American Journal of Water Resources, 2016, Vol. 4, No. 2, 35-43
Available online at http://pubs.sciepub.com/ajwr/4/2/2

DOI:10.12691/ajwr-4-2-2

Prediction of Missing Hydro-Meteorological Data Series Using Artificial Neural Networks (ANN) for Upper Tana River Basin, Kenya

Raphael M. Wambua[1,*], Benedict M. Mutua[1], James M. Raude[2]

[1]Egerton University Department of Agricultural Engineering, Kenya
[2]Jomo Kenyatta University of Agriculture and Technology, SWEED, Kenya
*Corresponding author: wambuarm@gmail.com

Abstract Accurate prediction of missing hydro-meteorological data is crucial in planning, design, development and management of water resources systems. In the present research, prediction of such data using Artificial Neural Networks (ANN) based on temporal and spatial auto-correlation has been conducted for upper Tana River basin in Kenya. Different ANN models were formulated using a combination of numerous data delays in the ANN input layer. The findings show that the best models comprise of a feed-forward neural network trained on Levenberg-Marquardt algorithm with single hidden layer. Additionally, the best ANN architecture model for predicting missing stream flow data was at gauge station 4CC03 with correlation coefficient and *MSE* of 0732 and 0.242 respectively during validation. Temporal auto-correlation of the observed and the predicted stream flow values were evaluated using a correlation coefficient R that resulted to highest value of 0.756 at gauge station 4AB05. The best ANN model for prediction of missing precipitation data was at station 9037112 with R value of 0.970. In both cases the best performance was at epochs 9 and 20 respectively. The spatial auto-correlation show that the best ANN architecture model for prediction of missing stream flow data was at gauge station 4CC03 with R value of 0.723, while the one for precipitation was at station 9037096 with *R* value of 0.712 during the validation. The results indicate that the spatial auto-correlation of hydro-meteorological data using ANN is better than the temporal auto-correlation in the data prediction in upper Tana River basin.

Keywords: *Prediction, hydro-meteorological data, ANN, data delay, auto-correlation, upper Tana River basin*

Cite This Article: Raphael M. Wambua, Benedict M. Mutua, and James M. Raude, “Prediction of Missing Hydro-Meteorological Data Series Using Artificial Neural Networks (ANN) for Upper Tana River Basin, Kenya.” *American Journal of Water Resources*, vol. 4, no. 2 (2016): 35-43. doi: 10.12691/ajwr-4-2-2.

1. Introduction

Among the hydro-meteorological data in a river basin, precipitation and stream flow are the most important. This is because of their application in water resource systems analysis, water balance computations [1], estimation of extreme events such as floods and drought that adversely affect socio-economic development [2] is vital. Scarcity of hydro-meteorological data in a river basin may be due lack of measuring instruments, damage of measuring instruments or negligence in measurements. It is not appropriate to discontinue or abandon a water resources project in any stage; planning, design or development due to incomplete hydro-meteorological data. Estimation of relevant missing hydro-meteorological data is inevitable for any river basin with specific hydrologic processes. Accurate estimation of the missing hydro-meteorological data increases the prospect and ensures success of a water resources project. According to [3] and [4], efficient prediction of any missing hydro-meteorological data require the spatial and temporal variation of their magnitudes and related physical processes at basin scales.

Some of the conventional techniques used in filling missing data include interpolation and extrapolation, Inverse Distance Weighting (IDW), differentials and formulation of kernel functions. These methods have successfully been applied by [4,5,6,7,8] respectively. The main challenge with these methods is that they are time consuming, sometimes giving errors in their results. The hydrological river basin processes are stochastic and non-linear in nature. These processes have always been explained using physically based methods which are complex. Numerous models have been developed to study stochastic hydrological processes. These include ANN models which have been extensively tested in modeling and forecasting of non-linear hydrological systems and extreme events such as droughts [9] and floods [10,11] and [12]. Numerous methods have previously been used to address data lack or limitation of data availability in different river basins. For instance, [13] and [14] estimated potential evapotranspiration and reference crop evapotranspiration for 31 provinces in Iran using average data from 181 synoptic stations and three temperature-based formulae. Spatial interpolation method was used to obtain an average value for different stations. In addition, [15] used multivariate fractional polynomial (MFP), Bayessian regression and robust regression to estimate reference evapotranspiration for three arid regions within which water crises is critical.

2. Artificial Neural Networks

An artificial neural network (ANN) model is an information processing system developed with a structure and operation similar to that of a human brain [16]. The ANN models have been improved over time by various calibration techniques. With sufficient amount of data and complexity, the ANN model can be adapted to establish any correlation between series of independent and dependent variables [17]. ANNs have some advantages [18,19] and [11] that attracted their use in this research. These advantages include:

i) the ability to process information based on their dynamic response to external input
ii) they can capture numerous kinds of relationships including non-linear functions which are not usually detected by other techniques
iii) they provide effective analytical techniques in modeling and forecasting systems
iv) the networks have the ability to model dynamic time series variables in Water Resources Engineering
v) the definition of physical processes need not be done and this. property makes it appropriate in processing large and complex data sets, including that of drought forecasting

The ANN model processes information through an elaborate network of neurons that are linked together. It predicts outputs based on defined inputs [20] by a working principle resembling that of human brain where the neuron receives a set of input signals and generates outputs. The nervous system of human beings is represented by a number of architectural structures that range from simple to complex structures. Whether the structures are simple or complex, the systems consist of neurons or neural cells as the chief building blocks.

An ANN model is similar to a biological neuron in that it has multiple input channels, data processing unit, and output channels called dendrites, cell body and the axon respectively as represented in Figure 1. The input signals (X_1, X_2, . . . , X_p) are passed to the neuron through the dendrites that represent different input channels. Each channel has its own weight referred to as connection weight that may be denoted as W_1, W_2,. . . ,W_p. The weights are very critical since they allow for collection and processing of signals based on their magnitude and effects on input functions. If a weight function gives a non-zero value at the synapse, it is allowed to pass through the cell body. Otherwise, if it has a value of zero, it is not allowed to pass the cell body. All the conveyed signals are normally integrated by summing up all the inputs [21].

