

Derivation of suspended sediment data for Al-Adhiam watershed-Iraq using artificial neural network model

Mahmoud Saleh Al-Khafaji¹, Mustafa Al-Mukhtar¹ and Ahmed Saud Mohena²

¹ Building and Construction Engineering Department, University of Technology, Baghdad, Iraq

² Ministry of Water Resources, Baghdad, Iraq

Abstract. The mean part of river sediments is suspended sediment load, its prediction and simulation has important significance to manage the water resources and environments. In Iraq, most researchers avoid to fighting in sediment researches when related with hydrological models spatially with that need enough observed sediment data for calibration and validation because the sediment data very limitation or scars. The aim of this study is employing the Artificial Neural Network (ANN) model to estimate the suspended sediment load of Al-Adhaim watershed in Iraq from available measured sediment data, identify the suitable pattern of input and target data sampling and obtaining the best nonlinear equation between the river discharge and suspended sediment load. To this end, the ANN model was training and tested with the available sediment data, which was for water year (1983-1984). Two modes were applied for input and target data sampling each mode has two cases, where in the first mode the time series data sampling was used with flow as an input for case one while flow and average precipitation in case two with used suspended sediment as a target variable. For second mode the supervise data sampling was used with the same input and target division in first mode. The performance of the model was evaluated by using Coefficient of determination (R^2) and the Nash- Sutcliffe efficiency (NS) and standardization of root mean square error (RSR), the statistical analysis model testing for Al-Adhiam watershed showed satisfactory agreement between observed and estimated daily values for Mode2- Case2. R^2 , NS and RSR of the testing period were 0.99 and 0.8 and 0.2 respectively. The result shows that the conducted ANN model can be used with the best net as a predictor for sediment yield in this watershed. The model was used to predict daily sediment load data for period from 1 Oct. 1984 to 31 Spt 1985. The predicted daily sediment data was plotted against daily measured flow. The correlation between predicted sediment and measured flow was in good agreement with $R^2=0.89$ and the best relation was polynomial equation from second degree.

1 Introduction

Sediment yield is defined as the total measured sediment load outflow from a watershed at a point of reference in a specified period. Processes of detachment, transportation, and deposition of soil materials induce the sediment outflow from the watershed by rainfall and runoff. Guesses of sediment yield are required in a wide range of problems such as transport of pollutants and sediment in rivers, design of reservoirs and dams, lakes and estuaries, undertaking clean up following floods, design of stable channels, Protection of fish and wildlife habitats, determination of the effects of watershed management [8].

The most reliable method in guessing sediment quantity is the use of its measured records, but sediment sampling is very complex and requires high experience because of its considerable fluctuation within the river section and user-unfriendly measurement tools. Moreover, it is consuming time and costly. These

constraints led to low frequency of sediment observation especially in the remote regions.

In Iraq the suspended sediment data is very scars or not available. Many models have been provided to simulate sediment load. Convention sediment rating curve and regression models, in which the system is supposed to be static, are often used to estimate suspended sediment load of a river. [10, 7] employed statistical approach and sediment rating curve to predict sediment load. The results of these studies show that the predicted sediment was underestimated. The hydrological models like deterministic models are useful for sediment yield prediction. The major drawback of these models are being require large amount of input data and computational time [9]. Physics-based increasingly being used to simulate complex water resource systems including simulation suspension sediment load but required enough measured sediment data for conducting the calibration and validation

processes [16]. Other researchers used Artificial Neural Network (ANN) model to prediction sediment load and compared the results with the sediment-rating curve [1, 14]. The results showed that the ability of ANN model for predicting sediment was better than other regression approaches due to their abilities in capturing nonlinear system among variables and used more than one input variable. [5] developed two different ANN methods to simulate relationship of suspended sediment load with river flow and precipitation by using hydro meteorological data. The obtained results showed that the provided models produced considerably better causes than Multi Linear Regression (MLR). [17] Predicted suspended sediment load by using ANN modeling based on the average of precipitations, precipitation intensity, temperature and flow discharge. This model was applied on the Long Chuanjiang River. It was capable to predict monthly suspended sediment load of flow accurately, [6] compared the estimated suspended sediment using the rating curves and ANN model for El- Kebir catchment. Daily suspended sediment and Daily water discharge data of Ain Assel gauging station, were used as inputs and targets in the models. The feed-forward and cascade-forward back-propagation was the ANN model based techniques by using Levenberg–Marquardt. The models results concluded that these models have the highest performance to re-estimate the global annual sediment yields and daily sediment load. Thus in the view of the superior performance of ANN represent suitable solution to predict sediment load with limitation suspension sediment data.

