden layer (Dawson *et al.* 2002). Also, in a study conducted by Dogan *et al.* (2007) on forecasting of streamflow comparing FFNN, RNN and radial basis neural

networks (RBNN) determined that the FFNN model yields the best result. Networks with fewer hidden nodes usually have better generalization capabilities and fewer over fitting problems hence are generally preferable.

Use of historical rainfall and water level is crucial in creating the ANNs' prediction model. The rainfall data with various return periods recorded with a rain gauge at Kabarnet meteorological station was used. Seo & Breidenbach (2002) stated that ground rain gauges and remote sensing with radars or satellites can be used to measure rainfall. Grecu & Krajewski (2000) emphasized the importance of remote sensing in capturing higher-resolution data in real time, which is more reliable compared to rain gauges, hence resulting in a prediction model with higher accuracy (Maddox *et al.* 2002).

Campolo Andreussi & Soldati (1999); Gizaw & Gan (2016) described in detail the principle and strategy for flood modelling using ML. To construct and evaluate the learning prediction models, individual sets of historical data are used for training/validation, verification and testing.

The architectural design of the FFNN model developed for this study consisted of three layers, which are an input layer, a hidden layer and an output layer as illustrated in Figure 2.

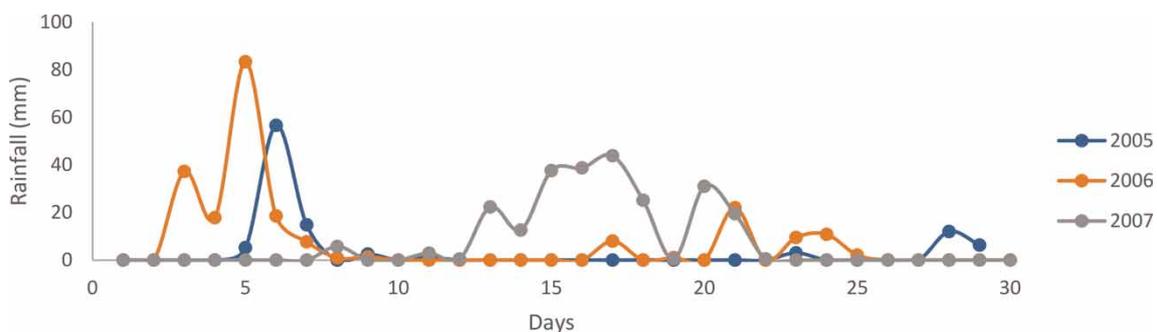


**Figure 2** | The design architecture of the three-layered multilayer FFNN.

### Model evaluation

The coefficient of correlation ( $R^2$ ), root mean squared error (RMSE), mean absolute error (MAE), relative root mean squared error (RRMSE) and Nash-Sutcliffe (NS) efficiency coefficient are the standard statistical performance evaluation criteria used to evaluate the performance of the model. The methodology of those criteria calculation references Chang *et al.* (2015) and Yaseen *et al.* (2015). These coefficients are independent of the scale of the data used and are useful in assessing the goodness of fit of the model (Dawson *et al.* 2002). In addition to the above statistical evaluation criteria (in this study, majorly  $R^2$  and RMSE are represented), interpretations of graphs and scatter plots of the simulated values versus the target values were used to evaluate the performance of the model.

The two popular statistical performance measures, RMSE and  $R^2$ , were used to evaluate the performance of the model for this study.  $R^2$  measures the proportion of the total variance of the observed data as explained by the predicted data while RMSE measures the concentration of data points and their standard deviation from the regression line to measure forecast exactness (it is a good overall performance indicator that punishes a model for not approximating peaks). The correlation of rainfall and runoff was obtained by performing a linear regression between the ANN's predicted values and the targets.



**Figure 3** | Rainfall for the month of April.

## RESULTS AND DISCUSSIONS

### Rainfall

Graphs of daily rainfall data for the months under study are shown in Figures 3–7.

As shown in Figures 3–7, the highest amount of rainfall recorded was 85 mm in the month of April 2007. Similarly high rainfall values in April and November were observed in the catchment during subsequent years 2008, 2015 and 2020. Generally, study of rainfall patterns helps to understand the possibilities of flood occurrences, hence mitigating irreparable damage to agricultural land and produce, damage to infrastructural facilities, loss of life and livestock.

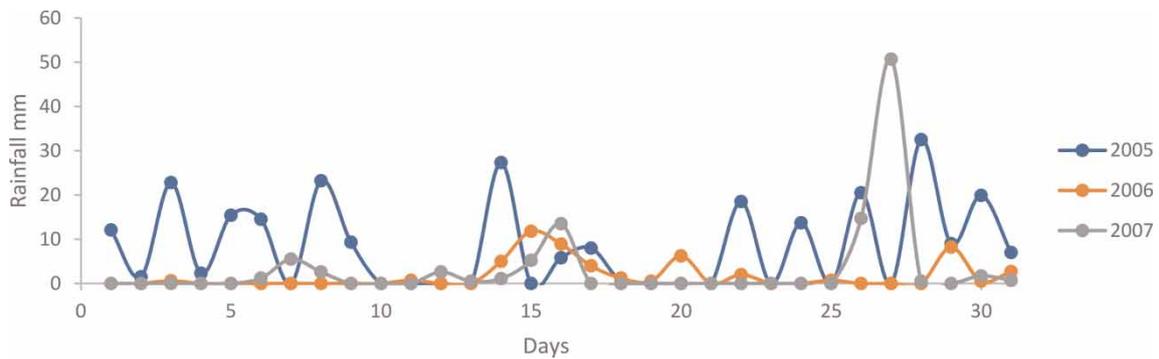


Figure 4 | Rainfall for the month of May.

