

Conjunction of ANN and threshold based wavelet de-noising approach for forecasting suspended sediment load

Vahid Nourani ¹, Aida Yahyavi Rahimi ² and Farzad Hassan Nejad ²

¹Associate Professor, ²M.Sc. Student

Department of Water Resources Engineering, Faculty of Civil Engineering,
University of Tabriz, 29 Bahman Ave., Tabriz, Iran.

ABSTRACT

Information on suspended sediment load (SSL) is fundamental for numerous water resources management and environmental protection projects. This phenomenon has the inherent complexity due to a large number of vague parameters and existence of both spatial variability of the basin characteristics and temporal climatic patterns. This complexity turns into a barrier to get accurate prediction by conventional linear methods. On the other hand, the extent of the noise on hydrological data reduces the performance of data-driven models like Artificial Neural Networks (ANNs). Although ANNs could capture the complex nonlinear relationship between input and output parameters, being data-driven method positioned it in a state of need to preprocessed data. In this paper, the application of ANN approach focusing on wavelet-based denoising method for modeling daily streamflow-sediment relationship was proposed. The daily streamflow and SSL data observed at outlet of the Potomac River in USA were used as the case study. Achieving this purpose, Daubechies (*db*) was used as mother wavelet to decompose both streamflow and sediment time series into detailed and approximation subseries. Decomposition at level ten via *db3* and at level eight via *db5* were examined for streamflow and SSL time series, respectively. At first, the appropriate input combination with raw data to estimate current SSL was determined and re-imposed to ANN with denoised data. The comparison of results reveals that in term of determination coefficient, the

obtained result by denoised data was improved up to 23.2% with raged to use noisy, raw data and this exhibits that denoised data can be employed productively in ANN-based daily SSL forecasting.

Keywords: Suspended Sediment Load (SSL), Artificial Neural Network (ANN), Wavelet denoising, Mother Wavelet, The Potomac River.

1 INTRODUCTION

River water quality is troubled by the existence of the suspended sediment load (SSL) transported by the streamflow. The accurate estimation of the volume of sediment being transported by a river is a considerable issue for hydrologists as it affects the design, management and operation of water resources projects as well as it is essential for the study of various problems of river improvement and utilization.

As stated by Jain and Ormsbee (2002), probably the most frequently used forecasting models for hydrological phenomenon are based on linear regression. Conventional time series models such as Auto Regressive Integrated Moving Average (ARIMA) models are widely used for hydrological time series forecasting (Salas et al., 1980). However, they are basically linear assuming that data are stationary, and have a limited ability to capture non-stationarities and non-linearities involved in the hydrologic data.

Artificial Neural Network (ANN) is one of the black box modeling tools which has been recently found good performance in modeling hydrological and environmental processes. Zhang et al. (1998), in a state-of-the-art survey, observed that although some contradicting reports exist in the literature,

ANN is a useful method for non-linear modeling whereas ARIMA approaches may be more proper for detecting linear relationships in the data. Currently, ANN as a self-learning and self-adaptive function approximator, has exposed great capability in modeling and forecasting non-linear hydrologic time series. ANNs offer an effective approach for handling large amounts of dynamic, non-linear and noisy data, especially when the underlying physical relationships are not fully understood, but there are enough data to train a network (Zhang et al., 1998). Several ANN configurations have been used for hydrological forecasting with good performance. In this respect, Jain and Ormsbee (2002), Bougadis et al. (2005), Adamowski (2008), Adamowski et al. (2010), Güldal and Tongal (2010) and Shirsath and Singh (2010) used ANNs to forecast various aspects of hydrological processes. Remarkable ANN models have been developed for modeling watersheds precipitation and rainfall-runoff processes (e.g., Hsu et al., 1995; Tokar and Johnson, 1999; Nourani et al., 2009a, b, 2012a; Rezaeian Zadeh et al., 2010). Abrahart et al. (2012) investigated studies about ANN applications on rainfall-runoff processes. Considering ANN success in modeling hydrological processes, some papers were also provided about SSL modeling. Jain (2001) used the ANN approach to develop an integrated stage-discharge-sediment concentration relation and showed the superiority of ANN over the conventional methods. Tayfur (2002) applied ANN for sheet sediment transport estimation. Agarwal et al. (2006) presented an ANN model for simulation of runoff and sediment yield and showed that daily time scale modeling performs well in both calibration and verification steps, and also pattern learning process of the model building is superior to batch learning process. Alp and Cigizoglu (2007) simulated SSL by two ANNs using rainfall, flow and sediment data. They used rainfall and water discharge as model inputs and SSL as output parameter. Rai and Mathur (2008) proposed a feed forward back propagation ANN for computing event-based temporal variation of sediment yield. Nourani (2009) used water discharge at current and previous time steps as input neurons for forecasting sediment load at

current time step by ANN. Rajaei et al. (2010) used conjunction of wavelet and neuro-fuzzy for suspended sediment prediction. Nourani et al. (2012b) developed two ANN models for semi-distributed modeling of the SSL the Eel River watershed at California. Efficiency of ANN method just like any other data-driven method largely depends on quantity and quality of the data. In physical systems, transmitted time series are usually distributed partially, or sometimes almost completely, by an additive noise from the transmitter, channel, and receiver. In general, there are two types of noise, measurement noise and dynamical noise. Measurement noise refers to the corruption of observations by errors, which are independent of the dynamics. Dynamical noise, in contrast to measurement noise, is a feedback process wherein the system is perturbed by a small random amount at each time step (Schouten et al., 1994). The approach investigated in this work is to consider dynamical noise reduction as a fundamentally required process to enhance the estimation process of the captured time series. Noise reduction is considered as a continuous mapping process of the noisy input data to a noise free output data. The resulted enhanced time series can be then imposed to the estimation process by the ANN which can improve the performance of the model.

Meantime, a diversity of noise reduction methods have been developed which are usually suitable for linear systems [(Wiener, 1949; Kalman, 1960). Regarding the fact that hydrological systems are nonlinear, these filters have limited applications in hydrological time series modeling. The extended Kalman filter, which can be used for nonlinear dynamics, becomes unstable when systemic nonlinearity is strong (Reichel et al., 2002; Hamid et al., 2004). Hence, the ensemble Kalman filter (EnKF), a Monte Carlo-based Kalman filter, was introduced and gained popularity in hydrology (Martyn et al., 2008). But in principle, the EnKF is suitable only for Gaussian error statistics and just propagates the first two moments of error statistics. Sivakumar et al. (1999) addressed some of the potential problems in applying such methods to chaotic hydrological data, and discussed the usefulness of estimating the noise level prior to noise reduction. Elshorbagy et al.

(2002) investigated commonly used algorithms for noise reduction in order to estimate chaotic hydrologic time series. Wiener and Kalman filtering methods are only capable of dealing with linear natural systems, and the results depend on the establishment of state space functions to a great extent (Sang, 2013). Compared with the so-called methods, the threshold based wavelet denoising method is more capable, since it can illuminate the localized characteristics of non-stationary time series both in the temporal and frequency domains (Janson, 2006). With specific regard to denoising methods based on wavelets, Cannas et al. (2006) and Nourani et al. (2009a) explored the multi-scaling property of wavelets for maximization of ANN forecasting accuracy in the context of flow forecasting. Guo et al. (2011) used the wavelet denoised method to reduce or eliminate the noise in runoff time series and improve the performance of Support Vector Machine (SVM) runoff prediction model. Nejad and Nourani (2012) applied global soft threshold based wavelet denoising method to denoise daily time series of river stream discharge. The denoised time series was then imposed into an ANN model to forecast flow discharge value on day ahead. Empirical results showed that networks trained with pre-processed data perform better than networks trained on un-decomposed, noisy raw time series. More recently, Nourani et al. (2013) used the wavelet transform to extract dynamic and multi-scale features of the non-stationary runoff time series and removed the observed data noise.