Al-Adhaim watershed is considered one of the most important catchment that feeding Tigris river in Iraq due to its size. Therefore, it effects remarkably on water quality of Tigris River. The watershed boundaries totally located inside Iraq in the northeast zone. In other words, the watershed is entirely managed by the Iraqi water resources ministry. Because of the lack in suspended sediment data, most of researchers avoid to use the hydrological model for simulating and study the impact of climate on sediment yield spatially in Iraq.

The aim of this study, modeling and driving suspended sediment data for Al-Adhaim watershed, from the available data scars data that was justly for water year 1983-1984, determine the best data input and target sampling and distribution that use in ANN model for this watershed and Deriving the best nonlinear correlation between flow and suspended sediment load.

2 Martials and methods

2.1 Descriptions of the study site

Al-Adhaim watershed is located on the northeastern border of Iraq, between 35° 42' 24" to 34° 33' 8" N and 43° 41' 9" to 45° 27' 31" E with a total drainage area of 11217 square kilometers Fig. 1. It lies entirely inside Iraqi territories, and it outstretches between the basin of

lower Zab northward, and Diyala river southward an, Shwan Secrema-Dagh eastward and Himreen mountains westward [2].

Al-Adhaim Dam was constructed on AL-Adhaim River In 2000 at coordinate 34° 32' 9"N and 44° 30'44" E approximately 134km north Baghdad City and about 65 km southeast of Kirkuk City all the drainage basin of Al-Adhiam is concentrated in Sallahadein, Diyala, Sulamaniya and Kirkuk provinces [4]. It is bounded between Himreen mountains westward and Kirkuk Hill eastward. AL_Adhaim Watershed is an arid area with limited rainfall and no snowfall. Thus, main flow occurs only during the rainy season. The major rainfall storm occurs from October to May. While the weather is relatively dry during the other months of the year. The minimum temperature is -4 °C in winter. It is raised to maximum 49 °C in summer. The annual rainfall of this region is 610 mm.

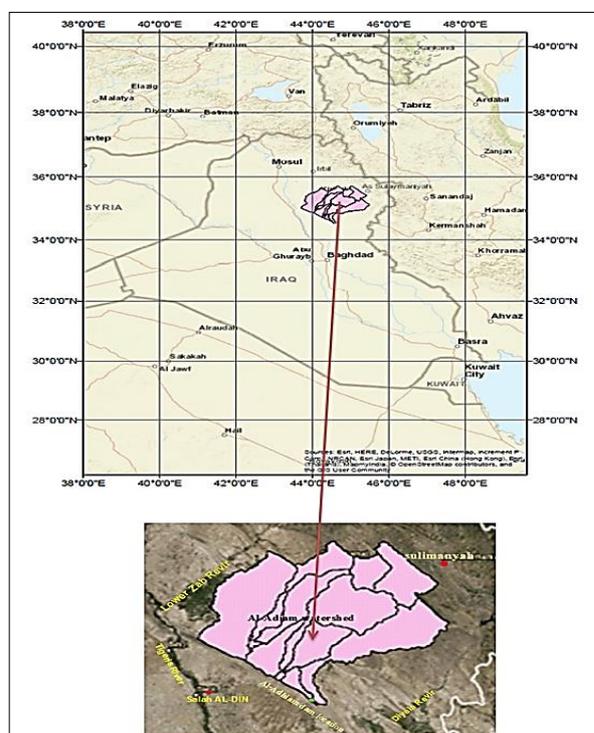


Fig. 1. Location map for Al-Adhiam watershed.

2.2 Data used

In this research, river discharges and average precipitation were used as input variables, while suspended sediment load was used as target variable .

2.3 Hydrological data

Considering the purpose of this study, ANN model needs for daily recorded flow. This data was obtained from the Iraqi Ministry of Water Resources (MOWR)/ National Center for Water Resources Management (NCWRM) for the period from 1983 to 1986.. The data were recorded at Enjana gauge station which located at coordinates 34° 32' 9"N and 44° 30'44" E at the current

outlet for the watershed. The daily observed sediment data for the water year 1983-1984 from October 1983 to October 1984 was collected from a study prepared by Al-Ansary et al. (1986) in which they calculated sediment discharge and soluble load for Al-Adhiam River.

