

# Artificial Neural Network Based Runoff and Sediment Yield Modeling of Maybar Watershed, Awash Basin, Ethiopia

Hussen Ali Hassen<sup>1,\*</sup>, Yonatan Tibebe<sup>2</sup>, Dagemawi Negashe<sup>1</sup>, Mehret Ayana<sup>1</sup>, Fikru Fentaw Abera<sup>1</sup>

<sup>1</sup>Department of Hydraulic and Water Resources Engineering, Kombolcha Institute of Technology, Wollo University, Kombolcha, Ethiopia

<sup>2</sup>Department of Hydraulic and Water Resources Engineering, Woldiya University, Woldiya, Ethiopia

## Email address:

hussenali1466@gmail.com (Hussen Ali Hassen)

\*Corresponding author

## To cite this article:

Hussen Ali Hassen, Yonatan Tibebe, Dagemawi Negashe, Mehret Ayana, Fikru Fentaw Abera. (2023). Artificial Neural Network Based Runoff and Sediment Yield Modeling of Maybar Watershed, Awash Basin, Ethiopia. *International Journal of Intelligent Information Systems*, 12(4), 63-75. <https://doi.org/10.11648/j.ijiis.20231204.12>

**Received:** November 27, 2023; **Accepted:** December 19, 2023; **Published:** December 28, 2023

---

**Abstract:** Runoff and sediment are important parameters to be understood and predict for managing land and water resource. So, understanding the dynamic process and prediction of the existing process by selecting suitable hydrological model is very essential. This study aims to test and evaluate the application of an artificial neural network (ANN) model for modeling runoff and sediment yield of Maybar watershed, Awash River basin. The ANN model was trained and cross validated using MATLAB, supported by the NN toolbox package. The main input for the ANN model was selected using correlation results from Statistical Packages for Social Science (SPSS). Present rainfall and previous one-day runoff up to four days of runoff were selected as inputs for runoff modeling, and present rainfall, present runoff, and previous one-day runoff were selected as inputs for sediment yield modeling. The proposed model was developed, trained, and cross validated by considering 7 years of data (2010–2016) for model training and 2 years of data for model testing (cross-validation), and their performance was evaluated using performance indicators ( $R^2$ , RMSE, and NSE). Adding lag of runoff as input results increase the model efficiency during training. Of the five proposed ANN runoff models, model B (2 inputs, 3 hidden neurons, 1 output) performed better than the other proposed runoff models. Similarly, of the three proposed ANN sediment models, model III (3 inputs, 6 hidden neurons, 1 output) performed better than the other proposed sediment models. In general, the ANN model was applicable for predicting runoff and sediment in the Maybar watershed in daily time steps.

**Keywords:** Artificial Neural Networks, Runoff Modeling, Sediment Yield Modeling, Maybar Watershed

---

## 1. Introduction

Soil erosion is process of the detachment and transportation of soil particles from their original place to further downstream by erosion agents such as water and wind. But rainfall is the major agent for soil erosion from watersheds. It is one of the normal aspects of landscape development and occurs in the form of sheet, rill and gully erosion. Deforestation, overgrazing, forest clearing, cultivating mountainous and steep slope and poor land management in general are major factors that accelerate the rate of soil erosion [26]. Today, due to inadequate consideration and protection, soil erosion is considered as one of the major land degradation, agricultural and environmental problems across the world [2]. Soil erosion

results in sedimentation of reservoir [5]. Reservoir Sedimentations is a series problem that affect the performance and suitability of reservoir by reducing the effectiveness of flood control, change water storage and ground water condition and by affecting operation of low level outlet gates and valves [19].

In Ethiopia, nearly 85% of the population's economy depends on subsistence agriculture [11]. Different types of crops are cultivated in the Ethiopian highland. One of the major problems affect agricultural productivity in the Ethiopian highlands is soil erosion. According to USAID study, Ethiopia loss billons of soil annually due to greater population and consequently more intensive cultivation. Many farmers in Ethiopian highlands cultivate sloped or hilly land. This process causes topsoil to be washed away and leaches much of the

fertile soil from highlands during the torrential rains of the rainy season [20]. Most of the Ethiopian highlands are extremely degraded because of intensive cultivation and in some cases mismanagement. The degradation of the Highlands affects directly or indirectly the lowlands through erosion and deposition. The Maybar watershed which found in high land of Awash River basin embarrasses most undulated and steep slope of the basin, which makes the watershed hot spot for flooding and sedimentation. Deforestation, human settlement and agricultural expansion in the watershed cases effect on the hydrology and sediment yield of the watershed. These effects on agricultural land where the redistribution of soil within a field, the loss of soil from a field, the breakdown of soil structure and the decline in organic matter and nutrients result in a reduction of cultivable soil depth and a decline in soil fertility. So, this type of problem needs careful prediction of runoff and sediment yield by selecting an appropriate hydrological model and effective watershed planning to reduce the effect of sediment on Maybar watershed. Currently, a number of hydrological models exist for predicting runoff and sediment yield of watershed using different time step [10]. Based on the degree of complexity, hydrological models are categorized into empirical (black-box) models, conceptual models, and physically based model [20]. However, many previous runoff and sediment yield modeling shows that the model were highly depend on catchment data (physical model) and highly sensitive to data quality. The physical based models consider the controlling physical process. These models are considered to be a better choice in rigorous theoretical sense, even though the data requirement is higher. But due to limited availability of data, such models do not perform satisfactorily. All most, all of the previous study on Maybar watershed uses physical based model [28, 27, 14, 12]. Runoff modeling on Maybar watershed has done by Habtamu Asrat et al. (2015) using soil and water assessment tool (SWAT) model to model the hydrology of Maybar watershed. An acceptable result was found by this modeling monthly basis but during the study there was a lack of reliable catchment characteristics data especially land use land cover and soil data. so if there is a lack of reliable catchment characteristics data using ANN model is preferable because ANN model give satisfactory result without using spatial data like land use/land cover data, soil data, topography etc. [28] have been applied SWAT model for sediment simulation on Maybar gauged watershed and showed an existence of an agreement between observations and model predictions of sediment yield at the watershed outlet within the performance ratings for recommended statistics at monthly basis, though SWAT exceptionally under-predicted peak sediment loads in both calibration and validation periods. [27] Apply SWAT-CN and SWAT-WB to compare the two models by simulating the hydrology of Maybar watershed. Based on statistical performance indicator he found SWAT-WB more effective then SWAT-CN. Even though the performance of SWAT-WB was good, the model run in monthly time steps it does not found the peak flow that occurred in the watershed. Since Maybar meteorological station is not class one station, meteorological data are incomplete that means only have

participation and temperature data. There is no relative humidity, sun shine hour and wind speed data. But previous study on Maybar uses Kombolcha station as an input for SWAT model during simulation of runoff and sediment yield. But ANN model work with incomplete data that means using the available data (precipitation & temperature) the model simulates runoff as well as sediment yield. Runoff and sediment yield modeling using SWAT model for Maybar gauged watershed was done before in monthly basis. But The SWAT model under-predicted peak sediment loads in both calibration and validation periods. The performance of ANN in daily basis is higher as compared with other models for daily runoff as well as daily sediment yield modeling. In general all previous study on runoff and sediment yield on Maybar watershed were done in monthly basis and simulation of model in monthly basis gives the average value of runoff or sediment that comes from the watershed, it does not give the peak value that occurred in the watershed. There was no any previous study in the study area using empirical model like artificial neural network (ANN) model. The use of artificial neural networks is becoming increasingly common in the analysis of hydrology and water resources problems [3, 6-8, 13, 22] *etc.*

