Optimizing Neural Network for Monthly Rainfall-Runoff Modeling with Denoised-Jittered Data

Afshin Partovian[1], Vahid Nourani[2*], Mohammad Taghi Aalami[2]

1- Faculty of Civil Engineering, Boukan Branch, Islamic Azad University, Boukan, Iran.

2- Faculty of Civil Engineering, University of Tabriz, 29 Bahman Ave, Tabriz, 51666, Iran.

* Corresponding Author: nourani@tabrizu.ac.ir

Received: 02 June 2016 / Accepted:22 December 2016 / Published online: 27 December 2016

Abstract

Successful modeling of hydro-environmental processes widely relies on quantity and quality of accessible data and noisy data might effect on the functioning of the modeling. On the other hand in training phase of any Artificial Intelligence (AI) based model, each training data set is usually a limited sample of possible patterns of the process and hence, might not show the behavior of whole population. Accordingly in the present article first, wavelet-based denoising method was used in order to smooth hydrological time series and then small normally distributed noises with the mean of zero and various standard deviations were generated and added to the smoothed time series to form different denoised-jittered training data sets, for Artificial Neural Network (ANN) modeling of monthly rainfall – runoff process of the Pole Saheb(Anyan) station in Zarrineh River watershed, which is a portion of orumiyeh lake drainage basin, that is located in Iran. To evaluate the modeling performance, the obtained results were compared with those of multi linear regression and Auto Regressive Integrated Moving Average models. Comparison of the obtained results via the trained ANN using denoised- jittered data showed that the proposed data pre-processing approach could improve performance of the ANN based rainfall-runoff modeling of the case study up to 38% in the verification phase.

Keywords: Rainfall-Runoff modeling; ANN; Wavelet denoising; Jittered data; Zarrineh river watershed.

1- Introduction

Nowadays water resources management is vitally important task and optimum planning of irrigation projects, development and exploitation of water resources especially during drought and flood events will be strictly dependent to the accuracy of the used rainfall-runoff modeling tool. Therefore different models have been already developed and employed for modeling rainfall-runoff process of the watersheds. Owing to the large number of vague physical parameters in the hydrological processes, black box (lumped) models are mostly applied, since they may have some benefits over fully distributed models (Nourani and Mano, 2007). For instance, successful hydro-environmental applications of auto regressive integrated moving average (ARIMA) and multi linear regression (MLR) models have been already reported by several researchers (e.g. see Wang et al., 2015; Salas et al., 1980; Zhang et al., 2011). Although these models are linear and may sometimes not be accurate due to their incapability to deal with non-stationary and non-linearity, they are still applied in practice because they can be easily used to compare and evaluate the effectiveness of novel methods. As such black box models, Artificial Neural Network (ANN) has recently indicated great ability for rainfall-runoff modeling (e.g., Nourani and Saeidifarzad, 2016; Nourani et al., 2014; Chau et al., 2015; Abrahart et al., 2012; ASCE, 2000). The efficiency of artificial intelligence (AI) techniques like ANN may be altered if noisy time series and data are used as inputs (Sang et al., 2009). Since the performance of any data-driven model is sensitive to the quality of the used data, different methods have been proposed for data denoising purpose, e.g. Wiener filter and Kalman filter (Wiener, 1949; Kalman, 1960), which are appropriate for linear systems but sometimes inappropriate for non-linear hydro-environmental processes. When classic methods for modeling hydrological time series do not meet the practical needs based on their limitations exposing to non-stationary characteristics and multi time scales, wavelet threshold denoising (WTD) method proposed by Donoho (1995) can be used as a reliable alternative. In hydrological practices, the WTD method is known more influential than conventional methods since it can contribute the illumination of the localized characteristics of non-stationary time series both in temporal and frequency domains (Jansen, 2006). There are a few studies on the application of wavelet denoising in hydrological modeling (e.g. see Nourani et al., 2014; Hassannejad and Nourani, 2012). On the other hand in training phase of an AI model, the training data set includes a limited sample of all data, so a set of selected data could not reflect all possible patterns of the process (Zhang, 2007). Jittered data for calibration of an AI model can enlarge the sample size of training data set by its supplementation using extra generated data which are similar to, but different from the original observed data. This can make it possible that the data are appeared more smoothly to an AI model and therefore enhance the model capability to learn the real patterns involved in the process (Zhang, 2007). Furthermore, it can prevent over fitting of model by supplying extra constraints, and imposing the jittered data into the training patterns can lead to improvements of the AI modeling. Therefore, the jittered data obtained by the noise injection method can be a useful pre-processing technique for AI-based model building (Zhang, 2007; Singh, 2000; Zur et al., 2004). The selection of a suitable noise size to be injected to the original time series to create jittering data has not been well described in technical literature. Obviously the appropriate variance of noise should be a problem reliant as distinct time series may have different inherent noise levels. Consideration of high levels of noise can deform the underlying pattern while small noises might not have sufficient influence on the jittering performance.