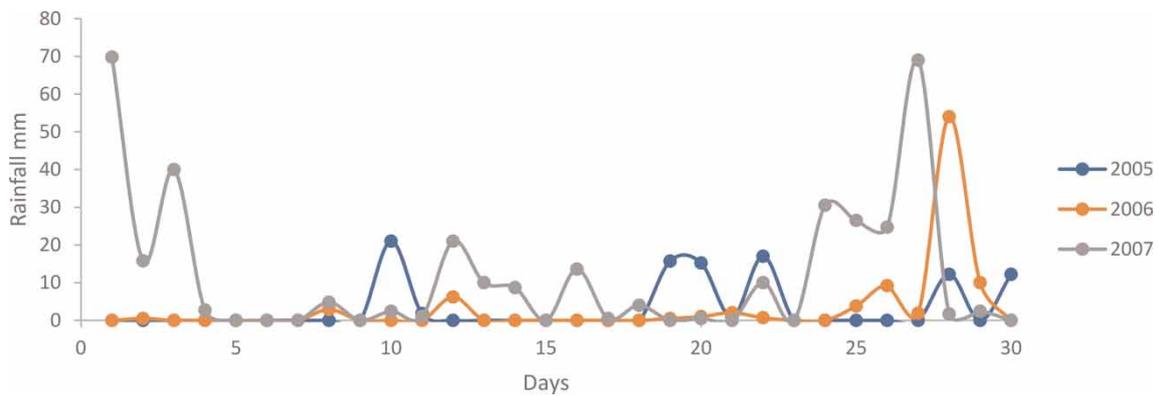


Figure 5 | Rainfall for the month of June.

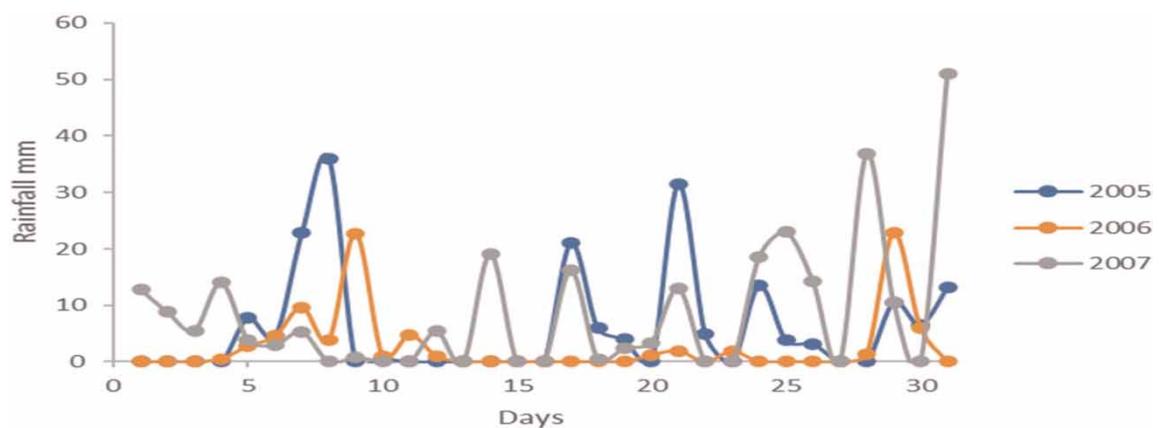
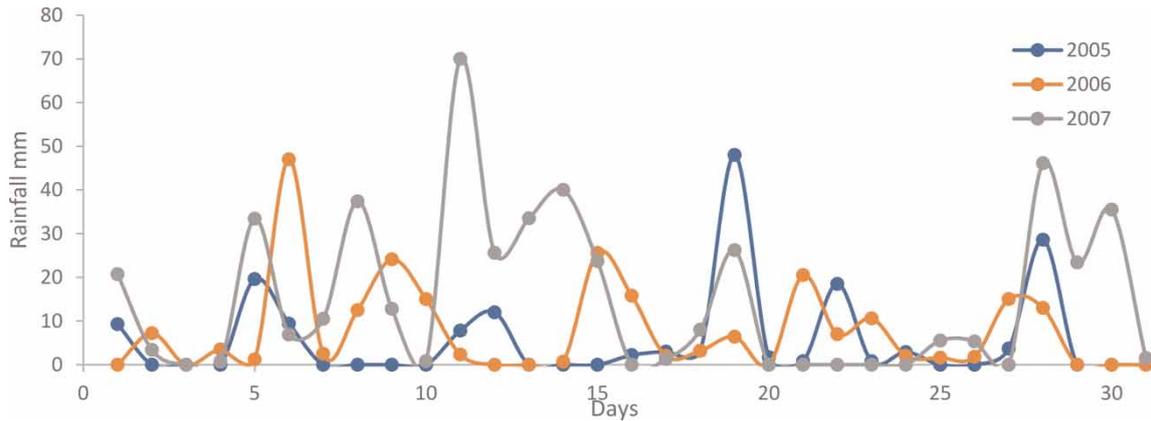


Figure 6 | Rainfall for the month of July.



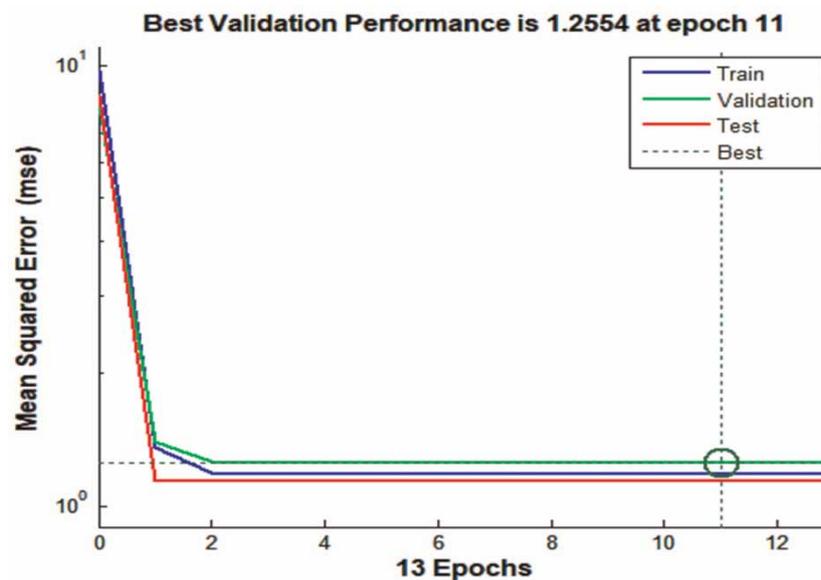
**Figure 7** | Rainfall for the month of August.

### Assessment of model performance

The FFNN model was used with variations of associated training rules until the highest level of efficiency of the model to predict runoff was achieved.