A number of researchers have investigated the potential of artificial neural network in modeling runoff and sediment yield. [1] Compare ANN model with traditional conceptual model in predicting watershed runoff as a function of rainfall, snow water equivalence and temperature. The ANN technique was applied to model watershed runoff in three basins with different climatic characteristic; Fraser river watershed, Raccoon River watershed and little Patuxent river basin. In Fraser River watershed, The ANN technique was applied to model monthly stream flow and was compared to conceptual water balance (watbal) model. The ANN technique was used to model the daily rainfall -runoff process and was compared with the sacramental soil moisture accounting (SAC-SMA) Model in Raccoon River watershed. Daily rainfall -runoff process was also modeled using the ANN Techniques little Patuxent River basin and compared with simple conceptual rainfall-runoff (SCRR). In all cases, the ANN model provided higher accuracy, a more systematic approach and shortened the time spent in training of the model. [9] Forecasted flow using artificial neural network (ANN) and soil assessment tool (SWAT). The authors used the daily flow data of the pracana basin in Portugal. Different combinations of rainfall and flow data with some lag periods were examined as input neurons in the input layer. The authors preferred to find the number of hidden neurons by using trial and error method. Sigmoid transfer function was used in the hidden neuron and for training the network, gradient descent with adaptive learning rate was used. From the result found that SWAT Model was unable to forecast peaks of flow data but ANN Model promisingly forecasted flow values at the peaks as well. The authors also conclude that the ANN Model as a fastest tool for flow forecasting. [21] Developed runoff and sediment yield modeling using ANN and support vector machines: a case study from Nepal watershed. The numerical performance indicators such as root mean square error, coefficient of

efficiency and correlation coefficient were considered to evaluate the performance of developed models. The results of the numerical performance indicators RMSE,  $R^2$ , CC for ANN were 103.67, 0.91, and 0.82 respectively and for SVM were 134.77, 0.85 and 0.68. Finally, they concluded that ANN being a computationally intensive method, SVM could be used as an efficient alternative for runoff and sediment yield predictions under comparable accuracy in predictions. [16] Compare Soil and Water Assessment Tool (SWAT) and Artificial Neural Network (ANN) models for Daily Runoff Simulation in Different Climatic Zones of Peninsular Spain. The performance SWAT and ANN Models have been evaluated with different periods of flows using regional flow duration curves (FDCs) such as very low, low, medium, high and very high flow period. The results indicate that SWAT and ANNs were generally good tools in daily stream flow modeling. However, SWAT was found to be more successful in relation to better simulation of lower flows, while ANNs were superior at estimating higher flows in all cases. [4] Apply ANN model for modeling stream flow, sediment transport and erosion rate of high-altitude river system in western Himalaya, uttarakhand. The study used Scaled Conjugate Gradient (SCG), Bayesian Regularization (BR), and Levenberg-Marquardt (LM) training algorithms to simulate the stream flow and Suspended Sediments Concentration (SSC). L-M based-ANN model result shows that the simulated results tracked the stream flow as well as SSC with the desired accuracy. [29] Used an ANN model to forecast suspended sediment in the upper Yangetze watershed of the Langchuanjiang River in China over a monthly time step. For the ANN network, average rainfall, temperature, rainfall intensity, and water discharge were used as inputs to forecast suspended sediment. [17] Used ANN to simulate daily suspended sediment concentration at two stations on the Tongue River in Montana, USA. In order to estimate the sediment concentration, he examined several

combinations of inputs, such as water discharges at both the current and past time steps, sediment concentrations at the station of interest at previous time steps, as well as data from the upstream station. This research showed that the application of ANN made it possible to model sediment, including its concentration in rivers or flux from slopes or watersheds.

The following factors also led to the ANN model being chosen for the maybar Watershed.

- 1) It requires less input data to run the model
- 2) Now a day, ANN model widely used to model a variety of non-linear hydrologic process. [3, 6-8, 22]
- 3) Since maybar meteorological station is not class one, only rainfall and temperature are present.

This study aims to address the following

- 1) To identify which hydrological and metrological parameter affects the runoff and sediment yield of the study area.
- 2) To develop the best ANN Network architecture to model Runoff and sediment yield of the study area.
- 3) To predict runoff and sediment yield at Maybar watershed outlet or at gauged site

## 2. Materials and Methods

### 2.1. Study Area Description

The Maybar watershed is located in the Amahra National Regional State, South Wollo Zone, Albuko district, Ethiopia. Maybar watershed part of Awash River basin (Figure 2) has an area of 113 ha, which is located in the highlands of wollo. The study area includes the Korisheleko River catchment. The altitude ranges from 2484 to 2850 meter above sea level. The gauging station located at the lower end of watershed of Korisheleko River and lies at  $39^{\circ}39'20''\text{E}$  and  $11^{\circ}0'51''\text{N}$ .

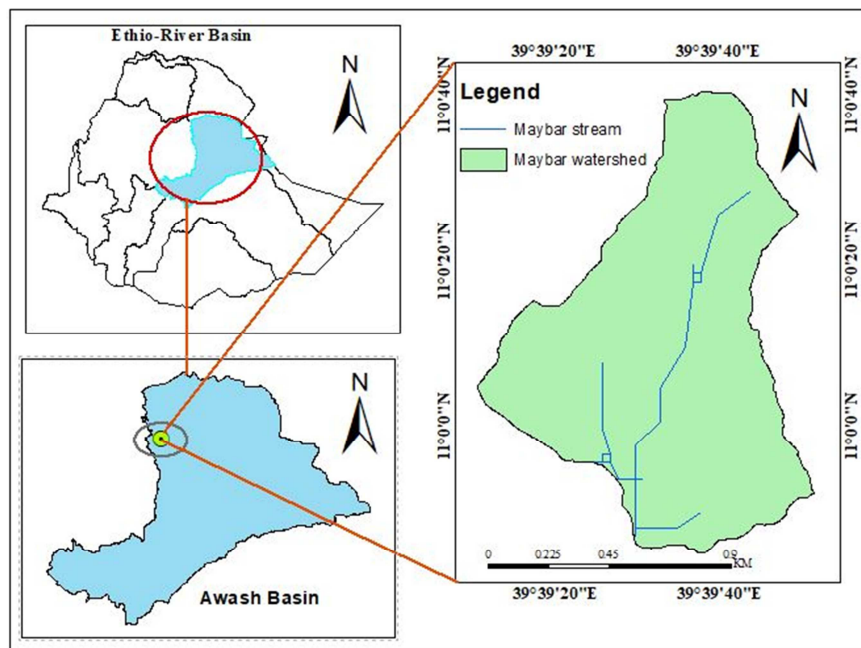


Figure 1. Location Map of study area.

The climate of Maybar watershed is categorized as *Dega* (cool) zone and dominated by two distinct periods. These are wet and dry periods. The wet season starts in June and ends in September whereas the dry season starts in November and ends in April. The remaining two months (May and October) are transition periods. May is the transition month from dry to wet season, while October is the transition month from wet

to dry season. The type of soil in Maybar watershed are Eutric Cambisols and Eutric Regosols. The most dominant soil type in the watershed is Eutric Cambisols encompassing the majority of the watershed. The land use land cover of the study area can be mainly categorized as cultivated land, open woody vegetation, grassland and open shrubs. But the dominant land use in Maybar is cultivated land.