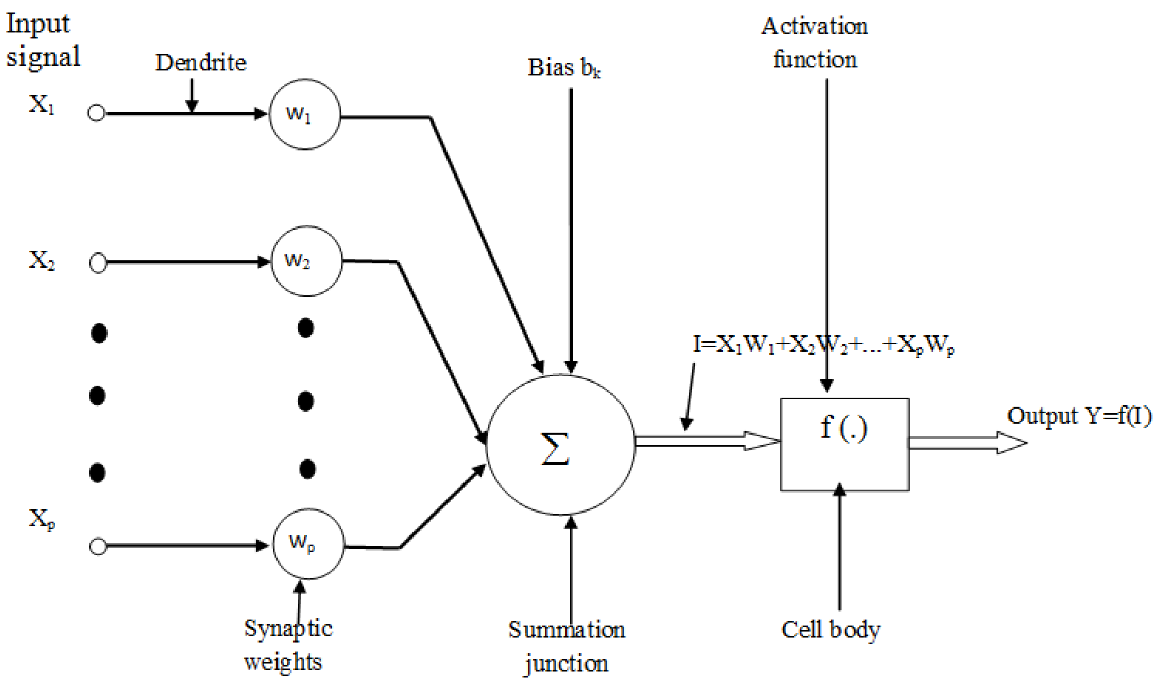

Figure 1. Fundamental parts of a typical neural network

The summation is achieved by application of a mathematical model referred to as activation function, within the cell body to generate an output signal. According to [21] and [3], the relationship between the input and output signal within an ANN model is represented using different data combinations and weight attached. This result into a function defined as:

$$Y = f(I) = f\left(\sum_{i}^{p} W_i X_i + b_k\right) \quad (1)$$

Where;

X_i = the input signal i
W_i =the weight attached to the input signal i
P =the number of input signals
b_k =the bias at the cell of the body
Y = the output
f=activation function

Numerous activation equations or functions can be used within the neurons. The most common functions used in the ANN models include; the step-function, non-linear sigmoidal, hyperbolic tangent and linear activation functions [22] and [16].

ANN models have previously been used for estimating missing data values. Some of the research that demonstrate the use of ANN in prediction and filling missing data include the work by [23-31] has showed that non-linear autoregressive neural network with exogenous input (NARNNX) was better than non-linear autoregressive neural network (NARNN) in forecasting precipitation in Gilan, Iran. However, the use of artificial neural networks in prediction of hydro-meteorological data in upper Tana River basin has not been explored. This basin is a key water resource in Kenya for hydro-electric power generation, water supply and agriculture. The main objective of this research was to predict and test the efficiency of estimation of missing hydro-meteorological data series using artificial neural networks (ANN), for upper Tana River basin.

3. Materials and Methods

3.1. The Study Area

The upper Tana River basin with an area of 17,420 km^2 (Figure 2) was the focus of the presented research. The upper Tana River basin lies between latitudes 00^0 05' and 01^0 30' south and longitudes 36^0 20' and 37^0 60' east. The basin is fundamental in influencing the ecosystem downstream [32]. It drains nine counties namely; Muranga, Nyandarua, Kiambu, Kirinyaga, Laikipia, Machakos, Nyeri, Embu and Kitui [33]. The basin was selected because it is located within a fragile ecosystem that represents all agro-ecological zones of Kenya where water resource systems, hydro-electric power generation and food security are negatively impacted by frequent drought occurrence. The basin area regulates the hydrology of the Kenya's largest river system called Tana River Basin with a total area of 126,000 km^2 [34]. Within the upper Tana River basin, numerous socio-economic activities such as hydro-power generation, agriculture, and irrigation take place.