2.4 Metrological data

Due to the lack in daily precipitation data, the monthly precipitation data was collected for the period from 1985 to 1989 from Iraqi Meteorological Organization and Seismology (MOS). Thus, the Climate Forecast System Reanalysis (CFSR) was used in this study. CFSR dataset consists of hourly weather forecasts produced by the National Centers for Environmental Prediction (NCEP). CFSR Provides weather requirement such as Precipitation, maximum and minimum temperatures [11], the monthly and yearly precipitation data were verified with those counterparts from the observed using statistical merits (R^2 , RMSE, NS). This data were downloaded on 23March 2017 from (<http://globalweather.tamu.edu/>) for AL_Adhaim watershed. 12 CFSR weather stations were for AL-Adhaim watershed.

2.5 Processing data in ANN

Neural networks are composed of simple components operating in parallel. These components are like that biological nervous system as in nature. The network form is determined largely by the linked between elements. It can train a neural network to behavior a particular form by modified the values of the weights (connections) between elements. The input values or signal (X_1, X_2, \dots, X_p) are transmitted to the neuron through the bifurcations that represent different input ways, Fig. 2. Each input way has its own linked weight (W_1, W_2, \dots, W_p). These input weights are very critical because they permit to sum and process the signals based on their values and effects on input vectors. An input weight vector is passed through the cell if it gives any real value except zero at the synapse. All the transmitted signals are ordinary integrated by collecting all the inputs. The mathematical model (referred to as activation form) is applied to complete the summation process within the cell body to produce an output signal. Different combinations of data and attached weight are used to obtain the correlation between the input and output within the ANN model using Equation (1), [15].

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

Where;

X_i = input value or signal i

W_i = weight of input value i

P = number of input values

b_k = bias at the cell of the body

Y = output

f = activation function

Many activation functions or equations can be used within the neurons. The Non-linear step function is the most commonly used function.

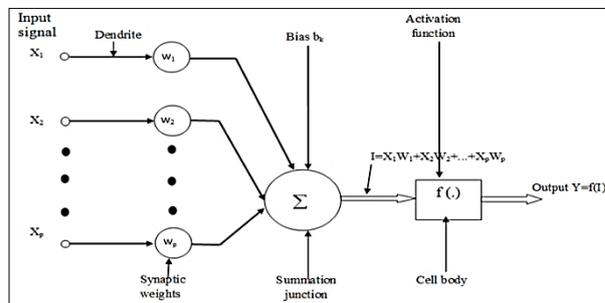


Fig. 2. Fundamental parts of a typical neural network [15].

2.6 Artificial neural networks structure

The basic neural network used in this study has a five-layer learning network. It consists of the input layer (introducing data to the ANN), the three hidden layers (processing data), and the output layer (producing the ANN results), Fig. 3.

The Levenberg–Marquardt algorithm of training the neural networks was used applied on the feed-forward backpropagation method (FFBP). In this study, the combinations of data partitioning approach, number of hidden layers, transfer function, number of neurons in each hidden layer, error goal, etc., were tried. The presentation of a newff function in relative to inputs, target, and number of neurons has produced a feed-forward network. This function is based on utilizing the units where a biased weighted summation of their inputs is performed. Then, these units transmit this activation level through a transfer function to produce their output, and the units were prepared in a layered feed-forward topology. Once the number of layers and number of elements in every layer, has been chosen, the network’s weights and levels have been set so as to minimize the estimate error made by the network. This was the role of the training algorithms [6].

To estimate daily-suspended sediment load applying ANN, at first, measured water discharge (Q_w), average precipitation and suspended sediment discharge (Q_s) data were collected for the period from 1Oct.1983 to 30 Sept.1984. Before entering the input and target data to the ANN model, the effect of extreme values in the data sets was prevented by normalizing this data between 0 and 1 by using Equation (2), [14].

$$Y_{norm} = \frac{Y_i - Y_{min}}{Y_{max} - Y_{min}} \tag{2}$$

Where, Y_{norm} is the normalized dimensionless variable; Y_i is the measured value of variable; Y_{min} and Y_{max} are

the minimum and maximum values of the variables, respectively [14].

The information was modeled utilizing Matlab R2015b software. Different structures of ANN model were built with a variation in hidden layers and joining different equations. Consequently, the optimal structure is found using Trial and Error method. Finally, the network of the obtained optimal structure and its training, was evaluated by using the test package data.

