A Wavelet-ANFIS Model to Estimate Sedimentation in Dam Reservoir

Mohamad Javad Alizdeh  
PhD Student, Faculty of Civil Engineering, K.N.Toosi University of Technology, Tehran, Iran

Pejman Mohammadian Joneyd  
MSc, Faculty of Civil Engineering, K.N.Toosi University of Technology, Tehran, Iran

Meysam Motahhari  
PhD Student, Department of Water Engineering, Imam Khomeini International University, Qazvin, Iran

Farid Ejali  
Assistant Professor, Payam Nour University, Tehran Branch, Tehran, Iran

Hamed Kiani  
MSc, Islamic Azad University, Nour Branch, Nour, Iran

ABSTRACT

An accurate prediction of sedimentation in dam reservoir is a challenging issue due to the complex and non-linear physics of the problem. Anyhow, soft-computing-based techniques showed great ability for predicting non-linear phenomena and have been used for different purposes. The main objective of this study is to estimate the volume of sedimentation in Karaj dam using a wavelet-ANFIS (WANFIS) and a wavelet-neural network (WANN) model. Monthly average flow is used to estimate monthly averaged suspended sediment load for a thirty-year period. The amount of bed load is computed based on the suspended sediment load and the river slope and the total volume of sedimentation in the reservoir is calculated with subtracting the upstream (Karaj River) and downstream (Beylaghan River) total sediment load. In WANFIS and WANN models, monthly average flow time series are decomposed to several sub-time series using different wavelet decomposition levels. The total volume of sedimentation in Karaj dam obtained from different techniques such as WANFIS, WANN, ANFIS, ANN and hydrography are compared together. The comparison demonstrates that WANFIS model is superior to the other techniques. For WANFIS and WANN models, the best model is obtained by two and three wavelet decomposition levels respectively. Findings of this study reveal that Wavelet-ANFIS models can be applied as a successful tool to predict the volume of sedimentation.

Keywords
Wavelet Transform, Neuro-Fuzzy System, Neural Networks, Suspended Sediment, Dam Reservoir

1. INTRODUCTION

Over a number of years, the sediment in the reservoir can build up, and reduce the space available for storing water. Some of the sediment held back in the reservoir would normally be carried downstream. If too much sediment is stored, the natural balance of the river downstream can be changed and subsequently it can affect people, wildlife and plants as far away as the river estuary [1]. As the sediments accumulate in the reservoir, so the dam gradually loses its ability to store water for the purposes for which it was built. Despite more than six decades of research, sedimentation is still probably the most serious technical problem faced by the dam industry [2]. Therefore, development of accurate models to estimate sedimentation in dam reservoir is a key element in the field of water resources and environmental engineering.

The amount of sedimentation in dam reservoir is a function of sediment enters and leaves dam by upstream and downstream flow. In order to estimate sedimentation in the reservoir, the upstream and downstream total sediment load should be calculated. Total sediment load consists of suspended sediment load (SSL) and bed load. Bed load can be calculated based on SSL as well as by other empirical methods. The prediction of SSL has many difficulties due to being a nonlinear and complex phenomenon. Many different formulas have been extended to estimate SSL based on empirical and physical models. However, due to existing large number of obscure parameters with SSL prediction, the theoretical governing equations may not be of much advantage for this purpose. Over the past years, Artificial intelligence techniques such as artificial neural network (ANN) and adaptive neuro fuzzy inference system (ANFIS) have been widely used to predict different non-linear parameters in the field of hydrology such as rainfall-runoff, ground water quality and sediment loads in rivers and the results demonstrated their efficiency [3-5]. Recently, a conjunction model of wavelet and artificial intelligence techniques has been applied successfully to predict SSL in rivers in which were superior to previous techniques [6-9]. Regarding sediment prediction by Wavelet-ANFIS (WANFIS) models, Rajaei (2010) proposed a model by combining the wavelet analysis and neuro-fuzzy (NF) approach to predict daily suspended sediment in a gauging station in the USA. Obtained results showed that the proposed model performs better than the NF and SRC models in prediction of suspended sediment [6]. Anyhow, applying daily prediction models for long term period prediction are not efficient and require high capacity of memory as well as being time consuming.