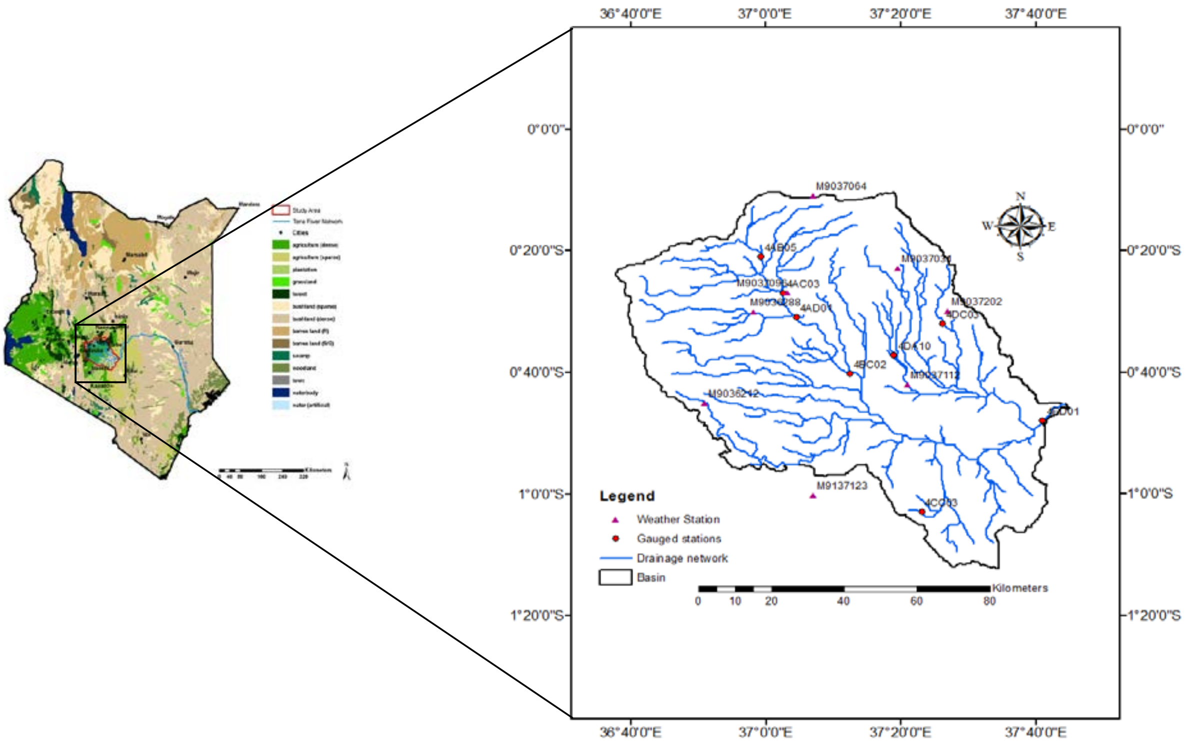

Figure 2. Map of Kenya showing the upper Tana River basin

3.2. Stream Flow Data

There were approximately fourteen gauge stations in the upper Tana River basin with complete and incomplete data records. However, only eight stations were selected for this study since they had sufficiently long and reliable data for the period 1970-2010 as required by this study. The stations were in addition considered a realistic representative of the basin as they are objectively located within low, lower middle, middle and higher elevations for different agro-ecological zones as required in the study. The names of the stations and gauge identification (ID) numbers, their spatial locations are shown in Figure 2 and Table 1 respectively.

Table 1. Stream flow gauge stations

S.No	Gauge Name	Gauge ID	Coordinates	
			Easting	Northing
1	Amboni	4AB05	36.989	-0.350
2	Sagana	4AC03	37.043	-0.449
3	Gura	4AD01	37.076	-0.517
4	Tana sagana	4BC02	37.207	-0.672
5	Yatta furrow	4CC03	37.361	-1.094
6	Nyamindi	4DA10	37.317	-0.621
7	Rupingazi	4DC03	37.438	-0.533
8	Kamburu	4ED01	37.683	-0.800

3.3. Precipitation Data

In the upper Tana River basin, data from twenty four meteorological stations were obtained from the Ministry of Environment, Water and Natural resources. Such stations provided precipitation data. The stations were also objectively located within the low, lower middle, middle and high elevations so as to study drought characteristics at different agro-ecological basin of the basin.

Table 2. Meteorological stations

S.No	Station name	Station ID	Coordinates		Elevation
			Longitude	Latitude	
1	MIAD	9037112	37.350	-0.700	1246
2	Embu	9037202	37.450	-0.500	1494
3	Kerugoya DWO	9037031	37.327	-0.382	1598
4	Sagana FCF	9037096	37.054	-0.448	1234
5	Nyeri	9036288	36.970	-0.500	1780
6	Muragua G. E. F.	9036212	36.850	-0.750	2296
7	Naro moru F.G.P.	9037064	37.117	-0.183	2296
8	Mangu HS	9137123	37.033	-1.100	1630

3.4. Temporal Auto-correlation Based on ANN Models

The hydro-meteorological data was partitioned into 70% and 30% training and validation both for stream flow data and precipitation data at each station. By considering different input neurons with different time delays; *t, t-1, t-2,…,t-n*, in the input layer, the ANN structure for each station was obtained. Hidden layer neurons initially $2n+1$, increase and decrease using trial and error technique. The output was is the predicted variable. Thus the neural network has three layers; input layer, hidden layer and the output layer (Figure 3a and Figure 3b). The ANN model at each station was trained using different structures (Table 3). The output layer comprise of neurons in all the networks that are equal to the next month's predicted value (I_{t+1}). For the present study, feed-forward neural network (FFN) and recursive neural network (RNN) were tested in the model training. Initially three different training algorithms were applied to train the structures; Back-propagation (BP), Conjugate Gradient (CG) and Levernberg-Marquardt (LM). Preliminary results indicated that a three-layer feed forward neural network with different input and hidden neurons was the best in performance, and that the superlative results were also obtained using the LM training algorithm. Thus the results presented are for best ANN structure of three-layer feed forward network based on LM training algorithm. The monthly data for different previous months was auto-correlated with the next month's predicted value. The performance of the auto-correlation was done using performance criteria of the correlation coefficient R, and mean square error (MSE) presented in Equations 3 and 4.