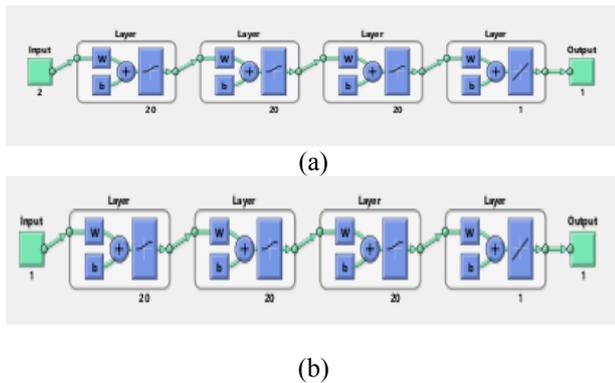


Fig. 3. An exemplary five-layer feed-forward ANN structure; (a): Two input variables feed-forward, (b): One input variables feed-forward.

2.7 Input and target data distribution

The obtained data was for one water year from 1 Oct. 1983 to 30 Sep. 1984 with 366 recorded points, Therefore, two modes were used for division and sampling the input variable and target variable as follows:

- i. Mode One:** Using time series sampling for input and target data with two cases as follows:
 - a. Case one: Using flow as an input variable and suspension sediment as a target variable, where the data from 1 Jan. 1984 to 30 Sep. 1984 for training and from 31Oct. 1983 to 31Dec. 1983 for testing.
 - b. Case two: Using flow and average precipitation as an input variable and suspended sediment as a target variable, where the data from 1 Jan. 1984 to 30 Sep. 1984 for training and from 31Oct. 1983 to 31 Dec. 1983 for testing.
- ii. Mode Two:** using supervised sampling for input and target data along the period from 1 Oct. 1983 to 30 Sep. 1984, with two cases as follows:
 - a. Case One: Using flow as an input variable and suspended sediment as a target variable, with 70% for training and 30% for testing.
 - b. Case Two: using flow and average precipitation as an input variable and suspended sediment as a target variable with 70% for training and 30% for testing.

2.8 Evolution of ANN model

ANN model efficiency was assessed utilizing coefficient of deterministic (R^2), standardizes of root mean squared error (RSR), and Nash–Sutcliffe coefficient of efficiency (Ns) [12]. The relationship between the measured and the predicated values within Al-Adhiam watershed was obtained using the correlation coefficient (R). The fundamental function was customized to the respective data input and outputs of the following general form:

$$R^2 = \frac{[\sum_{i=1}^n (S_{o,i} - \overline{S_o})(S_{o,i} - \overline{S_s})]^2}{[\sum_{i=1}^n (S_{o,i} - \overline{S_o})^2 \sum_{i=1}^n (S_{o,i} - \overline{S_s})^2]} \quad (3)$$

$$RSR = \sqrt{\frac{\sum_{i=1}^n [S_{s,i} - S_{o,i}]^2}{n}} \quad (4)$$

$$NS = 1 - \frac{\sum_{i=1}^n (S_o - S_s)_i^2}{\sum_{i=1}^n (S_{o,i} - \overline{S_o})_i^2} \quad (5)$$

R: coefficient of determination.

RSR: Standardizes of root mean squared error.

Ns: Nash–Sutcliffe coefficient of efficiency.

S_o : the measured value of the suspension sediment variable.

$\overline{S_o}$: mean of the measured values of the suspension sediment variable.

S_s : estimated value of suspension sediment variable.

$\overline{S_s}$: mean of the estimated values of suspension sediment variable.

n: number of the considered data points.

The value of R represents the strength of the relation between measured and estimated values. It value is ranged between 0 and 1. The low (high) values indicate a poor (perfect) forecasting capability of the model. The values of RSR is varies from 0, with perfect model simulation where there is no residual variation, to a large positive value. The best model performance corresponding to the lower RSR value. [15].

3 Results analysis

3.1 Structure characteristics of ANN

The number of neurons in the hidden layer was determined using the trial and error procedure to reach the best characteristics of optimal structure of ANN that were used in this research. Table 1 illustrates the characteristics of optimum structure of ANN.

Table 1. Characteristics of optimum structure of ANN.

Number of input layer neuron	1
Number of first hidden layer neurons	20
Number of second hidden layer neurons	20
Number of third hidden layer neurons	20
Number of output layer neuron	1
Transfer function	Gaussian-Sigmoid-Sigmoid-Sigmoid
Learning rate	0.001
Momentum coefficient	0.95
Iteration	1000