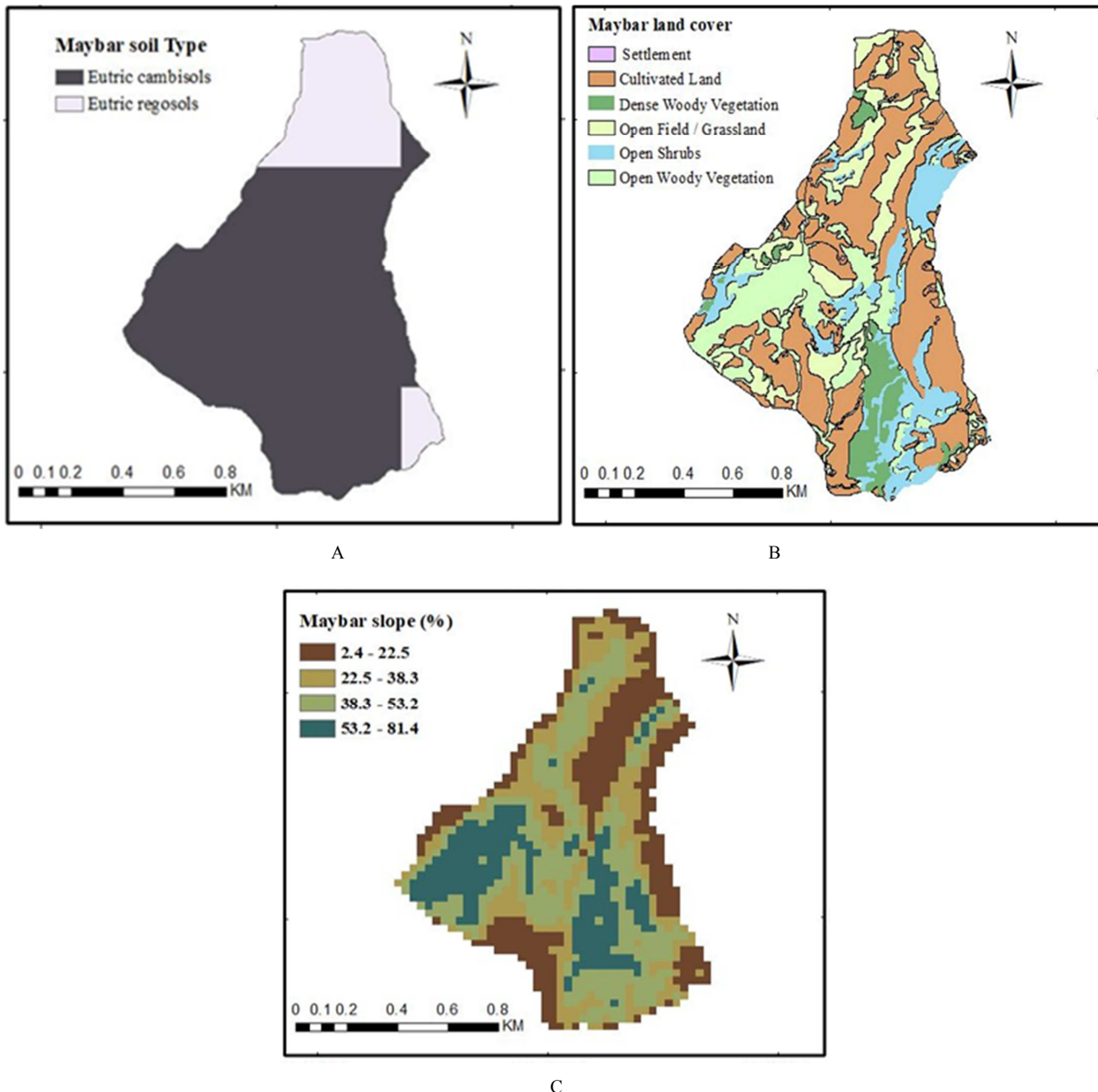


Figure 2. a) Soil type b) land use type c) slope map.

## 2.2. Data Description

### Observed hydro-meteorological

The main data set for runoff as well as for sediment modeling using ANN model are meteorological and hydrological data. The meteorological data include daily

precipitation (mm), maximum and minimum temperature ( $^{\circ}\text{C}$ ) of Maybar station which obtained from (WALRS). Hydrological data includes daily stream flow ( $\text{m}^3/\text{s}$ ) and sediment data (ton/day) of Maybar gauging station which obtained WALRS. Both

Meteorological and hydrological data has 9-year (2010-2018) record length. Maybar watershed have mean daily

minimum temperature of 9.88°C, mean daily minimum temperature 21.82°C and mean daily minimum temperature 15.8°C. Moreover, it has minimum mean monthly temperature at December and maximum mean monthly temperature at June as showed in figure 4. Based on the rainfall, runoff and sediment pattern shown in Figure 3, the watershed has higher rainfall, runoff and sediment in July and August and lower from December to February.

### 2.3. Methods

#### Observed data quality testing

After data collection, checking the quality of data is the first step in hydrological modeling. This process used to provide accurate and precise data to the model and to minimize false predicted result. In order to check the quality of data, outlier test, homogeneity, consistency test and filling missing data were carried out. For rainfall and temperature data, missing data were filled by normal ratio method. For checking the consistency and homogeneity of data, double mass curve and non-dimensional plot techniques were used respectively. Grubbs and Beck (1972) test (G-B) is used to detect outliers in the data.

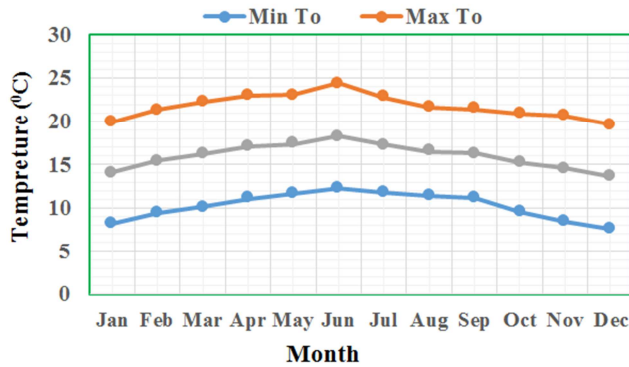


Figure 3. Mean Monthly Air Temperatures, and Mean Monthly Minimum and Maximum air Temperature of Maybar Watershed.

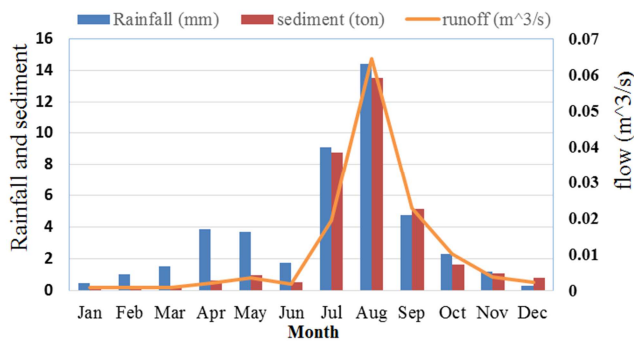


Figure 4. Mean Monthly Stream Flow and sediment at Maybar Gauging Station (2010-2018).

After the quality of data checked, Potential Evapotranspiration was calculated by using the Hargreaves method with minimum, maximum temperature and latitude of station. It is one of the potential inputs for ANN model during runoff as well as sediment modeling.

Hargreaves and samani developed a simplified equation

requiring only temperature, day of the year and latitude of station. It is calculated by the formula.

$$ET = 0.0023(T_{mean} + 17.8)(T_{max} - T_{min})^{0.5} * Ra$$

$$Ra = \frac{24(60)}{\pi} * G_{sc} * dr[\omega s \sin(\phi) \sin(\sigma) + \cos(\phi) \cos(\sigma) \sin(\omega s)]$$

$$\omega s = \arccos[-\tan(\phi) \tan(\sigma)]$$

$$dr = 1 + 0.033 \cos\left(\frac{2\pi}{365} * J\right)$$

$$\sigma = 0.409 * \sin\left(\frac{2\pi}{365} * J - 1.39\right)$$

Where Ra=extraterrestrial radiation in hour (short period) (MJ m<sup>-2</sup> h<sup>-1</sup>),

J=day of the year,

Φ=latitude of the station m<sup>-2</sup> h<sup>-1</sup>,

G<sub>sc</sub> is solar constant =0.082MJ and dr=inverse distance earth-sun.

### 2.4. Artificial Neural Network Model

A human brain is not capable of solving complex data and cannot extract information from compound structures. To overcome this lack of ability to resolve complex problems, Warren McCulloch and Walter Pitts came up with a mathematical model. This model is called Artificial Neural Networks (ANN), which falls under artificial intelligence.

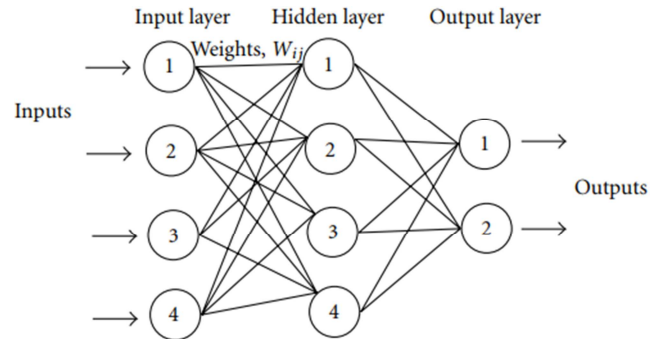


Figure 5. Three layered feed forward ANN [9].