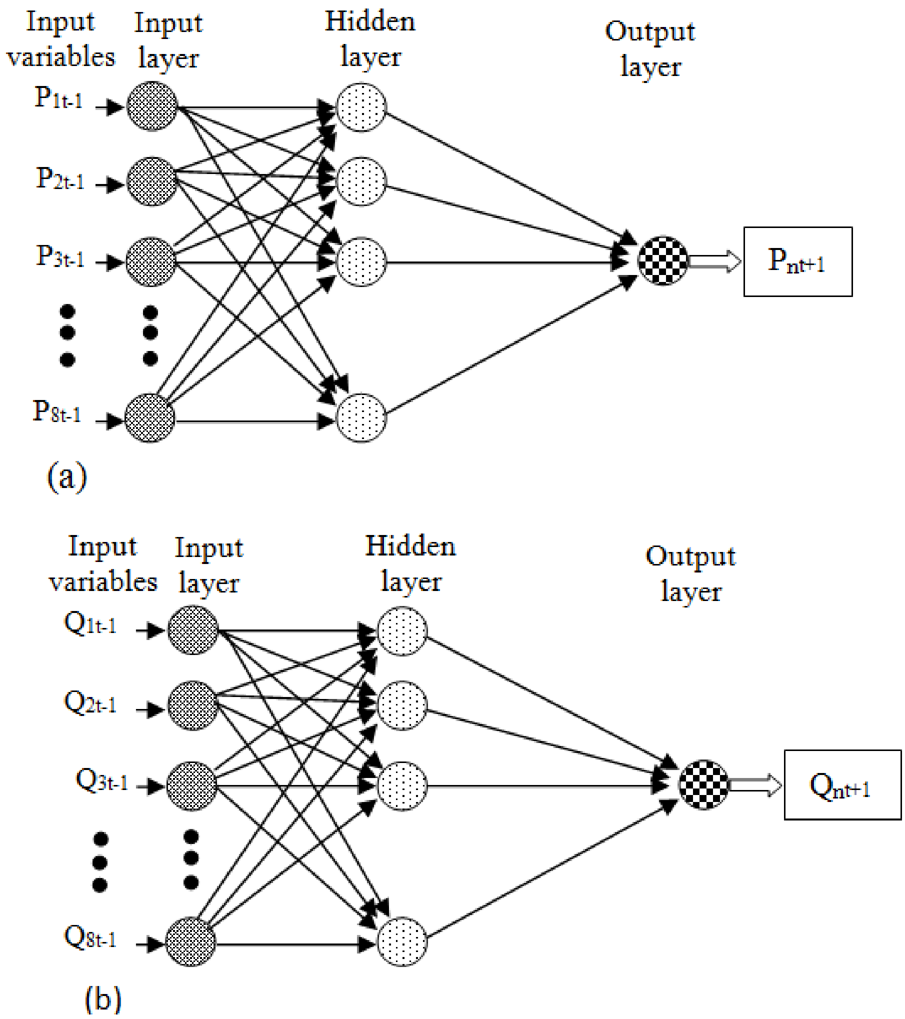

Figure 3. ANN model Architecture used for prediction of missing (a) meteorological and (b) stream flow data based on spatial auto-correlation

3.5. Spatial Auto-correlation Based on ANN Models

It was expected that there exist spatial-autocorrelation among the hydro-meteorological for stations at different time delays. Seven stations were used at a time to estimate and/predict the hydro-meteorological variables at one of the eight stations at a time. From each station, 70% of the data was used for training while 30% was used validation. Figure 3b shows the schematic diagram for the proposed spatially correlated ANN model. Different ANN models were formulated using different time delays and hidden neurons. The results of the spatially auto-correlated ANN structures and training epochs are shown in Table 5 and Table 6.

3.6. Performance Evaluation of the Models

In ANN the input and output variables used in the modelling were normalized by fitting their values between zero and one to speed up the forecasting process as stated by [18]. In the present study, the recorded stream flow and precipitation data values were scaled from a minimum to maximum value by applying the following function:

$$X_n = X_{\min} + \frac{(X_o - x_{\min})}{(x_{\max} - x_{\min})} \times (X_{\max} - X_{\min}) \qquad (2)$$

Where;

X_n= normalized/standardized value
X_{min} = the selected minimum value for standardization
X_{max} = the selected maximum value for standardization
X_o = original variable value
x_{min} = minimum value present in the original data set
x_{max} = maximum value present in the original data set

For the purpose of this research the X_{min} and X_{max} were chosen as 0.1 and 0.9 respectively as these have been found to give reliable results in hydrological studies [18]. After standardization, the prediction of data values for numerous stations was conducted. For illustration purpose, the results of selected stations are presented in the present paper. The normalized data was partitioned into training and validation data sets. At each hydro-meteorological station, 70% of continuous data was used for training while the remaining 30% was used for validation purpose. In this research, two criteria were used to determine the efficiency of the ANN models. These included the correlation coefficient (R) and the mean square error (MSE).

The statistical relationship between the observed and the predicted data values within the upper Tana River basin was done using the correlation coefficient (R). The fundamental function was customized to the respective data input and outputs of the following general form:

$$R = \frac{\sum_{i=1}^{n}\left(I_{Obs} - \overline{I_{Obs}}\right)\left(I_{For} - \overline{I_{For}}\right)}{\sum_{i=1}^{n}\left(I_{Obs} - \overline{DI_{Obs}}\right)\left(I_{For} - \overline{I_{For}}\right)^2} \qquad (3)$$

Where;

R = the correlation coefficient
I_{Obs} = the observed value of the hydro-meteorological variable
$\overline{I_{Obs}}$ = mean of the observed values the hydro-meteorological variable
I_{For}= the forecasted value of the variable
$\overline{I_{For}}$ = mean of the forecasted values of the hydro-meteorological variable
n=number of data points considered

The R is a measure of the strength of the relation between observed and forecasted DI values. It varies from 0 to 1. The values of 0 and 1 indicate a poor and perfect forecasting capability of the model respectively.