Most of the researches regarding the application of jittering data concentrate on classification problems and financial time series analysis and there is not any research in hydro-environmental modeling. Furthermore the impacts of denoising (smoothing) and noise injection (jittering) have not been simultaneously investigated neither in hydrology nor in any other engineering fields. Thus it is necessary to produce more researches on this filed and providing suitable solution to model hydro-environmental phenomena which is addressed in this article.

2- Material and Methods

2.1- The proposed hybrid model

In the proposed method in this study firstly by applying wavelet based denoising approach on raw data, the outliers and systematic noises of the series are identified and shrunk to produce smooth hydrological time series. The magnitude of the shrinkage is controlled according to a threshold value. Then to have several time series with similar pattern to the original smoothed time series, jittered time series are generated by adding normally distributed noise time series with specified standard deviations to the original smoothed time series of the hydrologic parameters. Finally, the produced jittered time series are imposed to the ANN forecasting model. In Figure 1, schematic diagram of the proposed method is shown.

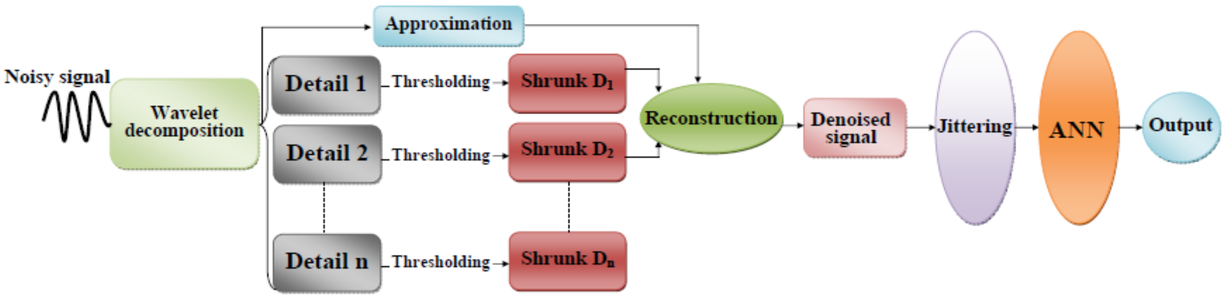

Figure 1) Schematic diagram of the proposed model.

2.1.1- Wavelet denoising procedure

Wavelet data denoising method based on the thresholding to obtain denoised signals has been introduced by Donoho (Donoho, 1995). In this method, first a signal is decomposed into different sub-signals at different resolutions through controlling scaling and shifting coefficients by the wavelet transform. By this way, reliable localization properties which are caught in both time and frequency domains can be provided. Second a thresholding rule is applied on the sub-signals. The basic factors that must be respected in this method include: selection of a mother wavelet, decomposition level, thresholding rule and accurate estimation of threshold rule. For a mother wavelet ψ (t), the wavelet basis function can be considered as follow (Nourani et al., 2014):

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}}\psi\left(\frac{t-b}{a}\right) \quad (a,b)\epsilon R; a \neq 0 \qquad (1)$$

In this equation a,b and R indicate respectively scale and shift factors and the real number domain and $\psi_{a,b}(t)$ is the successive wavelet. The wavelet transform of a signal f (t) ϵ $L^2(R)$ can be written as (Nourani *et al.*, 2014):

$$w_f(a,b) = \frac{1}{\sqrt{a}}\int_{-\infty}^{+\infty} f(t)\overline{\psi\left(\frac{t-b}{a}\right)}dt \qquad (2)$$

Which $\Psi(t)$ is complex conjugate of $\psi(t)$. As it is clear from equation (2), the wavelet transform of a time series like *f(t)* decomposes it under various resolution levels. By applying successive wavelet transform, the main signal of *f(t)* is reconstructed using inverse transform using the wavelet coefficients of w_f (a,b), as (Sang *et al.*, 2009):

$$f_{(t)} = \left[\int_{-\infty}^{+\infty}\frac{|\hat{\psi}(\omega)|^2}{|\omega|}d\omega\right]^{-1}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\frac{1}{a^2}w_f(a,b)\psi_{a,b}(t)\,da\,db \qquad (3)$$

Which $|\hat{\Psi}(\omega)|$ denotes to the Fourier transform of $|\psi(\omega)|$.