The aim of this study is to improve ANN modelling of runoff-sediment process by applying global soft threshold based wavelet denoising method to denoise daily time series of streamflow and suspended sediment, observed at outlet of the Potomac River in USA. In this way, four fundamental issues are addressed as i) the choice of appropriate wavelet, ii) the choice of decomposition level, iii) threshold value determination and iv) selection of thresholding rules.

The remaining part of the paper has been organized as follows. First, a brief description of wavelet transform and ANN along with information of study area and utilized data are

presented. Then, the proposed method is introduced and their performance results are presented and discussed in detail. The final section presents Conclusion.

2 MATERIALS & METHODS

2.1 Wavelet Denoising Procedure

The wavelet transform has increased in usage and popularity in recent years since its inception in the early 1980s, yet still does not enjoy the wide spread usage of the Fourier transform. Fourier analysis has a serious drawback. In transforming to the frequency domain, time information is lost. When looking at a Fourier transform of signal, it is impossible to tell when a particular event took place but wavelet analysis allows the use of long time intervals where more precise low-frequency information and shorter regions are necessary where high-frequency information is wanted. In the field of earth sciences, Grossmann and Morlet (1984), who worked especially on geophysical seismic signals, introduced the wavelet transform application. A comprehensive literature survey of wavelet in geosciences can be found in Foufoula-Georgiou and Kumar (1995) and the most recent hydrological contributions have been cited by Labat (2005) and Sang (2013). As there are many good books and articles introducing the wavelet transform, this paper will not delve into the theory behind wavelets and only the main concepts of the transform are briefly presented; recommended literature for the wavelet novice includes Mallat (1998) or Labat et al. (2000). The time-scale wavelet transform of a continuous time signal, $x(t)$, is defined as (Mallat, 1998) :

$$T(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} g^* \left(\frac{t-b}{a} \right) x(t) . dt \quad (1)$$

where * corresponds to the complex conjugate and $g(t)$ is called wavelet function or mother wavelet. The parameter a acts as a dilation factor, while b corresponds to a temporal translation of the function $g(t)$, which allows the study of the signal around b . The main property of wavelet transform is to provide a time-scale localization of process, which derives from the compact support of its basic function. This is opposed to the classical

trigonometric function of Fourier analysis. The wavelet transform searches for correlations between the signal and wavelet function. This calculation is done at different scales of a and locally around the time of b .

For practical applications, the hydrologist does not have at his or her disposal a continuous – time signal process but rather a discrete – time signal. A discretization of Eq. (1) based on the trapezoidal rule maybe is the simplest discretization of the continuous wavelet transform. This transform produces N^2 coefficients from a data set of length N ; hence redundant information is locked up within the coefficients, which may or may not be a desirable property (Addison et al., 2001).

To overcome the mentioned redundancy, logarithmic uniform spacing can be used for the scale discretization with correspondingly coarser resolution of the b locations, which allows for N transform coefficients to completely describe a signal of length N . Such a discrete wavelet has the form (Mallat, 1998):

$$g_{m,n}(t) = \frac{1}{\sqrt{a_0^m}} g^* \left(\frac{t - nb_0 a_0^m}{a_0^m} \right) \quad (2)$$

where m and n are integers that control the wavelet dilation and translation respectively; a_0 is a specified fixed dilation step greater than 1; and b_0 is the location parameter and must be greater than zero. The most common and simplest choice for parameters are $a_0 = 2$ and $b_0 = 1$. This power-of-two logarithmic scaling of the dilation and translation is known as the dyadic grid arrangement. The dyadic wavelet can be written in more compact notation as (Mallat, 1998):

$$g_{m,n}(t) = 2^{-m/2} g(2^{-m}t - n) \quad (3)$$

Discrete dyadic wavelets of this form are commonly chosen to be orthonormal; i.e. (Mallat, 1998):

$$\int_{-\infty}^{+\infty} g_{m,n}(t) g_{m',n'}(t) dt = \delta_{m,m'} \delta_{n,n'} \quad (4)$$

which δ is Kronecker delta.

This allows for the complete regeneration of the original signal as an expansion of a linear

combination of translates and dilates orthonormal wavelets.

For a discrete time series, x_i , the dyadic wavelet transform becomes (Mallat, 1998):

$$T_{m,n} = 2^{-m/2} \sum_{i=0}^{N-1} g(2^{-m}i - n)x_i \quad (5)$$

where $T_{m,n}$ is wavelet coefficient for the discrete wavelet of scale $a=2^m$ and location $b=2^m n$. Eq. (5) considers a finite time series, x_i , $i=0,1,2, \dots, N-1$; and N is an integer power of 2: $N=2^M$. This gives the ranges of m and n as, respectively, $0 < n < 2^{M-m} - 1$ and $1 < m < M$. At the largest wavelet scale (i.e., 2^m where $m=M$) only one wavelet is required to cover the time interval, and only one coefficient is produced. At the next scale (2^{m-1}), two wavelets cover the time interval, hence two coefficients are produced, and so on down to $m=1$. At $m=1$, the a scale is 2^1 , i.e., 2^{M-1} or $N/2$ coefficients are required to describe the signal at this scale. The total number of wavelet coefficients for a discrete time series of length $N=2^M$ is then $1 + 2 + 4 + 8 + \dots + 2^{M-1} = N - 1$.

In addition to this, a signal smoothed component, \bar{T} is left, which is the signal mean. Thus, a time series of length N is broken into N components, i.e., with zero redundancy. The inverse discrete transform is given by (Mallat, 1998):

$$x_i = \bar{T} + \sum_{m=1}^M \sum_{n=0}^{2^{M-m-1}} T_{m,n} 2^{-m/2} g(2^{-m}i - n) \quad (6)$$

or in a simple format as (Mallat, 1998):

$$x_i = \bar{T}(t) + \sum_{m=1}^M W_m(t) \quad (7)$$

which $\bar{T}(t)$ is called approximation sub-signal at level M and $W_m(t)$ are details sub-signals at levels $m=1, 2, \dots, M$.

The wavelet coefficients, $W_m(t)$ ($m=1, 2, \dots, M$), provide the detail signals, which can capture small features of interpretational value in the data; the residual term, $\bar{T}(t)$, represents the background information of data. Because of simplicity of $W_1(t)$, $W_2(t)$, ..., $W_M(t)$, $\bar{T}(t)$, some interesting characteristics, such as period, hidden period, dependence and jump can be diagnosed easily through wavelet components.

Wavelet denoising method based on threshold application was proposed for acquiring correct denoised results by Donoho (1995). This method, which is now the most common method of wavelet denoising, is performed as follows:

(I) Choose an appropriate mother wavelet and number of resolution level l . The original one-dimensional time series x_i is decomposed into an approximation at resolution level l and detailed signals at various resolution levels up to level l , using the wavelet transform.

(II) As a soft thresholding process, below certain threshold (T'), the absolute values of detailed signals, $d_j(t)$ ($j= 1, 2, \dots, l$), are set to zero at each resolution level. The subscript j represents the j th resolution level and sgn shows Signum function. The absolute values of detailed signals that exceed certain threshold are treated as the difference between the values of detailed signals and threshold by Eq. (8):

$$\hat{d}_j(t) = \begin{cases} sgn(d_j(t)) (|d_j(t)| - T') & |d_j(t)| > T' \\ 0 & |d_j(t)| \leq T' \end{cases} \quad (8)$$

Eq. (8) gives the threshold quantifications used to obtain the processed detailed signals at each resolution level during wavelet denoising. The approximation usually does not perform threshold quantifications.

(III) Wavelet reconstruction can derive the denoised time series data from the approximation at resolution level l and processed detailed signals ($\hat{d}_j(t)$) at all resolution levels.

