International Journal of Engineering, Mathematical and Physical Sciences
ISSN: 2517-9934
Vol:8, No:8, 2014

Fractal - Wavelet Based Techniques for Improving the Artificial Neural Network Models

Reza Bazargan Lari, Mohammad H. Fattahi

***Abstract*—**Natural resources management including water resources requires reliable estimations of time variant environmental parameters. Small improvements in the estimation of environmental parameters would result in grate effects on managing decisions. Noise reduction using wavelet techniques is an effective approach for preprocessing of practical data sets. Predictability enhancement of the river flow time series are assessed using fractal approaches before and after applying wavelet based preprocessing. Time series correlation and persistency, the minimum sufficient length for training the predicting model and the maximum valid length of predictions were also investigated through a fractal assessment.

***Keywords*—**Wavelet, de-noising, predictability, time series fractal analysis, valid length, ANN.

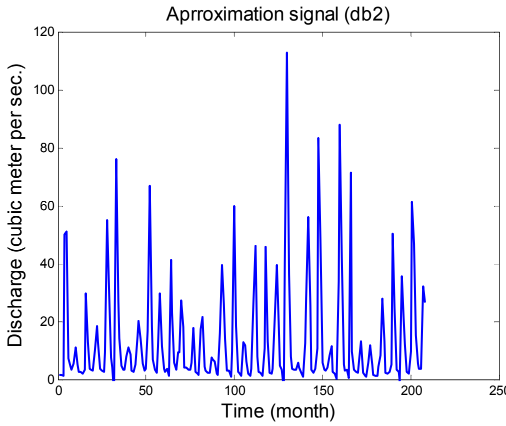
I. Introduction

TIME series trend, persistency, correlation, long-term memory and fractal properties are some important keys to judge on the future variations of the natural temporal events as well as their past history. Practical time series always contain some noise due to random influences (dynamical noise) and inaccuracies (additive noise) [1]. The performance of many techniques of modeling, prediction and control of hydrological systems is significantly affected by noise. Noise removing or noise reduction is a non-negligible stage to reach the basic trend and geometry of hydrological time series. Moving average, low-pass filters, nonlinear smoothing and wavelet-based de-noising algorithms are the common noise reduction techniques. The efficiency and suitability of the wavelet based de-noising techniques in order to enhance the predictability of river flow time series is investigated through fractal analysis approaches in this study.

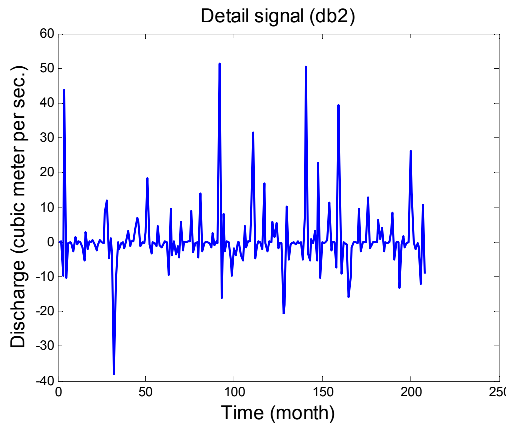
Wavelets have been shown to be an indispensable tool for scale variant representation and analysis of temporal data [2]. Fourier Transform (FT) and Fast Fourier Transform (FFT) have commonly been used for time series analysis. Fourier transform is not fully efficient in analyzing the frequency contents of fBm signals [3]. Another insufficiency of the Fourier transform is its incapability to determine the occurring time of each frequency beside its amplitude 2.

WT decompose the signals into coefficients of coarser scales which define as approximation signals and coefficients of finer scales that are called the detail signals (Fig 1). Continuous Wavelet Transform (CWT) and Discrete Wavelet Transform (DWT) are two wavelet approaches that have been used in this study.

R. Bazargan Lari is with the Department of Mechanical Engineering, Marvdasht Branch, Islamic Azad University, Marvdasht, Iran (Corresponding Author e-mail: rbazarganlari@gmail.com).
M.H.Fattahi is with the Department of Civil Engineering, Marvdasht Branch, Islamic Azad University, Marvdasht, Iran.



(a)



(b)

Fig. 1 Signal decomposing into the approximation and the detail signals using discrete WT (a) Band-e- Bahman approximation signal (b) Band-e-Bahman detail signal

As the preprocessing stage the noise components of the original signals are reduced using WT. A feed forward Multi-Layer Perceptron (MLP) artificial neural network (ANN) with Levenberg-Marquardt (L-M) training algorithm has been used as the predicting model. Based on types of the inputs to the ANN three scenarios are developed to assess the model's performance under the influence of the data processing.