The wavelet based thresholding technique as a widely used data denoising approach is conducted through three steps as (Donoho, 1995):

a) First a proper mother wavelet and a reasonable resolution level of N are chosen for the specified period of the study process to decompose the main time series to an approximation sub-series at level N and N detailed sub-series via wavelet transform.

b) In the second step, the absolute values of the detailed sub-series in resolution level of i $d_i(t)$ (i = 1, 2,..., N) which are less than a specified threshold of T, will be changed to zero, but if the values of detailed sub-series at the same resolution level exceed this specified threshold, their difference with the threshold value are considered as the modified values of detailed sub-series. Which this thresholding procedure can be mathematically shown by (Donoho, 1995):

$$\hat{d}_i(t) = \begin{cases} sgn(d_i(t))(|d_i(t)| - T) & |d_i(t)| > T \\ 0 & |d_i(t)| \leq T \end{cases} \qquad (4)$$

Where i refers to i th resolution level. Equation 4 applies the thresholding at all resolution levels on detailed sub-series, but the approximation sub-series is not included in this thresholding procedure.

Donoho and Johnstone (1995) proposed a formula to determine a general optimal threshold value for signals which are included white Gaussian noises as (Donoho and Johnstone, 1995):

$$T = \hat{\sigma}\sqrt{2\log_e(n)} \qquad (5)$$

Where the number of samples in the noisy signal is n and $\hat{\sigma}$ is the standard deviation of noises which may be obtained as (Donoho and Johnstone, 1995):

$$\hat{\sigma} = \left[\frac{\text{median}(|d_i(t)|)}{0.6745}\right] \qquad (6)$$

Therefore, the $|d_i(t)|$ represent detailed wavelet coefficient of main time series of first level.

c) At the third step, the denoised (smoothed) sub-series can be reconstructed by modified detailed sub-series at all resolution levels and approximation sub-series at resolution level N through the inverse wavelet transform (Eq. 3).

2.1.2- Jittered data generation

Generating random data usually consists of two steps. First, random data with uniform distribution are generated thereafter, these random numbers with uniform distribution are used to produce random numbers with arbitrary distribution. After generating random numbers with uniform distribution, some of methods e.g. reverse conversion method could be used in order to generate random numbers with arbitrary distribution. In this approach whenever x random variable has cumulative distribution, in this case u=F(x) has uniform distribution of u(0,1) and vice versa if u~u(0,1), in this case, $x = F^{-1}(u)$ has an F cumulative distribution function and if x has F distribution, for generating y random variable with G distribution function we will have (Bowker and Lieberman, 1972):

$$y = \underbrace{G^{-1}(\overbrace{F(x)}^{\text{uniform distribution}})}_{\text{includes G cumulative distribution}} \qquad (7)$$

Random numbers based on different distributions could be generated by software. In this study NORMRND toolbox of MATLAB, was used to produce normally distributed random time series of jittered noises with mean of zero and several small standard deviations consistent with the original time series of the hydrological parameters.

2.1.3- Artificial Neural Networks (ANNs)

Artificial neural networks (ANNs) are widely used for modeling and prediction of hydro-environmental processes. In this regard, feed forward ANN trained by the back-propagation algorithm including one input, one hidden and one output layers are more suitable option in compared to other ANN types in most engineering disciplines (ASCE, 2000; Hornik, 1988). This network has great ability to learn involved patterns within non-linear systems through only three layers. Neurons (nodes) in each layer are connected to all nodes in previous layer. Due to the feed forward framework, the path of signals is in forward direction and the outputs of input layer, create the input vector for hidden layer and similarly the outputs of the hidden layer make inputs for the output layer. The output value of a feed forward neural network with three layers can be obtained through the following equation (Kim and Valdes, 2003):

$$\hat{y}_k = f_o\left[\sum_{j=1}^{M_N} G_{kj} \cdot f_h\left(\sum_{i=1}^{N_N} G_{ji}x_i + G_{jo}\right) + G_{ko}\right] \qquad (8)$$

Where Eq. 8 applies weight of G_{ji} on a node in hidden layer which connects *i* th node of the input layer to the *j* th node of the hidden layer and bias of G_{jo} on the *j* th hidden node. f_h is the activation function for all nodes of hidden layer, weight G_{kj} is applied on the output layer to the path where connects the *j* th node in hidden layer to the *k* th node of the output layer, G_{k0} is the bias of the *k* th output node, f_0 is the activation function for the output node, x_i denotes to the input value of *i* th node in input layer and $\hat{y}_k$, y show respectively calculated and observed values for target (output) node. Finally, N_N and M_N indicate respectively the number of input and hidden layers' nodes. The different bias and weights applied on the nodes of hidden and output layers are tuned through the calibration phase of modeling.

2- Study area

The data used in this paper are from Pole Saheb (Anyan) station in Jighatu River Watershed that is a sub basin of Zarrineh River watershed, which is a portion of orumiyeh lake drainage basin, that is located in Iran (Latitude 36°12', Longitude 46°26') (Figure2). The monthly mean and maximum runoff are 18.08 m^3/s and 213 m^3/s, respectively in the study duration. The monthly runoff data for 21 years (from 1992 to 2012, 252 months) were used in this research. The statistical parameters of the monthly average rainfall and runoff data such as the mean, standard deviation, maximum and minimum values (i.e., Xmean, Sd, Xmax and Xmin, respectively) are given in Table 1. Due to the training and verification goals, data set was divided into two parts. The first division as 70% of total data included the training set and the rest 30% data set was used for the verification purpose.