Donoho (1995) derived a general optimal universal threshold for the white Gaussian noise under a mean square error criterion and its side condition that with high probability, the enhanced signal is at least as smooth as the clean signal. In this method, threshold is selected as Donoho (1995):

$$T' = \hat{\sigma} \sqrt{2 \ln(n)} \quad (9)$$

where n is number of samples in the noisy signal and $\hat{\sigma}$ is the standard deviation of noise that is estimated by Donoho (1995):

$$\hat{\sigma} = \left[\frac{\text{median}(|d_j(t)|)}{0.6745} \right] \quad (10)$$

in which $|d_j(t)|$ is the first level detail coefficients of wavelet transform of the signal.

2.2 Artificial Neural Network (ANN) and Efficiency Criteria

ANN is widely applied in hydrology and water resource studies as a forecasting tool. In ANN, feed forward (FF) back-propagation (BP) network models are common to engineers. It has proved that BP network model with three-layer is satisfied for the forecasting and simulating any engineering problem (Hornik, 1988; Nourani et al., 2008). Three-layered feed forward neural networks (FFNNs), which have been usually used for forecasting hydrologic time series, provide a general framework for representing nonlinear functional mapping between a set of input and output variables. Three-layered FFNNs are based on a linear combination of the input variables, which are transformed by a non-linear activation function as expressed by Eq. (11) where i, j and k denote input layer, hidden layer and output layer neurons, respectively and w is the applied weight by the neuron. The term “feed forward” means that a neuron connection only exists from a neuron in the input layer to other neurons in the hidden layer or from a neuron in the hidden layer to neurons in the output layer and the neurons within a layer are not interconnected to each other. The explicit expression for an output value of a three-layered FFNN is given by Nourani et al. (2008):

$$\hat{y}_k = f_0 \left[\sum_{j=1}^{M_N} w_{kj} \cdot f_h \left(\sum_{i=1}^{N_N} w_{ji} x_i + w_{j0} \right) + w_{k0} \right] \quad (11)$$

where $w_{i,j}$ is a weight in the hidden layer connecting the i th neuron in the input layer and the j th neuron in the hidden layer, w_{j0} is the bias for the j th hidden neuron, f_h is the activation function of the hidden neuron, w_{kj} is a weight in the output layer connecting the j th neuron in the hidden layer and the k th neuron in the output layer, w_{k0} is the bias for the k th output neuron, f_0 is the activation function for the output neuron, x_i is i th input variable for input layer and \hat{y}_k, y are

computed and observed output variables, respectively. N_N and M_N are the number of the neurons in the input and hidden layers, respectively. The weights are different in the hidden and output layers, and their values can be changed during the process of the network training.

The model that yields the best results in terms of determination coefficient (R^2) as Eq. (12) and root mean squared error ($RMSE$) as Eq. (13) in the training and verifying steps can be determined through trial and error process (Nourani et al., 2009a).

$$R^2 = 1 - \frac{\sum_{i=1}^N (O_{obs_i} - O_{com_i})^2}{\sum_{i=1}^N (O_{obs_i} - \bar{O}_{obs})^2} \quad (12)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (O_{obs_i} - O_{com_i})^2}{N}} \quad (13)$$

where R^2 , $RMSE$, N , O_{obs_i} , O_{com_i} and \bar{O}_{obs} are determination coefficient, Root Mean Squared Error, number of observations, observed data, computed values and mean of observed data, respectively.

The $RMSE$ is used to measure forecast accuracy, which produces a positive value by squaring the errors. The $RMSE$ increases from zero for perfect forecasts through large positive values as the discrepancies between forecasts and observations become increasingly large. Obviously high value for R^2 (up to one) and small value for $RMSE$ indicate high efficiency of the model.

2.3 Case study and data

Time series data of daily streamflow and SSL at outlet of the Potomac River located in Frederick

County of Maryland State, USA (Latitude 39°16'24.9", Longitude 77°32'35.2") were used in this study (Fig.1). The river forms part of the borders between Washington D.C., and Virginia on the left descending bank and Maryland and Virginia on the river's right descending bank, 33% is farmland and pasture and an estimated 27% is urban. Approximately, 40% of the Potomac River basin is forested; gently sloping hills and valleys from Harpers Ferry to approximately 72.42 kilometer downriver characterize the topography of the upper Piedmont region of the Potomac River sub-basin. In the central Piedmont area, the profile is rather flat until it nears the fall line at Great Falls, where the stream elevation rapidly descends from over 61 meter to sea level. Tributaries in the central Piedmont exhibit moderate and near constant profiles. Their flat slope largely characterizes streams in the Coastal Plain area. Approximately 40% of the Potomac River basin is forested 33% is farmland and pasture and an estimated 27% is urban. The average flow of river observed at the outlet is 306m³/s. The majority of the lower Potomac River is part of the State of Maryland. The upstream drainage area is 25996 Km². The historical daily streamflow and SSL data for 20 years (from 1960 to 1980, 7333 days) which were used in this research are available at the United States Geological Survey website (USGS, <http://waterdata.usgs.gov/usa/nwis/uv?01638500>). The time series are presented in Figs. 2 and 3 for streamflow and SSL, respectively. The 75% part of total data (from 1960 to 1975) were used as training set and the rest 25% (from 1975 to 1980) were used to determine how well the training model performed.

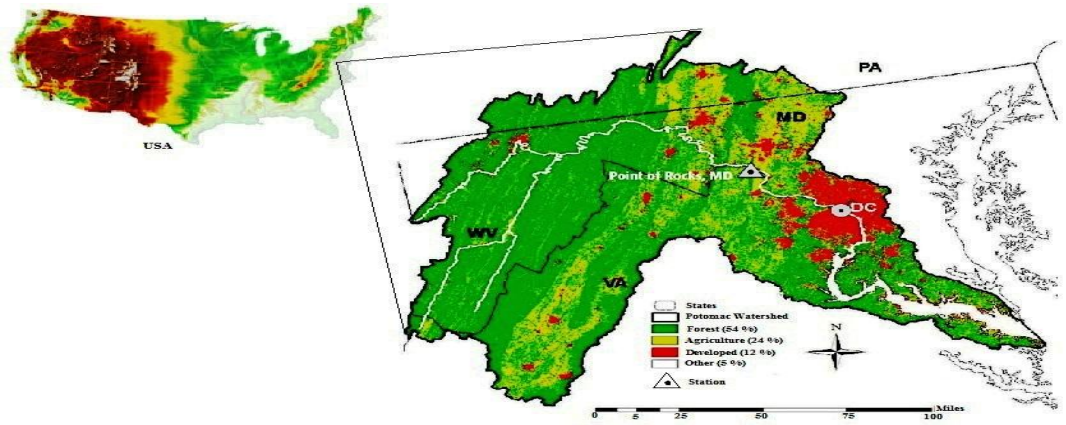


Fig. 1 The Potomac River at Point of Rocks at Frederick County.