The Mean Square Error (MSE) is a measure of the difference between the observed and forecasted drought values from different indices. It measures the average of the squares of the errors between two values being compared. For the purpose of drought, the following function was adopted for calculating MSE associated with drought forecasting:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(I_{iFor} - I_{iObs}\right)^2 \qquad (4)$$

Where;

MSE = the mean square error
I_{Obs} = the observed value of the variable
I_{for} = the forecasted value of the ANN model
n = number of data points

The MSE ranges from 0 to 1. The smaller the MSE value the better the forecasting capability of the model.

4. Results and Discussions

Table 3. The best ANN structure used in filling of missing stream flow data for different gauging stations

Gauge station	Gauge ID	Epoch No	Coordinates Longitude	Coordinates Latitude	Input function into the neurons	ANN architecture	R Train	R validation	MSE train	MSE validation
Amboni	4AB05	9	36.989	-0.350	$Q_{(i+1)} = f(Q_i, Q_{i-1}, Q_{i-2})$	3-9-1	0.545	0.735	0.214	0.105
Sagana	4AC03	14	37.043	-0.449	$Q_{(i+1)} = f(Q_i, Q_{i-1})$	2-9-1	0.621	0.724	0.312	0.213
Gura	4AD01	20	37.076	-0.517	$Q_{(i+1)} = f(Q_i, Q_{i-1})$	2-6-1	0.583	0.655	0.362	0.351
Tana sagana	4BC02	9	37.207	-0.672	$Q_{(i+1)} = f(Q_i, Q_{i-1})$	2-5-1	0.643	0.563	0.384	0.346
Yatta furrow	4CC03	20	37.361	-1.094	$Q_{(i+1)} = f(Q_i, Q_{i-1}, Q_{i-2})$	3-9-1	0.701	0.732	0.276	0.242
Nyamindi	4DA10	8	37.317	-0.621	$Q_{(i+1)} = f(Q_i, Q_{i-1})$	2-2-1	0.684	0.654	0.332	0.344
Rupingazi	4DC03	9	37.438	-0.533	$Q_{(i+1)} = f(Q_i, Q_{i-1}, Q_{i-2})$	3-6-1	0.595	0.673	0.346	0.318
Kamburu	4ED01	10	37.683	-0.800	$Q_{(i+1)} = f(Q_i, Q_{i-1})$	2-2-1	0.643	0.686	0.278	0.229
Mean values							0.627	0.678	0.313	0.269

Table 4. The best ANN structure used in filling of missing precipitation data for different meteorological stations

Gauge station	Gauge ID	Epoch No	Coordinates		Input neurons	ANN architecture	R Train	R validation	MSE train	MSE validation
			Longitude	Latitude						
MIAD	9037112	20	37.350	-0.700	$P_{(i+1)} = f(P_i, P_{i-1}, P_{i-2})$	3-10-1	0.758	0.797	0.325	0.056
Embu	9037202	22	37.450	-0.500	$P_{(i+1)} = f(P_i, P_{i-1}, P_{i-2})$	3-9-1	0.654	0.752	0.287	0.172
Kerugoya DWO	9037031	30	37.327	-0.382	$P_{(i+1)} = f(P_i, P_{i-1})$	2-5-1	0.664	0.718	0.361	0.123
Sagana FCF	9037096	40	37.054	-0.448	$P_{(i+1)} = f(P_i, P_{i-1}, P_{i-2})$	2-4-1	0.587	0.653	0.294	0.195
Nyeri	9036288	20	36.970	-0.500	$P_{(i+1)} = f(P_i, P_{i-1}, P_{i-2})$	3-10-1	0.625	0.694	0.302	0.237
Muragua G. E. F.	9036212	15	36.850	-0.750	$P_{(t+1)} = f(P_t, P_{t-1})$	2-4-1	0.581	0.635	0.275	0.228
Naro moru F.G.P.	9037064	25	37.117	-0.183	$P_{(i+1)} = f(P_i, P_{i-1})$	2-3-1	0.596	0.736	0.311	0.279
Mangu HS	9137123	28	37.033	-1.100	$P_{(i+1)} = f(P_i, P_{i-1})$	2-5-1	0.603	0.729	0.273	0.234
Mean values							0.634	0.727	0.304	0.191

Temporal auto-correlation of the observed and the predicted stream flow values was evaluated using a correlation coefficient R, giving 0.756 and 0.731 at gauge stations 4AB05 and 4AC03 respectively. The architecture for the best ANNs models at these stations are 3-9-1 and 2-9-1. The best ANN model for filling missing precipitation data was at station 9037112 with R and *MSE* values of 0.970 and 0.056 respectively. In both cases the best performance was at epochs 9 and 20 respectively (Table 3 and Table 4). The prediction of stream flow was done for one month and or more to represent medium and long-term range that are critical to a hydrological risk such as drought (Morid *et al.*, 2007) [18].

4.1. Spatial Auto-correlation of Hydro-meteorological Data

According to the results of spatial auto-correlation the performance of ANN changed by decreasing or increasing the number of hidden neurons from the initial 2n+1. From the results, the best ANN structure obtained was 7-8-1 which defines ANN structure with seven, eight and one neurons in the input, hidden and output layers respectively trained at epoch 16 for station 4CC03. Comparing Table 1 and Table 2 shows that the spatial auto-correlation is superior to temporal auto-correlation in estimating the missing values based on the R and MSE. This high performance in spatial auto-correlation is attributed to the fact that hydrological processes exhibit spatial variability.

