ARCE ADVANCE RESEARCHES IN CIVIL ENGINEERING

Volume 1, No.1, pages:1-8

Application of Wavelet Denoising and Artificial Intelligence Models for Stream Flow Forecasting

Gholamreza Andalib[1]✉, Vahid Nourani[1, 2]

[1] *Department of Water Resources Engineering, Faculty of Civil Engineering, University of Tabriz,Tabriz, Iran.*
[2] *Department of Civil Engineering, Near East University, P.O.Box: 99138, Lefkosa, North Cyprus, Mersin 10 Turkey.*
✉Corresponding author's Email: Gholamreza.andalib@gmail.com

ORIGINAL ARTICLE
Received 20 Aug. 2018
Accepted 19 Jan. 2019

ABSTRACT:

In this study, the ability of threshold based wavelet denoising Least Square Support Vector Machine (LSSVM) and Artificial Neural Network (ANN) models were evaluated for forecasting daily Multi-Station (MS) streamflow of the Snoqualmie watershed. For this aim, at first step, outflow of the watershed was forecasted via ad hoc LSSVM and ANN models just by one station individually. Therefore, MS-LSSVM and MS-ANN were employed to use entire information of all sub-basins synchronously. Finally, the streamflow of sub-basins were denoised via wavelet based thresholding method, then the purified signals were imposed into the LSSVM and ANN models in a MS framework. The results showed the superiority of ANN to the LSSVM, MS model to the individual sub-basin model, using denoised data with regard to the noisy data, e.g., DCLSSVM=0.82, DCANN=0.85, DCMS-ANN=0.91, DCdenoised-MS-ANN=0.94.

Key words: Stream flow; denoising; artificial neural network; least square support vector machine; multi-station; Snoqualmie watershed.

1- Introduction

Astute stream flow forecasting ability will guide river managers and water authorities with better management decisions. In this way, there is a need for forecasts of stream flow events in order to: a concomitant reduction in water losses and deficits in irrigation orders, better targeting of environmental flows, basin wide consistency in management operations based on a thorough knowledge of variation in inflows, an improved capability for predicting and monitoring flood events. Due to the complexity of stream flow process in a river, the black box (lumped) modelling may have some avails over the modelling by the theoretical ruling (white box) equations, so Artificial Intelligence (AI) approaches as new generation of robust tools have been developed for stream flow time series forecasting [1,2]. Among such AI models, the Least Square Support Vector Machine (LSSVM) and Artificial Neural Network (ANN), former one as a persuasive forecasting tool and the latter as a novel neural network technique were employed for the simulation step of this paper [3, 4, and 5].

In spite of suitable flexibility of LSSVM and ANN in modelling hydrologic time series such as stream flow, sometimes, there is a shortage in appropriate forecasting results while data consist of noises. This happens because the efficiency of data-driven models is highly dependent on the available data in context of quantity and quality. Recent studies have shown that the noise limits the performance of many techniques used for identification and prediction of deterministic systems. Therefore, noise reduction is considered as a continuous mapping process of the noisy input data to a noise free output data. Focusing on hydrological processes that are nonlinear, the classic denoising filters may not behave effectively. But the threshold based wavelet denoising, which illuminates the localized characteristics of non-stationary time series both in temporal and frequency domains, is a potential filter in comparison to the other denoising methods [6]. With specific regard

To cite this paper: Gholamreza Andalib, Vahid Nourani (2019). Application of Wavelet Denoising and Artificial Intelligence Models for Stream Flow Forecasting. *Advance Researches in Civil Engineering* 1(1): 1-8. www.arce.ir

to denoising methods based on wavelets, the multi-scaling property of wavelets was explored for maximization the AI forecasting accuracy in the context of hydrological time series forecasting [7, 8, and 9].

Altogether, based on the importance of model's inputs in order to obtain the general pattern of stream flow dynamical process, participation of all part of watershed's effective data has the paramount importance. To this end, spatio-temporal investigation, identification and using all sub-basins records as a Multi-Station (MS) system can improve prediction of hydro-environmental process as stream flow. Recently, MS models have employed in some fields of hydrology [10, 11, 12].
In this paper, a novel methodology was proposed that considers purification of all sub-basins information via wavelet denoising for MS stream flow forecasting by robust LSSVM and ANN models. Whereas, in the first part, the watershed's flow was forecasted by sub-basin's discharge individually. Then, stream flow time series of sub-basins were denoised by threshold based wavelet denoising tool. Finally after removing redundant information, denoised time series of all sub-basins were imposed to the models via a MS form for outlet streamflow forecasting of the Snoqualmie watershed.