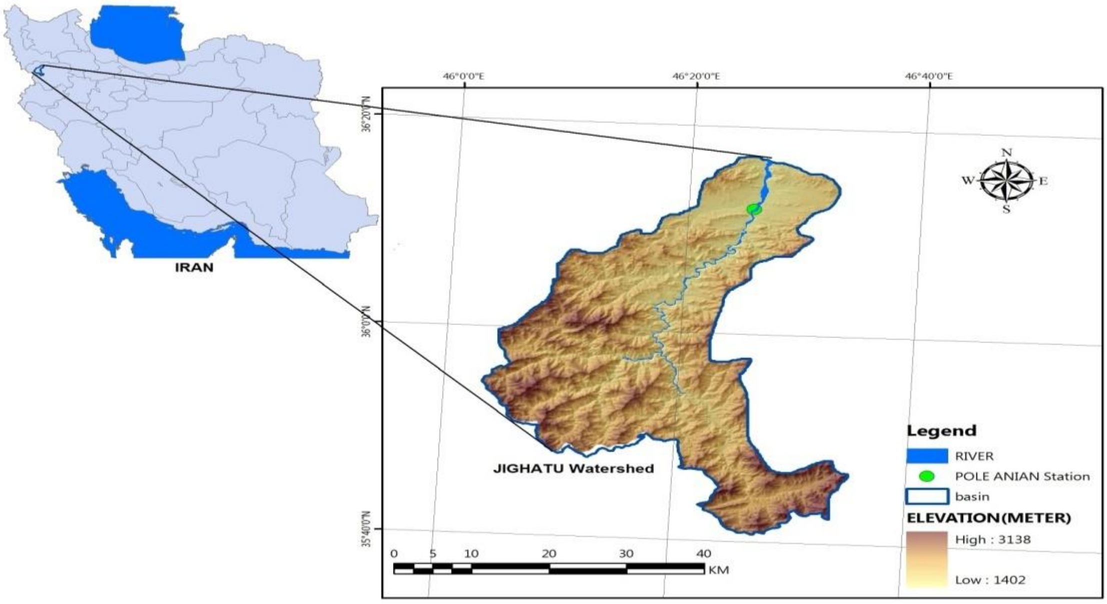

Figure 2) Jighatu River Watershed.

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

Statistical parameters	All Data		Training Data		Verifying Data	
	Runoff (m^3/S)	Rainfall (mm)	Runoff (m^3/S)	Rainfall (mm)	Runoff (m^3/S)	Rainfall (mm)
X_{mean}	18.08	1.012	19.55	1.094	17.44	0.822
X_{max}	213	7.033	213	7.033	150.68	3.6
X_{min}	0	0	0	0	0	0
Sd	31.49	1.168	37.95	1.25	27.93	0.926

2-3- Efficiency criteria

The model that yields the best results in terms of coefficient of determination (DC) as equation (9) and root mean square error (RMSE) as equation (10) in the training and verifying steps can be determined through trial and error process (Nourani et al., 2009).

$$DC = 1 - \frac{\sum_{i=1}^{n}\left(O_{obs_i} - O_{com_i}\right)^2}{\sum_{i=1}^{n}\left(O_{obs_i} - \bar{O}_{obs}\right)^2} \qquad (9)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(O_{obs_i} - O_{com_i}\right)^2}{n}} \qquad (10)$$

Where DC, RMSE, n, O_{obsi} , O_{comi} and $\bar{O}_{obs}$ are determination coefficient, root mean squared error, number of observations, observed data, computed values and mean of observed data, respectively. Clearly small value for RMSE and high value for DC (up to one) show high efficiency of the model. The generated noise time series may include negative quantities, therefore in order to prevent the producing of negative hydrological data , in this study the input and target data were normalized to scale the data between 0.1 and 0.9 by the equation (11) (Rajurkar *et al.*, 2002):

$$N_i = 0.8\left[\frac{(X_i - X_{min})}{X_{max} - X_{min}}\right] + 0.1 \qquad (11)$$

Regarding the equation (11), x_i is the desired variable value, x_{min} and x_{max} are the minimum and maximum values, respectively. N_i is the normalized variable.