The main objective of this study is to estimate sedimentation in Karaj Dam based on wavelet and intelligence techniques. In this regard, different models including ANFIS, WANN and WANFIS models have been developed to forecast SSL in the upstream and downstream of Karaj dam. Monthly average flow discharge has been used as model input in order to predict monthly SSL for both the upstream and downstream rivers. Discrete wavelet transform (DWT) with ‘dme’y wavelet function was applied to decomposed original time series of monthly average discharge. Bed load was calculated based on SSL and the river slope. Different wavelet decomposition levels were used to get the most accurate models. Data were used to estimate sedimentation in the reservoir for a thirty-year period. Finally to calculate the
volume of sedimentation in the dam reservoir, the total output sediment load is subtracted from the total input sediment load. Results related to this study are compared with the results of hydrographic surveying method.

2. MATERIALS AND METHODS

2.1 Feed Forward Neural Networks

The feed forward neural network (FFNN) was the first and simplest type of artificial neural network devised. A usual FFNN consists of three layers, input, output and hidden layer. The input layer consists of just the inputs to the network. Then follows a hidden layer consists of any number of neurons, or hidden units placed in parallel. Each neuron performs a weighted summation of the inputs, then passes a nonlinear activation function $\sigma$, also called the neuron function [10]. Mathematically the functionality of a hidden neuron is described by

$$\sigma(\sum_{j=1}^{n} w_j x_j + b)$$ (1)

where the weights $w_j$ are symbolized with the arrows feeding into the neuron.

The network output is formed by another weighted summation of the outputs of the neurons in the hidden layer. This summation on the output is called the output layer. Generally, the number of output neurons equals the number of outputs of the approximation problem.

Different back propagation algorithm can be employed to train a FFNN model. Here, the Levenberg–Marquardt back propagation algorithm, which is a simplified version of Newton method, has been used in training of the FFNN.

2.2 Adaptive Neuro Fuzzy Inference System (ANFIS)

Fuzzy inference system (FIS) is a rule based system consists of three conceptual components. These are: (1) a rule-base, containing fuzzy if-then rules, (2) a data-base, defining the Membership Function (MF) and (3) an inference system, combining the fuzzy rules and produces the system results [11]. The first phase of FL modeling is the determination of MFs of input – output variables, the second phase is the construction of fuzzy rules and the last phase is the determination of output characteristics, output MF and system results.

ANFIS consisting of the combination of the ANN and the FL uses the learning ability of ANN to define the input-output relationship and construct the fuzzy rules by determining the input structure. Let $x$ and $y$ be the two typical input values fed at the two input nodes, which will then transform those values to the membership functions (say bell-shaped) and give the output as follows. (Note in general, $w$ is the output from a node, $m$ is the membership function, and $M_i$ and $N_i$ are fuzzy sets associated with nodes $x$, $y$)

$$\mu_M(x_i) = \frac{1}{1+[(x-c_i)/a_i]^{2b_i}}$$ (2)

where, $a_i$, $b_i$ and $c_i$ are changeable premise parameters. Similar computations are carried out for the input of $y$ to obtain $\mu_N(y)$. The membership functions are then multiplied in the second layer, e.g.

$$w_i = \mu_M(x_i) \cdot \mu_N(y_i) \quad (i = 1, 2)$$ (3)

where, $x$ (or $y$) is the input to the node; $M_i$ (or $N_i$) is a linguistic label (such as ‘low’ or ‘high’) associated with this node, characterized by the form of the membership functions in this node and can be any suitable function that is continuous and piecewise differentiable such as Gaussian, trapezoidal shaped, generalized bell shaped and triangular shaped functions. Such products or firing strengths are then averaged:

$$\bar{w_i} = w_i / \sum w_i$$ (4)

Nodes of the fourth layer use the above ratio as a weighting factor. Furthermore, using fuzzy if-then rules produces the following output: (an example of an if-then rule is:

If $x$ is $M_1$ and $y$ is $N_1$, then $f_1 = plx+qlx+r1$

$$\bar{w_i}f_i = \bar{w_i}(p_i x + q_i y + r_i), \quad (i = 1, 2)$$ (5)

Where, $p$, $q$ and $r$ are changeable consequent parameters. The final network output $f$ was produced by the node of the fifth layer as a summation of all incoming signals, which is exemplified in Eq. (5).