Based on network architecture, there are three fundamental classes of networks. These are single layer feed-forward, multi-layer feed forward and recurrent network [15]. The most popular neural network in a Variety of studies is the multi-layer feed forward that is trained with back-propagation [25]. In this study, a multi-layer feed forward that was trained with a back-propagation was used to train the ANN model for both runoff and sediment. In multi-layer feed forward network the information flow and processing is from input layer to hidden layer and from hidden layer to output layer. The number of neuron in input layer and output layer depends on the type of problem where as the number of hidden layer and neuron in hidden layer is decided by trial and error approach.

A synaptic weight is assigned to each link to represent the



relative connection strength of two nodes at both ends in predicting the input output relationship. The output,  $y_j$  of any node

$$Y_i = f(\sum_{i=1}^m W_i X_i + b_i) \quad (1)$$

Where  $X_i$  is the input received at node  $j$   
 $W_i$  is the weight assigned to each neuron  
 $m$  is the total number of inputs to node  $j$  and  
 $b_j$  is the node threshold.

Function  $f$  is an activation function which converts the weighted sum in to output. In the majority of studies, the logistic sigmoid function is used [7]. So in this study, logistic sigmoid function was used to convert input data from the input layer to the output layer through a hidden layer in which this processing taken place. Sigmoid function is given by

$$f(x) = \frac{1}{1+e^{-x}} \quad (2)$$

#### ANN Model Development

Neural net tool (nntool) package available in MATLAB 2019 b was used to train as well also validate ANN model for predicting runoff and sediment yield of Maybar watershed. There are no fixed rules for developing an ANN model. For this study the following steps were used to develop ANN model for predicting runoff and sediment yield of Maybar watershed.

#### Model Input Selection

One of the main steps in the ANN model development is the determination of an appropriate set of inputs because the overall output of the model is highly dependent on input selection. The variables that have influence on runoff and

sediment yield of catchment are numerous. But for present study, to select appropriate input vector for runoff and sediment yield modeling, auto correlation and cross correlation analysis between input and output parameters were done using SPSS. Autocorrelation is correlation between the same parameter with its antecedent value, whereas Cross correlations between the different parameter with its antecedent value.

#### Data Normalization

Data normalization refers to the rescaling of data to a standard normal distribution. Since the input and target data are different in SI unit, data normalization is used to treat equally and change this data in to numeric value (0, 1). It is used to avoid computational problem and to facilitate learning of network. It is given by

$$N_i = \frac{R_i - \text{Mini}}{\text{Maxi} - \text{Mini}} \quad (3)$$

Where  $R_i$  - is the real value applied to node  $i$ ;  $N_i$  - is the subsequent standardized value calculated for node  $i$ ;  $\text{Mini}$  is the minimum value of all values applied to node  $i$  and  $\text{Maxi}$  is the maximum value of all values applied to node  $i$ .

After the data is trained and test the result is denormalized in to original unit of measurement.

#### Data Division

The available data set is generally divided in to two parts. These are training set for training the network and validation set for validation of the model. So, the first 7 years (2010-2016) data is selected for model training and the remaining 2 years (2017 and 2018) data is used for testing the performance of the trained model. The training set has input and target (output) data whereas the testing set have only input data.

**Table 1.** Statistical analysis of observed data at maybar watershed.

Data	Training set				Testing set			
	min	Max	mean	SD	min	max	mean	SD
Rainfall (mm)	0	72.5	3.54	8.15	0	59.69	4.37	8.39
Flow (m <sup>3</sup> /s)	0	0.35	0.01	0.0289	0.00029	0.23	0.013	0.024
Sediment (ton/day)	0	681.9	2.41	19.52	0.15	61.86	4.47	7.258

#### Model Architecture Selection

Number of input and output nodes depends on training set in hand. The study of Marcoulides, (2005) argued that choosing number of nodes in hidden layer could be challenging task. The number of layers in the hidden layer and number of hidden neurons within these layers are obtained through a trial –and- error approach. However, the study of Shu and Oudarda, (2007) recommended that number of hidden nodes in hidden layer should be less than twice the number of input nodes. But for this study, the number of hidden nodes in hidden layer was determined through a trial –and- error approach from one to ten hidden nodes.

#### Selection of Training Algorithm

In majority of studies the fed forward neural network (FFNN) is trained using the error back propagation algorithm [7]. So, in this study ANN model was trained using FFNN for both runoff and sediment yield modeling.

#### Model Training (Calibration) and testing (validation)

After the data is divided, neural network is trained by adjusting the weights that link its neurons. To start training, first the ANN architecture and training parameters are adjusted. During training, the value of parameters was varied iteratively within an allowable range until the simulated value was as close as observed value. The training phase of the ANN model was terminated (stopped) when the mean square error was minimal and the R (all) value was maximum on the regression plot. At the start of the training, R (all) is small and gradual, increasing to a certain level and remaining constant. Then the training process stopped when R (all) unchanged. Therefore, the ANN model found with R approaching to one and MSE approaching zero among the selected epoch was adopted for further testing (cross validation).

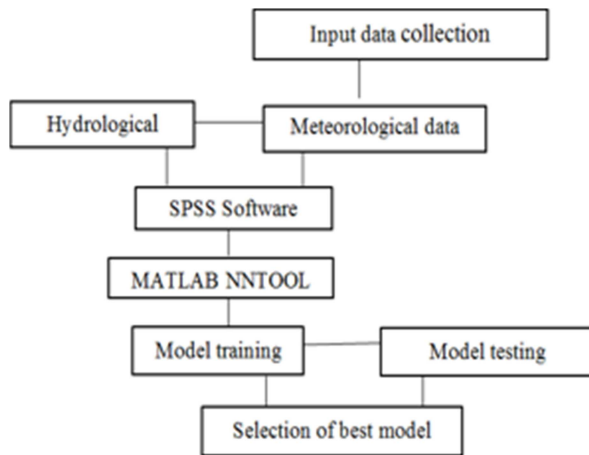


Figure 6. General methodology of the study.

## 2.5. Evaluation of Model Performance

There are various measures to evaluate the model performance during the calibration and validation periods. For this study, Nash and Sutcliffe simulation efficiency (NSE) Root mean square error (RMSE) and Correlation of determination ( $R^2$ ) were used to assess the prediction ability of the model.

$$NSE = 1 - \frac{\sum_{i=1}^{\infty} (O_i - P_i)^2}{\sum_{i=1}^{\infty} (O_i - P_{avg})^2} \quad (4)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N [(Q_o)_i - (Q_p)_i]^2} \quad (5)$$

$$R^2 = \left[ \frac{\sum_{i=1}^N (O_i - O_{avg})(P_i - P_{avg})}{(\sum_{i=1}^N (O_i - O_{avg})^2 \sum_{i=1}^N (P_i - P_{avg})^2)^{0.5}} \right]^2 \quad (6)$$

Where:  $O_i$  is observed flow or sediment,  $P_i$  is predicted flow or sediment,  $O_{avg}$  is average observed flow or sediment,  $P_{avg}$  is average predicted flow or sediment.

## 3. Result and Discussion

### 3.1. Input Selection

Appropriate Input variable for runoff and sediment yield modeling were selected using autocorrelation and cross correlation result of SPSS. A Pearson correlation has a value between -1 and 1 that indicate the linear relation extent of two quantitative variables. A Pearson correlation of 0 means there is no linear relation b/n two variable, a value of approach to 1 indicates a strong ascending linear relationship and a value of -1 indicates a perfect descending linear relation.

Table 2. SPSS correlation result for input selection of Maybar watershed runoff modeling.