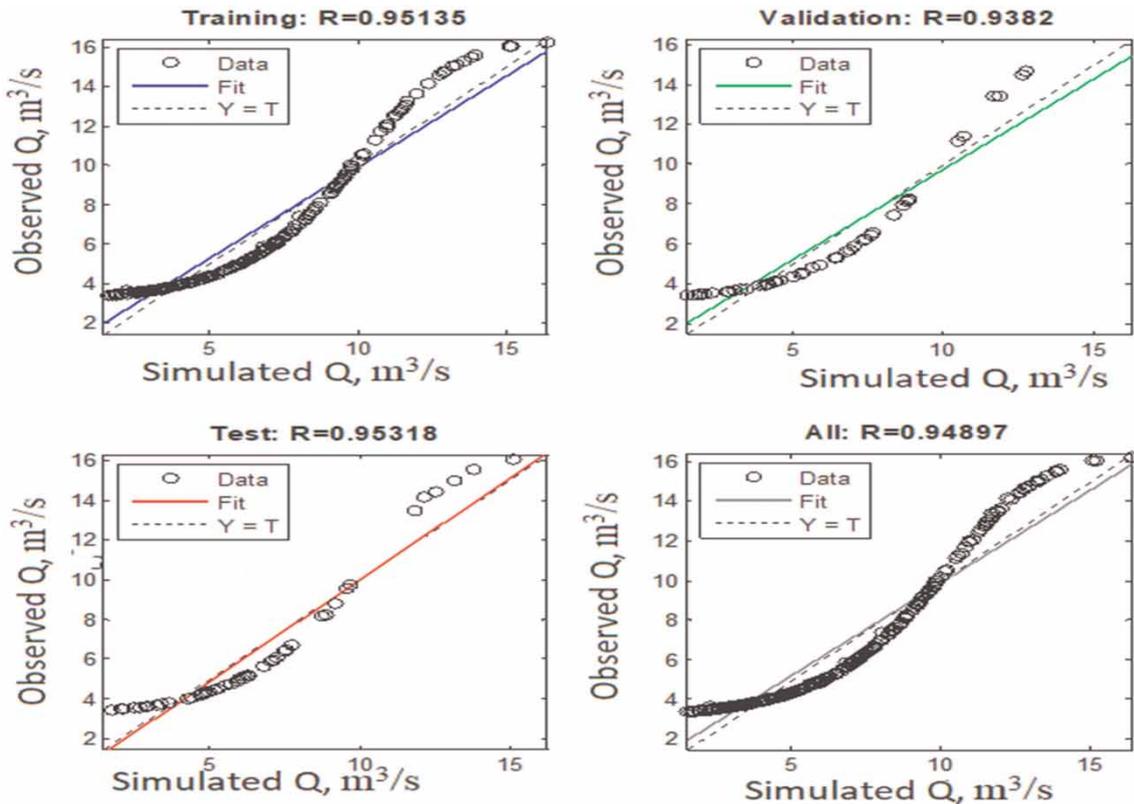
*The model accuracy.* Data pre/post-processing are proven to be an efficient way to improve the prediction precision in most models (Alizadeh 2018). The training of the FFNN was done using the two input variables (observed  $Q$  and rainfall) against only one output, the river flow. Variations in the number of hidden nodes, standardization and for different numbers of epochs were put into consideration during training.

Figure 8 shows the results of the ANN with 10 hidden nodes, trained for 11 epochs, with input and output values standardized by range. The RMSE was calculated as  $0.9204 \text{ m}^3/\text{s}$ . The low value of RMSE shows that the hidden layer of the FFNN was able to learn the pattern of training resulting in a better predictive accuracy of the developed model.



**Figure 8** | The model accuracy plot.

*Coefficient of correlation  $R^2$ .* The strength of a linear relationship between the values that have been predicted and those observed is shown by the correlation coefficient plot Figure 9. Analyzing the linear scatter plots visually, the plots for the training, validation and testing period are very well distributed around the 1:1 line, i.e., the regression line represents all the plotted data, hence a better goodness of fit for the observations such that if more samples were added, they are more likely to fall within the line.



**Figure 9** | Perkerra catchment regression plot.

The model produced a best fit regression with a coefficient of correlation of 0.951 for training, 0.938 for validation and 0.953 for testing. With the average correlation coefficient of 0.949, it indicates that there is a close relationship between the input and output values, hence ANN was very good for rainfall runoff simulation for flood prediction since according to Singh *et al.* (2014), a high predictive accuracy is indicated by a correlation coefficient value greater than 0.67, a moderate effect is indicated by a range of 0.33–0.67, a low effect is a value between 0.19 and 0.33 while a value below 0.19 is considered unacceptable.

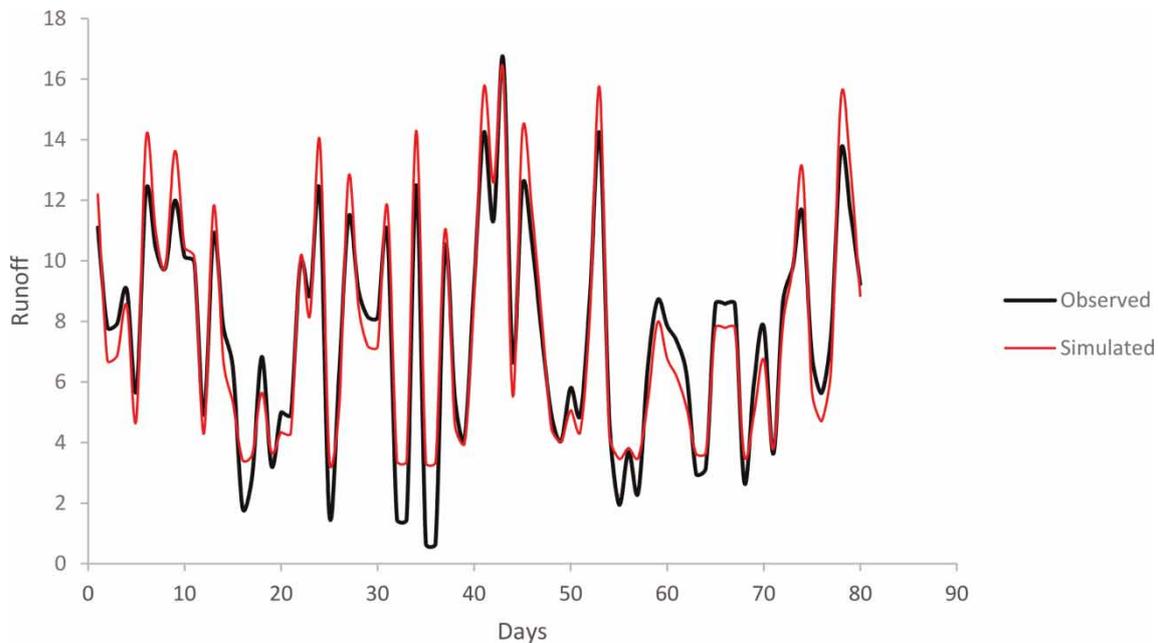
*Comparison between observed and simulated flows.* A good performance measure is essential in evaluation of complex hydrographs. There is no universally accepted measure of ANN skill (Dawson *et al.* 2002). The good visual evaluation of the graph shows the general appropriateness of the ANN for the prediction of long-term streamflow for ungauged catchments, rather than relying blindly on performance measures. The ANN reproduced the largest flow (when the highest rainfall amount was recorded at 87 mm) and intervening low flows.