3.2 ANN model performance

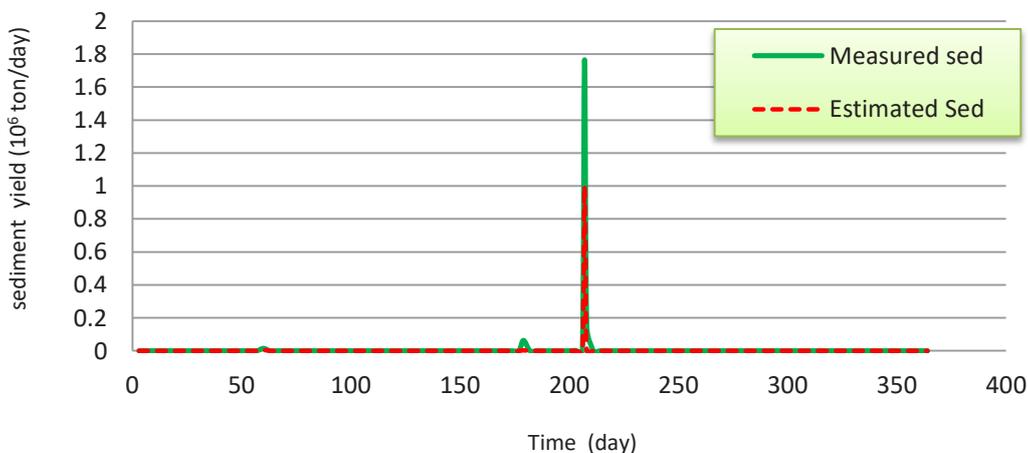
Two modes were applied for data sampling and data distribution for training and validation, to get the optimum distribution with best net. Each Mode has two cases of data sampling, as it discussed earlier. When analyzing the descriptive statistical results that listed in Table 2. R², NS and RSR were relied upon to evaluating the ANN model performance to determine the best

sampling mode for data distributing for training and validation. Also to identifying the optimum ANN model, which can be used for sediment load prediction depend on model performance in training and validation processes. From Table 2, it can be noticed that Mode2, when supervised distribution data used, has the best values of statistical measurements. It has the highest value of R² and NS with lowest value of RMSE and RSR for training and validation. The model performance in Case2, when used average precipitation (ave pcp) and flow (Q) as an input variable, is better than that of Case1 when only flow was used as an input variable. Fig. 4 shows the comparison between the estimated and measured suspended sediment load for validation period. Fig. 5 shows the correlation between the measured and estimated suspended sediment load of ANN model for the validation period Mode2 Case1 and Case2 respectively.

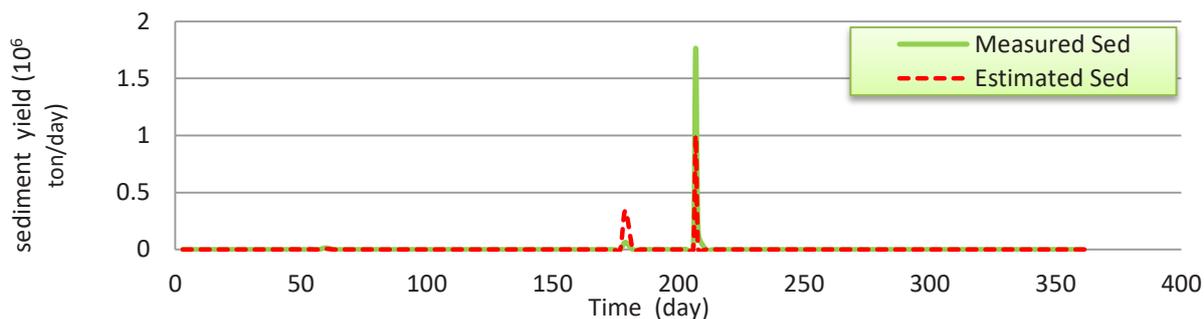
Table 2 shows the best performance of ANN model was with Mode2 for the two cases when using the supervised sampling for the data used in training and validation with some vantage for Case2. Therefore, the model that obtained from the two cases was applied to predict sediment suspended sediment yield of Al-Adhiam watershed.

Table 2. The best data sampling and their related performance criteria for the two Modes.

Mode (M)	Input (v)	Targ-et (v)	Training				Validation			
			R ²	NS	RMSE	RSR	R ²	NS	RMSE	RSR
M1-cas1	Q	Sed	0.68	4E-05	69264	1	0.005	-591718	1187155	591719
M1-cas2	ave pcp+Q	Sed	1	0.99	723	0.01	0.038	-104360	5.00E+05	104361
M2-cas1	Q	Sed	0.99	0.99	622	0.01	0.984	0.78	74580	0.22
M 2-cas2	ave pcp+Q	Sed	1	0.99	250.4	0.01	0.993	0.8	70480	0.2