#### a. Autocorrelation of runoff with antecedent runoff

		Correlations					
		Q(t)	Q(t-1)	Q(t-2)	Q(t-3)	Q(t-4)	Q(t-5)
Q(t)	Pearson Correlation	1	.768**	.707**	.643**	.604**	.528**
	N	3287	3287	3287	3287	3287	3287

#### b. Cross correlation of runoff with rainfall

		correlations					
		Q(t)	P(t)	P(t-1)	P(t-2)	P(t-3)	P(t-4)
Q(t)	Pearson Correlation	1	.597**	.540**	.491**	.477**	.449**
	N	3287	3287	3287	3287	3287	3287

#### c. Cross correlation of runoff with mean temperature

		Correlations					
		Q(t)	Tm(t)	Tm(t-1)	Tm(t-2)	Tm(t-3)	Tm(t-4)
Q(t)	Pearson Correlation	1	.036**	.047**	.055**	.062**	.066**
	N	3287	3287	3287	3287	3287	3287

#### d. Cross correlation of runoff with evapotranspiration

		Correlations					
		Q(t)	Eto(t)	Eto(t-1)	Eto(t-2)	Eto(t-3)	Eto(t-4)
Q(t)	Pearson Correlation	1	-.118**	-.122**	-.103**	-.098**	-.095**
	N	3287	3287	3287	3287	3287	3287

For runoff modeling, the inputs were selected based on their Pearson correlation value. As shown in the table 2, pt, Qt-1, Qt-2, Qt-3 and Qt-4 have a Pearson correlation value that approaches one with runoff. Mean temperature has a Pearson correlation value approaching zero and evapotranspiration has a negative Pearson correlation with runoff. As a result, pt, Qt-1, Qt-2, Qt-3 and Qt-4 are input

variables that have Significant influence on the runoff of Maybar watershed, whereas mean temperature and evapotranspiration has no significant influence on runoff, so they are not considered as input variables for the ANN model during runoff modeling. Therefore,

$$Q_t = f(pt, Q_{t-1}, Q_{t-2}, Q_{t-3} \text{ and } Q_{t-4}) \quad (7)$$

Where  $Q(t)$  = Discharge at time  $t$ ,  $P_t$  = Precipitation at time  $t$ ,  $Q_{t-1}$ ,  $Q_{t-2}$ ,  $Q_{t-3}$  and  $Q_{t-4}$  are discharge at  $t-1$ ,  $t-2$ ,  $t-3$  and  $t-4$  respectively. Not that  $t$  is in days.

**Table 3.** SPSS correlation result for input selection of Maybar Sediment yield modeling.

a). Autocorrelation of runoff with antecedent runoff

		Correlations					
		Ss(t)	Ss(t-1)	Ss(t-2)	Ss(t-3)	Ss(t-4)	Ss(t-5)
Ss(t)	Pearson Correlation	1	.128**	.107**	.144**	.158**	.041**
	N	3287	3287	3287	3287	3287	3287

b). Cross correlation of sediment with runoff

		Correlations					
		Ss(t)	Q(t)	Q(t-1)	Q(t-2)	Q(t-3)	Q(t-4)
Ss(t)	Pearson Correlation	1	.398**	.266**	.244**	.234**	.204**
	N	3287	3287	3287	3287	3287	3287

c). Cross correlation of sediment with mean Rainfall

		Correlations					
		Ss(t)	P(t)	P(t-1)	P(t-2)	P(t-3)	P(t-4)
Ss(t)	Pearson Correlation	1	.253**	.186**	.187**	.194**	.150**
	N	3287	3287	3287	3287	3287	3287

d). Cross correlation of sediment with mean temperature

		Correlations					
		Ss(t)	Tm(t)	Tm(t-1)	Tm(t-2)	Tm(t-3)	Tm(t-4)
Ss(t)	Pearson Correlation	1	.048**	.053**	.064**	.048*	.052**
	N	3287	3287	3287	3287	3287	3287

e). Cross correlation of sediment with evapotranspiration

		Correlations					
		Ss(t)	Eto(t)	Eto(t-1)	Eto(t-2)	Eto(t-3)	Eto(t-4)
Ss(t)	Pearson Correlation	1	-.03	-.031	-.019	-.031	-.034
	N	3287	3287	3287	3287	3287	3287

From the above Pearson correlation results table 3, observed sediment has a high Pearson correlation with current rainfall ( $P_t$ ), current runoff ( $Q_t$ ), and previous one-day runoff ( $Q(t-1)$ ) when compared to other input variables. As a result, current rainfall ( $P_t$ ), current runoff ( $Q_t$ ), and previous one-day runoff ( $Q(t-1)$ ) are chosen as input variables with a significant influence on the sediment yield of the Maybar watershed. On the other hand, observed sediment has poor correlation with previous sediment, mean temperature and Negative correlation with evapotranspiration. So, it has no significant influence and is not considered an input variable for sediment yield modeling.

Therefore

$$Ss(t) = f(P_t, Q(t) \& Q(t-1)) \quad (8)$$

Where  $Ss(t)$  =suspended sediment at time  $t$ .

### 3.2. Input Combination

The selected inputs are combined from one input to five inputs for runoff modeling and from one to three inputs for sediment yield modeling. In general, a total five and three combination of input variables are investigated for runoff modeling and sediment yield modeling respectively.

**Table 4.** Proposed ANN model for Maybar runoff modeling.

Model	Inputs	Output
A	$P(t)$	$Q(t)$
B	$p(t)$ $Q(t-1)$	$Q(t)$
C	$p(t)$ $Q(t-1)$ $Q(t-2)$	$Q(t)$
D	$P(t)$ $Q(t-1)$ $Q(t-2)$ $Q(t-3)$	$Q(t)$
E	$P(t)$ $Q(t-1)$ $Q(t-2)$ $Q(t-3)$ $Q(t-4)$	$Q(t)$

**Table 5.** Proposed ANN model for Maybar sediment yield modeling.

Model	Inputs	Output
I	$p(t)$	$Ss(t)$
II	$p(t)$ $Q(t)$	$Ss(t)$
III	$p(t)$ $Q(t)$ $Q(t-1)$	$Ss(t)$

#### ANN model training

For modeling runoff, the multi-layered feed forward perceptron back propagation (MLFFP) algorithm with sigmoid transfer function was chosen to create the artificial neural network model. Three-layer network architecture consisting of input, hidden and output layers was selected. As showed in table 6, the input layer was combined from one to five inputs of daily areal rain fall ( $P_t$ ), one-day stream flow lag ( $Q_{t-1}$ ), two-day stream flow lag ( $Q_{t-2}$ ), three-day stream flow lag ( $Q_{t-3}$ ) and four-day stream flow lag ( $Q_{t-4}$ ). The ANN model



performance for Maybar watershed was checked by observing the R and MSE values at training testing and validation stages to select the appropriate neurons in hidden layers with different input parameters. The appropriate numbers of hidden neurons were selected on a trial- and- error basis. The individuals with different combinations of inputs were evaluated for a different number of neurons (1 up to 10) in the hidden.

Like Runoff modeling, sediment modeling of Maybar watershed was simulated using MATLAB 2029 b run with nn tool package, and it follows the steps and procedures when training and testing the model. As showed in table 7, the input was selected varying from single to three in combination of daily rainfall (P(t)), present runoff (Q(t) and runoff lag by one day (Q(t-1)).