From Figure 10, the FFNN model was able to predict the runoff values with high accuracy.

The comparison of present results with the previous literature was made efficiently in terms of the accuracy level of predictive tool. The statistical performance of the ANN model demonstrated a permissible accuracy level (RMSE = 0.9204 and  $R^2 = 0.949$ ). When compared to the investigation of Kanda *et al.* (2016), which aimed to assess the ability of ANN to predict dissolved oxygen as an indicator of water pollution in River Nzoia in Kenya, the results obtained during training and testing were acceptable with  $R^2$  varying from 0.79 to 0.94 and RMSE values ranging from 0.34 to 0.64 mg/l implying that an ANN can be used as a monitoring tool considering the non-correlational relationship of the input and output variables.

In the Kim & Seo (2015) study, streamflow forecasting in the Nakdong River, Korea was done using an ANN where management of water use and flood control has been an important national project. The performance of the ANN was evaluated using coefficient of correlation and RMSE. According to the evaluation results, ANN produced more accurate and reliable forecasts than other models with several combinations of input variables. Coefficient of correlation results in the validation were 0.92, 0.95 and 0.97 for the three gauging stations.

Kotlarski *et al.* (2012) studied river flow modelling using an ANN in the Elbe River in the Czech Republic. Statistical analysis of August 2002 flooding was done yielding a value of coefficient of determination ( $R^2$ ) of 0.56.



**Figure 10** | Plot of a comparison between the ANN simulated flows vs the observed daily streamflow over a period of April-August 2005, April-August 2006 and April-August 2007.

Therefore, by performing analyses and examining the performance of the hydrological model, these studies have shown that ANN is useful as a substitute to the traditional methods including the empirical and statistical methods in the evaluation of the diverse physical theories including flood prediction.

### FLOOD PREDICTION

The integrated hydrologic responses of the Perkerra catchment during the 2004 to 2010 period has been reflected in the flow duration curve (FDC), [Figure 11](#), plotted from the simulated streamflow data against percentile flow exceedance percentages. Flood prediction performance of the FDC can be evaluated by drawing an FDC using serially correlated data. However, to stochastically evaluate the severity of high, ordinary and low flow regimes of streams, an FDC was drawn using an arbitrary flow return period for successive years at suitable time intervals rather than drawing an FDC using serially correlated data ([Sugiyama et al. 2003](#)).

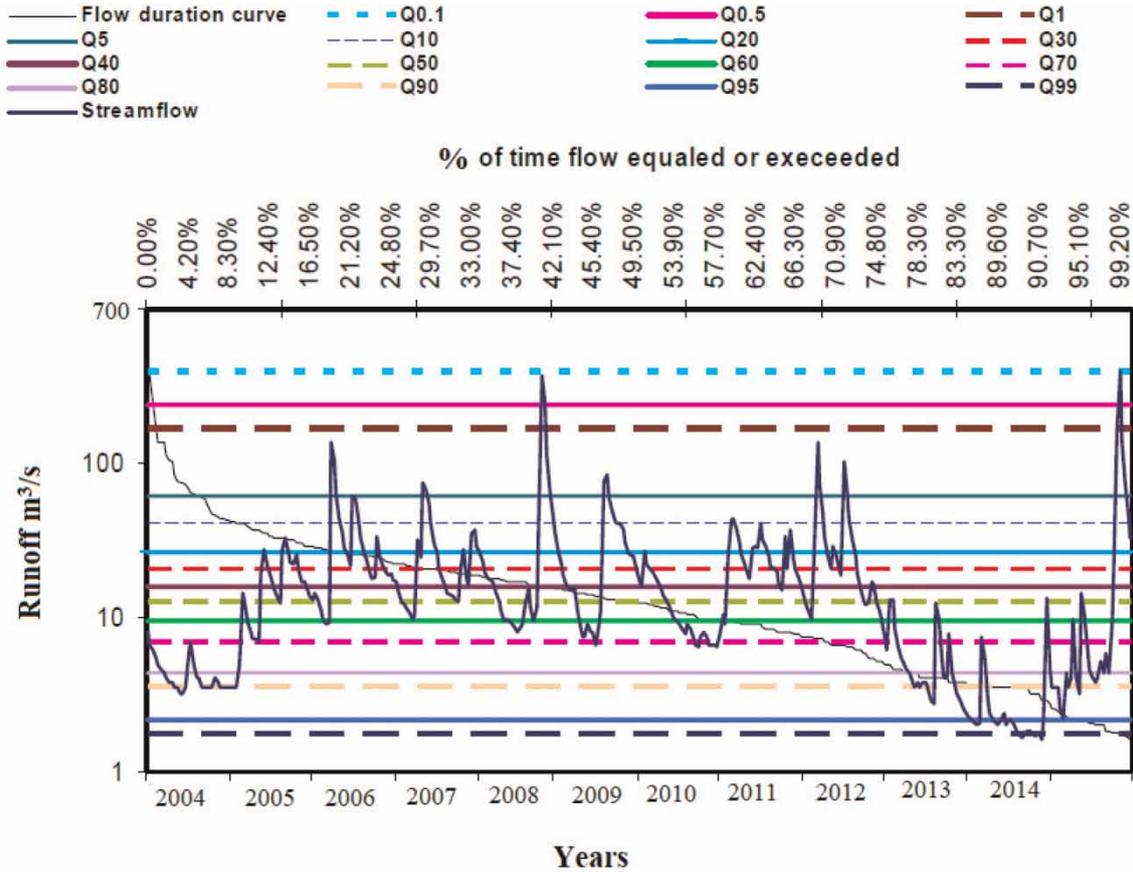
From [Figure 12](#), flow rates between  $Q_1$  and  $Q_{10}$  are considered high flow rates and are most often caused by short-term intense rainfall which normally occurs in the months of April and November. Flow greater than  $14 \text{ m}^3/\text{s}$  would be extreme flood events. With a return period of 1 year,  $Q_{10}$  is equalled or exceeded 10% of the time, 90% of the time the flows are less than this value.

