

INVESTIGATION OF SNOW COVER DATA ON ANN AND ANFIS MODELS FOR STREAM FLOW FORECASTS

Sedigheh ANVARI TAFTI¹, Bahram SAGHAFIAN² & Saeid MORID^{3*}

- (1) Department of Water Structure Engineering, Tarbiat Modares University, Iran, e-mail: anvari@modares.ac.ir
(2) Soil and Water Conservation Research Institute, Iran, email: Saghafian@scwmri.ac.ir
(3) Department of Water Resources Engineering, Tarbiat Modares University, Iran, e-mail: morid_sa@modares.ac.ir

Abstract

In recent decades, population growth and increasing water demand, on one hand, and climate change and drought episodes, on the other, underline the importance and sensitivity of water resources management. Streamflow forecast is one of the fundamental tools in water resources management. Demand for increased accuracy of flow forecasts is particularly raised in large mountainous basins, such as Karoon basin in Iran. In this study, 1- to 3-months ahead forecasts of Karoon river flow were conducted using nonlinear and intelligent techniques including Artificial Neural Networks (ANNs) and Adaptive Neuro-Fuzzy Inference System (ANFIS) models. In addition, the effect of snow cover (SC) data, extracted from the National Oceanic and Atmospheric Administration (NOAA) images on the performance of these models was investigated. Relying on cross correlation and PCA techniques, three input patterns were constructed and their performances as inputs of ANN and ANFIS models were evaluated. Then, the most accurate model was considered as the selected pattern and the effect of SC data on the performance of the selected pattern was investigated. In order to investigate the effectiveness of SC data during both snowfall and snowmelt periods, the performance of selected patterns for the ANN and ANFIS models were further analyzed separately using SC data. Results showed that decreasing the number of input variables through the PCA technique proved efficient, while the effect of SC data on the performance of forecast models was positive. For ANN models, the improvements in streamflow forecasts for all lead times were about 7.8%, 5.1% and 7.9% in Root Mean Square Error (RMSE), determination coefficient (R²) and mean absolute error (MAE) values, respectively. Inclusion of SC as input data also improved the performance of the selected model in both snowfall and snowmelt periods, but the degree of improvement in the snowmelt period was considerable.

1. INTRODUCTION

In the last decades, forecast and modeling of the hydrological processes is quite important for water resources management. In order to estimate hydrological processes such as river flow, application of nonlinear intelligent techniques like as Artificial Neural Networks (ANNs) and Adaptive Neuro-Fuzzy Inference System (ANFIS) have

presented more accurate results than traditional models (ASCE Task Committee, 2000 a,b; Zealand et al., 1999; Coulibaly et al., 2000). A wide range of ANN and ANFIS model applications for the prediction, forecast and simulation of hydrological process, could be found in the findings of the Bazartseren et al. (2003); Nayak et al., (2004); Kisi (2005); Jain and Kumar (2007); Zounemat and Teshnehlab, (2007); Sedki et al.; (2009); Demirel et al., (2009); Shiri and Kisi (2010); Besaw et al. (2010). For example, Nayak et al. (2004) investigated the capability of ANFIS for river flow modeling in India. They found ANFIS outperforms traditional ARMA models. Aqile et al. (2006) applied ANFIS model to forecast weekly flow for Citarum river in Indonesia and compared results with multiple linear regression (MLR) model. Demirel et al. (2009) applied ANN and SWAT models to predict daily flow of Pracana basin in Portugal. They found ANN model is more successful than SWAT in the peak flow forecast. Besaw et al. (2010) developed and tested two ANN models to forecast streamflow in ungauged basins. Results showed that ANNs always converged and were applicable in small ungauged basins. Shiri and Kisi (2010) made a hybrid model including wavelet and neuro-fuzzy (WNF) to simulate daily, monthly and yearly Filyos River flows. It was found that WNF model increased the accuracy of the NF models.