**Table 6.** Input parameter and ANN Structure for runoff modeling for Maybar watershed.

Model	Model input parameter	No of input parameter	No neurons in hidden layer	Output layer	Model structure
A	$Q(t)=f(Pt)$	1	1	1	1 1 1
B	$Q(t)=f(Pt, Qt-1)$	2	3	1	2 3 1
C	$Q(t)=f(Pt, Qt-1, Qt-2)$	3	5	1	3 5 1
D	$Q(t)=f(Pt, Qt-1, Qt-2, Qt-3)$	4	7	1	4 7 1
E	$Q(t)=f(Pt, Qt-1, Qt-2, Qt-3, Qt-4)$	5	9	1	5 9 1

**Table 7.** Input variable and selected ANN Structure for sediment modeling of Maybar watershed.

Model	Model input parameter	No of input parameter	No neurons in hidden layer	Output layer	Model structure
I	$Ss(t)=f(P(t))$	1	6	1	1 6 1
II	$Ss(t)=f(P(t), Q(t))$	2	3	1	2 3 1
III	$Ss(t)=f(P(t), Q(t), Q(t-1))$	3	6	1	3 6 1

### 3.3. Comparison of Proposed Model Result

Runoff modeling: From performance result showed in table 8, it was observed that adding lag of runoff as input increases the model efficiency up to a certain level and decreases it again (model B to model E) during the training period. During the testing period, adding one day lag of runoff as input increased model efficiency while adding more than one day lag of runoff decreased model efficiency (model B to model E). The network configuration with two inputs, three hidden neurons in the hidden layer and a unique output represented by model D provides the best performance during the training period with the highest  $R^2$  (0.818) compared with another proposed model. But during the testing period, the best performance was achieved for model B. Model D resulted in a Nash Sutcliff Efficiency of 81.61% during the training period and 65.04% during the testing period. It was also observed that model B resulted in a Nash Sutcliff efficiency of 75.4% during the training period and 76% during the testing period. Since testing periods are used to select the best performing network, it is considered the performance of a testing period to compare one model to another model. During the testing period, Model B outperformed Model D and the other proposed models in terms of Nash Sutcliff efficiency. Therefore, model B (2 inputs, 1 hidden layer with 3 neurons and 1 output) was

better as compared to another proposed model. Therefore, model B was selected as the best performing ANN model for the prediction of daily runoff in Maybar watershed.

Sediment yield modeling: The three proposed models for predicting sediment yield were evaluated their performance using performance measures ( $R^2$ , RMSE and NSE) for both training and Testing period. The higher the value of  $R^2$  and NSE and the smaller the value of RMSE indicates the performance of the model is good and vice versa. As showed in table 9 below, model I and II have an  $R^2$  and NSE value of between 0.45 and 0.55 in training and less than 0.4 in testing period. So, model I and II gives satisfactory result in training and unsatisfactory result in model testing. Even if model I and II gives satisfactory result in training period, the model is not applicable for modeling sediment yield because models predicting capability is highly dependent on performance of testing period. Model III (3 6 1) have  $R^2$  (0.584), RMSE (12.59) and NSE (0.58) in training and  $R^2$  (0.588), RMSE (4.9) and NSE (0.58) value in testing. Model III gives an  $R^2$  and NSE value of between 0.55 and 0.65 in both training and testing period. Based on performance rating for daily time step, model III considered as good. So, for this study model III (3 inputs, 1 hidden layer with 6 neurons and 1 output) was better as compared to another proposed model. Therefore, model III was selected as the batter performing ANN model for the prediction of daily sediment yield of Maybar watershed.

**Table 8.** Performance value of proposed Maybar ANN runoff model during training and testing period.

Model No.	Training (2010-2016)			Testing (2017-2018)		
	$R^2$	RMSE ( $m^3/s$ )	NSE	$R^2$	RMSE ( $m^3/s$ )	NSE
A	0.41	0.023	40.06	0.188	0.023	11.09
B	0.755	0.014	75.4	0.76	0.012	76
C	0.785	0.0133	78.45	0.69	0.014	68.74
D	0.818	0.012	81.7	0.67	0.015	65.04
E	0.761	0.014	76.04	0.716	0.013	70.94

**Table 9.** Performance value of proposed Maybar ANN sediment model for both training and testing period.

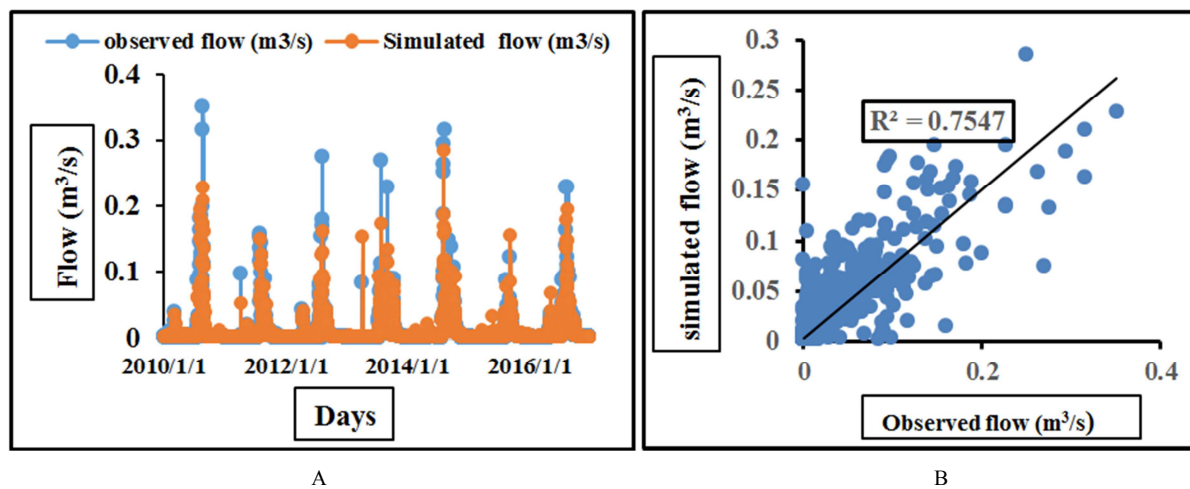
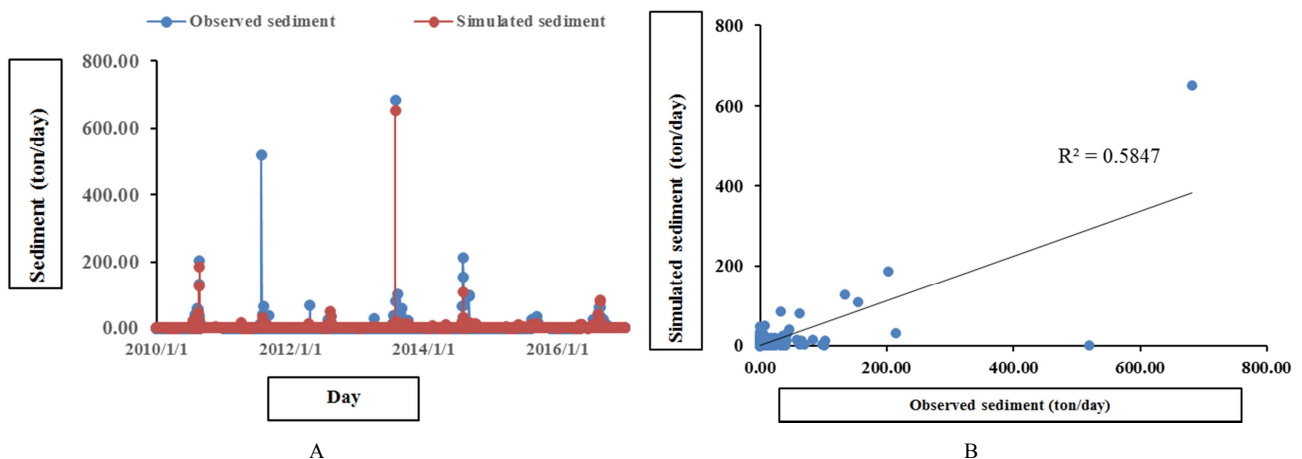
Model No	Training (2010-2016)			Testing (2017-2018)		
	R <sup>2</sup>	RMSE (ton/day)	NSE	R <sup>2</sup>	RMSE (ton/day)	NSE
I	0.5	14.19	0.45	0.21	6.7	0.19
II	0.53	13.37	0.53	0.13	6.9	0.09
III	0.584	12.59	0.58	0.588	4.9	0.58

### 3.4. Comparison of Observed and Simulated Value

#### Training phase

Runoff modeling: 2557 data set of rainfall and previous one day runoff data was used and the performance of the model was compared between the observed and simulated runoff. As showed in figure 7, model B attains very good agreement between observed and simulated runoff on a daily basis. The statistical indicator R<sup>2</sup> value is equal to 0.755, which indicates that this model can be used for simulating runoff of Maybar watershed.