II. Materials and Methods

A. Wavelet Analysis

Studying Fourier transform shows its strong limitation to deal with the real data. Sine waves which are the basis of Fourier analysis do not have limited duration. They extend from $-\infty$ to $+\infty$ with a smooth and predictable pattern. However most natural signals have sharp changes and tend to be irregular and asymmetric. Wavelet concept provides a method to overcome these limitations. Fourier analysis consist of breaking up a signal in to sine waves of various frequencies while wavelet analysis is the breaking up of a signal into shifted and scaled version of the original wavelet which is called the mother wavelet [4]. The mother wavelet should have properties that lead to a meaningful interpretation of the decomposition [2]. It should be able to split the signal recursively in to the actual trend called approximation and the residuals or fluctuations which are the detail signals.

The CWT of a function f is an integral transformation of the form

$$(w_\psi f)(a,b) = \int_R f(x)\psi_{a,b}(x)dx, \tag{1}$$

In which a is scale and b is the translation parameter that depict the location. $\psi_{a,b}(x)$ defines as

$$\psi_{a,b}(x) = \frac{1}{\sqrt{|a|}}\psi\left(\frac{x-b}{a}\right) \tag{2}$$

Applying the CWT to a function f for different pair of parameters a and b make different information on f. Therefore, after the wavelet decomposition the signal frequency can be obtained at different scales.

The CWT is useful for theoretical purposes but for practically analyzing signals it is not the suitable choice. It needs a great amount of computation time and resources. The DWT which corresponds to the transform "(1)" for discrete values of a and b, reduce the computation time significantly and is simpler to implement.

$$f(t) = \sum_{j=-\infty}^{\infty}\sum_{k=-\infty}^{\infty} d_k^j \psi_{j,h}(t) \tag{3}$$

In which the coefficients are determined as

$$d_k^j = \int_R f(t)\psi_{j,k}(t)dt \tag{4}$$

This could be achieved if we represent the wavelet as

$$\psi_{j,k}(t) = \frac{1}{\sqrt{|a_0^j|}}\psi\left(\frac{t-kb_0a_0^j}{a_0^j}\right) \tag{5}$$

where $a_0 > 1$ is a fixed dilation step and the translation parameter b_0 depends on the dilation step. The signals are conducted through high pass and low pass filters of different frequencies. The signal is then decomposed to one containing the high frequencies which indicate the signal's noise and the rest is the actual trend. This filtering procedure continues and in the next level the approximation signal is decomposed into noise and trend signals through the same procedure. Depending on the level of decomposing, the procedure repeats till a defined frequency is removed from a specific part of the signal.

B. Fractal Analysis

Fractal processes have similar conducts when studied at different scales. Fractal analysis is a mathematical set of methods to reveal the fractal properties of such processes. By using the context of self-affinity, a fractal property, Mandelbrot & Van Ness [5] extended the concept of fractal analysis to time series. Since then, fractal analysis has become a valuable tool in studying the time series behavior of natural processes. To evaluate the fractal scaling of a process, two indices are commonly employed. Fractal dimension denoted by D_f and Hurst exponent H are the mentioned criteria. A variety of methods have been proposed according to the time series class and length to measure D_f and H [6], [7]. Two main classes of signals are the fractional Gaussian noise (fGn) signals which are the stationary signals and the fractional Brownian motion (fBm) signals which are the non-stationary signals. In the present study, the Detrended fluctuation analysis (DFA) method is used to analyze the fractal properties of both the fGn and fBm time series. Time series correlation, predictability index and persistency along with the minimum length of the time series needed for prediction and the maximum valid predicted length of the time series are then calculated according to the fractal dimension based relations.