Feed-forward and RNN with different training algorithms; Back-propagation (BP), Conjugate Gradient (CG) and Levernberg-Marquardt (LM), were used to compare network performance using R MSE. The best structure was selected based on the best values of R and MSE during the training and validation periods as shown in Table 5 and Table 6. The results show that the best model was found to be a three-layer feed-forward neural network trained by the LM algorithm. This best model structure was then used to estimate the missing monthly hydro-meteorological data values. The simulated monthly time series of missing data values for the gauging stations 4AB05 and 4DC03 are as shown in Figure 3 and Figure 4, while those for meteorological stations 9037096 and 9037064 are given in Figure 5 and Figure 6 respectively. A plot of the observed and the predicted values show that the models are efficient on prediction capabilities. For instance the scatter plot obtained by plotting the predicted and observed values of stream flow for stations 4AB05 and 4AC03 show that the points are consistently distributed on the 1:1 line, with correlation coefficient R values of 0.756 and 0.731 in that order (Figure 8).

Table 5. The results of the best spatially auto-correlated ANN models for prediction of stream flow

Station ID	Epoch No	ANN architecture	R Train	R validation	MSE train	MSE validation
4AB05	10	7-10-1	0.798	0.674	0.033	0.041
4AC03	14	7-9-1	0.682	0.702	0.025	0.046
4AD01	12	7-6-1	0.784	0.675	0.038	0.048
4BC02	11	7-10-1	0.690	0.712	0.042	0.047
4CC03	16	7-8-1	0.821	0.723	0.044	0.329
4DA10	14	7-6-1	0.756	0.686	0.039	0.043
4DC03	12	7-7-1	0.654	0.632	0.042	0.246
4ED01	15	7-10-1	0.676	0.634	0.038	0.044

Table 6. The results of the best spatially auto-correlated ANN for prediction of precipitation

Station ID	Epoch No	ANN architecture	R Train	R validation	MSE train	MSE validation
9037112	8	7-8-1	0.721	0.613	0.043	0. 406
9037202	6	7-8-1	0.673	0.626	0.025	0.033
9037031	10	7-10-1	0.684	0.682	0.037	0.039
9037096	5	7-12-1	0.726	0.712	0.026	0.326
9036288	6	7-11-1	0.721	0.653	0.039	0.045
9036212	12	7-9-1	0.675	0.697	0.038	0.042
9037064	14	7-6-1	0.732	0.656	0.046	0.048
9137123	20	7-12-1	0.702	0.681	0.029	0.236

The spatial auto-correlation show that the best ANN model for filling missing stream flow data was at gauge station 4CC03 with *R* and *MSE* values of 0.723 and 0.326 respectively, while the best ANN architecture model for filling missing precipitation data was at station 9037096 with *R* and *MSE* values of 0.712 and 0.329 respectively during the validation (Table 5 and Table 6).

Monthly time series hydro-meteorological data were predicted for the period without data records. For instance at gauging stations 4AB05 and 4DC03, data was predicted for the period of years 2004 to 2010, and 2000 and 2010. The highest predicted stream flow values for the period for the stations were 12.44 and 15.65 m^3/s (Figure 4 and Figure 5). On the other hand, the highest predicted precipitation values are 550 mm and 10 mm at meteorological stations 9037096 and 9037064, for the period 2002 to 2007 and 2000 to 2010 respectively (Figure 6 and Figure 7). Based on the validation *R* values, the prediction capability of ANNs models at gauged stations can be arranged from the largest to the smallest in the order 4CC03, 4BC02, 4AC03, 4DA10, 4AD01, 4AB05, 4ED01 and 4DC03 respectively. On the other hand, the ANNs models prediction capability at the meteorological stations decline from the largest to the smallest in the order of 9037096, 9036212, 9037031, 9137123, 9037064, 9036288, 9037202 and 9037112 respectively. The best ANN models from this study can be applied in other river basins with appropriate calibration and validation.

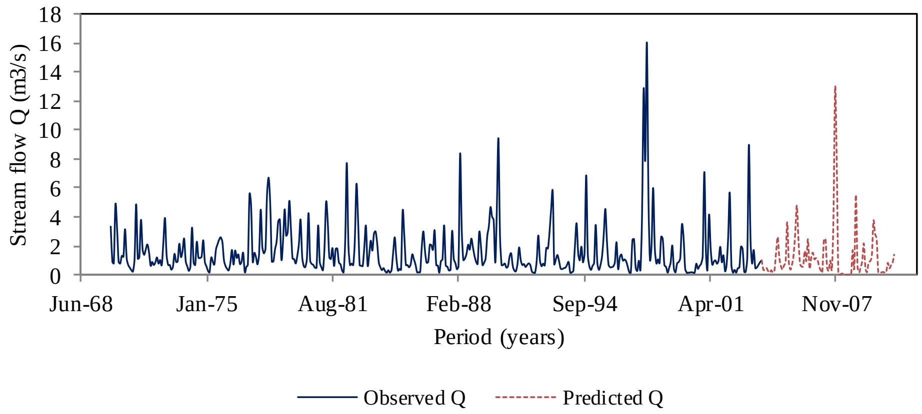

Figure 4. Monthly time series of predicted stream flow values at gauging station 4AB05

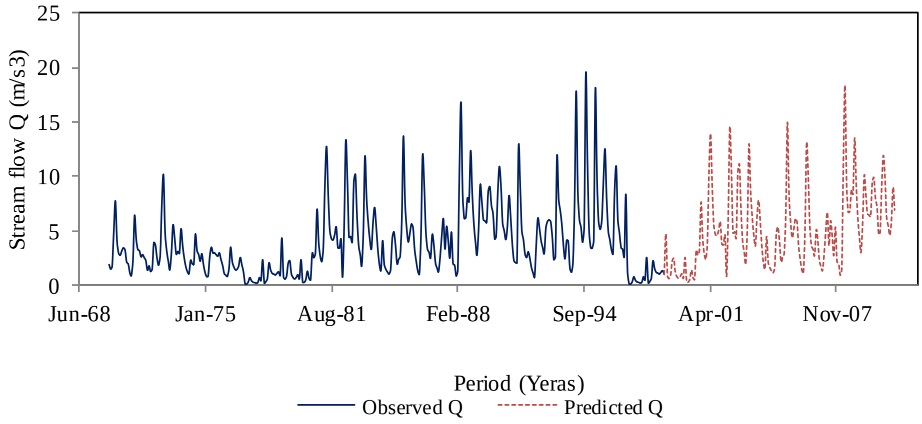

Figure 5. Monthly time series of the predicted stream flow values at gauging station 4DC03