Sediment yield modeling: the model was trained using 7 years of data from 2010-2016 and a total of 2557 p(t), Q(t) and Q(t-1) data sets were used as input and sediment data as target. the comparisons of daily observed sediment with daily simulated sediment are shown in figure 8 in both time series and scatter plot. From the figure, the model shows good agreement between daily observed and simulated sediment in training (calibration). It gives coefficient of determination (R<sup>2</sup>) value of 0.584. So based on performance criteria, ANN model can be applicable for Maybar watershed to simulate daily sediment yield.

**Figure 7.** Comparison of daily simulated flow with observed flow of model B during training A) Time series plot B) Scatter plot.**Figure 8.** Comparison of simulated sediment with observed sediment of model III during period A) Time series plot B) Scatter plot.

#### Testing phase

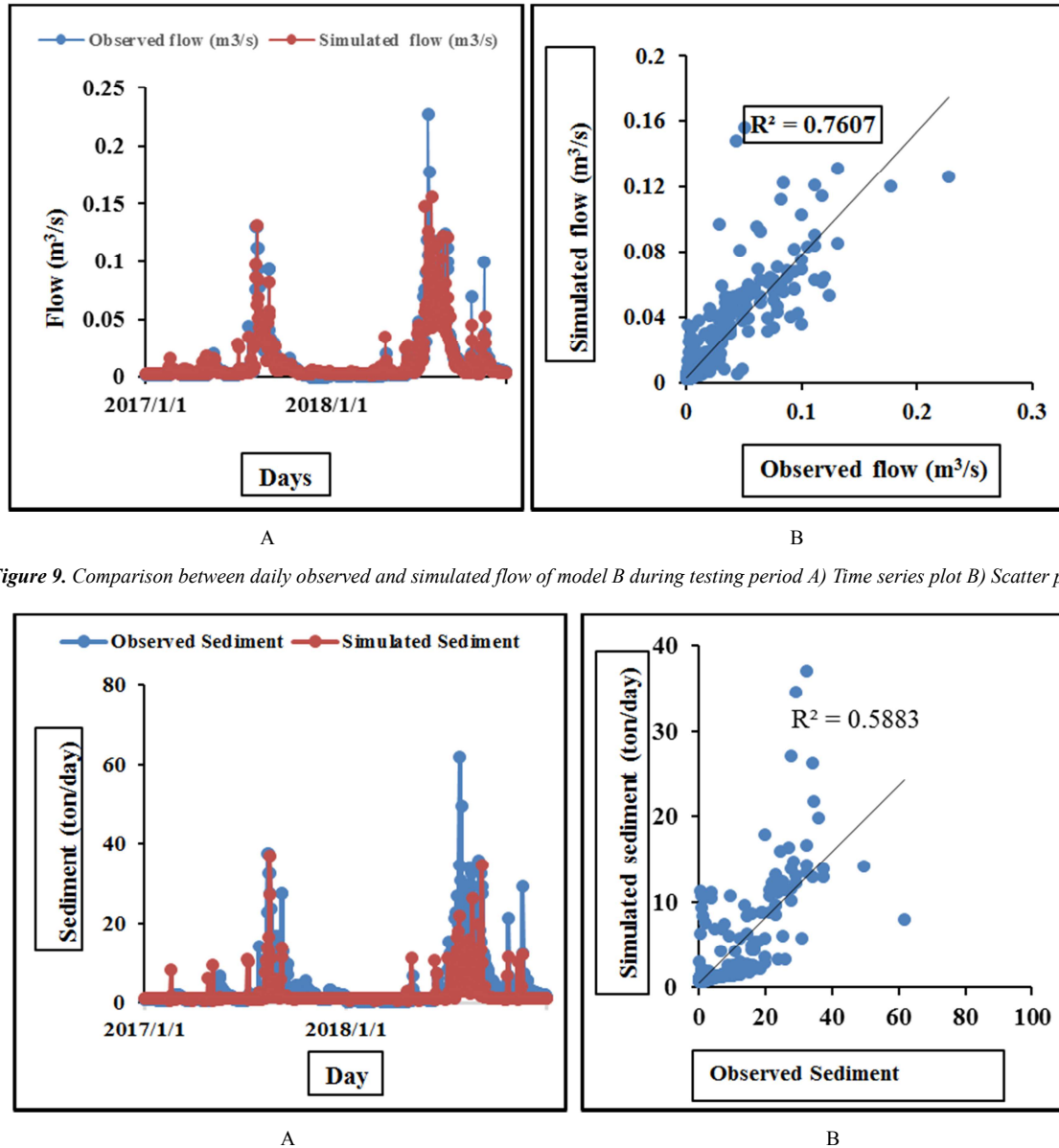
Runoff modeling: Testing of the trained ANN model was done using an independent data set of two years from 2017 to 2018 without adjustment of parameters. A total of 730 data set of p(t) and Q(t-1) were used as input and not have target.

It can be seen from the time series plot that, there is very good agreement between the observed and simulated flow. And also, from the scatter graph between daily simulated and observed flow, the statistical indicator R<sup>2</sup> value is equal to 0.76, which indicates that this model can be used for

simulating runoff values of the Koresheleko River.

Sediment yield modeling: Testing of the trained ANN model was done using two years data from 2017 to 2018 without adjustment of parameters. It can be seen from the time series plot that, there is agreement between the observed

and simulated sediment yield. Moreover, from the scatter graph between daily simulated and observed sediment yield, the statistical indicator  $R^2$  value is equal to 0.588, which indicates that this model can be used for simulating sediment of Korisheleko River.



**Figure 9.** Comparison between daily observed and simulated flow of model B during testing period A) Time series plot B) Scatter plot.

**Figure 10.** Comparison of simulated sediment with observed sediment of model III during testing period A) Time series plot B) Scatter plot.

The peak observed and simulated runoff in the Maybar watershed were  $0.35m^3/s$  and  $0.285 m^3/s$  in the training period and  $0.2287m^3/s$  and  $0.156 m^3/s$  in the testing period, respectively. The model produced good simulation results for the Maybar watershed in daily time steps. From the hydrograph, it has been observed that the model has a tendency to some underestimation in the high flow period and overestimation in the low flow period during model training and testing.

The peak observed and simulated sediment of Maybar watershed were 689.9 ton/day and 651.4 ton/day in training period and 61.86 ton/day and 37.1 ton/day in testing period

respectively. The mean annual observed and simulated suspended sediment Rates were 1042 and 850 tons/year, which is approximately 9.2 and 7.53 tons/ha/year respectively. Similarly as shown in training and testing hydrograph of daily observed and simulated sediment load, the model under estimated the high sediment load and overestimated the low sediment load or no sediment load. Even if the result is within acceptable evaluation rating, there were a weak goodness fit. The weak goodness fit is resulted from errors in measured data. Errors in measured sediment data mainly result from sampling errors. In general, the weak performance of the model may result from human and

instrumental errors during sediment data measurement. In general ANN model produced good simulation results for maybar watershed in daily time steps. So ANN mode is applicable model for runoff and sediment simulation of

maybar watershed. The model also applicable for watershed which have limitted data availability even if only precipitation present in the watershed.