2- Methods and materials

2.1. Case study

The data used in this study are from hydrometric stations on the Snoqualmie watershed which is in the U.S. state of Washington from 2000 to 2014 and are available at the United States Geological Survey website (USGS, Fig. 1) (http:// http://waterdata.usgs.gov//). Due to the training and verification goals, data set was divided into two parts. The first part as 75% of total data used for the training and the rest 25% data were used for the verification purpose. As it can be seen in Table 1, X_{max} and standard deviation (S_d) values of calibration data set are higher than the verification data set which denote to the heterogeneity of calibration data set with regard to the verification data.

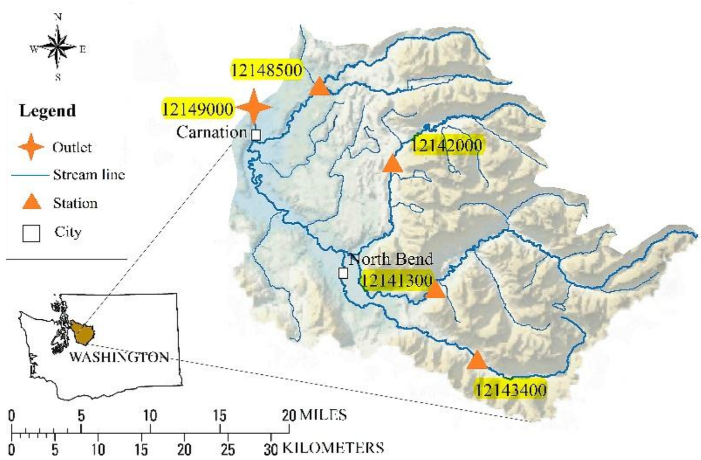

Fig. 1. The Snoqualmie watershed.

Table 1. Statistics of streamflow time series (m^3/sec)

Statistical Parameters	USGS Stations ID of sub-basins				
	12143400	12141300	12142000	12148500	12149000 Outlet
X_{mean}	8	35	15	16	102
X_{max}	165	728	374	286	2161
X_{min}	1	3	1	3	13
S_d	9	40	18	15	102

2.2. Proposed wavelet based denoised LSSVM and ANN models

The proposed wavelet based denoised LSSVM and ANN models consist a two-stage framework, wavelet-based denoising and one-time step ahead forecasting stage. The schematic diagram of developed proposed modeling is shown in Fig. 2. In the first stage using wavelet transform, the stream flow *Q(t)* time series are decomposed into sub-signals at different scales, i.e., a large-scale sub-signal and several small-scale sub-signals. Dyadic discrete wavelet transformation of a signal at level L yield L+1 sub-signals, one approximation (at level L) which denotes to the general trend of time series and L detailed sub-signals each representing a specific periodicity and seasonality of process, e.g., 2^1*-mode*, 2^2*-mode*, ... and 2^L*-mode*. Then sub-signals are shrunk via the fixed threshold as Eq. 1. Then shrunk sub-signals are reconstructed with approximation to create denoised signal. Finally the forecasting models are fed by pure utile information. The ANN with a logarithmic sigmoid activation function was used in this study. For the LSSVM modelling, parameters should be determined for catch input data set (For this purpose in this study, γ and σ were determined through a grid search trial-error process [13]. The grid search algorithm performs an exhaustive search through the parameter space of a learning algorithm to solve the problem of model selection i.e., finding the optimal parameters for a dataset).

2.3. Threshold based wavelet denoising

Threshold based wavelet denoising method was proposed for acquiring correct denoised results [14]. This method, which is now the most common method of wavelet denoising, is performed as: An appropriate mother wavelet and number of resolution level are chosen. The original one-dimensional time series is decomposed into an approximation at resolution level *M* and detailed signals at various resolution levels up to level *M*. The absolute values of detailed signals that exceed certain threshold are treated as the difference between the values of detailed sub-signals and threshold by Eq. 1 otherwise, are set to zero, which 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.