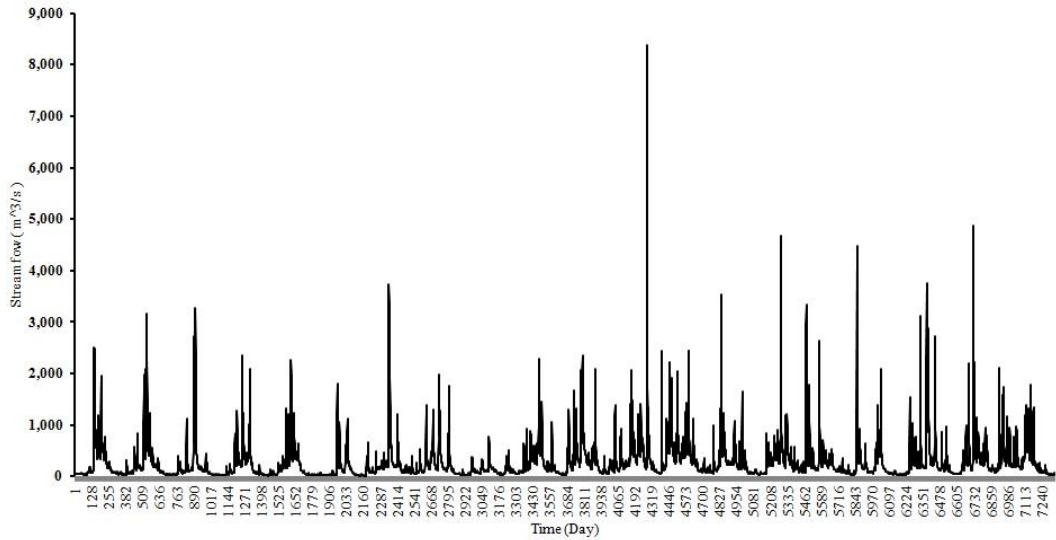


Fig. 2 Streamflow time series observed at outlet of the Potomac River from 1960 to 1980

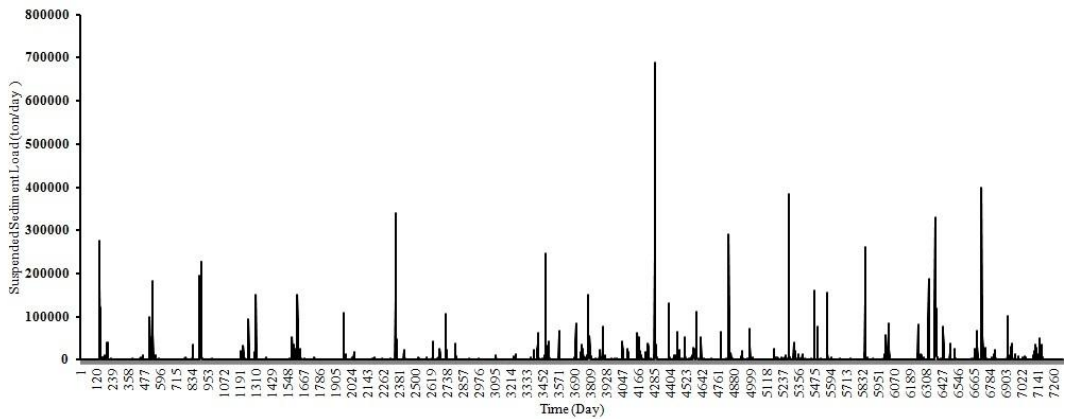


Fig. 3 Sediment time series observed at outlet of the Potomac River from 1960 to 1980

Statistical Parameters	All Data		Training Data		Verifying Data	
	Q(m ³ /s)	SSL (ton/d)	Q(m ³ /s)	SSL (ton/d)	Q(m ³ /s)	SSL (ton/d)
X _{mean}	276.42	3144	266.94	3182.03	304.84	41885.46
X _{max}	8288	689000	8288	689000	4816	400000
X _{min}	15.316	2	15.316	2	30.8	3.7
S _d	2324.3	19082.3	2271	18541	2469.8	20622

Table 1. Statistics of time series for calibration, verification and all data sets.

The statistical parameters of the streamflow and sediment data are given in Table 1. In Table 1, the X_{mean}, S_d, X_{max} and X_{min} denote the mean, standard deviation, maximum and minimum values, respectively.

3. RESULTS & DISCUSSION

At first, the SSL was modeled via ANN with raw, noisy data. Afterward, inputs were denoised by threshold based wavelet denoising method and re-imposed into ANN. The Sediment- streamflow process usually behaves as a Markov chain process, so that the value of the parameter in the current time may be related to the previous time step condition. In this study, seven combinations of streamflow and SSL values as inputs for the ANN were examined as follows:

- Comb. (1): SSL_{t-1}
- Comb. (2): SSL_{t-1}, SSL_{t-2}
- Comb. (3): Q_{t-1}, Q_{t-2}
- Comb. (4): SSL_{t-1}, Q_{t-1}
- Comb. (5): SSL_{t-1}, SSL_{t-2}, Q_{t-1}
- Comb. (6): SSL_{t-1}, SSL_{t-2}, Q_{t-1}, Q_{t-2}

Comb. (7): SSL_{t-1}, SSL_{t-2}, SSL_{t-3}, Q_{t-1}, Q_{t-2}, Q_{t-3} in all cases, *t* represent time step and the output layer was contained only one neuron, as the SSL at time *t* (SSL_{*t*}).

The network training process could be speeded up by normalizing the input and target data before training Rogers (1996). In this study, the input and target data were normalized to scale the data between the range 0 and 1 as:

$$y_i = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad (14)$$

to normalized *x_i* variable with minimum and maximum values of *x_{min}* and *x_{max}*, respectively. *y_i* is the normalized variable.

In this application, two sets of data were used. The first data set was 75% part of total data (from 1960 to 1975) as training set. The rest 25% (from 1975 to 1980) was used for verification purpose. According to Table 1, maximum values of the data were appeared in the training set because such data division scheme helps ANN, as a data interpolator, to learn the pattern of process much

better and leads to more accurate predictions in the validation step, so when data set is transferred between $[0,1]$, a small change in y_{max} is the upper bound of the normalized interval, a normalized input in the mentioned range, has a greater influence on the output which makes the training of the ANN more quickly (Nourani et al., 2012c). SSL predicting was performed with different inputs to determine the best input combination. Because of some financial and technical problems the exact measurement of the SSL is a very difficult matter and usually after several years measuring will be stopped. In such condition, a model which employs available streamflow data in order to estimate the SSL values can be reliable choice. So, Comb.3 is essential among other combinations. The values of R^2 obtained using combs.1 and 4 inputs show that for achieving to good performance of model data with more than one time step lag are required. On the other hand, in combinations with one type variable (streamflow or SSL) combs.2 and 3 are suitable for SSL and streamflow, respectively. These combinations will be used in choosing parameters for threshold based wavelet denoising of SSL and discharge time series. However, when both variables take place in the input neurons, comb.6 leads to the best performance.

Table2. Results of ANNs for different input variables

Input Combination	Network Structure ^a	RMSE (Normalized)		R ²	
		Calibration	Verification	Calibration	Verification
Comb. 1	(1-5-1) ^b	0.0209	0.022	0.396	0.365
Comb. 2	(2-10-1)	0.0149	0.0189	0.692	0.601
Comb. 3	(2-10-1)	0.0162	0.0202	0.635	0.543
Comb. 4	(2-10-1)	0.0199	0.022	0.449	0.374
Comb. 5	(3-7-1)	0.0150	0.0181	0.685	0.630
Comb. 6	(4-4-1)	0.0153	0.0176	0.674	0.651
Comb. 7	(6-6-1)	0.0153	0.0185	0.676	0.618

^a The results have been presented for the best structure

^b First number from left represents number of the input neurons, second one is the number of the hidden neurons and third shows the number of output neuron.

In this study, the river daily suspended sediment was modeled via the multi-layer perceptron FF ANN without any data pre-processing. This kind of ANN model accompanied by BP training algorithm is extremely used in hydrologic modeling (ASCE, 2000). Selection of an appropriate mother wavelet is a substantial challenge in wavelet denoising issue, since the type of used mother wavelet can affect the modeling results remarkably (Nejad and Nourani 2012). The essence of wavelet transform is to discover the similarity between the analyzed series and the wavelet used however in practice it can't be done easily. In discrete wavelet analysis, several ways to decomposition level choice have been suggested by Nourani et al. (2009a) and Sang(2009). The threshold based wavelet denoising has three important parameters: decomposition level, mother wavelet and threshold value. This paper's effort is to introduce a procedure to show how to choose appropriate mother wavelet, decomposition level and threshold value for denoising two different hydrological time series (i.e., streamflow and SSL) in predicting SSL.