C. Detrended Fluctuation Analysis (DFA)

Peng et al. [8] suggested the DFA method for the first time. Based on root mean square analysis, non-stationary trends from long range correlated time series are removed. The signal is integrated and the mean is subtracted:

$$X(k) = \sum_{i=1}^{k}\left[x(i) - \bar{x}\right] \tag{6}$$

This integrated series is divided into non overlapping intervals of length n. In each interval, a least square line is fit on the data (representing the trend in the interval). The series $X(t)$ is then locally detrended by subtracting given interval length n, fluctuation is determined as a variance upon the local trend and the $X_n(t)$ theoretical values. For a characteristic size of fluctuation for this integrated and detrended series is calculated by:

$$F = \sqrt{\frac{1}{N}\sum_{k=1}^{N}\left[X(k) - X_n(k)\right]^2} \tag{7}$$

This computation is repeated over all possible interval lengths. Typically, F increases with interval length n. A power law relationship is expected as $F \propto n^{\alpha}$, in which α is expressed as the slope of a double logarithmic plot of F as a function of n. Detrended fluctuation analysis is also a classifying tool for the time series and can be used to distinguish between fBm and fGn series. fGn corresponds to α exponent ranging from 0 to 1, and fBm to α exponents from 1 to 2. α can be converted in to H according to the following equations:

$$H = \alpha \quad \text{for} \quad \text{fGn} \tag{8}$$

$$H = \alpha - 1 \quad \text{for} \quad \text{fBm} \tag{9}$$

The fractal dimension (D_f) and the Hurst coefficient (H) can be transformed to each other using the relation bellow:

$$D_f = 2 - H \tag{10}$$

D.Fractal Analysis of Correlation

To identify a random process $B_H(t)$ as fractional Brownian motion (fBm) [5], [9], [10], the increment $B_H(t) - B_H(t_0)$ should have a Gaussian distribution. The graph $B_H(t)$ is a geometrical object with a fractal dimension D_f [11] which range as $1 < D_f < 2$.

The past $\left[B_H(0) - B_H(-t)\right]$ and the future $\left[B_H(t) - B_H(0)\right]$ increments are correlated as [12]

To simplify the calculations and by applying the identity

$$\left\langle \left[B_H(t) - B_H(-t)\right]^2 \right\rangle \equiv 2\left\langle \left[B_H(t)\right]^2 \right\rangle - 2\left\langle B_H(-t)B_H(t) \right\rangle \tag{11}$$

and assuming $t_0 \equiv -t$ in calculating the variance we get

$$C(t) = \frac{|2t|^{2H}}{2|t|^{2H}} - 1 = 2^{2H-1} - 1 \tag{12}$$

The correlation of past and future increments in $B_H(t)$ lasts independently of temporal severance between records in time [13]. Clearly greater values of $C(t)$ indicate more significant correlation in the time series. According to H values one can determine the persistency or anti-persistency of a process. $0.5 < H < 1$ indicates the persistent processes and $0 < H < 0.5$ characterizes anti-persistent time series.

E. Fractal Based Predictability Indices

Autocorrelation is a quality of time series to predict their future values based on their past values. Any time series that posses this feature is known as long-memory processes. Most hydrologic time series including river flow time series have a long-range dependency in their time sequences. Fractal analysis provides an approach for calculating the dependency of the future values in a time series based on its past history. The grater values of dependency rate yields more predictable time series. The fractal based predictability indices of time series can be derived from the relation presented by Rangrajan and Sant [14].

$$PI_x = 2\left|D_x - 1.5\right| \tag{13}$$

In which D_x indicate the fractal dimension of time series and PI_x depict the predictability indices of the series. Some researchers have developed this concept to some of hydrological processes like pressure, temperature and precipitation time series [15]. We have applied the concept of predictability index to stream flow time series as bellow:

$$PI_f = 2\left|D_f - 1.5\right| \tag{14}$$

where the predictability index of the river flow time series is PI_f and D_f indicates the series fractal dimension.

The Minimum Sufficient Data Length for Prediction and the Maximum Valid Length of Predictions

Through a fractal analysis the minimum length of time series which is sufficient to be used as input to the predicting model and the maximum valid length of time series predicted were derived. Concerning the fact that $H = 0.5$ indicate a normal independent time series one can conclude that $H > 0.5$ depict increases in the long memory of the time series. The closer the H value to 1 the more predictable the series are. To determine the minimum length needed for prediction the variations of Hurst number in time series have been studied. Hurst number is calculated for decreasing lengths of time series. According to the cross validation approach the dedicated length for training was 240 months. Based on the chosen technique a decreasing step of 12 months is utilized from the 240th month towards the 1st month. The H values are calculated for each series length and the variations are recorded. Calculations are stopped in the first series with a significant difference in the H values trend or the first interval with $H \le 0.5$ and the series length is assumed to be the minimum length of time series which is sufficient for the training phase. Moreover, a windowing technique is employed to determine the maximum valid length of predictions. A window with the initial length of τ is chosen. The second window length τ_2 would be one month wider. The fractal dimension is calculated for each length of the window. The selected window is gotten wider in each trial and a different fractal dimension is estimated. The window widening and the fractal dimension calculation would be continued until the two sequential window lengths, τ_n and τ_{n+1}, show a significant difference in their fractal dimensions. Constant fractal dimension in a length of a time series indicate a defined trend in that length which leads a more reliable prediction. As the data set within the length of τ and τ_n has constant fractal dimension and consequently constant trend and geometry, the window length τ_n is assumed to be the maximum valid length of the predictions.