3- Results And Discussion

At first, the multi-layer perceptron (MLP) feed forward ANN model without any data pre-processing were used to model the watershed monthly rainfall-runoff process. This kind of ANN model accompanied by back propagation training algorithm is widely used in hydrologic modeling (ASCE, 2000). Each MLP was trained with 2 to 10 hidden neurons in a single hidden layer and scheme of the Levenberg–Marquardt back propagation was used as the training algorithm. No great improvement in model performance was found when the number of hidden neurons was increased from a threshold, which is similar to the outcome reported by several researchers (Abrahart and See, 2000; Danandeh Mehr *et al.*, 2015). In this study, five combinations of input data for runoff prediction were consumed as:

Comb. 1: R_t, Q_{t-12}, Q_t

Comb. 2: R_t, Q_{t-12}, Q_{t-1}, Q_t

Comb. 3: R_t, R_{t-1}, Q_{t-12}, Q_{t-1},Q_t

Comb. 4: R_t, Q_{t-12},Q_{t-2}, Q_{t-1},Q_t

Comb. 5: R_{t-1}, R_t, Q_{t-12}, Q_{t-2}, Q_{t-1}, Q_t

In all cases the output was the discharge at the next time step Q_{t+1} where R_t presents rainfall value at time step t. In order to get appropriate 1-month-ahead prediction of Q, the input layer should be arranged in a way that could enjoy all pertinent information on the target data. Based on sensitivity analysis, the input layer was optimized with only the most important time memories. In this regard, in all combinations Q_{t-12} was considered as model input for predictions. The results of ANN model with noisy data are shown in Table 2.

Table- 2. Results and structures of ANN model for the different input combinations

Input variables (rainfall and runoff)	Network Structure	RMSE (normalized)		DC	
		Calibration	Verification	Calibration	Verification
Comb.1	(3-5-1)	0.072	0.108	0.522	0.421
Comb.2	(4-6-1)	0.058	0.095	0.685	0.556
Comb.3	(5-3-1)	0.066	0.103	0.593	0.475
Comb.4	(5-3-1)	0.061	0.099	0.652	0.519
Comb.5	(6-3-1)	0.066	0.108	0.601	0.427

Based on the efficiency criteria, it is clear that input Comb no. 2 could lead to better performance in ANN modeling and thereafter used for ANN modeling.

In the next step of modeling, in order to eliminate the outliers and systematic large noises of the observed data, wavelet-based denoising approach was applied on raw data. Since the type of used mother wavelet and decomposition level can alter denoising performance, wavelet denoising was performed and compared using Daubechies mother wavelets (Haar or Db1, Db2, Db3 and Db4) at three different resolution levels of 3,4 and 5 (Walker, 1999). The reason of choosing these three resolution levels is that one year includes 12 months performed between two modes of 23 and 24, therefore these three possibilities focus on annual period intensity. The denoising procedure of hydrological time series was performed using different mother wavelets and decomposition levels of 3, 4 and 5 and specified threshold obtained through equation 5, then the models were trained using such smoothed input combination set determined in sensitivity analysis step (Comb. 2). The results of ANN modeling using denoised input data have been summarized in Table 3. As it can be seen in Table 3, the obtained results indicate improvement of about 25% in ANN modeling in verification phase when using smooth time series as inputs. The results show that Db4 mother wavelet could lead to reliable results according to level of removed noise and to similarity of mother wavelet shape with main time series formation.

Table 3) Results and structures of ANN modeling using denoised data (using Comb.2).

Mother Wavelet	Decomposition Level	Threshold (Normalized)	Network Structure	RMSE (normalized)		DC	
				Calibration	Verification	Calibration	Verification
Haar	5	0.175	4-5-1	0.062	0.087	0.648	0.626
Db2	4	0.175	4-9-1	0.059	0.087	0.678	0.625
Db3	4	0.175	4-6-1	0.057	0.084	0.698	0.652
Db4	4	0.175	4-4-1	0.055	0.078	0.722	0.697

In the third step of modeling, several jittered input time series with similar pattern to the original time series were produced by adding normally distributed generated noises with zero mean and different standard deviations to the smoothed time series of the hydrologic parameters (obtained in second step of modeling). In this manner the time series would have unique and similar trend (approximation) to the original time series but with different stochastic terms represented by the added small generated noises. Therefore via the training phase of AI modeling, the AI model (ANN in this study) could see and learn different stochastic situations of process which in turn this could enhance the performance of modeling in the verification step (for the unseen data). For this purpose, normally distributed noise time series with mean of zero and standard deviations of 0.0001, 0.001, 0.003, 0.005 and 0.01 (normalized value) were generated and injected to the smoothed hydrological time series (obtained in second step of modeling) and the ANN modeling was performed by these jittered input time series. In this stage, according to the best input combination set (with appropriate lag) determined in the first step of modeling, the input combinations in ANN modeling were considered as:

Comb. 1: R_t , Q'_{1t}, , $Q_{t-12D,}$ Q_{t-1D},Q_{tD}

Comb. 2: R_t , $Q'_{2t,}$ Q'_{1t}, $Q_{t-12D,}$ Q_{t-1D},Q_{tD}

Comb. 3: R_t ,Q'_{3t} ,$Q'_{2t,}$ Q'_{1t}, $Q_{t-12D,}$ Q_{t-1D},Q_{tD}

Comb. 4: $R_{t,}$ Q'_{4t} ,Q'_{3t} ,$Q'_{2t,}$ Q'_{1t}, $Q_{t-12D,}$ Q_{t-1D},Q_{tD}

Where, Q_{tD} represents value of smooth time series at time step t, and Q'_t indicates the denoised-jittered time series.The indexes 1,2,3 and 4 indicate different generated noise (with same standard deviation) added to smoothed time series at time step t. For instance, the original (Q_t) and three samples of jittred time series generated by noises with standard deviation of 0.01($Q'_{1,}$ Q'_2 and Q'_3) are depicted in Figure 3. The obtained results of modeling are shown in Table 4. It should be noticed that for each of noise time series with a specified standard deviation, different time series(up to four) were generated and different combinations (Comb.,1,2,3,4) were produced but only the results of the input combination which lead to best results have been presented in the Table.

Table 4) Results of ANN model using the denoised-jittered data.

Standard deviation of noise	Input structure	Network Structure	RMSE (normalized)		DC	
			Calibration	Verification	Calibration	Verification
0.0001	Comb. 2	6-4-1	0.046	0.076	0.807	0.716
0.001	Comb. 1	5-7-1	0.041	0.069	0.844	0.766
0.003	Comb. 2	6-3-1	0.051	0.071	0.758	0.752
0.005	Comb. 4	8-3-1	0.048	0.074	0.781	0.725
0.01	Comb. 2	6-4-1	0.052	0.078	0.750	0.697

Based on the efficiency criteria, it is clear that input Comb. (1) could lead to better performance in modeling, including generated noise with standard deviation 0f 0.001**,** so that the proposed methodology, in comparison to the situation in which the modeling was done by un-preprocessed data, indicates an improvement of 38 percent in testing phase. The scatter plot of optimum ANN model in training and verification phase are shown in Figures (4) and (5).

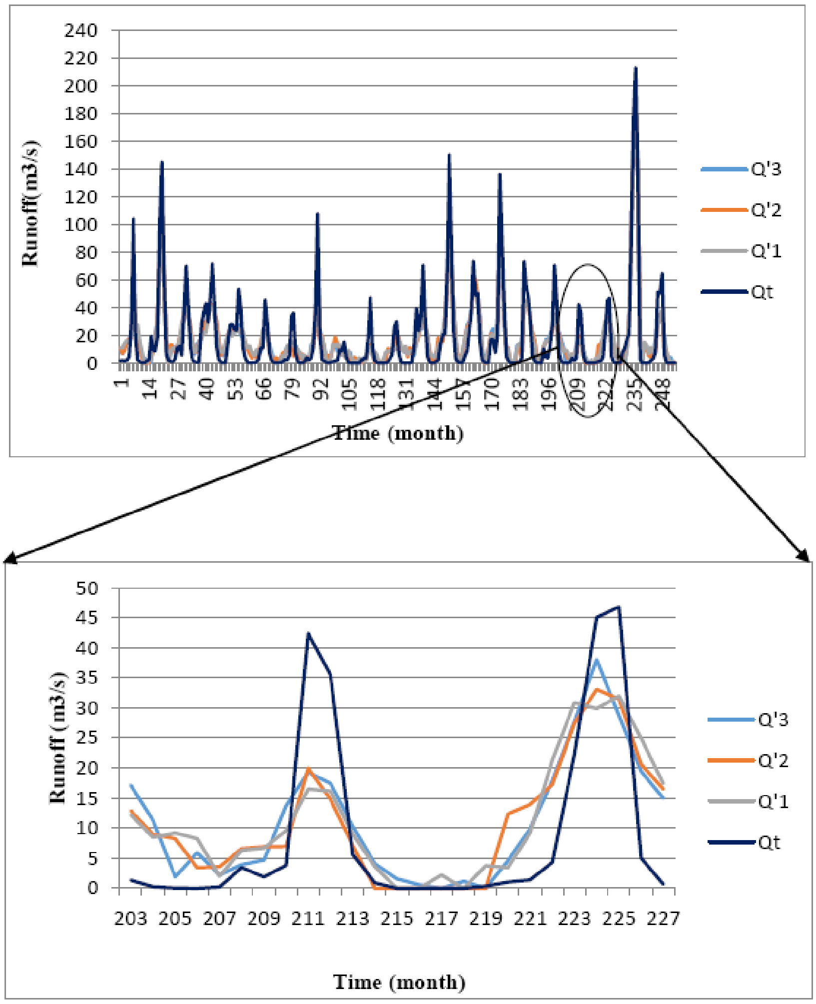

Figure 3) The original and three samples of generated jittred time series with noise standard deviation of 0.01.