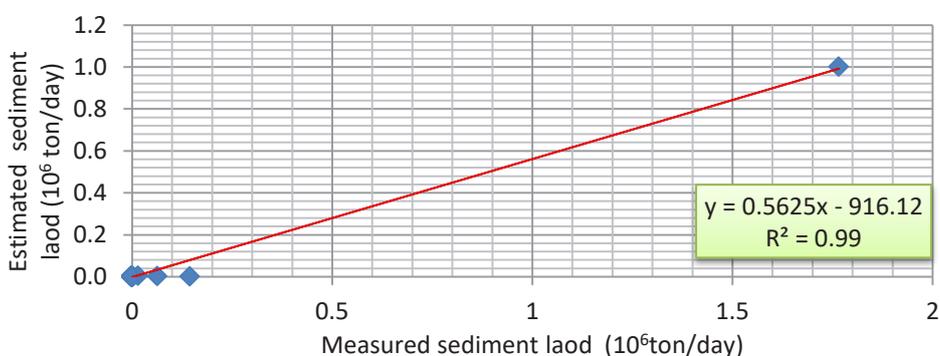


(a)

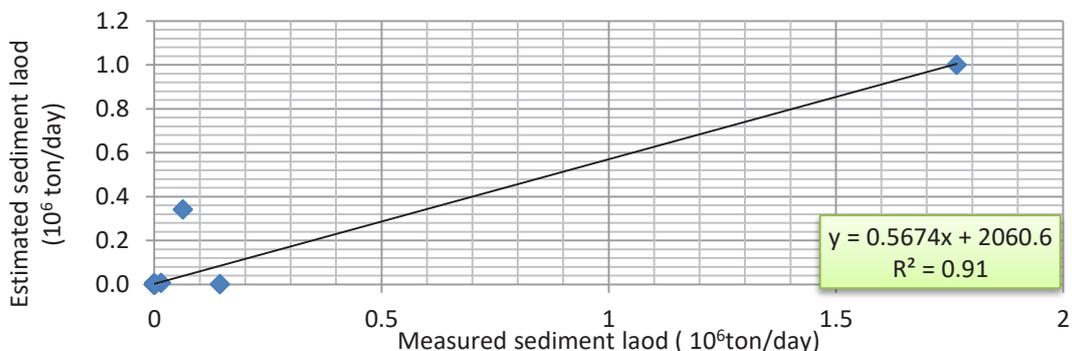


(b):

Fig. 4. Measured and estimated suspended sediment of ANN model. for the validation period; (a): Mode2-Case1 (for water year 1983-1984 (122 day)), (b): Mode2-Case2 (for water year 1983-1984 (122 day)).



(a)



(b)

Fig. 5. Correlation between the measured and estimated suspended sediment of ANN model for validation period; (a):Mode2 -Case1, (b):Mod2-Case2.

3.3 Prediction of suspended sediment load using ANN model

The two best ANN models that obtained using Mode2 with Case1 and Case2 from training and validating processes were applied to predict suspended sediment yield from Al-Adhiam watershed. These models were applied on the watershed for period from 1 Oct. 1984 to 31 Dec. 1985. Figs 6 and 7 show the daily time series of the measured flow and estimated

suspended sediment for the period 1 Oct. 1984 to 31 Dec. 1985. As shown in Figs 6a and 7b, the relationship between the predicted suspended sediment load and the measured flow for the two Cases can be represented by a nonlinear relationship. The value of R^2 for Case2 (using precipitation and flow as input variables) with ANN model was higher than that of Case1 (using only flow as input variable). In addition, results of regression analysis of these cases show the effect of precipitation on the values of predicted

suspended sediment load when it used with flow as an input variable to ANN model. In addition, these results show the sensitivity of ANN model to the type and number of input variables. The predicted daily sediment data was plotted against daily measured flow.

The correlation between predicted sediment and measured flow was in good agreement with $R^2 = 0.89$ and the best relation was polynomial equation from second degree.