2.3 Discrete Wavelet Transform (DWT)

Discussing about the theory behind wavelet transform (WT) is beyond the scope of this study and just the main concepts of the discrete wavelet transform (DWT) are briefly presented. A mathematical overview of WT and a review of applications are presented by [12]. The WT performs the decomposition of a signal into a group of functions [13]:

$$\psi_{j,k}(x) = 2^{j/2} \psi_{j,k} (2^j x - k)$$ (6)

where $\psi_{j,k}(x)$ is produced from a mother wavelet $\psi(x)$ which is dilated by $j$ and translated by $k$. The mother wavelet has to satisfy the condition.

$$\int \psi(x)dx = 0$$ (7)

The discrete wavelet function of a signal $f(x)$ can be calculated as follows:

$$c_{j,k} = \int_{-\infty}^{\infty} f(x) \psi_{j,k} (x)dx$$ (8)

$$f(x) = \sum_{j,k} c_{j,k} \psi_{j,k} (x)$$ (9)

where $c_{j,k}$ is the approximate coefficient of a signal. The mother wavelet is formulated from the scaling function $\varphi(x)$ as:

$$\varphi(x) = \sqrt{2} \sum h_0(n) \varphi(2x - n)$$ (10)

$$\psi(x) = \sqrt{2} \sum h_1(n) \varphi(2x - n)$$ (11)

Where $h_0(n) = (-1)^n h_0(1-n)$. Different sets of coefficients $h_0(n)$ can be found corresponding to wavelet bases with various characteristics. In the DWT, coefficients $h_0(n)$ play a critical role [14].

2.4 Study Area and used Data

The study area includes the Karaj dam watershed in which is located between 51° 2’ to 51° 32’ east longitudes and 35° 53’ to 36° 11’ northern latitudes. It has an area of 850 Km2 and circumference of 146Km, and is located on the southern slope of the Alborz Mountains (Figure 1) [15]. The upstream river enters the dam reservoir is called Karaj River. The hydrometric station for the upstream river (Seera) is located in the dam entrance. Also, the downstream hydrometric station of dam reservoir is called Beylughan.
Data used in this study includes monthly averaged flow discharge as input variable and suspended sediment load (SSL) as output. A thirty-year period data range from 1960 to 1990 was used in the forecasting models. To obtain monthly averaged flow discharge, daily flow discharge is averaged for each month. These data obtained for the upstream and downstream rivers. The Upstream and downstream data obtained from Seera and Beylaghan hydrometric stations respectively. These data (flow discharge and SSL) are recorded daily at both stations. In this study, monthly average flow for a thirty year period has been used to predict monthly suspended sediment load for the upstream and downstream rivers of the reservoir.

2.5 Statistical Analysis
To begin with, homogeneity tests including Run-Test and other statistical tests were carried out on the data. Too small and too large values of suspended sediment load (SSL) were eliminated to increase the accuracy of the forecasting models. Afterwards, the data statistical analysis related to the both hydrometric stations have been carried out and the results are tabulated in Table 1. It should be noticed that in Table 1, Sd denotes the standard deviation of data.

Data normalization is a common procedure dealing with large number of data. Generally it happens by scaling data between 0 and 1 in order to eliminate their dimension. In this study, equation (13) has been used to normalize the data.

\[
X_{\text{norm}} = \frac{X_i}{X_{\text{max}}}
\]  

(12)

where \(X_{\text{norm}}\), \(X_i\), and \(X_{\text{max}}\) represent the normalized, observed and maximum of data respectively.