In this paper, a hydrological hypothesis was considered to pick out the minimum decomposition level. In this manner, the signal approximation which is the general aspect of SSL or discharge time series should be positive and doesn't include any negative data (Fig. 5). In view of the fact that negative data don't have any physical meaning and can't represent considered hydrological processes (i.e., streamflow and SSL). Due to such a strict condition about values of approximation signal, decomposition levels less than eight were not chosen. Three decomposition levels (8, 9 and 10) were used to investigate decomposition level effect on denoising performance (Table 3). On the other hand, Daubechies family of wavelet (Haar, *db2*, *db3*, *db4*, *db5*) was examined as the mother wavelets (Fig. 4). Efficiency of the mother wavelets at different decomposition levels was compared using the universal thresholding method (see

Table 3). According to Table 3, it is noticeable that universal threshold based denoising value was effected by varying the mother wavelet i.e. each mother wavelet requires its own best universal threshold value. This procedure was performed for SSL time series by Comb.2 and for streamflow

time series by Comb.3. The results in Table 3 show that based on R^2 , mother wavelet *db5* with $l = 8$ and mother wavelet *db3* with $l = 10$ are appropriate for global denoising of SSL and river discharge time series, respectively.

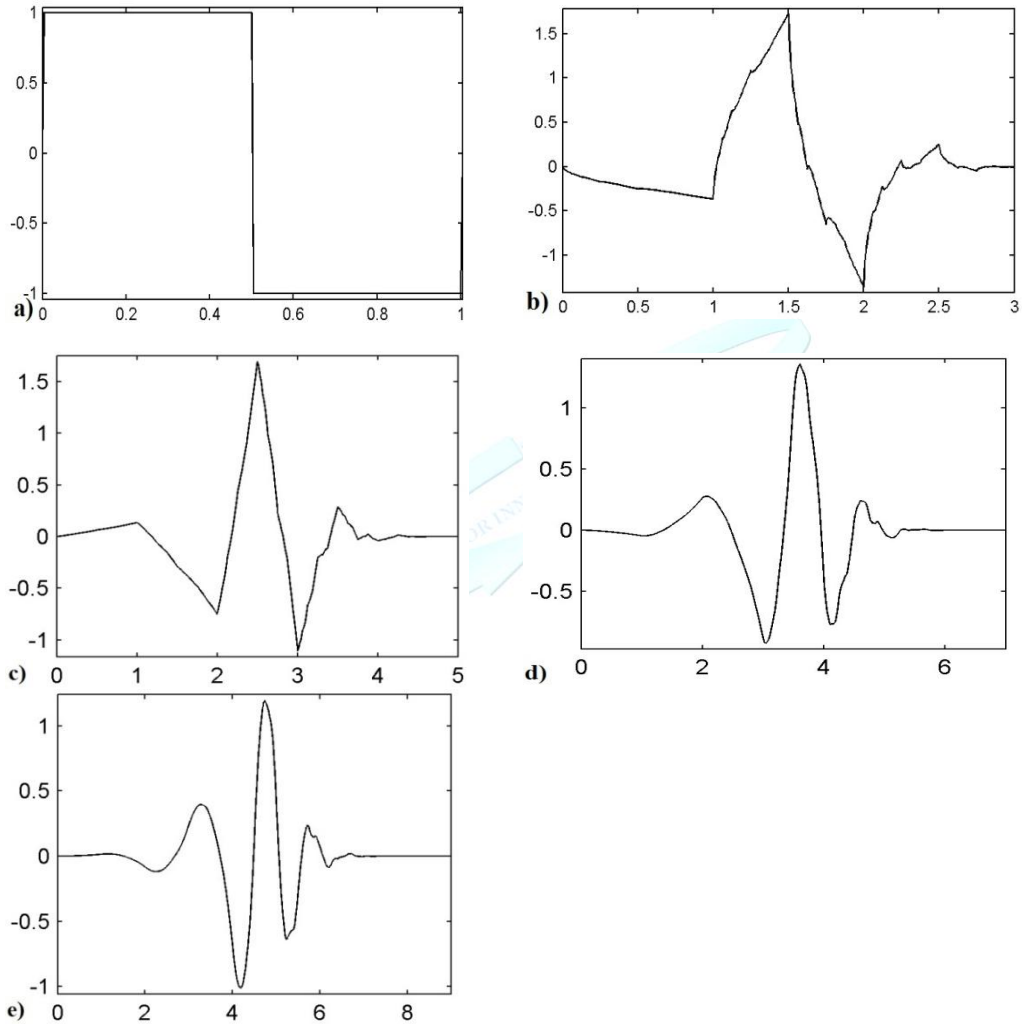


Fig.4 a)Haar b)db2 c)db3 d)db4 e)db5 mother wavelets

Table3. Results of ANNs for investigation of mother wavelet and resolution level impacts

Mother Wavelet	Input comb.	Decomposition Level	Network Structure	RMSE (Normalized)	R^2
----------------	-------------	---------------------	-------------------	-------------------	-------

				Calibration	Verification	Calibration	Verification
<i>Haar</i>	(Q_{t-1}, Q_{t-2})	8	(2-4-1)	0.0165	0.0191	0.619	0.590
		9	(2-8-1)	0.0163	0.0191	0.663	0.593
		10	(2-10-1)	0.0151	0.0187	0.682	0.607
	(SSL_{t-1}, SSL_{t-2})	8	(2-5-1)	0.0161	0.0182	0.639	0.628
		9	(2-7-1)	0.0151	0.0185	0.684	0.616
		10	(2-9-1)	0.0144	0.0182	0.709	0.630
<i>db2</i>	(Q_{t-1}, Q_{t-2})	8	(2-4-1)	0.0165	0.0193	0.621	0.584
		9	(2-4-1)	0.0165	0.0190	0.621	0.594
		10	(2-4-1)	0.0162	0.0185	0.635	0.615
	(SSL_{t-1}, SSL_{t-2})	8	(2-10-1)	0.0161	0.0219	0.639	0.603
		9	(2-10-1)	0.0147	0.0184	0.699	0.620
		10	(2-6-1)	0.0154	0.0187	0.668	0.608
<i>db3</i>	(Q_{t-1}, Q_{t-2})	8	(2-7-1)	0.0161	0.0195	0.641	0.572
		9	(2-10-1)	0.0156	0.0189	0.662	0.597
		10	(2-8-1)	0.0153	0.0172	0.676	0.669
	(SSL_{t-1}, SSL_{t-2})	8	(2-10-1)	0.0154	0.0182	0.671	0.667
		9	(2-6-1)	0.0165	0.0189	0.620	0.601
		10	(2-7-1)	0.0164	0.0185	0.625	0.614
<i>db4</i>	(Q_{t-1}, Q_{t-2})	8	(2-7-1)	0.0168	0.0189	0.610	0.600
		9	(2-5-1)	0.0155	0.0180	0.668	0.636
		10	(2-7-1)	0.0166	0.0191	0.625	0.591
	(SSL_{t-1}, SSL_{t-2})	8	(2-6-1)	0.0168	0.0187	0.608	0.608
		9	(2-7-1)	0.0151	0.0185	0.684	0.617
		10	(2-7-1)	0.0159	0.0182	0.648	0.626
<i>db5</i>	(Q_{t-1}, Q_{t-2})	8	(2-4-1)	0.0169	0.0189	0.602	0.599
		9	(2-7-1)	0.0159	0.0193	0.650	0.583
		10	(2-6-1)	0.0166	0.0189	0.615	0.601
	(SSL_{t-1}, SSL_{t-2})	8	(2-4-1)	0.0145	0.0180	0.651	0.632
		9	(2-7-1)	0.0147	0.0186	0.698	0.611
		10	(2-6-1)	0.0168	0.0187	0.608	0.607