F. Artificial Neural Network (ANN) Model

In the present research, a feed forward MLP type ANN model [16] is improved. Efficient performance of ANN is truly dependent on the selected training algorithm. As stated by many researchers, the Levenberg-Marquardt (L-M) algorithm has superiority to other training algorithms because of its efficiency and high convergence speed [17]-[19]. The ANN architecture in this study contains three layers: an input layer which consist the explanatory variables, one hidden layer and an output layer consisting of a single neuron representing the flow to be modeled at time $t+1$. A cross validation approach is employed to divide the time series into training, cross validation and testing portions. All the forecasts have been done for the stream flow of one month ahead.

III. Results and Discussion

A. Results of the ANN Model Performance in the Three Scenarios

Three scenarios were developed to investigate the influence of wavelet based preprocessing on the ANN model's performance. Using two sets of approximation signals derived from CWT and DWT and the original river flow time series which were utilized as three types of inputs to the ANN model made the three possible scenarios. The model was trained with each of these data sets individually and the predictions were achieved (Fig. 2). Results indicated that the model's performance for CWT and DWT preprocessed inputs (Scenario 1 & 2) is more efficient than the noisy non-processed original data as in scenario 3. Three different evaluation criteria are used to compare the performance of the three models developed in this study. Correlation coefficient (R), Root Mean Square Error $(RMSE)$ and the Mean Absolute Error (MAE) are the mentioned evaluation criteria. Clearly the lower $RMSE$ and MAE indicate better performance of the ANN model and correlation coefficient values close to 1 depict the models efficiency.

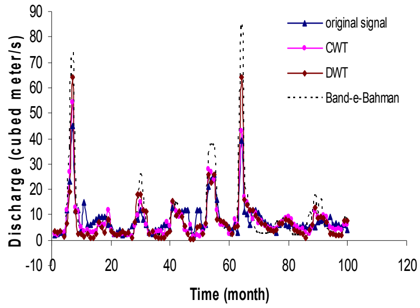


Fig. 2 The actual and predicted stream flow time series for the three scenarios. The approximation signal based on CWT coefficients, DWT coefficients and the original signal is used as input.

Results indicated the superiority of preprocessed data to the original data for the ANN model. The RMSE and MAE values, which indicate the model's performance, were the least for the DWT coefficients of the second order Daubechies [20] wavelet (db2) and the most for the original data series (Table I). The correlation coefficient (R) also depicted the nearest values to 1 for the DWT coefficients. The results indicating the models' efficiency in each case have been depicted in Table I.

TABLE I
Results Indicating the Performance Evaluation Criteria of the ANN Model for the three Scenarios in the Three Stations

Station	Scenario	RMSE	MAE	R
Band-e-Bahman	DWT approximation signal. (db2)	6.47	3.59	0.75
	CWT approximation signal. (db2)	8.94	4.93	0.68
	Original signal	12.0189	7.0035	0.38
Aliabad	DWT approximation signal. (db2)	7.5765	4.1876	0.71
	CWT approximation signal. (db2)	9.3211	5.24	0.64
	Original signal	15.2623	8.546	0.34
Tang-e-Karzin	DWT approximation signal. (db2)	11.863	6.7413	0.42
	CWT approximation signal. (db2)	14.512	8.332	0.34
	Original signal	24.6981	14.1736	0.28

B. Fractal Analysis Results

The fractal analysis results indicating the time series correlation in two states of original and wavelet preprocessed time series depicted a remarkable increase in the Hurst number and the fractal dimension of the preprocessed signals. According to [13] the time series correlation has exponential relation with Hurst number. As shown in Table II, the correlation values of the stream flow series in the three gauge stations increase while the signals are decomposed.