<table>
<thead>
<tr>
<th></th>
<th>Q</th>
<th>SSL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Min</td>
<td>Max</td>
</tr>
<tr>
<td>Seera</td>
<td>2.4</td>
<td>54</td>
</tr>
<tr>
<td>Beylaghan</td>
<td>3.2</td>
<td>61</td>
</tr>
</tbody>
</table>

Table 1. Statistical analysis of the data

Fig 1: The Catchment under consideration
2.6 WANN and WANFIS Models
In the current study, WANN and WANFIS models were developed to predict monthly averaged SSL for Karaj River and Beylaghan River. Firstly, the original time series related to monthly averaged flow discharge are decomposed into sub-time series by discrete wavelet transform (DWT). Next, different combinations of sub-time series are examined in the input structures of ANN and ANFIS to get the effective sub-time series. Selected sub-time series are imposed as new input of ANN and ANFIS models to predict monthly SSL. Different decomposition levels of wavelet have been applied and results have been compared with each other to get the best model performances. The comparisons demonstrated that the wavelet decomposition levels of 3 and 2 lead to most accurate models for WANN and WANFIS models respectively. In the ANN model development, FFNN approach with nine neurons in the hidden layer was used. For the WANFIS models, ‘Gauss2” was used as membership function in FIS structure with 3 membership function. In both WANN and WANFIS models, 60% data were used to train the network. 30% remains for validation and testing sets of WANN models. Also, 40% of data were used in testing set of WANFIS models. A schematic layout of the wavelet-ANFIS model is predicted in Figure 2.

To assess the performances of forecasting models, different methods such as coefficient of determination ($R^2$), sum of square error (SSE), root mean square error (RMSE), etc. can be employed. Here, in this study, the two indices of determination coefficient and RMSE have been used to evaluate the performances of models. In brief, the models’ predictions are optimum if $R^2$ and RMSE are found to be close to 1 and 0, respectively. These indices are defined as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_{i\text{measured}} - Y_{i\text{predicted}})^2}{\sum_{i=1}^{n}(Y_{i\text{measured}} - \bar{Y}_{\text{measured}})^2}$$

(13)

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Y_{i\text{measured}} - Y_{i\text{predicted}})^2}{n}}$$

(14)

Where $n$ is the number of data, $y$ denotes the output variable in which in this study includes SSL.

3. RESULTS AND COMPARISONS
In this section, different models are developed for both hydrometric stations based on intelligence techniques (Seera and Beylaghan). The results related to $R^2$ and RMSE for the best models of ANN, ANFIS, WANN and WANFIS are presented in Tables 2 and 3 for Seera and Beylaghan hydrometric stations respectively. It should be noticed that the best ANN model has 8 neurons in its hidden layer and the best ANFIS model obtained for ‘gauss’ membership function for 3 numbers of membership function. Also, the best WANN and WANFIS models have 3 and 2 decomposition levels respectively. By applying DWT, it was found that the best performance is drawn for ‘dmey’ wavelet function. Generally, it can be concluded that applying monthly average flow discharge and a combined model of wavelet and artificial neural network can provide an acceptable prediction of suspended sediment load. From Tables 2 and 3, it is inferred that for both hydrometric stations the WANN and WANFIS models are efficient models for SSL prediction. The high values of $R^2$ (close to 1) and small values of RMSE (close to 0) prove this assertion.

According to Tables 2 and 3, it can be concluded that the forecasting models used in this study provide a more accurate prediction for Seera hydrometric station than Beylaghan hydrometric station. The best model for Seera station (WANFIS model) during testing period has a $R^2$ and RMSE of 0.84 and 0.0076 while outperforms the best model for Beylaghan station (WANFIS model) with corresponding $R^2$ and RMSE values of 0.786 and 0.0078 respectively.