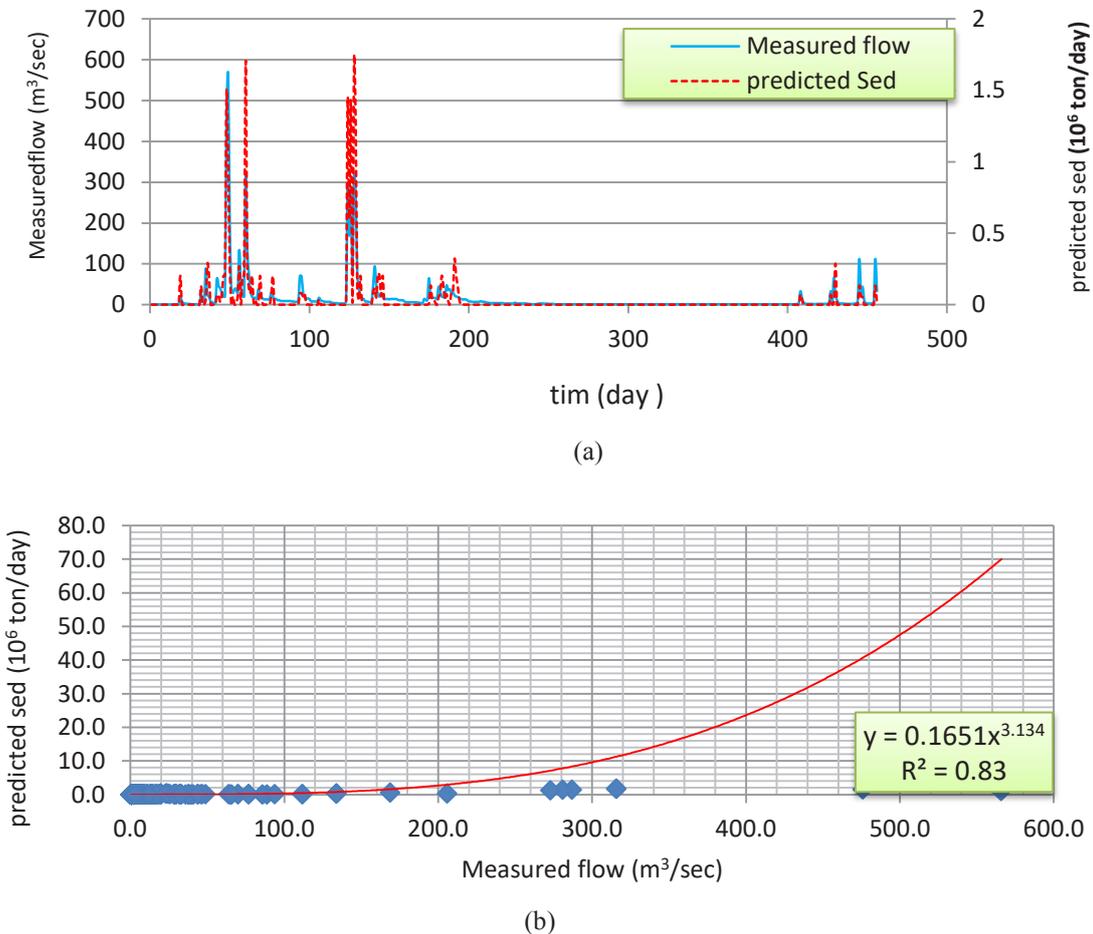
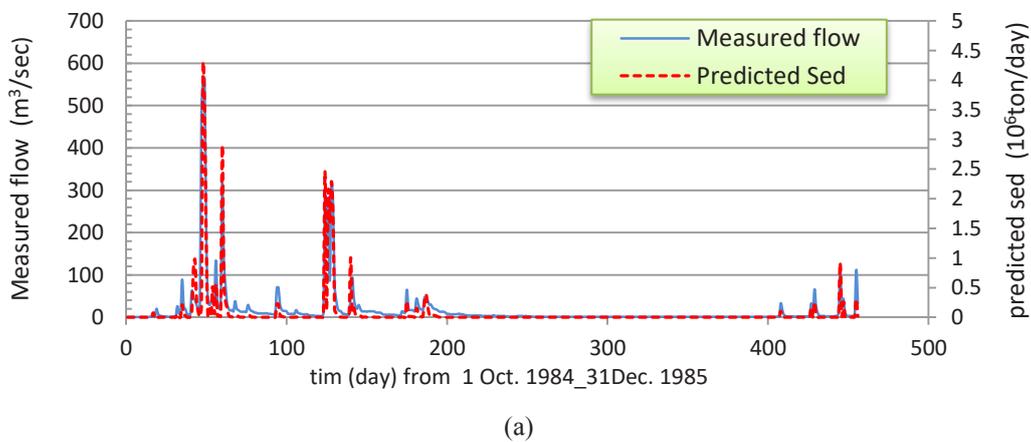
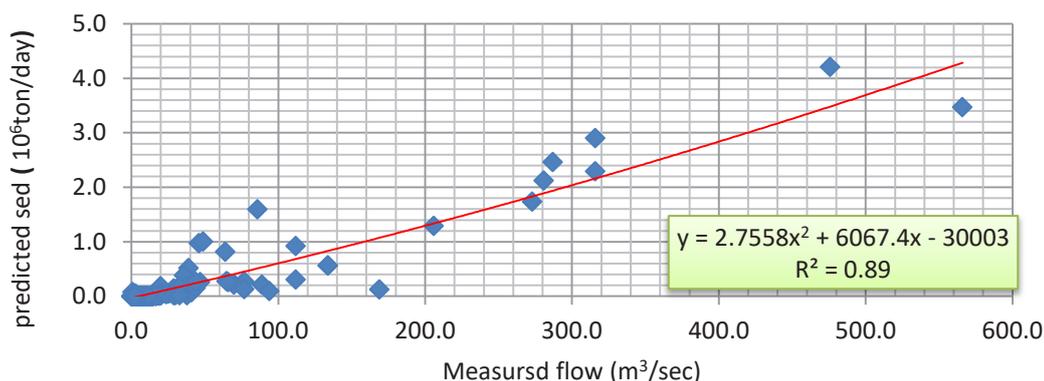