TABLE II
CORRELATION VARIATIONS OF THE RIVER FLOW TIME SERIES FOR THE THREE STATIONS

	H			$C(t)$		
Station	Original signal	DWT app. signal	CWT app. Signal	Original signal	DWT app. signal	CWT app. signal
Band-e-Bahman	0.4417	0.81287	0.7841	-0.0776	0.543	0.4827
Aliabad	0.4341	0.78696	0.7521	-0.0871	0.488	0.4184
Tang-e-Karzin	0.4875	0.88979	0.834	-0.0171	0.7166	0.5888

Time series persistency is also increased after excluding the noise contents from the series. As stated before, the Hurst values in the range of $0.5 < H < 1$ indicate the persistency of the time series. As seen in Table III the preprocessed time series possess higher H values and consequently more persistency. Results of the fractal based predictability indices of time series clearly indicated the time series predictability enhancement after the de-noising process.

TABLE III
PREDICTABILITY INDICES OF THE RIVER FLOW TIME SERIES FOR THE THREE STATIONS

	Df			PIf		
Station	Original signal	DWT app. signal	CWT app. Signal	Original signal	DWT app. signal	CWT app. Signal
Band-e-Bahman	1.5582	1.1817	1.2158	0.1165	0.6365	0.5682
Aliabad	1.5658	1.2130	1.247	0.13162	0.5043	0.5043
Tang-e-Karzin	1.51426	1.1102	1.157	0.02852	0.77958	0.686

As the fractal dimension of a persistent hydrologic time series vary in the range of $1 < D_f < 1.5$ the predictability indices vary as $0 < PI_f < 1$. The more persistent the series are, the higher the predictability indices of the series would be. Accordingly, the wavelet preprocessing of the river flow time series has increased the predictability of series as well as their correlation and persistency. Results of the predictability analysis have been presented in Table III.

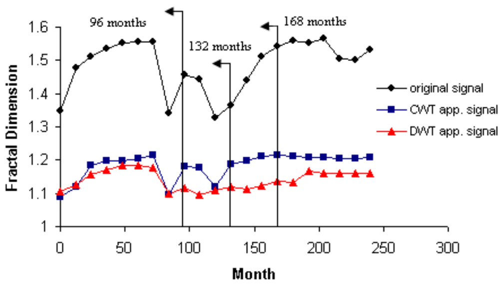
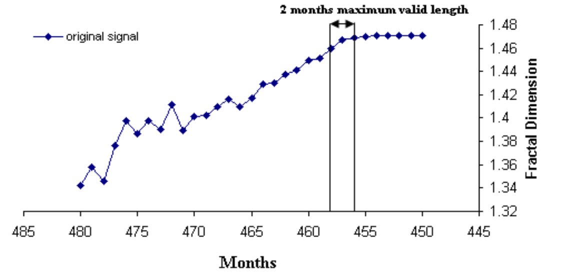
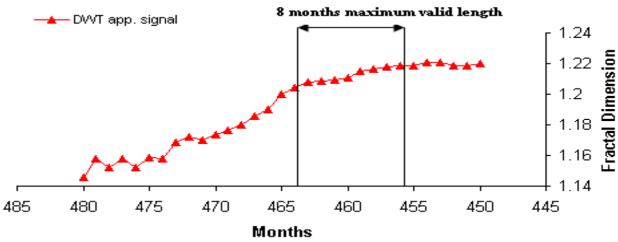


Fig. 3 The minimum sufficient length for prediction in the three scenarios

As can be seen in Figs. 3 and 4 the minimum sufficient length of the time series which is needed for the training phase of the prediction model is influenced by the wavelet preprocessing as well as the maximum valid length of the predictions. Results depicted that while the minimum needed length of data series got shorter the maximum valid length of the predictions increased.



(a)



(b)

Fig. 4 The maximum valid length in Band-e-Bahman station (a) the maximum valid length for the original signal (b) the maximum valid length for the DWT app. Signal

According to Fig. 3 while the minimum length of the series needed to train the neural network model is 168 months for the original signal, this length would reduce to 132 months for CWT approximation signals and 96 months for the DWT approximation signals. Studying Fig. 4 also depict that while the maximum valid length of the prediction is only 2 months for the original signals, the maximum valid length exceed 8 months for the preprocessed signals. The accessible data were 456 months and the predictions were done for 24 months. The trend of the fractal dimension variations was considered from the 450th month to the last predicted month.