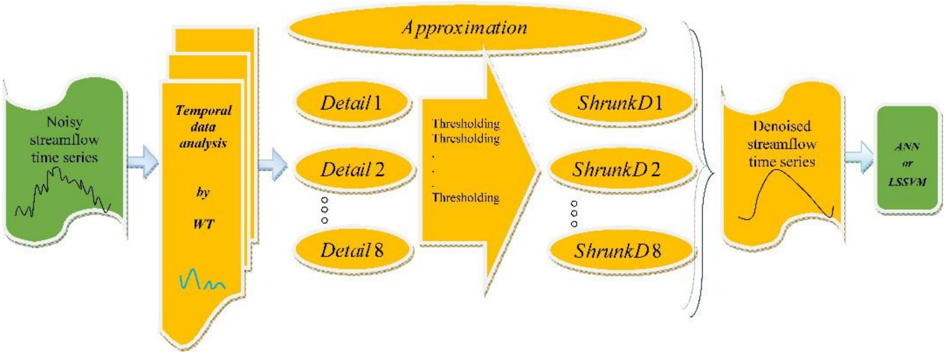

Fig. 2. Schematic of proposed wavelet based denoised LSSVM and ANN models.

$$dj(t)=\begin{cases} sgn(dj(t))(|dj(t)|-T') & |dj(t)|>T' \\ 0 & |dj(t)|\leq T' \end{cases} \tag{1}$$

In Eq. 1, T' and $dj(t)$ $(j= 1, 2,..., M)$ denote certain threshold and the absolute values of detailed signals for the j th resolution level, respectively. A general optimal universal threshold for the white Gaussian noise under a mean square error criterion and its side condition was derived by [14] that with high probability, the enhanced signal is at least as smooth as the clean signal. In this method, threshold is selected as:

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

Where n is number of samples in the noisy signal and $\hat{\sigma}$ is the standard deviation of noise that is estimated:

$$\hat{\sigma}=\left[\frac{median(|d_j(t)|)}{0.6745}\right] \tag{3}$$

in which $dj(t)$ is the first level detail coefficients of wavelet transform of the signal. In the current study, the soft threshold wavelet based denoising was done by global method (i.e., all the detailed signals shrunk just with the same threshold value).

2.4. Artificial neural network

The FFNN is widely applied in hydro-environmental studies as a prediction tool. It has already been demonstrated that an FFNN model trained by the back-propagation (BP) algorithm with three layers is satisfactory for forecasting and simulating hydrological problems. Three-layered FFNNs, provide a general framework for representing nonlinear functional mapping between a set of input and output variables. The explicit expression for an output value of a three layered FFNN is given by [13]:

$$\hat{y}_k = f_0\left[\sum_{j=1}^{M_N} w_{kj}\cdot f_h\left(\sum_{i=1}^{N_N} w_{ji}x_i + w_{j0}\right)+w_{k0}\right] \tag{4}$$

Where i, j and k denote the input layer, hidden layer and output layer neurons, respectively. w_{ji} 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_{jo} 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_{ko} is the bias for the k *th* output neuron, f_o is the activation function for the output neuron, x_i is i *th* input variable for input layer and y_k and 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 network training process.

2.5. Least square support vector machine

LSSVM was first proposed by [15] (Eq. 5) which several kernels could be used in its structure but the Radial Basis Function (RBF) kernel of LSSVM is commonly used in regression problems. The RBF kernel function is used in current study as Eq. 6 that γ and σ should determine as LSSVM's parameters by [15]:

$$f(x) = \sum_{i=1}^{m} \gamma e_i K(X, X_i) + b \tag{5}$$

$$K(X, X_i) = exp(-\frac{\|X - X_i\|^2}{2\sigma^2} \tag{6}$$

Where b is the bias term and e_i is the slack variable for X_i.

2.6. Evaluation of model precision

The Determination Coefficient (DC) and Root Mean Square Error *(*RMSE*)* as two different criteria were employed to evaluate the performance of the MS nitrate loads prediction. The DC and RMSE can be utilized to indicate discrepancies between predictions and recorded values [16]. Showed that hydro-environmental models could be adequately assessed by Eqs. 7 and 8.

$$DC = 1 - \frac{\sum_{i=1}^{T} (O_{obs_i} - O_{com_i})^2}{\sum_{i=1}^{T} (O_{obs_i} - \overline{O}_{obs})^2} \tag{7}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{T} (O_{obs_i} - O_{com_i})^2}{N}} \tag{8}$$

Where DC, RMSE, N, O_{obs_i}, O_{com_i} and $\overline{O_{obs}}$ are determination coefficient, root mean squared error, number of observations, observed data, computed values and mean of observed data, respectively.