Fig. 6. Correlation between the measured flow and predicted suspended sediment load using ANN model for Mode2-Case1 (for the period from 1 Oct. 1984 to 31 Dec. 1985); (a): Daily time series, (b): Regression model.





(b)

Fig. 7. Correlation between the measured flow and predicted suspended sediment load using ANN model Mode2-Case2 for period from 1 Oct. 1984 to 31 Dec. 1985; (a): Daily time series, (b): Regression model.

4 Conclusion

ANN model was training and validated with two modes each mode have two cases. The best performance for the model where the higher value of R^2 and NS with lower value of RSR was when applying supervised sampling and using precipitation and flow as input variables and sediment as a target variable. Consequently, ANN model is sensitive to the type and number of input variable. In addition, the predicted daily sediment data was plotted against daily measured flow. The correlation between predicted sediment and measured flow was in good agreement with $R^2 = 0.89$ and the best relation was polynomial equation from second degree, it can be concluded the relationship between suspended sediment and flow mostly be nonlinear.

Acknowledgment. The authors would like to thank Ministry of Water Resources/ National Center for Water Resources Management in Iraq and to the staff of Building and Construction Engineering Department / Water and Hydraulic Structures Engineering Branch for their efforts to achieve this work.

1. R. J Abrahart and S. M. White, *Hydrology, Oceans and Atmosphere* **26.1**, 19-24 (2001).
2. N. Al-Ansari A., F. K Al-Shami, H. A Hassan and S. Ali, H, 5th Iraqi Geological Congress (1978).
3. NA Al-Ansari, A Sayfy, GT Al-Sinawi, AA Ovanessian, *Water Res*, 5.2, 173-187 (1986).
4. A. Al-Kadhimi, M., L. Ahmed, and R. Al-Mphergee, *Jordan Journal of Civil Engineering* **5.2**, 229-244, (2011).

5. M. Alp, and H. Kerem Cigizoglu, *Environmental Modelling & Software* **22.1**, 2-13. (2007).
6. Alp, M., and Cigizoglu, H. K, *Environmental Modelling & Software*, 22(1), 2-13 (2007).
7. Z. Boukhrissa, A., et al., *Journal of earth system science* **122.5**, 1303-1312 (2013).
8. J. Córdova. R., and M. Gonzalez, *Soil technology* **11.1**, 57-65 (1997).
9. H. CIĞIZOĞLU KEREM, *Turkish Journal of Engineering and Environmental Sciences* **26.1**, 16-26 (2001).
10. J. De Vente, , and P. Jean, *Earth-Science Reviews* **71.1**, 95-125 (2005).
11. R. I. Ferguson, *Water Resources Research* **22.1**, 74-76 (1986).
12. D. R. Fuka, et al., *Hydrological Processes* **28.22**, 5613-5623 (2014).
13. J. E. Nash and J. V. Sutcliffe, *Journal of hydrology*, **10(3)**, 282-290 (1970).
14. National Centers for Environmental Prediction (NCEP), (<http://globalweather.tamu.edu/>)
15. M. Shabani and N. Shabani, *Australian Journal of Basic and Applied Sciences*, **6(12)**, 157-164 (2012).
16. R. M. Wambua, B. M. Mutua, and J. M. Raude, *American Journal of Water Resources*, **4(2)**, 35-43 (2016).
17. Z. X. Xu, J. P. Pang, C. M. Liu and J. Y. Li., *Hydrological Processes*, 23(25), 3619-3630 (2009).
18. Y. Zhu, X. X. Lu, and Z. Yue, *Geomorphology* **84**, no. 1, 111-125 (2007).