The presented values of threshold in Table 3 have been computed by Eq. (9). It is important to mention that Eq. (9) has been developed on the basis of the Gaussian noise involved in the signal. Moreover, the noise included in a complex hydrological process may not obey Gaussian noise. So, more trial-error procedure is needed to determine the appropriate threshold value. The

threshold value influences denoising the hydrological time series in two different aspects. First, if threshold value be less, then noise reduction will be skimpy, inversely if threshold value is high, as a result, noise will be removed but on the other hand, main information of the time series will be also removed. Nevertheless, the appropriate threshold not only removes the noise

but also keeps the main part of the time series. So determination of the appropriate threshold value leads to achieve higher performance in the forecasting phase. The different threshold values around the threshold value computed by Eq. (9), were also examined for both streamflow and SSL time series to find the appropriate threshold value (See Tables 4 and 5). As it depicted in Tables 4 and 5, by increasing the threshold value of the denoising, overall likewise R^2 is increased but after a special threshold, R^2 goes down. This

special threshold can be considered as the "appropriate threshold" which in SSL and streamflow time series are 9000 (ton/day) and 453.1 (m^3/s), respectively. As it is clear, R^2 for the noisy time series with the best combination is 0.670 but just with streamflow time series denoising and using Comb.3 it is elevated to 0.708. So, in the modeling, using only streamflow values as the inputs when there is a gap for SSL data, SSL could be predicted with acceptable accuracy (Table 4)

Table4. Results of ANNs after denoising by db3 at level 10 for comb.3*

Threshold Value(m^3/s)	Network Structure(Inputs= Q_{t-1}, Q_{t-2})	RMSE (Normalized)		R^2	
		Calibration	Verification	Calibration	Verification
2.832	(2-9-1)	0.0162	0.0205	0.635	0.529
19.822	(2-10-1)	0.0151	0.0176	0.685	0.653
28.316	(2-10-1)	0.0154	0.0194	0.672	0.581
42.475	(2-5-1)	0.0163	0.0182	0.633	0.628
56.633	(2-7-1)	0.0164	0.0187	0.628	0.609
84.950	(2-10-1)	0.0154	0.085	0.669	0.614
113.267	(2-8-1)	0.0156	0.0174	0.663	0.662
169.900	(2-4-1)	0.0149	0.0166	0.691	0.689
226.534	(2-4-1)	0.0147	0.0164	0.699	0.698
283.168	(2-10-1)	0.0144	0.0164	0.712	0.698
339.802	(2-7-1)	0.0144	0.0164	0.713	0.698
424.752	(2-10-1)	0.0147	0.0163	0.710	0.700
453.068	(2-9-1)	0.0144	0.0162	0.713	0.708
458.732	(2-7-1)	0.0144	0.0163	0.710	0.702
481.385	(2-8-1)	0.0146	0.0163	0.704	0.702
509.702	(2-9-1)	0.0146	0.0164	0.702	0.700
566.336	(2-4-1)	0.0150	0.0168	0.688	0.683
707.92	(2-7-1)	0.0155	0.0174	0.665	0.659
849.504	(2-5-1)	0.0162	0.0182	0.636	0.631

Table5. Results of ANNs after denoising by db5 at level 8 for comb.2*

Threshold Value (ton/day)	Network Structure (Inputs= S_{t-1}, S_{t-2})	RMSE (Normalized)		R^2	
		Calibration	Verification	Calibration	Verification
50	(2-4-1)	0.0152	0.0182	0.681	0.632

4000	(2-6-1)	0.0149	0.0170	0.695	0.679
5000	(2-8-1)	0.0145	0.0167	0.708	0.686
5500	(2-9-1)	0.0146	0.0167	0.704	0.687
6000	(2-7-1)	0.0145	0.0169	0.706	0.678
6500	(2-8-1)	0.0146	0.0168	0.704	0.686
7500	(2-9-1)	0.0147	0.0167	0.701	0.685
8500	(2-7-1)	0.0147	0.0167	0.702	0.689
9000	(2-9-1)	0.0148	0.0166	0.699	0.694
10500	(2-8-1)	0.0151	0.0171	0.685	0.671
11000	(2-6-1)	0.0148	0.0168	0.697	0.684
12000	(2-9-1)	0.0148	0.0168	0.694	0.682
13000	(2-8-1)	0.0149	0.0166	0.693	0.693
13500	(2-6-1)	0.0147	0.0169	0.699	0.681
14500	(2-8-1)	0.0153	0.0177	0.672	0.648
16000	(2-7-1)	0.0145	0.0169	0.709	0.679
16500	(2-7-1)	0.0148	0.0172	0.696	0.672
17000	(2-5-1)	0.0145	0.0169	0.708	0.686

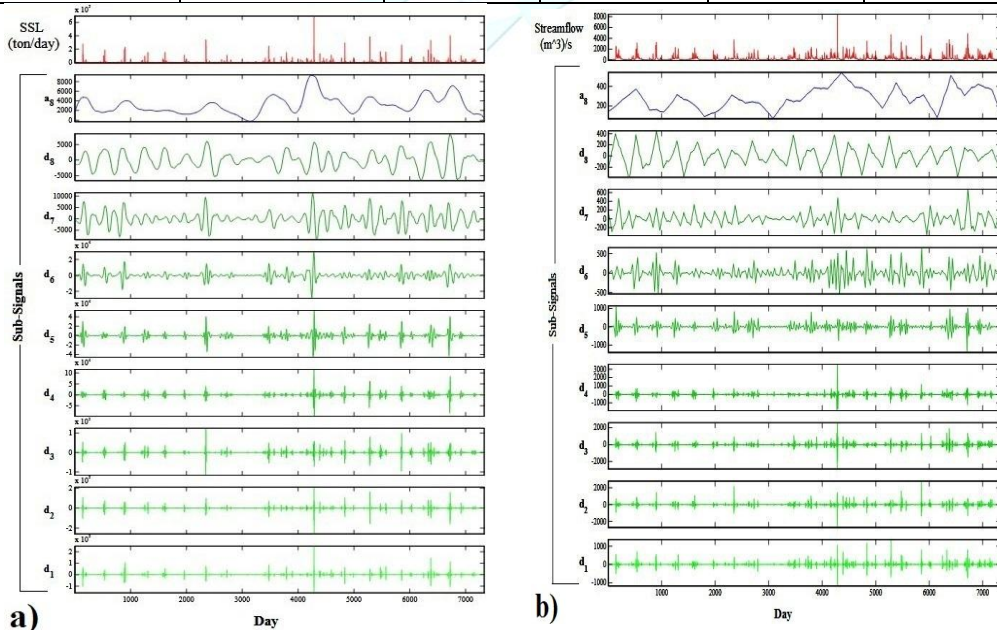


Fig. 5 a) decomposed SSL time series at level 8 via *db5* (a_s : approximation signal, d_1, d_2, \dots, d_8 : detailed signals)
b) decomposed streamflow time series at level 8 via *db3* (a_s : approximation signal, d_1, d_2, \dots, d_8 : detailed signals)

The proposed denoising scheme was also applied for input comb.6 (this input comb. was led to best performance when the raw, noisy data were used as inputs of ANN, Table 2) and results have been tabulated in Table 6.