3- Results and discussion

In this part, discharge values of the stations (12141300, 1214200, 12143400, 12148500) associated to four sub-basins, were imposed to LSSVM and ANN individually, in order to forecast one-day-ahead stream flow for outflow of the watershed (12149000). LSSVM model via RBF kernel was employed for this stage. As well as LSSVM, ANN model without any data pre-processing was also used to model the river's stream flow process. The architecture of LSSVM and ANN models were arranged according to the antecedents of stream flow processes; whereas, stream flow time series usually behave as Markovian process, so that the value of parameter at the current time step could be related to the conditions at the previous time steps. Therefore, 3 combinations of stream flow values were used as inputs of ANN and LSSVM models to forecast daily out flow of the watershed.

Input combinations for streamflow forecasting were consumed as:
Comb. 1: Q_{t-1}
Comb. 2: Q_{t-1}, Q_{t-2}
Comb. 3: Q_{t-1}, Q_{t-2}, Q_{t-3}

in all cases, *t* demonstrates the current time step. The output layer was comprised of only one variable, i.e., Q at current time step Q_t at outlet of the watershed. In order to get appropriate forecasting of stream flow, 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. For each input combination, the RBF-kernel's parameters in LSSVMs were adapted to achieve highest performance. For this purpose, various pairs of (γ, σ) values were tried with a grid search procedure and the one with the best accuracy was chosen. Then, the trained LSSVM was just applied to verify the model in stream flow forecasting. The arrangement of ANN model was done as a three layered model including input, hidden, and output layers. After determination of proper ANN architecture in terms of performance criteria, training was terminated and the weights were saved in order to be used in the verification step. Stream flow forecasting results of the basin via individual sub-basin's discharge results are presented in Table 2. It is inferred from the results that nearest sub-basin has a short memory and the farthest has the longest one. It should be noticed that ANN in all of the input combinations of all sub-basins behave more accurate than LSSVM model. Since the most important objective of such modelling is to have appropriate outlet forecasting, it is of prime importance to have appropriate forecasting results for outflow station. On the other hand, the fact is that the discharge at outlet station is affected by stream flow from entire watershed in a cumulative form. In order to obtain accurate model for outlet discharge station, it is necessary to consider effects between the sub-basins in a unique model, thus, the MS-LSSVM and MS-ANN were proposed.

As it was mentioned using all the information of sub-basins in the unique model could help to the exact forecasting of the basin's flow. Based on this, the LSSVM and ANN models could enjoy all the information just synchronously and weighted themselves to the sub-basins discharge according to their systems. Table 3 shows the results of MS-LSSVM and MS-ANN models which enhanced the accuracy of streamflow forecasting in comparison to the individual sub-basin's input, where best input combinations of each sub-basin were imposed to the models.

Table 2. Results of LSSVMs and ANNs for different input variables in daily stream flow forecasting

Stations ID	Structure	Model	RMSE (Normalized)		DC	
			Calibration	Verification	Calibration	Verification
12143400	(3,2)[a]	LSSVM	0.020	0.024	0.83	0.74
(Comb. 3)	(3,8,1)[b]	ANN	0.022	0.022	0.81	0.79
12141300	(5,7)	LSSVM	0.018	0.021	0.86	0.79
(Comb. 2)	(2,5,1)	ANN	0.019	0.019	0.85	0.83
12142000	(6,2)	LSSVM	0.019	0.019	0.84	0.82
(Comb. 2)	(2,4,1)	ANN	0.017	0.018	0.87	0.85
12148500	(4,9)	LSSVM	0.020	0.023	0.83	0.76
(Comb. 1)	(1,10,1)	ANN	0.019	0.020	0.84	0.81

[a] (γ, σ)

[b] The first, second and third numbers represent input variable, hide neurons and output variable, respectively.