$Q_{45}$  is the  $Q_{\text{mean}}$  ( $7.66 \text{ m}^3/\text{s}$ ). It is equalled or exceeded 45% of the time, and 55% of the time the flows are less than this value.

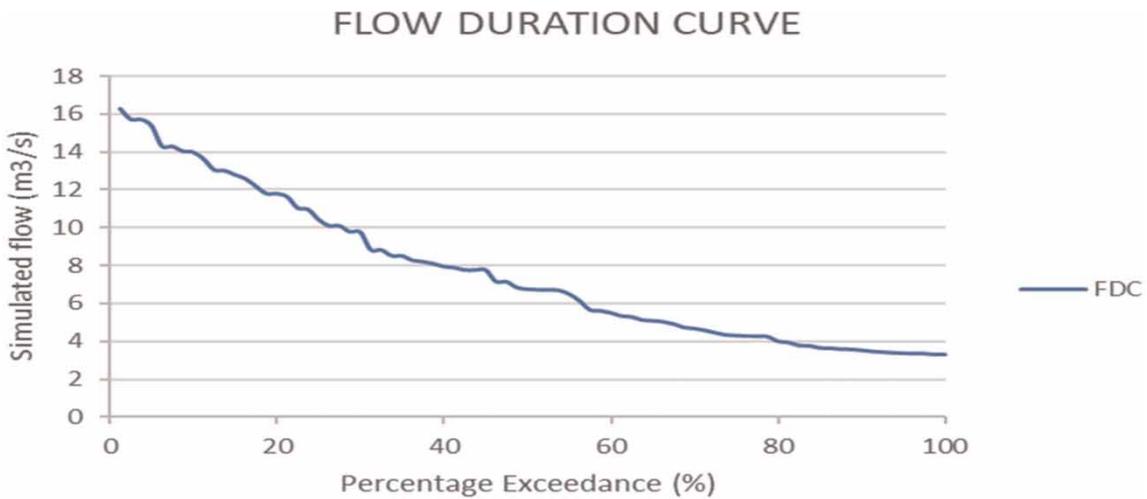
Flows from  $Q_{20}$  and  $Q_{70}$  ( $14 \text{ m}^3/\text{s}$  to  $5 \text{ m}^3/\text{s}$ ) are the median flows for River Perkerra. The Perkerra Irrigation Scheme will draw water from the river efficiently right across these flow rates.

Flow rates between  $Q_{80}$  and  $Q_{99}$  are the 'low flows' in the river, and as you move further to the right on the FDC, the scheme will not benefit from the river due to low flows. As flow rates move from  $Q_{95}$  towards  $Q_{99}$  you move into the low-flow drought flows and the river will eventually dry up.

Based on the estimates of recurrence intervals of floods, flood frequency analysis aids in designing of structures for flood protection such as waterworks, sewage disposal plants, and infrastructures design such as dams, bridges, culverts, levees, and highways. The flood frequency graph gives the estimate for the design flow values corresponding to specific return periods. As indicated in [Figure 13](#), the estimated floods in Perkerra catchment for 1-, 2-, 5-, 10- and 50-year return period has been found from the ANN as 24, 61.57, 128, 159.47, 234.88 and  $476.17 \text{ m}^3/\text{s}$  respectively.



**Figure 11** | Streamflow, duration curve and percentile flows drawn using the daily discharge values observed for Perkerra between 2004–2014. (streamflow data obtained from Water Resources Management Authority, Nakuru Regional office).



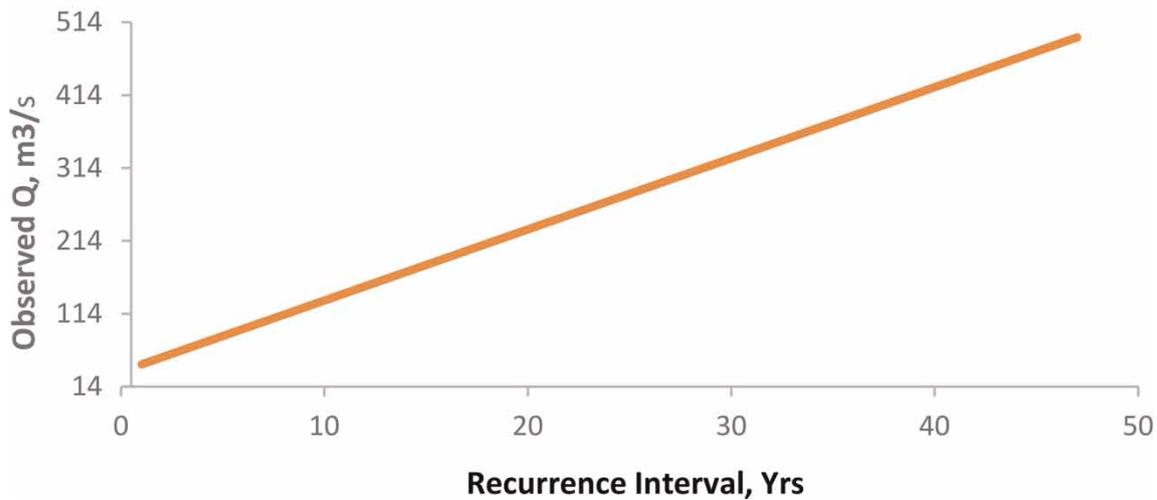
**Figure 12** | FDC drawn using the daily discharge values simulated for Perkerra for the period April-August 2005, April-August 2006, and April-August 2007.