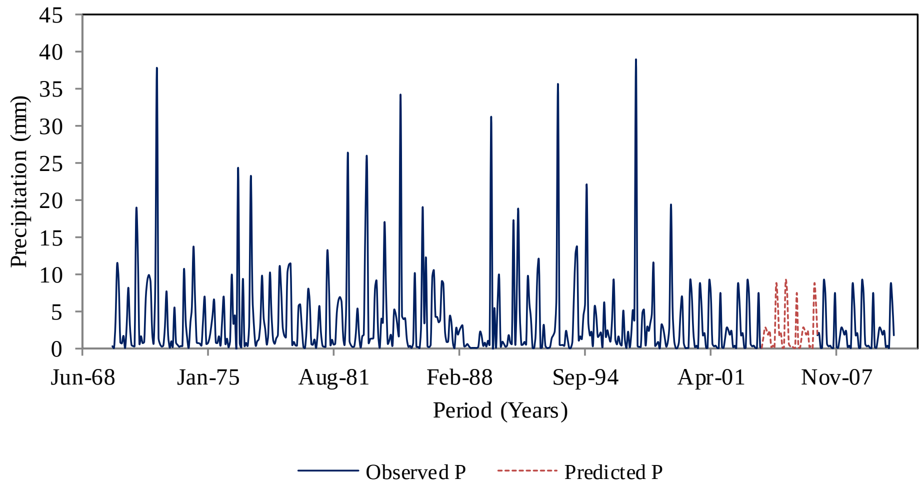

Figure 6. Monthly time series of the forecasted values of missing data at meteorological station 9037096

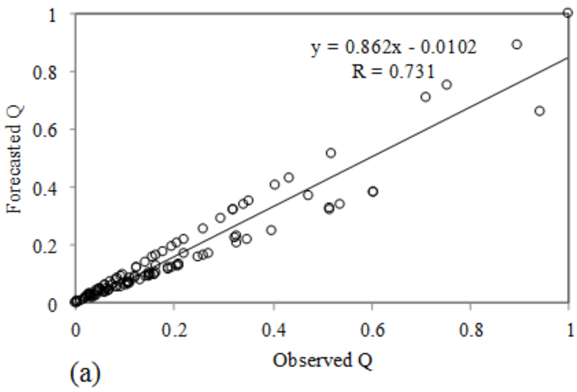

Figure 7. Scatter plot for observed against predicted stream flow at gauge station 4AB05)

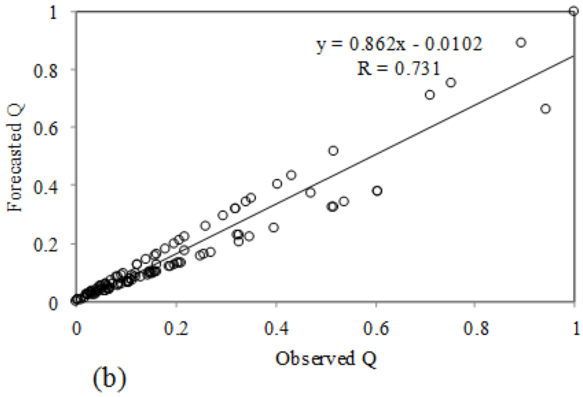

Figure 8. Scatter plot for observed against predicted stream flow at gauge station 4AC03

5. Conclusions

In the presented study the artificial neural networks (ANN) was used to establish the relationship between the input and output of specified hydro-meteorological data combination. The ANN was used to formulate the temporal and spatial auto-correlation functions for prediction of missing hydro-meteorological data in the upper Tana River basin. As shown by the R and MSE values, the results indicate that the spatial auto-correlation of station data using ANN is better than the temporal auto-correlation in predicting missing hydro-metric data for upper Tana River basin. From the three neural networks tested in this study, the most efficient network was found to be feed-forward neural network (FFN) which was trained using the Levernberg-Marquardt (LM) algorithm

6. Recommendations

It is recommended form this investigation that the methods and results of the research be applied to investigate temporal and spatial auto-correlation between numerous parameters for prediction of other hydrological processes such as soil erosion, sediment yield and runoff. Such prediction would be very useful in prioritized planning of the land use practices, agricultural practices, soil conservation and water supply within the upper Tana River basin.

Acknowledgement

The authors of this paper acknowledge the scholars whose articles are cited and included in the references of this article. In addition, authors are grateful to the editors and publishers of journals and books from where the literature has been referred. The authors are grateful to the editorial board and the reviewers of the American Journal Water Resources who assisted in improving the quality of this article.

References

[1] Dreps, C., James, A. L., Sun G and Boggs, J. (2014). Water balance of the two piedmont headwater catchments; implications for regional hydrologic landscape classification, *J. of American water resources association*, 50(4): 1063-1079.

[2] Botai, C. M., Botai, J. O., Muchuru, S. and Ngwana, I. (2015). Hydro-meteorological research in South Africa, A review, *Water journal*, 7: 1580-1594.

[3] Ismail, M. I. S, Okamoto, Y. and Okada, A. (2013). Neural network modeling for prediction of weld bead geometry in laser microwelding , *Advances in optical technologies*, 2013: 1-7.

[4] Lee, H. and Kang, K. (2015). Interpolation of missing precipitation data using Kernel estimations for hydrologic modeling, *Advances in Meteorology*, 2015: 1-12.

[5] Jang, D., Park, H. and Choi, J. T. (2015a). Create a missing precipitation data base on spatial interpolation methods not covered by a region climate change scenario, *Advanced Science and Technology L*, 99: 109-112.