Without shadow of a doubt, efficiency of LSSVM and ANN models as a data driven method depends on a quality and quantity of data. Indeed, all hydrological time series like stream flow consists of noise. Therefore, the soft threshold wavelet based denoising approach was employed as a applicable and powerful noise reduction method, which could perform more efficiently if reasonable Mother Wavelet (MW), proper time scale levels and eventually appropriate threshold value could be picked out. So, it was tried to investigate the effects of the used MW on the efficiency of MS-LSSVM and MS-ANN models. In this way, Daubechies family of wavelets (*db3*, *db4*, *db5*) were examined as the MWs in current research. Also, the Decomposition Level (DL) 8 was chosen as a yearly mode. Moreover, appropriate threshold was determined via Eq. 2 which guide to the favourable threshold without consuming time. After pre-processing step, the denoised stream flow time series of sub-basins were imposed to the LSSVM and ANN models in a MS form. The results of wavelet based denoised thresholding MS-LSSVM and MS-ANN are shown in Table 3, where MW *db5* showed acceptable behaviour than other MWs that it might be because of similarity between *db5* and observed stream flow time series of the Snoqualmie sub-basins. Also, evaluation criteria (verification DC) was elevated after utilizing denoised method in both MS-LSSVM and MS-ANN models. Computed stream flow time series of verification step (by denoised MS-ANN) at outlet of the Snoqualmie watershed is illustrated in Fig. 3.

Table 3. Results of noisy and denoised MS-LSSVMs and MS-ANNs for daily stream flow forecasting

Pre-processing	Structure	Model	RMSE (Normalized)		DC	
			Calibration	Verification	Calibration	Verification
ad hoc	(7,9)	LSSVM	0.015	0.016	0.90	0.88
	(8,10,1)	ANN	0.014	0.014	0.91	0.91
denoised	(9,2)	LSSVM	0.012	0.015	0.91	0.89
db3, DL8	(8,8,1)	ANN	0.011	0.015	0.94	0.90
denoised	(2,2)	LSSVM	0.010	0.013	0.95	0.92
db4, DL8	(8,4,1)	ANN	0.011	0.012	0.95	0.94
denoised	(6,6)	LSSVM	0.013	0.013	0.93	0.92
db5, DL8	(8,4,1)	ANN	0.012	0.012	0.94	0.93

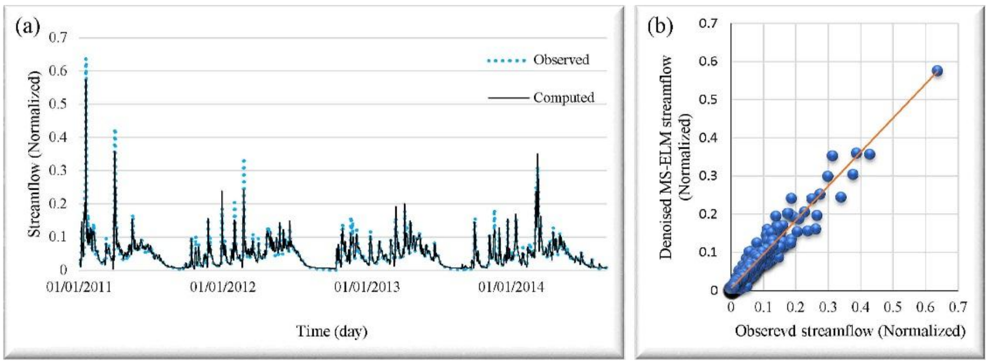

Fig. 3. (a) Computed (by denoised MS-ANN) versus observed stream flow time series of the Snoqualmie watershed; (b) Scatter plot.

4- Conclusion

Stream flow forecasting has a critical role in all aspects of water resources planning and management. According to the stream flow importance as a complex process, employment of intelligent black box models could lead to accurate forecasting of stream flow. The ANN model was used to forecast 1-step-ahead the Snoqualmie basin's streamflow in daily time scale via individual sub-basins. Also, ANN's ability was assayed in comparison to the LSSVM, and finally the efficacy of wavelet denoising was investigated as proposed hybrid model in the MS form.