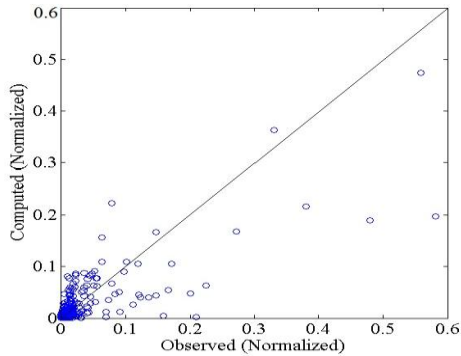


Fig. 6 Scatter plot for observed and computed sediment using ANN with noisy data at verification step

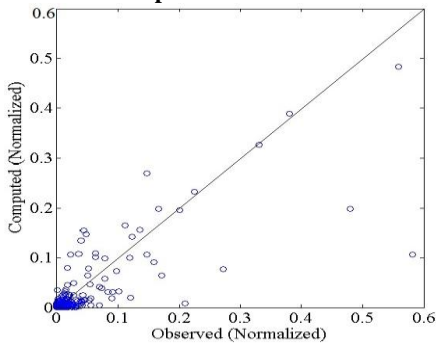


Fig. 7 Scatter plot for observed and computed sediment using ANN with denoised data at verification step

Table 6 shows the values of R^2 and $RMSE$ for three different input sets. i) Both denoised streamflow and SSL time series were considered as inputs. ii) Only one denoised time series (streamflow or SSL) was considered as input (the other time series were imposed without pre-processing). iii) Both raw streamflow and SSL time series were imposed without denoising. With the respect to the result obtained in Table 6, it is obvious that denoised streamflow time series has fundamental effect on final result compared with denoised sediment time series and this shows that streamflow time series contains much noise than sediment time series. The absolute value for the R^2 criterion obtained by applying both denoised signals simultaneously into ANN (0.802) is significantly better than obtained by raw data (0.651). In term of the $RMSE$, the ANN result with denoised data (0.0133) outperforms the obtained result with raw data (0.0176). Figs 6 and 7 present scatter plots between observed and computed SSL using ANN method with raw and denoised data, respectively, at verification step. In Fig.8 observed noisy data was compared with computed SSL with denoised data.

Table6. Final results and structures of ANN model for noisy and denoised inputs

Input combination	Mother wavelet		Decomposition	Network structure	RMSE (Normalized)		R^2	
					Calibration	Verification	Calibration	Verification
$S_{t-1}, S_{t-2}, Q_{t-1}, Q_{t-2}$	raw data imposed		-	(4-4-1)	0.0153	0.0176	0.674	0.651
			-					
$S_{t-1}, S_{t-2}, Q_{t-1}, Q_{t-2}$	Sed ^a	<i>Db5</i>	8	(4-4-1)	0.0132	0.0162	0.756	0.704
	Str ^b	- ^c	-					
$S_{t-1}, S_{t-2}, Q_{t-1}, Q_{t-2}$	Sed	-	-	(4-9-1)	0.0126	0.0145	0.764	0.726
	Str	<i>Db3</i>	10					
$S_{t-1}, S_{t-2}, Q_{t-1}, Q_{t-2}$	Sed	<i>Db5</i>	8	(4-4-1)	0.0119	0.0133	0.802	0.802
	Str	<i>Db3</i>	10					

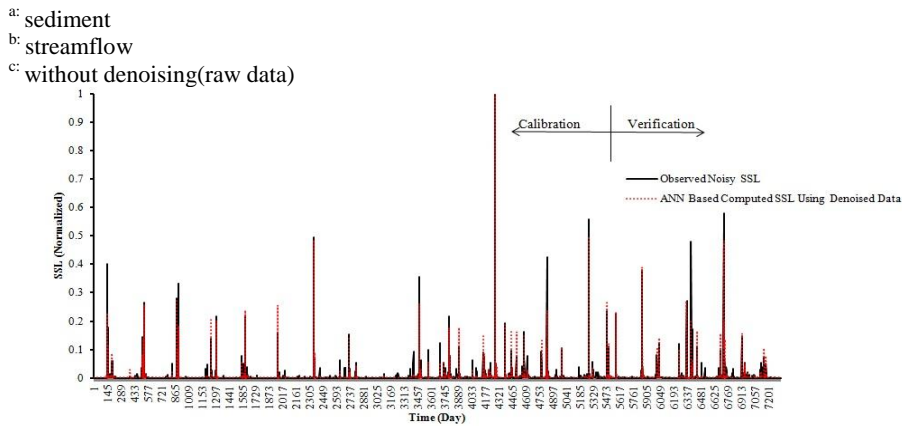


Fig. 8 Computed and observed data decomposed at db5 with L = 8 for SSL and db3 decomposed at L =10 for streamflow time series

4. CONCLUSIONS

Accurate prediction of suspended sediment load (SSL) is a significant key for computing and designing of the costly projects in both water resource management and environmental aspects. An attempt was made in this paper to investigate the use of wavelet-based denoising technique as a preprocessing method for daily SSL prediction by Artificial Neural Network for the Potomac River at Maryland, USA. Opting the best combination was done for starting the denoising procedure. Through this procedure a suitable wavelet function and proper decomposition level for both streamflow and sediment time series was determined. Then decomposed time series was denoised by the universal thresholding method. The comparison results reveal that chosen mother wavelet and resolution level directly affect the prediction result. On the other hand, the threshold value, as well as so-called factors, is another challenging issue. According to obtained result high values of threshold didn't mean accurate result and after a specific threshold value the reduction in performance of the model was occurred. So, this specific threshold value was called "appropriate threshold" (Nejad and Nourani 2012). Meanwhile, the same mother wavelet and resolution level can be employed for denoising

streamflow and sediment time series and got the capable result with slightly difference with the best result. At the end, result presented that this procedure extensively enhanced accuracy when modeling streamflow-SSL process. In order to complete current study, it is suggested to use level-dependent soft threshold based wavelet denoising and also to do this work for monthly time series and the results to be compared with other modeling methods as Auto Regressive Integrated Moving Average and Support Vector Machine.

REFERENCES

- [1] Jain, A. and Ormsbee, L.E., (2002). "Evaluation of short-term water demand forecast modeling techniques: Conventional methods versus AI." *J. American Water Works Assoc.*, 94(7): 64-72.
- [2] Salas, J.D., Delleur, J.W., Yevjevich, V. and Lane, W.L., (1980). "Applied modeling of hydrologic time series." *Water Resour. Publ., Highlands Ranch, Colo.*
- [3] Zhang, G., Patuwo, E. and Hu, M., (1998). "Forecasting with artificial neural networks: The state of the art." *Int. J. Forecast.*, 14: 35-62.