## CONCLUSION AND RECOMMENDATION

### Conclusion

The aim of this study was to predict floods in the Perkerra catchment using ANN. First, the hydrological modeling approach were evaluated and compared with respect to observed runoff in the ungauged Perkerra river catchment. The scatter plots for the training, validation and testing period produced a best fit regression with

## FLOOD FREQUENCY CURVE



**Figure 13** | Flood frequency curve.

an average coefficient of correlation of 0.949 implying a close relationship between the input and output values. The RMSE was calculated as 0.9204 m<sup>3</sup>/s, showing that the hidden layer of the model was able to learn the pattern of training resulting in a better predictive accuracy of the developed model. Without the need for resource-intensive and complex additional data in a catchment, an ANN can be used to accurately and reliably predict river runoff to solve a typical water problem in many other hydrological fields. It provides useful information on possible floods in the Perkerra catchment while helping in understanding the river flow patterns to enhance sustainment of the Perkerra Irrigation Scheme during critical flow levels of the river.

Finally, flood frequency analysis was performed using the 10 years flood series. Analyzing the predicted discharge statistically, the extreme flood events in the catchment were determined and at least twice every year, flooding of different magnitudes has been experienced. Flow rates between  $Q_1$  and  $Q_{10}$  are considered high flow rates, which are closely associated with short-term rainfall.  $Q_1$  (flow greater than 14 m<sup>3</sup>/s) is the extreme flood event, which causes irrigation infrastructure damages, loss of lives and livelihood. With a return period of 5 years,  $Q_1$  is equalled or exceeded 5% of the time, 95% of the time the flows are less than this value. The extreme floods of 2009 occurred in 2015 and 2020 and is also likely to occur in 2025. Therefore, this flood magnitude ( $Q_1$ ) needs to be considered by the county policy and decision makers for planning purposes, creating flooding awareness, design and maintenance of irrigation infrastructures/hydraulic structures, etc. hence reducing the negative impacts of floods.

### Recommendation

Understanding and modelling of ungauged river catchments such as Perkerra using an ANN provides crucial information for analysis of water resource management and planning. Use of hydrological models such as ANN to predict floods can be used by the county policy and decision makers as a possible measure to mitigate flood damages. Through flooding awareness, water resource managers and policy makers can prepare effectively by making informed decisions like designing of an off-stream reservoir for storing excess water during high flows for sustainability of Perkerra irrigation during low flows and aiding in flood protection.

Daily rainfall and runoff data was used in this study as an input in the flood prediction model. Consideration of other input variables or combination of other influential variables with different prediction time frames of few hours, or short time intervals of days will be interesting to study the extent of influence on the flood prediction model.

### DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

## REFERENCES

- Adamowski, J., Chan, H. F., Prasher, S. O., Ozga-Zielinski, B. & Sliusarieva, A. 2012 Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. *Water Resources Research* **48**(1), 1–14. doi: 10.1029/2010WR009945.
- Aichouri, I., Hani, A., Bougherira, N., Djabri, L., Chaffai, H. & Lallahem, S. 2015 River flow model using artificial neural networks. *Energy Procedia* **74**, 1007–1014.
- Akivaga, E. M., Otieno, F. A. O., Kipkorir, E. C., Kibiyi, J. & Shitote, S. 2010 Impact of introducing reserve flows on abstractive uses in water stressed catchment in Kenya: application of WEAP21 model. *International Journal of Physical Sciences* **5**(16), 2441–2449.
- Alizadeh, J. F. 2018 A hybrid linear – nonlinear approach to predict the monthly rainfall over the Urmia Lake watershed using wavelet- SARIMAX-LSSVM conjugated model. *Journal of Hydroinformatics* 246–262. doi: 10.2166/hydro.2017.013.
- Anusree, K. & Varghese, K. O. 2016 Streamflow prediction of karuvannur river basin using ANFIS, ANN and MNLN models. *Procedia Technology* **24**, 101–108. Elsevier B.V. doi:10.1016/j.protcy.2016.05.015.
- Araghinejad, S. 2014 *Data-Driven Modeling: Using MATLAB® in Water Resources and Environmental Engineering*. Springer Science & Business Media, New York, USA, London, UK, Dordrecht, Netherlands. doi: 10.1007/978-94-007-7506-0.
- Asati, S. R. & Rathore, S. 2012 Stream flow prediction using soft computing technique. *Int J Lakes River* **5**(1), 7–22.
- Augusto, Juan Carlos & Nugent, C. D. 2006 *Designing smart homes: the role of artificial intelligence*. Edited by J. C. Augusto and C. D. Nugent. New York London: Springer, Berlin, Heidelberg. doi: <https://doi.org/10.1007/11788485>.
- Blöschl, G. 2005 Rainfall-runoff modeling of ungauged catchments. *Encyclopedia of Hydrological Sciences*. doi: 10.1002/0470848944.hsa140.
- Brath, A., Montanari, A. & Toth, E. 2002 'Neural networks and non-parametric methods for improving real-time flood forecasting through conceptual hydrological models'. *Hydrology and Earth System Sciences Discussions* **4**(6), pp. 627–639. doi: <https://hal.archives-ouvertes.fr/hal-00305217>.
- Brown, R. A. 2009 *Where the Great River Rises: An Atlas of the Connecticut River Watershed in Vermont and New Hampshire*. Edited by R. A. Brown. New England, One Court Street, Lebanon, NH 03766: Dartmouth College Press. doi: <https://books.google.co.ke/books?id=D65Zy8kk4bEC>.
- Campolo, M., Andreussi, P. & Soldati, A. 1999 River flood forecasting with a neural network model. *Water Resources Research* **35**(4), 1191–1197. doi: 10.1029/1998WR900086.
- Chang, J., Wang, G. & Mao, T. 2015 Simulation and prediction of suprapermafrost groundwater level variation in response to climate change using a neural network model. *Journal of Hydrology* **529**. doi: 10.1016/j.jhydrol.2015.09.038.
- Collier, C. G. 2007 Flash flood forecasting: what are the limits of predictability? *Quarterly Journal of the Royal Meteorological Society* **133**(622), 3–23. <https://doi.org/10.1002/qj.29>.
- Dawson, C. W., Harpham, C., Wilby, R. L. & Chen, Y. 2002 'Evaluation of artificial neural network techniques for flow forecasting in the River Yangtze, China'. *Hydrology and Earth System Sciences* **6**(4). doi: 10.5194/hess-6-619-2002.
- Diamantopoulou, M. J., Papamichail, D. M. & Antonopoulos, V. Z. 2005 The use of a neural network technique for the prediction of water quality parameters. *Operational Research* **5**(1). doi: 10.1007/bf02944165.
- Dibike, Y. B. 2002 *Model Induction from Data: Towards the next generation of computational engines in hydraulics and hydrology*. The Netherlands: CRC Press, 2002.
- Dogan, E., Isik, S., Toluk, T. & Sandalci, M. 2007 'Daily streamflow forecasting using artificial neural networks', *International Congress on River Basin Management*, (November 2014), pp. 449–459.
- Elsafi, S. H. 2014 Artificial neural networks (ANNs) for flood forecasting at dongola station in the River Nile, Sudan. *Alexandria Engineering Journal* 655–662. Elsevier B.V. doi: 10.1016/j.aej.2014.06.010.
- Farhadi, H. & Najafzadeh, M. 2021 Flood risk mapping by remote sensing data and random forest technique. doi: 10.3390/w13213115.
- Gizaw, M. S. & Gan, T. Y. 2016 Regional flood frequency analysis using support vector regression under historical and future climate. *Journal of Hydrology* **538**, 387–398. Elsevier. doi: 10.1016/J.JHYDROL.2016.04.041.
- Govindaraju, R. S. 2000 Artificial neural networks in hydrology. I: preliminary concepts. *Journal of Hydrologic Engineering* **5**, 115–123. American Society of Civil Engineers.
- Greco, M. & Krajewski, W. F. 2000 A large-sample investigation of statistical procedures for radar-based short-term quantitative precipitation forecasting. *Journal of Hydrology* **239**(1–4), 69–84. Elsevier. doi: 10.1016/S0022-1694(00)00360-7.
- Hadush, M. 2018 Modelling of river flow in ungauged catchment using remote sensing data: application of the empirical (SCS-CN), artificial neural network (ANN) and hydrological model (HEC-HMS). *Modeling Earth Systems and Environment* **0**(0), 0. Springer International Publishing, doi: 10.1007/s40808-018-0532-z.
- He, Y., Bárdossy, A. & Zehe, E. 2011 A review of regionalisation for continuous streamflow simulation. *Hydrology and Earth System Sciences* **15**(11), 3539–3553. doi: 10.5194/hess-15-3539-2011.
- Kanda, E. K., Kipkorir, E. C. & Kosgei, J. R. 2016 Dissolved oxygen modelling using artificial neural network: a case of River Nzoia, Lake Victoria basin, Kenya. *Journal of Water Security* **2**(1). doi: 10.15544/jws.2016.004.
- Kenya Meteorological Department 2020 THE WEATHER OUTLOOK FOR THE OCTOBER-NOVEMBER-DECEMBER (OND) 2020 'SHORT – RAINS ' SEASON', 254(0), pp. 1–11.