At first step, basin's outflow was forecasted via sub-basin's discharge individually with LSSVM and ANN models by three input combinations. It was seen that near sub-basin has a short time delay and the farthest one has a long time delay. Also, ANN led to better outcomes than LSSVM because of LSSVM's. In the second step, the MS-LSSVM and MS-ANN were employed due to using entire information of the stations in a same time and in a unique model. For this purpose, the LSSVM and ANN models were fed synchronously the all sub-basins discharges. The results showed MS-based models superiority to the individual discharge inputs. Finally, because of noise effect in hydrological time series, the sub-basins discharge were purified via threshold based wavelet denoising method. Then, denoised discharge data of sub-basins were imposed to the LSSVM and ANN models in a proposed MS proposed form. The results revealed that data denoising could enhance accuracy of MS-LSSVM and MS-ANN in a favorable manner and guide hydrological engineers to access reliable streamflow forecasting. It is suggested to apply the presented method in other hydrological processes and also assess the ability of denoised MS models in water quality management.

5- References

1. Tongal, H., Booij, M. J., (2018). Simulation and forecasting of streamflows using machine learning models coupled with base flow separation. Journal of Hydrology 564, 266–282.

2. Danandeh Mehr, A., (2018). An improved gene expression programming model for streamflow forecasting in intermittent streams. Journal of Hydrology 563, 669–678.

3. Kalteh, A., (2016). Improving Forecasting Accuracy of Streamflow Time Series Using Least Squares Support Vector Machine Coupled with Data-Preprocessing Techniques. Water Resources Management 2, 747–766.

4. Prasad, R., Deo, R. C., Li, Y., Maraseni, T., (2017). Input selection and performance optimization of ANN-based streamflow forecasts in the drought-prone Murray Darling Basin region using IIS and MODWT algorithm. Atmospheric Research 197, 42–63.

5. Adnan, R. M., Yuan, X., Kisi, O., Adnan, M., Mehmood, A., (2018). Stream Flow Forecasting of Poorly Gauged Mountainous Watershed by Least Square Support Vector Machine, Fuzzy Genetic Algorithm and M5 Model Tree Using Climatic Data from Nearby Station. Water Resources Management 14, 4469–4486.

6. Jansen, M., (2006). Minimum Risk Thresholds for Data with Heavy Noise. IEEE Signal Processing Letters 13, 296–299.

7. Guo, J., Zhou, J., Qin, H., Zou, Q., Li, Q., 2011. Monthly Streamflow Forecasting Based on Improved Support Vector Machine Model. Expert Systems with Applications 38, 13073–13081.

8. Nejad, F. H., Nourani, V., (2012). Elevation of Wavelet Denoising Performance via an ANN-Based Streamflow Forecasting Model. International Journal of Computer Science and Management Research 1, 764–770.

9. Nourani, V., Mousavi, S., (2016). Spatiotemporal Groundwater Level Modeling using Hybrid Artificial Intelligence-Meshless Method. Journal of Hydrology 536, 10–25.

10. Nourani, V., Komasi, M. (2013). A Geomorphology-Based ANFIS Model for Multi-Station Modeling of Rainfall–Streamflow Process. Journal of Hydrology 490, 41–55.

11. Lee, W. K., Resdi, T. A. T., (2016). Simultaneous Hydrological Prediction at Multiple Gauging Stations using the NARX Network for Kemaman Catchment, Terengganu, Malaysia. Hydrological Science Journal 61, 2930–2945.

12. Nourani, V., Andalib, G., Sadikoglu, F., Sharghi, E., (2017). Cascade-based multi-scale AI approach for modelling rainfall-runoff process. Hydrology Research 49, 1191-1207.

13. Nourani, V., Andalib, G., (2015). Wavelet Based Artificial Intelligence Approaches for Prediction of Hydrological Time Series. In: Chalup S.K., Blair A. D., Randall M. (eds) Artificial Life and Computational Intelligence. ACALCI 2015. Lecture Notes in Computer Science, vol 8955. Springer, Cham

14. Donoho, D. H., (1995). Denoising by Soft-Thresholding. IEEE Transactions on Information Theory 41, 613–617.

15. Suykens, J. A. K., Vandewalle, J., (1999). Least Square Support Vector Machine Classifiers. Neural Processing Letters 9, 293–300.

16. Legates, D. R., McCabe, Jr. G. J., (1999). Evaluating the Use of Goodness-of-Fit Measures in Hydrologic and Hydroclimatic Model Validation. Water Resources Research 35, 233–241.