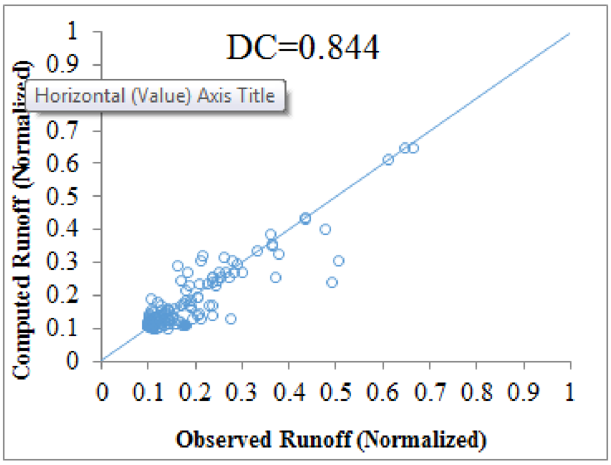

Figure 4) The scatter plot of ANN results in training phase.

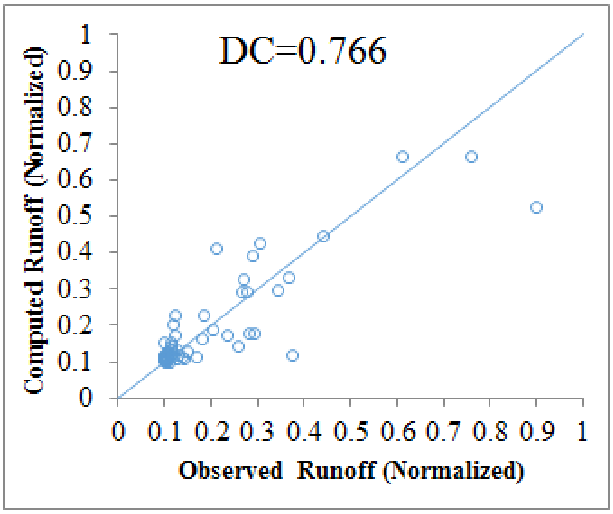

Figure 5) The scatter plot of ANN results in testing phase.

In order to evaluate the ability of proposed modeling, some comparisons with classic linear models of ARIMA (Salas et al., 1980) and MLR (Snedecor, 1981) were also conducted in modeling the watershed rainfall–runoff process. ARIMA and MLR modeling have been done by denoised-jittered time series as well. The comparison results are presented in Table 5.

The results indicate poor outcomes of ARIMA and MLR models with regard to the proposed model. This is due to the limited ability of linear models in modeling non-linearity and non-stationary time series and on the other hand, high dependence of data-driven models to quantity and quality of the used data.

Table 5) Comparison of different rainfall–runoff modeling approaches.

Model	RMSE (normalized)		DC	
	Calibration	Verification	Calibration	Verification
ARIMA	0.057	0.088	0.699	0.612
MLR	0.061	0.093	0.673	0.571
ANN	0.041	0.069	0.844	0.766

4- Conclusions

In this study via data pre-processing techniques, the input of wavelet-based denoised-jittered data was employed in AI-based rainfall-runoff modeling. Accordingly, first it was tried to smooth the hydrological time series by eliminating the outliers and large noises of raw observed time series, which may be due to human or tool measurement error or systematic error. Then different training time series were generated by noise injection to the smoothed time series, and uesd to train ANN model for monthly rainfall-runoff modeling. The comparison of obtained results using processed and unprocessed data, indicates the merit of applied data pre-processing approaches due to robust identification of hidden patterns in data, so that the developed models could simulate and predict runoff values with lower margin of error and higher confidence and the best results were achieved by employing the denoised-jittered data via producing more different training time series with the same pattern of original time series.

For future study, it is recommended to examine the efficiency of the proposed data-pre-processing method in rainfall-runoff modeling of other watersheds. Since it is expected that the merit of the method is more highlighted where the quality of the gathered data is due to the technical limitations, then it is worth to examine the performance of the proposed data pre-processing linked to other data driven methods.

Furthermore, it is suggested to evaluate the efficiency of the proposed method in modeling the process at other time scales and also for modeling other hydrological processes which may involve distinct noise level and pattern regarding to the type of process.

References

Abrahart, R. J., Anctil, F., Coulibaly, P., Dawson, C. W., Mount, N. J., See, L. M., Shamseldin, A. Y., Solomatine, D. P., Toth, E., Wilby, R. L. 2012. Two decades of anarchy? Emerging themes and outstanding challenges for neural network river forecasting. Progress in Physical Geography: 36, 480–513.