Having a comparison among different models applied in this study reveals that WANFIS model is the most efficient model. Moreover, WANN model can provide an acceptable prediction for SSL in these two hydrometric stations. For Seera station and for testing period, the WANFIS model has a $R^2$ of 0.84 and RMSE of 0.0076 in which the corresponding values for WANN model are equivalent to 0.782 and 0.008 respectively. In a similar way, the superiority of WANFIS model to WANN model can be seen for Beylaghan hydrometric station.
Table 2. Models’ performances for Seera hydrometric station

<table>
<thead>
<tr>
<th></th>
<th>Train</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANN</td>
<td>(0.542)</td>
<td>(0.13)</td>
</tr>
<tr>
<td>ANFIS</td>
<td>(0.578)</td>
<td>(0.113)</td>
</tr>
<tr>
<td>WANN</td>
<td>(0.826)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>WANFIS</td>
<td>(0.882)</td>
<td>(0.0038)</td>
</tr>
</tbody>
</table>

Table 3. Models’ performances for Beylaghan hydrometric station

<table>
<thead>
<tr>
<th></th>
<th>Train</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANN</td>
<td>(0.512)</td>
<td>(0.144)</td>
</tr>
<tr>
<td>ANFIS</td>
<td>(0.564)</td>
<td>(0.126)</td>
</tr>
<tr>
<td>WANN</td>
<td>(0.793)</td>
<td>(0.0058)</td>
</tr>
<tr>
<td>WANFIS</td>
<td>(0.826)</td>
<td>(0.005)</td>
</tr>
</tbody>
</table>

It can be found from Tables 2 and 3 that WANN and WANFIS models in both hydrometric stations have smaller values of RMSE rather than ANN and ANFIS models. This finding is valid for both train and test data sets. The smallest value of RMSE for Seera station and for testing period is 0.0076 that obtained for WANFIS model. Similarly, for Beylaghan station, the smallest value of RMSE equals to 0.0078 that obtained for WANFIS model. Comparing the training and testing sets performance in terms of \(R^2\) and RMSE reveals that for WANN and WANFIS models, these values are close to each other while for ANN and ANFIS models, there is relatively a big difference between the values of \(R^2\) and also RMSE of training and testing sets.

A comparison between the performance of WANN and WANFIS models in terms of \(R^2\) and RMSE demonstrate that WANFIS model outperform WANN model for both hydrometric stations. During testing period, applying WANFIS model instead of WANN model and for Seera station improves the model accuracy in terms of \(R^2\) and RMSE about 6% and 5% respectively.

More comparisons between the performance of WANN and WANFIS models is given by Figure 2. Figure 2 presents the predicted SSL versus observed SSL for all data set in Seera station. As it can be seen from Figure 2, the WANFIS model has a higher value of \(R^2\) and a better fitted line compared to WANN model. Anyhow, in both WANN and WANFIS models, a good correlation between the predicted and real values of suspended sediment load can be observed. All in all, for this study, the WANN and WANFIS models can be applied to provide a successful estimation of the volume of sedimentation in the reservoir.

Figure 3 presents results related to time series of WANN and WANFIS models of whole data set in Seera station. This figure shows the normalized values of observed and predicted SSL in descending order.
Regarding Table 4, it can be obtained that the total volume of sedimentation in Karaj dam for WANN and WANFIS models are about 10.5 and 11 million m$^3$ respectively. These values contain for a thirty-year period (1960-1990). The total volume of sedimentation in Karaj dam has been calculated by using hydrographic surveying method in which the total sedimentation for this thirty-year period is equivalent to 12.5 million m$^3$. Comparison the results derived from WANN and WANFIS models with hydrographic surveying method indicate that WANN and WANFIS models give a reasonable prediction of the volume of sedimentation in Karaj dam. Moreover, the amount of sedimentation estimated by WANFIS model is closer to the real volume of sedimentation than WANN model.

### 4. CONCLUSIONS

In the current research, an attempt was made to estimate the volume of sedimentation in Karaj dam using a conjunction model of wavelet and artificial intelligence techniques (WANN and WANFIS). The monthly average flow discharge recorded in the upstream and downstream hydrometric stations of the reservoir (Seera and Beylaghan) was used as input variables to predict the suspended sediment load for the rivers. Using theoretical methods as well as SSL, the bed load and subsequently the total sediment load and the total volume of sedimentation in the reservoir were calculated. According to Tables 2 and 3 it can be concluded that DWT as a pre-processing technique can improve the model performances remarkably. Results of this study demonstrated that WANN and WANFIS models can give an acceptable prediction of SSL. The best WANN and WANFIS models were obtained for 3 and 2 wavelet decomposition levels respectively. Moreover the best WANFIS model outperforms the best WANN model.