- [4] Bougadis, J., Adamowski, J.F. and Diduch, R., (2005). "Short-term municipal water demand forecasting." *Hydrol. Process*, 19: 137–148.
- [5] Adamowski, J.F., (2008). "River flow forecasting using wavelet and cross-wavelet transform models." *Hydrol. Process*, 22: 4877–4891.
- [6] Adamowski, J.F. and Sun, K., (2010). "Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds." *J. Hydrol.*, 390: 85–91.
- [7] Guldal, V. and Tongal, H., (2010). "Comparison of recurrent neural network, adaptive neuro-fuzzy inference system and stochastic models in Egirdir Lake level forecasting." *Water Resour. Manag.*, 24(1): 105–128.
- [8] Shirsath, P.B. and Singh, A.K., (2010). "A comparative study of daily pan evaporation estimation using ANN, regression and climate based models." *Water Resour. Manag.*, 24(8): 1571-1581(11).
- [9] Hsu, K.L., Gupta, H.V. and Sorooshian, S., (1995). "Artificial neural network modeling of the rainfall-runoff process." *Water Resour. Res.*, 31: 2517-2530.
- [10] Tokar, A.S. and Johnson, P.A., (1999). "Rainfall-runoff modeling using artificial neural network." *J. Hydrol. Eng.*, 3: 232-239.
- [11] Nourani, V., Komasi, M. and Mano, A., (2009a). "A multivariate ANN-wavelet approach for rainfall-runoff modeling." *Water Resour. Manag.*, 23: 2877–2894.
- [12] Nourani, V., Alami, M.T. and Aminfar, M.H., (2009b). "A combined neural-wavelet model for prediction of Ligvanchai watershed precipitation." *Eng. Appl. Artif. Intel.*, 22: 466-472.
- [13] Nourani, V., Komasi, M. and Alami, M.T. (2012a). "A hybrid wavelet-genetic programming approach to optimize ANN modeling of rainfall-runoff process." *J. Hydrol. Eng.*, 17(6): 724-741.
- [14] Rezaeian Zadeh, M., Amin, S., Khalili D. and Singh, V.P., (2010). "Daily outflow prediction by multi layer perceptron with Logistic Sigmoid and Tangent Sigmoid activation functions." *Water Resour. Manag.*, 24(11): 2673-2688.
- [15] Abrahart, J.R., Anctil, F., Coulibaly, P., Dawson, Ch., Mount, N., See, L., Shamseldin, A., Solomatine, D., Toth, E., L.Wilby, R., (2012). "Two decades of anarchy? Emerging themes and outstanding challenges for neural network river forecasting." *Prog. Phys. Geog.* 36(4): 480-513.
- [16] Jain, S.K., (2001). "Development of integrated sediment rating curves using ANNs." *J. Hydraul. Eng.* 127 (1): 30–37.
- [17] Tayfur, G., (2002). "Artificial neural networks for sheet sediment transport." *Hydrol. Sci. J.*, 47 (3): 879–892.
- [18] Agarwal, A., Mishra, S.K., Ran, S. and Singh, J. K., (2006). "Simulation of runoff and sediment yield using artificial neural networks." *Biosys. Eng.*, 94(4): 597-613.
- [19] Alp, M. and Cigizoglu, HK., (2007). "Suspended sediment load simulation by two artificial neural network methods using hydrometeorological data." *Environ. Model Softw.*, 22: 2-13.
- [20] Rai, R.K. and Mathur, B.S., (2008). "Event-based sediment yield modeling using artificial neural network." *Water Resour. Manag.*, 22: 423-441.
- [21] Nourani, V. (2009). "Using artificial neural networks (ANNs) for sediment load forecasting of Talkherood River mouth." *J. Urban Environ. Eng.*, 3(1): 1-6.
- [22] Rajaei, T., Mirbagheri, S. A., Nourani, V. and Alikhani, A., (2010). "Prediction of daily suspended sediment load using wavelet and neuro-fuzzy combined model." *Int. J. Environ. Sci. Tech.*, 7 (1): 93-110.
- [23] Nourani, V., Kalantari, O. and Baghanam, A., (2012b). "Two semi-distributed ANN-based models for estimation of suspended sediment load." *J. Hydrol. Eng.*, 17(12): 1368–1380.
- [24] Schouten, J.C., Takens, F. and Van den Bleek, C.M., (1994). "Estimation of the dimension of a noisy attractor." *Phys. Rev. E.*, 50(3): 1851–1861.

- [25] Wiener, N., (1949). "Extrapolation, interpolation, and smoothing of stationary time series." *J. Am. Stat. Assoc.*, 47(258): 319-321.
- [26] Kalman, R.E., (1960). "A new approach to linear filtering and prediction problems." *J. Basic Eng-T*, 82(1) :35-45.
- [27] Reichel, R.H., Walker, J.P., Koster, R.D. and Houser, P.R., (2002). "Extended versus ensemble Kalman filtering for land data assimilation." *J. Hydrometeorol.*, 3: 728-740.
- [28] Hamid, M., Soroosh, S., Hoshin, V.G. and Paul, R.H., (2004). "Dual state-parameter estimation of hydrological models using ensemble Kalman filter." *Adv. Water Resour.*, 28(2): 135-147.
- [29] Martyn, P.C., David, E.R., Ross, A.W., Xiaogu, Z., Richard, P.I., Andrew, G.S., Jochen, S. and Michael J.U., (2008). "Hydrological data assimilation with the ensemble Kalman filter: Use of streamflow observations to update states in a distributed hydrological model." *Adv. Water Resour.*, 31: 1309-1324.
- [30] Sivakumar, B., Phoon, k.k., Liong, S.Y. and Liaw C.Y., (1999). "A systematic approach to noise reduction in chaotic hydrological time series." *J. Hydrol.* 219: 103-135.
- [31] Elshorbagy, A., Simonovic, S.P. and Panu, U.S., (2002). "Noise reduction in chaotic hydrologic time series: facts and doubts." *J. Hydrol.*, 256: 147-165.
- [32] Sang, Y.F., (2013). "A review on the applications of wavelet transform in hydrology time series analysis." *Atmos. Res.*, 122: 8-15.
- [33] Jansen, M., (2006). "Minimum risk thresholds for data with heavy noise." *IEEE Signal Process Lett.* , 13: 296-299.
- [34] Cannas, B., Fanni, A., Seeand, L. and Sias, G., (2006). "Data preprocessing for river flow forecasting using neural networks: wavelet transforms and data partitioning." *Phys. Chem. Earth*, 31(18):1164-1171.
- [35] Guo, J., Zhou, J., Qin, H., Zou, Q. and Li, Q., (2011). "Monthly streamflow forecasting based on improved support vector machine model." *Expert Syst. Appl.*, 38:13073-13081.
- [36] Nejad, F.H. and Nourani, V., (2012) "Elevation of wavelet denoising performance via an ANN-based streamflow forecasting model." *Int. J. Comput. Sci. Manage. Res.*, 1(4): 764-770.
- [37] Nourani, V., Baghanam, A., Adamowski, J.F. and Gebremichael, M., (2013). "Using self-organizing maps and wavelet transforms for space-time pre-processing of satellite precipitation and runoff data in neural network based rainfall-runoff modeling." *J. Hydrol.*, 476: 228-243.
- [38] Grossman, A. and Morlet, J., (1984). "Decomposition of Hardy functions into square integrable wavelets of constant shape." *J. Math. Anal.*, 15(4): 723-736.
- [39] Foufoula-Georgiou, E. and Kumar, P., (1995). "Wavelets in geophysics." Academic, San Diego, CA, 4: 1-43.
- [40] Labat, D., (2005). "Recent advances in wavelet analyses. Part 1: a review of concepts." *J. Hydrol.*, 314:275 - 288.
- [41] Mallat, S.G., (1998). "A Wavelet tour of signal processing." Academic, San Diego, CA.
- [42] Labat, D., Ababou, R. and Mangin, A., (2000). "Rainfall-runoff relation for Karstic Spring. Part 2: continuous wavelet and discrete orthogonal multi resolution analyses." *J. Hydrol.*, 238: 149 - 178.
- [43] Addison, P.S., Murrar, K.B. and Watson, J.N., (2001). "Wavelet transforms analysis of open channel wake flows." *J. Eng. Mech.*, 127(1) :58 - 70.
- [44] Donoho, D.H., (1995). "Denoising by soft-thresholding." *IEEE T. Inform. Theory.*, 41(3): 613-617.
- [45] Hornik, K., (1988) "Multilayer feed-forward networks are universal approximators." *Neural Networks.*, 2(5), pp. 359-366.
- [46] Nourani V., Mogaddam A.A. and Nadiri A.O., (2008). "An ANN based model for spatiotemporal groundwater level forecasting." *Hydrol. Process*, 22: 5054-5066.
- [47] ASCE Task Committee on Application of Artificial Neural Networks in Hydrology

- (2000) Artificial Neural Networks in hydrology 2: hydrologic applications, *J. Hydrol. Eng.*, 5(2): 124–137.
- [48] Sang, Y.F., Wang, D., Wu, J.C., Zhu, Q.P. and Wang, L., (2009). “Entropy-based wavelet de-noising method for time series analysis.” *Entropy*, 11(4): 1123–1147.
- [49] Rogers, R., (1996). “Neural Networks: A Systematic Introduction.” *Springer Verlag, Berlin*.
- [50] Nourani, V. and Fard, M.S., (2012c). “Sensitivity analysis of the artificial neural network outputs in simulation of the evaporation process at different climatologicr.” *Adv. Eng. Softw.*, 47(1): 127-146.