Abrahart, R. J., See, L. 2000. Comparing neural network (NN) and auto regressive moving average (ARMA) techniques for the provision of continuous river flow forecasts in two contrasting catchments. Hydrological Process: 14, 2157–2172.

ASCE task Committee on Application of Artificial Neural Networks in hydrology. 2000. Artificial Neural Networks in hydrology 1: hydrology applications. Journal of Hydrologic Engineering: 5, 124–137.

Bowker, A. H., Lieberman, G. J. 1972. Engineering statistics. Prentic-Hall, 608 p.

Chau, K., Wang, W., Chen, X. 2015. A novel hybrid neural network based on continuity equation and fuzzy pattern-recognition for downstream daily river discharge forecasting. Journal of Hydroinformatics: 17, 733–744.

Danandeh Mehr, A., Kahya, E., Şahin, A., Nazemosadat, M. J. 2015. Successive-station monthly streamflow prediction using different ANN algorithms. International Journal of Environmental Science and Technology: 12, 2191–2200.

Donoho, D. H. 1995. Denoising by soft-thresholding. IEEE Transactions on Information Theory: 41, 613–617.

Donoho, D. L., Johnstone, I. M. 1995. Ideal spatial adaptation via wavelet shrinkage. Biometrik: 81, 425–455.

Hassannejad, F., Nourani, V. 2012. Elevation of wavelet de-noising performance via an ANN-based streamflow forecasting model. International Journal of Computer Science and Management Research: 1, 764–770.

Hornik, K. 1988. Multilayer feed-forward networks are universal approximators. Neural Networks: 2, 359–366.

Jansen, M. 2006. Minimum risk thresholds for data with heavy noise. IEEE Signal Processing Letters: 13, 296–299.

Kalman, R. E. 1960. A new approach to linear filtering and prediction problems. Journal of Basic Engineering -T ASME: 82, 35-45.

Kim, T., Valdes, J. B. 2003. Nonlinear model for drought forecasting based on a conjunction of wavelet transforms and neural networks. Journal of Hydrologic Engineering: 8, 319–328.

Nourani, V., Saeidifarzad, B. 2016. Detection of land use/cover change effect on watershed’s response in generating runoff using computational intelligence approaches. Stochastic Environmental Research and Risk Assessment: published online,1–17.

Nourani, V., Hosseini-Baghanam, A., Yahyavi-Rahimi, A., HassanNejad, F. 2014. Evaluation of wavelet-based de-noising approach in hydrological models linked to artificial neural networks, in : Islam et al. (Eds.), Computational Intelligence Techniques in Earth and Environmental Sciences. Springer, pp. 209-241.

Nourani, V., Komasi, M., Mano, A. 2009. A multivariate ANN–wavelet approach for rainfall–runoff modeling. Water Resources Management: 23, 2877–2894.

Nourani, V., Mano, A. 2007. Semi-distributed flood runoff model at the subcontinental scale for southwestern Iran. Hydrological Processes: 21, 3173–3180.

Rajurkar, M. P., Kothyari, U. C., Chaube, U. C. 2002. Artificial neural networks for daily rainfall-runoff modeling. Hydrological Sciences Journal: 47, 865–877.

Salas, J. D., Delleur, J. W., Yevjevich, V., Lan, W. L. 1980. Applied Modeling of Hydrological Time Series, first ed., Water Resources Publications, Littleto, 484 p.

Sang, Y. F., Wang, D., Wu, J. C., Zhu, Q. P., Wang, L. 2009. Entropy-based wavelet de-noising method for time series analysis. Entropy: 11, 1123–1147.

Singh, S. 2000. Noise time-series prediction using pattern recognition techniques. Computational Intelligence: 16, 114–133.

Snedecor, G.W., Cochran, W.G. 1981. Statistical methods. seventh ed, Iowa: Iowa State University Press, 503 p.

Walker, J.S. 1999. A primer on wavelets and their scientific applications. New York, Chapman and Hall, CRC, 320 p.

Wang, W. C. h., Chau, K., Xu, D., Chen, X. Y. 2015. Improving Forecasting Accuracy of Annual Runoff Time Series Using ARIMA Based on EEMD Decomposition. Water Resources Management: 29, 2655–2675.

Wiener N. 1949. Extrapolation, Interpolation, and Smoothing of Stationary Time Series, New York, Wiley, 173 p.

Zhang, Q., Wang, B. D., He, B., Peng, Y., Ren, M.L. 2011. Singular spectrum analysis and ARIMA hybrid model for annual runoff forecasting. Water Resources Management: 25, 2683–2703.

Zhang, G. P. 2007. A neural network ensemble method with jittered training data for time series forecasting. Information Sciences: 177, 5329–5346.

Zur, R. M., Jiang, Y., Metz, C. E. 2004. Comparison of two methods of adding jitter to artificial neural network training. International Congress Series: 1268, 886–889.