According to the result of this study, a good correlation between predicted and real values of suspended sediment load in the upstream and downstream rivers can be achieved when WANN and WANFIS models are used. High values of $R^2$ and calculated based on SSL and the river slopes. Previous studies indicated that for Iran Rivers, the bed load is a function of the river slope and applying Karashef’s theory provides a reasonable result [17]. In this study, the bed load for both upstream and downstream rivers were obtained from previous study [16]. The slopes of the upstream and downstream rivers are 1.8 and 0.92 in which gives the ratio of bed load to SSL for Seera and Beylaghan stations are equal to 2 and 0.45 respectively [16]. Having SSL and bed load, the total sediment load for both stations and subsequently the sedimentation in dam reservoir can be determined. The density of sediment was assumed to be 1.4 (ton/m$^3$) to change the mass of the sediment (ton) into the volume of sedimentation (m$^3$). Finally, the volume of sedimentation for a 30-year period derived by the WANN and WANFIS models compared with the amount of sedimentation obtained by hydrographic surveying method for Karaj dam (Table 4).

### Table 4. Total volume of sedimentation in dam reservoir obtained by WANN and WANFIS models

<table>
<thead>
<tr>
<th>Model</th>
<th>Seera</th>
<th></th>
<th>Beylaghan</th>
<th></th>
<th>Sedimentation in dam (m$^3$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SSL (ton)</td>
<td>Bed Load (ton)</td>
<td>Total (m$^3$)</td>
<td>SSL (ton)</td>
<td>Bed Load (ton)</td>
</tr>
<tr>
<td>WANN</td>
<td>77271255</td>
<td>14542511</td>
<td>15581261</td>
<td>4964512</td>
<td>2234030</td>
</tr>
<tr>
<td>WANFIS</td>
<td>7304766</td>
<td>14609532</td>
<td>15653070</td>
<td>2312259</td>
<td>1040517</td>
</tr>
</tbody>
</table>

The time-series plot of observed and predicted data for Seera station (Figure 3) reveals that WANFIS model can give an accurate prediction for extreme values of SSL in the river. This finding is interesting to apply such an accurate model for predicting of SSL for peak discharges in rivers. Generally, it is difficult to measure SSL in rivers during high flow conditions. The problem will be more critical during flood period when direct measurement is not possible. Therefore, WANFIS models in daily or hourly scale can be developed to give a sound prediction of SSL in rivers during flood or high flow conditions.

The total volume of sedimentation estimated by this study was compared with the amount of sedimentation proposed by hydrographic surveying method to evaluate the accuracy of the forecasting models. The comparison reveals the efficiency of the applied models. The total volume of sedimentation assessed by the hydrographic method is about 12.5 million m$^3$ which are close to the predicted values derived from WANFIS and WANN models (11 and 10.5 million m$^3$ respectively).

The great advantage of the method used in this study in comparison with other models in which apply daily data for prediction is that in this way the amount of data are reduced to 1/30 times. In this way, it will be less time consuming and requires a smaller storage capacity. Moreover, the model gives a reasonable prediction of SSL. Result of this study is encouraging the link of the data pre-processing techniques such as wavelet to existent intelligent models such as ANN and ANFIS in order to enhance the performances of such forecasting models. Anyhow, for short-term prediction or in cases that more accuracy is required, smaller time scales can be considered (e.g. weekly, daily or hourly models). These models can be helpful in prediction of SSL during flood, rainfall with high intensity and high flow discharge.
Consequently, they can lead to a successful assessment of the amount of sediments in the reservoir and computation of the amount of water in the reservoir which can be helpful in water resources and management planning.

5. REFERENCES