IV. CONCLUSION

A feed forward MLP with L-M training algorithm was used to predict the stream flow time series. Wavelet de-noising approach using CWT and DWT techniques was employed in order to preprocess the data series. As derived from the evaluation criteria the ANN model performance, in terms of predictions, was remarkably better for preprocessed time series. DWT based approximation signals (db2) led to the best results. CWT based approximation signals also seemed more appropriate than the raw noisy original data series. Results of the fractal assessment emphasized on the major influence of signal de-noising on increasing the correlation rate and the long-term memory of the preprocessed time series. The fractal analysis of the data series showed an increase in time series persistency after the preprocessing. The fractal based predictability indices also depicted a notable improvement for preprocessed time series. Results of the utilized fractal windowing technique depicted decreases in the estimates of the minimum sufficient lengths of data for training the prediction model and increases in the maximum valid lengths of the predicted data series which indicated on the disclosure of the inherent trend of the time series and the reliability of the predictions when using wavelet based preprocessing techniques.

References

[1] A. Elshorbagy, S.P. Simonovic, U.S.Panu, "Noise reduction in chaotic hydrologic time series: facts and doubts", Journal of Hydrology 256 pp.147-165, 2002.

[2] T. Gerstner, H. P. Helfrich, A. Kunoth "Wavelet Analysis of Geoscientific Data Dynamics of multiscale earth systems", Springer, pp.70-88. 2003.

[3] B. Cannas, A. Fanni, L. See, G. Sias, "Data preprocessing for river flow forecasting using neural networks: Wavelet transforms and data partitioning", Physics and Chemistry of the Earth 31 pp. 1164–1171, 2006.

[4] G.P. Nason, R. Von Sachs, "Wavelets in time series analysis", Philosophical Transactions of the Royal Society of London, series A 357 pp. 2511–2526, 1999.

[5] B.B. Mandelbrot, J.W. Van Ness, "Fractional Brownian motions", fractional noises and applications, SIAM Rev. 10 pp. 422-437, 1968.

[6] A. Eke, P. Hermann, L. Kocsis, L.R. Kozak, "Fractal characterization of complexity in temporal physiological signals", Physiological Measurement 23 pp.1–38, 2002.

[7] D. Delignieres, S. Ramdani, L. Lemoine, K. Torre, M. Fortes, G. Ninot "Fractal analyses for short time series: A re-assessment of classical methods", Journal of Mathematical Psychology 50 pp. 525-544, 2006.

[8] C. K.Peng, J. Mietus, J. M. Hausdorff, S Havlin, H. E Stanley, A.L. Goldberger, "Long-range anti-correlations and non-Gaussian behavior of the heartbeat" Physical Review Letter; 70 pp. 1343-1346, 1993.

[9] B. B. Mandelbrot, "The Fractal Geometry of Nature", W H Freeman publication, New York, 1982.

[10] B.B. Mandelbrot, "Multi fractals and $1/f$ noise", Wild self-affinity in physics, Springer, New York, 1999.

[11] B. B. Mandelbrot, "Fractals: form, change and dimension", San Francisco, CA: Freeman publication, New York, 1977.

[12] J. Feder, "Fractal", Plenum Press, New York, 1988.

[13] F. Pallikari, "A study of the fractal character in electronic noise process", Chaos, Solitons and Fractals 12 pp. 1499-1507, 2001.

[14] G. Rangrajan, D.A. Sant, "Fractal dimension analysis of Indian climate dynamics", Chaos,Solitons and Fractals 19 pp.285-291, 2004.

[15] S. Rehman, M. El-Gebeily, "A study of Saudi climate parameters using climate predictability indices", Chaos, Solitons and Fractals 41(3) pp. 1055-1069, 2009.

[16] M. T. Hagan, H. B. Demuth, M. Beale, "Neural network design", (2nd Edition), PWS publishing company, Boston, USA, 1996.

[17] M. Nayebi, D. Khalili, S. Amin, Sh. Zand-Parsa, "Daily stream flow prediction capability of artificial neural networks as influenced by minimum air temperature data", Biosystems and S. Sorooshian. Artificial neural network modeling of the rainfall runoff process Engineering 95(4) pp. 557-567, 2006.

[18] F. Antcil, N. Lauzon, "Generalization for neural networks through data sampling and training procedure, with applications to stream flow predictions", Hydrology and earth system science 8(5) pp. 940-958, 2004.

[19] N. J. De Vos, T. H. M. Rientjes, "Constraints of artificial neural networks for rainfall-runoff modeling: trade-offs in hydrological state representation and model evaluation", Hydrology and earth system science 7(5) pp. 693-706, 2005.

[20] I. Daubechies, "Ten Lectures on Wavelets", CSBM–NSF Series Application Mathematics, vol.61, SIAM publication, Philadelphia, PA. 1992.