Upgrading the accuracy of forecast models through the input variables also is important especially in the mountainous basins such as Karoon in Iran, which the main part of runoff is formed by snow melting in spring. Review of literature showed no research has been reported on the investigation of Snow Cover (SC) data on the performance of forecasting models e.g. ANN and ANFIS. Regarding to use of snow satellite images, Lee et al (2003) used SC maps obtained from MODIS images as input for snowmelt models for runoff simulation. Tekeli et al (2005) used MODIS-8 daily snow cover products as input of Snowmelt Runoff Model (SRM) for runoff modeling. Parjka and Bloschl (2008), Tested the potential of snow cover data from the MODIS satellite sensor for calibration and validation of a conceptual semi-spatial hydrological model. Result of these researches showed that the application of snow satellite images cause to improve snowmelt model's performance. So this paper aims to study whether SC data extracted from National Oceanic and Atmospheric Administration (NOAA) satellite images, improve the forecast accuracy.

2. CASE STUDY DATA

Karoon basin, in southwestern of Iran, is a mountainous basin. It is limited to 49° 30' to 52° Eastern Longitude and 30° 30' - 32° Northern Latitudes and the weighted average height in this region is 2400 m. The basin has an area of 24,202 km² at Poleshaloo hydrometric station just upstream of Karoon III Dam, which used as a case study. Most of precipitation in this region occurs in the cold season and a good portion is in the form of snow. Therefore, snow plays a key role in surface water supply.

In this study precipitation, river flow, minimum and average temperature as well as SC data for a period of 13 years (1991-2003) was applied to forecast Karoon river flow. Figure 1 shows the location of the study area.

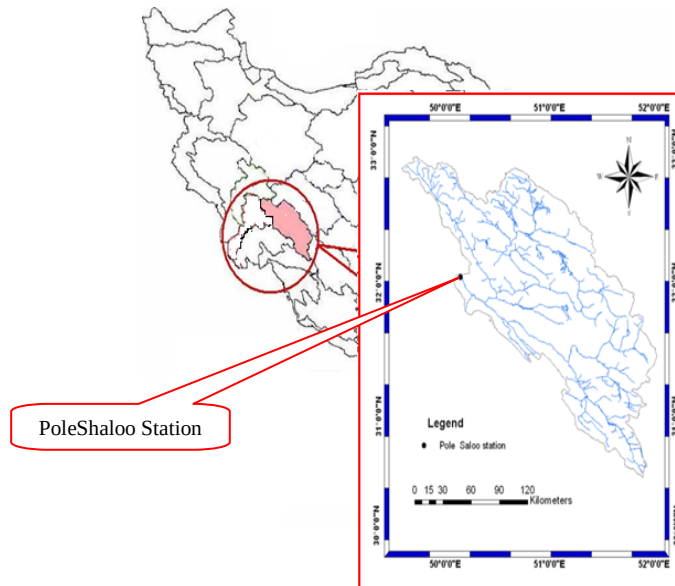


Figure 1. Location of Karoon basin and its selected area, in Iran

3. SNOW COVER EXTRACTION FROM NOAA SATELLITE IMAGES

Use of satellite images is currently one of the most widely practiced SC estimation methods. The application of remote sensing data has significantly increased in snow mapping because of accuracy, ease of access, and coverage of large areas by satellite images. The selection of suitable satellite images is based on the length of image time series, temporal and spatial resolution, size of the area and cost. Overall, for the region of study, National Oceanic and Atmospheric Administration/Advanced Very High Resolution Radiometer (NOAA/AVHRR) are the most appropriate.

The NOAA image resolution is about 1.1 km, which is generally acceptable for extracting snow cover maps in large areas (Saghafian and Davtalab, 2007). SC maps had been derived by Iran Space Agency after discarding images with poor quality owing to atmospheric effects or excessive cloud cover. The image processing procedure followed that proposed by Simpson et al. (1998). In each snow map, zero represents no-snow and one represents snow-covered.