**Table 10.** Summery statistical properties of Maybar ANN Runoff modeling result.

Statistical Property	Training		Testing	
	Observed (m <sup>3</sup> /s)	Simulated (m <sup>3</sup> /s)	Observed (m <sup>3</sup> /s)	Simulated (m <sup>3</sup> /s)
Minimum	0	0.019	0	0.002
Maximum	0.35	0.285	0.227	0.156
Average	0.01	0.011	0.013	0.0126
Standard Deviation	0.029	0.0245	0.025	0.021
Coefficient of Variation	2.9	2.28	1.92	1.66

**Table 11.** Summery statistical properties of Maybar ANN sediment modeling result.

statistical property	Training		Testing	
	Observed (ton/day)	Simulated (ton/day)	Observed (ton/day)	Simulated (ton/day)
Minimum	0	0.008	0.15	0.64
Maximum	691.9	651.4	61.86	37.1
Average	2.363	2.41	4.42	2.21
Standard Deviation	14.31	19.52	7.258	3.62
Coefficient of Variation	6.055	8.1	1.64	1.64

## 4. Conclusion

Accurate runoff and sediment yield estimation is crucial for sustainable watershed management. In this paper the applicability artificial neural network model approach to simulate the amount of runoff and sediment yield of Maybar watershed was investigated. From SPSS correlation result, present rainfall and previous one up to four-day runoff have strong relation with maybar runoff and similarly, present runoff, previous one-and two-day runoff have strong relation with maybar sediment yield. Temperature and evapotranspiration have little influence for maybar runoff and sediment. Adding more lag of runoff as input increases the model efficiency up to a certain level and decreases it again during the training period. During testing period, adding one day lag of runoff as input increased model efficiency while adding more lags of runoff decreased model Efficiency. From the five-runoff proposed model, model B (2 inputs, 3 hidden neurons, 1 output) gives a high coefficient of determination ( $R^2$ ) and Nash Sutcliff efficiency (NSE) and low root mean square error (RMSE). Its coefficient of determination ( $R^2$ ), root mean square error (RMSE) and Nash Sutcliff efficiency (NSE) are found to be 0.755, 0.014 and 75.4% in training and 0.76, 0.012 and 76% in testing, respectively. From the three proposed sediment models, model III (3 inputs, 6 hidden neurons, 1 output) selected as the best ANN model for modeling Maybar sediment. Its coefficient of determination ( $R^2$ ), root mean square error (RMSE) and Nash Sutcliff efficiency (NSE) are found to be 0.584, 12.59 and 58% in training and 0.588, 4.9 and 5% in testing respectively The total mean annual sediment yield loading from Maybar watershed simulated by ANN was 7.53 ton/ha/year. Generally, the study confirmed that ANN model is able to simulate the runoff and sediment yield of Maybar watershed.

## Data Availability

The data sets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

## Author Contributions

Hussen Ali Hassen and Fikru Fentaw Abera designed the technical route of the study, analyzed the data and wrote the manuscript. Dagmawi Negash, Mehret Ayana and Yonatan Tibebe proposed suggestions to improve the quality of the manuscript. The author has read and agreed to the published version of the manuscript.

## Aknowledgments

The author would like to express his gratitude to Kombolcha Institute of Technology, Ethiopian Ministry of Water and Energy, Ethiopian Meteorological Service Agency and land and water resource center for providing stream flow and rainfall data.

## Conflicts of Interest

The authors declare no conflict of interest.

## References

- [1] A. sezin tokar and momcilo markus. 2015. "Precipitation Runoff Modeling Using Artificial Neural Network and Conceptual PRECIPITATION-RUNOFF MODEL." 0699 (November).
- [2] Abirendra Bharti, Ashish Pandey, S. K. Tripathi and Dheeraj Kumar. 2017. "Modelling of Runoff and Sediment Yield Using ANN, LS-SVR, REPTree and M5 Models." 1489–1507.

- [3] Agarwal, Avinash, R K Rai, and Alka Upadhyay. 2014. "Forecasting of Runoff and Sediment Yield Using Artificial Neural Networks." (May).
- [4] Alegre, Porto et al. 2022. "Application of ANNs for the Modeling of Streamflow, Sediment Transport, and Erosion Rate of a High-Altitude River System in Western Himalaya, Uttarakhand.": 1–15.
- [5] Andredaki, M., A. Georgoulas, V. Hrisanthou, and N. Kotsovinos. 2014. "Assessment of Reservoir Sedimentation Effect on Coastal Erosion in the Case of Nestos River, Greece." *International Journal of Sediment Research* 29(1): 34–48.
- [6] Chakravarti, Ankit, Nitin Joshi, and Himanshu Panjiar. 2015. "Rainfall Runoff Analysis Using Artificial Neural Network." 8(July).
- [7] Dawson, C W, and R L Wilby. 2001. "Hydrological Modelling Using Artificial Neural Networks." 1: 80–108.
- [8] Dawson, Christian W. 1998. "An Artificial Neural Network Approach to Rainfall- Runoff Modelling." 43(February): 47–66.
- [9] Demirel, Mehmet C., Anabela Venancio, and Ercan Kahya. 2009. "Flow Forecast by SWAT Model and ANN in Pracana Basin, Portugal." *Advances in Engineering Software* 40(7): 467–73.
- [10] Devia, Gayathri K., B. P. Ganasri, and G. S. Dwarakish. 2015. "A Review on Hydrological Models." *Aquatic Procedia* 4(Icwrcoe): 1001–7.
- [11] FAO. 2016. "AQUASTAT Country Profile – Ethiopia. Food and Agriculture Organization of the United Nations (FAO). Rome, Italy." *FAO, AQUASTAT reports*: 11–12.
- [12] Gebremedhin Tesfayea, Abera Assefab and Desale Kidanec. 2016. "Runoff, Sediment Load and Land Use / Cover Change Relationship: The Case of Maybar Sub-Watershed, South Wollo, Ethiopia." *Intl. J. River Basin Management* 0(0): 1–13.
- [13] Govindaraju, Rao S. 2000. "A RTIFICIAL N EURAL N ETWORKS IN H YDROLOGY. By the ASCE Task Committee on Application of Artificial Neural Networks in Hydrology 1." (April).
- [14] Habtamu Asrat and Wasihun Gedefaw. 1975. "Maybar Watershed Run off Modeling Habtamu Asrat and Gedefaw Wasihun, Debre Marko s University.": 148–76.
- [15] Haykin, Simon. 2001. "Neural Networks - A Comprehensive Foundation."
- [16] Jimeno-s, Patricia, Javier Senent-aparicio, and David Pulido-velazquez. "A Comparison of SWAT and ANN Models for Daily Runoff Simulation in Different Climatic Zones of Peninsular Spain.
- [17] Kişi, Özgür. 2004. "Multi-Layer Perceptrons with Levenberg-Marquardt Training Algorithm for Suspended Sediment Concentration Prediction and Estimation." *Hydrological Sciences Journal* 49(6): 1025–40.
- [18] Marcoulides, George A. 2005. 100 Journal of the American Statistical Association *Discovering Knowledge in Data: An Introduction to Data Mining*.
- [19] Mohammed, Jemal. 2015. "Stream Flow and Sediment Yield Modeling, The Case Of Gidabo Watershed."
- [20] Nina, Kemal. 2016. "Megech Dam Reservoir Seidement Estimation (Using SWAT Model)." (September).
- [21] Nivesh, Shreya, and Pravendra Kumar. 2017. "Modelling River Suspended Sediment Load Using Artificial Neural Network and Multiple Linear Regression: Vamsadhara River Basin, India." 5(5): 337–44.
- [22] Raju, M Mohan et al. 2011. "Development of Artificial Neural-Network-Based Models for the Simulation of Spring Discharge." 2011.
- [23] Shu, C., and T. B. M. J. Ouarda. 2007. "Flood Frequency Analysis at Ungauged Sites Using Artificial Neural Networks in Canonical Correlation Analysis Physiographic Space." *Water Resources Research* 43(7).
- [24] Singh, Aparajita. 2018. "A Concise Review on Introduction to Hydrological Models." *Global Research and Development Journal for Engineering* 3(10): 14–19.
- [25] Svozil, Daniel, Vladimír Kvasnička, and Jiří Pospíchal. 1997. "Introduction to Multi-Layer Feed-Forward Neural Networks." *Chemometrics and Intelligent Laboratory Systems* 39(1): 43–62.
- [26] Tamene, Lulseged, and Paul L. G. Vlek. 2008. "Soil Erosion Studies in Northern Ethiopia." *Land Use and Soil Resources*: 73–100.
- [27] Tesfa, Yimer. 2015. "Comparative Hydrological Modeling of Maybar Watershed Using Swat-WB and Swat-CN, Awash Basin, Amhara Region." (May).
- [28] Yesuf, Hassen M, Mohammed Assen, Tena Alamirew, and Assefa M Melesse. 2015. "Modeling of Sediment Yield in Maybar Gauged Watershed Using SWAT, Northeast Ethiopia." 127: 191–205.
- [29] Zhu, Yun-mei, X X Lu, and Yue Zhou. 2007. "Suspended Sediment Flux Modeling with Artificial Neural Network: An Example of the Longchuanjiang River in the Upper Yangtze." 84: 111–25.