- Keitany, S. B. 2016 *A History of the Perkerra Irrigation Scheme in Baringo County, Kenya, 1954-2013*. Egerton University, Njoro, Kenya.
- Kim, S. E. & Seo, I. W. 2015 Artificial neural network ensemble modeling with exploratory factor analysis for stream flow forecasting. *Journal of Hydroinformatics*. doi: 10.2166/hydro.2015.033.
- Kostić, S., Stojković, M., Prohaska, S. & Vasović, N. 2016 Modeling of river flow rate as a function of rainfall and temperature using response surface methodology based on historical time series. 651–665. doi: 10.2166/hydro.2016.153.
- Kotlarski, S., Hagemann, S., Krahe, P., Podzun, R. & Jacob, D. 2012 'The Elbe river flooding 2002 as seen by an extended regional climate model'. *Journal of Hydrology*, 472–473. doi: 10.1016/j.jhydrol.2012.09.020.
- Kramer, O. 2013 'Dimensionality Reduction with Unsupervised Nearest Neighbors'. *Intelligent Systems Reference Library* 51, 13–14. doi: 10.1007/978-3-642-38652-7.
- Laaha, G. & Blöschl, G. 2007 A national low flow estimation procedure for Austria. *Hydrological Sciences Journal* 52(4), 625–644. doi:10.1623/hysj.52.4.625.
- Lafdani, E. K., Nia, A. M., Pahlavanravi, A., Ahmadi, A. & Jajarmizadeh, M., (2013) 'Daily Rainfall-Runoff Prediction and Simulation Using ANN, ANFIS and Conceptual Hydrological MIKE11/NAM Models'. *International Journal of Engineering & Technology* 1(1), pp. 32–50. Available from: <http://www.aropub.org/journals/international-journal-of-engineering-technology-sciences-ijets/ijets-current-issue/abstractdaily-rainfall-runoff-prediction-and-simulation-using-ann-anfis-and-conceptual-hydrological-mike11nam-models/>.
- Liong, S., Khu, S. & Babovic, V. 2007 *Genetic Programming: A new Paradigm in Rainfall Runoff Modeling GENETIC PROGRAMMING: A NEW PARADIGM*. doi: 10.1111/j.1752-1688.2002.tb00991.x.
- Lohani, A. K., Goel, N. K. & Bhatia, K. K. S. 2014 Improving real time flood forecasting using fuzzy inference system. *Journal of Hydrology* 509, 25–41. Elsevier. doi: 10.1016/J.JHYDROL.2013.11.021.
- Maddox, R. A., Zhang, J., Gourley, J. J. & Howard, K. W. 2002 'Weather radar coverage over the contiguous United States'. *Weather and Forecasting* 17(4), pp. 927–934. doi: 10.1175/1520-0434(2002)017 < 0927:WRCOTC > 2.0.CO;2.
- Mahdavinjad, M. S., Rezvan, M., Barekatin, M., Adibi, P., Barnaghi, P. & Sheth, A. P. 2018 'Machine learning for internet of things data analysis: a survey'. *Digital Communications and Networks*. doi: 10.1016/j.dcan.2017.10.002.
- Maier, H. R., Jain, A., Dandy, G. C. & Sudheer, K. P. 2010 'Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions'. *Environmental Modelling and Software* 25(8), pp. 891–909. doi: 10.1016/j.envsoft.2010.02.003.
- Meresa, H. K., Osuch, M. & Romanowicz, R. 2016 Hydro-meteorological drought projections into the 21-st century for selected Polish catchments. *Water (Switzerland)* 8(5). doi: 10.3390/w8050206.
- Michael, E. & Patience, O. 2018 Flood prediction In Nigeria using artificial neural network. *American Journal of Engineering Research (AJER)*. 9, 15–21.
- Mosavi, A., Bathla, Y. & Varkonyi-Koczy, A. 2018a 'Predicting the future using web knowledge: State of the art survey'. *Advances in Intelligent Systems and Computing* 660(2017), pp. 341–349. doi: 10.1007/978-3-319-67459-9\_42.
- Mosavi, A., Ozturk, P. & Chau, K. 2018b 'Flood Prediction Using Machine Learning Models'. *Water*, pp. 1–40. doi: 10.3390/w10111536.
- Mosavi, A., Ozturk, P. & Chau, K.-W. 2018b Flood prediction using machine learning models. *Water*, 1–40. doi: 10.3390/w10111536.
- Nagesh Kumar, D., Srinivasa Raju, K. & Sathish, T. 2004 River flow forecasting using recurrent neural networks. *Water Resources Management* 18(2). doi: 10.1023/B:WARM.0000024727.94701.12.
- Nasri, M., Modarres, R., Dastorani, M. T. & Branch, A. 2008 'Rainfall-Runoff Relationship of Zayandehrud Dam Basin by Multi Layer Perceptron Neural Network'. *Environmental Research And Development* 2(4), pp. 790–795.
- Ortiz-García, E. G., Salcedo-Sanz, S. & Casanova-Mateo, C. 2014 Accurate precipitation prediction with support vector classifiers: a study including novel predictive variables and observational data. *Atmospheric Research. Elsevier* 139, 128–136. doi: 10.1016/J.ATMOSRES.2014.01.012.
- Parasuraman, K. 2007 *Hydrologic Prediction Using Pattern Recognition and Soft Computing Technique*. University of Saskatchewan, Saskatoon, Canada.
- Ren, L., Xiang, X.-Y. & Ni, J.-J. 2013 Forecast modeling of monthly runoff with adaptive neural fuzzy inference system and wavelet analysis. *Journal of Hydrologic Engineering* 18(9). doi: 10.1061/(asce)he.1943-5584.0000514.
- Riad, S. et al. 2004 Rainfall-runoff model usingan artificial neural network approach. *Mathematical and Computer Modelling* 40(7–8). doi: 10.1016/j.mcm.2004.10.012.
- Roy, P. & Husain, Z. 2014 Magnitude of floods and its consequences in Puthimari River Basin of Assam, India. *European Ecademic Research* II(11), 2665–2685. Available from: <http://www.euacademic.org/UploadArticle/579.pdf>.
- Sarkar, A. & Kumar, R. 2012 Artificial Neural Networks for Event Based', 2012(October), pp. 891–897.
- Sattari, M. T., Rezazadeh-Joudi, A. & Kusiak, A. 2017 Assessment of different methods for estimation of missing data in precipitation studies. *Hydrology Research* 48(4). doi: 10.2166/nh.2016.364.
- Seo, D. J. & Breidenbach, J. P. 2002 Real-time correction of spatially nonuniform bias in radar rainfall data using rain gauge measurements. *Journal of Hydrometeorology* 3(2), 93–111. doi: 10.1175/1525-7541(2002)003 < 0093:RTCOSN > 2.0.CO;2.
- Shamseldin, A. Y. & O'Connor, K. M. 2001 A non-linear neural network technique for updating of river flow forecasts. *Hydrology and Earth System Sciences* 5(4), 577–597. doi: 10.5194/hess-5-577-2001.