[6] Jang, D., Park, H. and Choi, J. T. (2015b). Applicability of the kriging method in data missing area using region climate change, E-proceedings of the 36th IAHR world congress, 28th June -3rd July, 2015, the Hague, the Netherlands.

[7] Getahun, Y. S. and Gebre, S. L. (2015). Flood hazard assessment and mapping of flood inundation area of the Awash River Basin in Ethiopia using GIS and HEC-GEORAS/HEC-RAS Model, *J. of Civil and Environmental Engineering*, 5(4): 1-12.

[8] Rivero, C. R., Pucheta, J.,Laboret, S., Partino, D. and Sauchelli, V. (2015). Forecasting short time series with missing data by means of energy associated to series, *Applied mathematics*, 2015(6): 1611-1619.

[9] Belayneh, A and Adamowski, J. (2013). Drought forecasting using new machine learning methods, *Journal of Water and Land development*, 18(9): 3-12.

[10] Ghumman, A. R., Ghazaw, Y. M., Sohail, A. R. and Watanabe, K. (2011). Runoff forecasting by artificial neural networks and conventional model, *Alexandria Engineering Journal*, 50(4): 345-350.

[11] Mustafa, M. R., Isa, M. H., Rezaur, R. B. (2012). Artificial neural networks modeling in water resources engineering; infrastructure, and applications, *Int. J. of Civil, Environmental, Structural, Construction and Architecture Engineering*, 6(2) 128-136.

[12] Elsafi, S. H. (2014). Artificial neural networks (ANNs) for flood forecasting at Dongola station in the River Nile Sudan, *Alexandria Engineering Journals*, 53(3): 655-662.

[13] Valipour M. K (2014). Analysis of potential evapotranspiration using limited weather data, *Appl Water Sci*..

[14] Valipour, M. (2015). Calibration of mass transfer-based models to predict reference crop evapotranspiration, *Appl Water Sci*..

[15] Khoshravesh, M., Sefidkouhi M. A. G. and Valipour M. (2015). Estimation of reference evapotranspiration using multivariate fractional polynomial, Bayesian regression and robust regression models in three arid environments, *Appl Water Sci*..

[16] Maier, A. R., Jain, A., Dandy, G. C. and Sudheer, K. P. (2010). Methods used for development of neural networks for the prediction of water resource variables in river systems: current status and future directions, *Journal of Environmental modeling*, 25(8): 891-909.

[17] Luk K. C., Ball, J. C. and Sharma, A. (2000). A study of optimal model lag and spatial inputs to artificial neural network for rainfall forecasting, *Journal of Hydrology*, 227: 56-65.

[18] Morid, S. Smakhtin, V. and Bagherzadeh, K. (2007). Drought forecasting using artificial neural networks and time series of

drought indices *International Journal of climatology*, 27 (15): 2103-2111.

[19] ran, H. D., Muttil, N. and Perera, B. J C. (2009). Investigation of artificial neural network models for stream flow forecasting, 19th international congress on modeling and simulation, Perth, Australia, 12th-16th December 2011.

[20] Zali, M. A,, Retnam, A., Juahir, H. and Zain, S. M. (2011). Sensitivity analysis for water quality index (WQI) prediction for Kinta River, Malaysia, *World Applied Sciences Journal*, 14: 60-65.

[21] Barua, S. (2010). Drought assessment and forecasting using a non-linear aggregated drought index, PhD thesis, Victoria University, Australia.

[22] Mishra, A. K. and Desai, V. R. (2006). Drought forecasting using feed-forward recursive models, *Journal of Ecological Modeling*, 198: 127-138.

[23] Campolo, M., Andreussi, P., Soldati, A. (1999). River flow forecasting with a neural network model, *Water resources research*, 35(34): 1191-1198.

[24] Anmala, J., Zhang, B., Govindaraju, R. S. (2000). Comparison of neural networks and empirical approaches for predicting water shed runoff, *Journal of water resources, planning and management*, 126(3): 156-166.

[25] Bodri, L., Cermak, V. (2000). Prediction of extreme precipitation using neural networks, application to summer and flood occurrence in Maravia, *Advances in engineering software*, 31: 311-321.

[26] Grimes, D. I. F., Coppola, E., Verdecchia, M., Visconti, G. (2003). A neural network approach to real time rainfall estimation for Africa using satellite data, *Journal of Hydrometeorology*, 4: 1119-1133.

[27] Lliunga, M., Stephenson, D. (2005). Infilling stream flow data using feed forward back-propagation (BP) artificial neural networks, applications of standard BP and pseudo Mac Laurin power series BP techniques, *Water journal*, 31(2): 171-176.

[28] De Silva, R. P., Daya-wansa, N. D. K., Ratnasiri, M. D. (2007). A comparison of methods used in estimating missing rainfall data, *Journal of agricultural sciences*, 3(2): 101-108.

[29] Sacchi, R., Ozturk, M. C., Principe, J. C.,Carneiro, A. A. F. M., Silva, I. N. D. (2007). Water inflow forecasting using the echo state network, a Brazilian case study, IEEN proceedings of international joint conference on neural network, 12-17th August 2007, Orlando FL, USA.

[30] Starrett, S. K., Heir, T., Su, Y. (2010). Filling in missing peak flow data using artificial neural networks, *Journal of Engineering and Applied Sciences*, 5(1) 49-55.

[31] Valipour, M. (2016). Optimization of neural networks for precipitation analysis in a humid region to detect drought and wet year alarms, *Meteorological Applications Journal*, 23(2016): 91-100.

[32] NEMA. (2004). Kenya state of environment report: Chapter 7, fresh water, coastal and marine resources, Nairobi, Government printer.

[33] World Resources Institute (WRI). (2011). Kenya GIS data –world resources institute, www.wri.org/resources/data-sets/kenya-gis-data.

[34] GoK. (2012). Tana River Delta strategic environmental assessment scoping report.