- Singh, S. K., Jain, S. K. & Bárdossy, A. 2014 Training of artificial neural networks using information-rich data. 40–62. doi: 10.3390/hydrology1010040.
- Sketchley, H. R., Scilley, F. M., Mbuvi, P. J. & Wokabi, M. S. 1978 'Reconnaissance Soil Survey of Lake Baringo-Kerio Valley area Republic of Kenya'. Ministry of Agriculture, Kenya: United States department of Agriculture.
- Solomatine, D. P. & Ostfeld, A. 2008 **Data-driven modelling: some past experiences and new approaches**. *Journal of Hydroinformatics* 3–22. doi: 10.2166/hydro.2008.015.
- Sugiyama, H., Vudhivanich, V., Whitaker, A. C. & Lorsirirat, K. 2003 'Stochastic flow duration curves for evaluation of flow regimes in rivers'. *Journal of the American Water Resources Association* 39(1), pp. 47–58. doi: 10.1111/j.1752-1688.2003.tb01560.x.
- Tsakiri, K. G., Marsellos, A. E. & Zurbenko, I. G. 2014 **An efficient prediction model for water discharge in schoharie creek, NY**. *Journal of Climatology* 2014, 1–10. doi: 10.1155/2014/284137.
- Valipour, M., Banihabib, M. E. & Behbahani, S. M. R. 2012 **Parameters estimate of autoregressive moving average and autoregressive integrated moving average models and compare their ability for inflow forecasting**. *Journal of Mathematics and Statistics* 8(3), 330–338. doi: 10.3844/jmssp.2012.330.338.
- WMO 2011 Manual on Flood Forecasting and Warning. 2011th edn. CH-1211 Geneva 2, Switzerland: World Meteorological Organization.
- Wu, C. L., Chau, K. W. & Fan, C. 2010 'Prediction of rainfall time series using modular artificial neural networks coupled with data-preprocessing techniques', 23(8), pp. 1350–1367.
- Xu, Z. X. & Li, J. Y. 2002 **Short-term inflow forecasting using an artificial neural network model**. *Hydrological Processes* 16(12), 2423–2439. <https://doi.org/10.1002/hyp.1013>.
- Yaseen, Z. M., El-shafie, A., Jaafar, O., Afan, H. A. & Sayl, K. N. 2015 'Artificial intelligence based models for stream-flow forecasting: 2000-2015'. *Journal of Hydrology*. doi: 10.1016/j.jhydrol.2015.10.038.

First received 1 November 2021; accepted in revised form 30 March 2022. Available online 11 April 2022