Inaccessibility to regular series of satellite images cause to create gaps between snow map's time series. One method to overcome this problem is use of compound maximum SC maps. These are based on the fact that if a cell is snowy in whole selected period, it is known as snowy in the compound maximum maps. Maximum SC is more suitable than its average and minimum because average and minimum SC maps ignore some information of the snow cover. In addition, compound technique minimizes the errors of snow cover extraction (Hall et al., 2000). In this study, Maximum SC data derived from NOAA satellite images have been used for 13-year period.

4. INPUT SELECTION

In this section, the input selection process is described. At first, the lag number of each input variable was determined. Next, using input selection methods as described later, the most important variables were selected as the input patterns to the forecast models.

4.1 lag number identification for each variable

River flow, precipitation, minimum and average temperature as well as SC data are variables which used in this study. For each variable, the lag number corresponding to different antecedents extracted using Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF) techniques. The ACF and PACF showed a significant correlation at the 95% confidence level interval up to 2 months of snow cover, flow, precipitation and other variables lags. So five primary variables with two lags for them, i.e. 15 variables, are total variables used in this study. Using these variables exception of SC data, 3 inputs configurations were prepared which called input patterns. Then ANN and ANFIS models were run using these patterns and the most accurate model was considered as the selected model. Finally the effect of SC data, including different lags, on the performance of selected pattern was investigated. These input patterns and their results have been presented as follow.

4.2.1 Input patterns

After determination of the lag numbers, using input selection methods three configurations of inputs were prepared and called input patterns. Then, these Input patterns are labeled for easy referencing following this scheme: "*model type –number of forecast months ahead - input selection method*". So, for example, "NN-1- PCA" represents neural network model for 1 month ahead forecast where the main input variables have been selected by the PCA method.

4.2.2 The first input pattern:

In this pattern the most important variables were selected by Principal Component Analysis (PCA) technique. PCA is regarded as one of the main feature extraction methods that transform a set of P dependent variables to the smaller and new set of m independent components. These components are included in variance of original data. Assume that X vector has P variables, N observations, average vector μ and covariance matrix Σ . For obtaining principal components (PC) of this matrix, input variables are first standardized and covariance matrix (Σ) is derived. Then eigenvalues (λ) and eigenvectors (A) of Σ are computed. The eigenvalue for a PC show its variance proportion to the total variance. Then, PCs are calculated as follow:

$$PC_i = a_{i1}X_1 + a_{i2}X_2 + \dots + a_{ip}X_p \quad (1)$$

where PC_i is the i th PC ; a_{ij} is the element of given eigenvectors and X_i is the original variables. So, among of dependent variables i.e. river flow, precipitation,

minimum and average temperature and their corresponding lags, by PCA method, the first five components having 95.2% of total variance and no dependency, were selected as the first input pattern.

4.2.3 The second input pattern

After computation of correlation coefficients between each PC and the input variables, the variable which has maximum correlation with the PC , is selected as input variable (Morrison, 1976). So, the second input pattern include $T_{mean(t-1)}$, $R_{(t-2)}$, $R_{(t-1)}$, $Q_{(t-2)}$ and $Q_{(t)}$ were selected as the second input pattern.

4.2.4 The third input pattern

For this pattern, Cross Correlation (CC) technique was used to choose input variables. In CC method, Pearson correlation coefficient between predictor and predicted variables is computed. Then, variables having significant level lower than 0.05 are selected as input variables for the forecast models.

5 Forecasting Models

5.1 Artificial Neural Networks (ANNs)

The ANNs were introduced as an efficient tool for modeling and forecasting about two decades ago. These networks consist of processing units, called neurons, which can map available linear and nonlinear relations between input –output pairs in one phenomena (ASCE, 2000 a,b). The ANNs used in this paper are three layer feed forward networks with sigmoid function and back propagation training algorithm (for details, see for example Coulibaly et al., 2000). Number of neurons in the hidden layer was obtained with trial and error procedure (ASCE, 2000a). In addition, to avoid network weight minimization, input and output data were rescaled in [0.1-0.9] range (Wang et al., 2006).

5.2 Adaptive neuro-fuzzy inference system (ANFIS)

Neuro-fuzzy modeling refers to the coupled ANNs with a fuzzy inference system (FIS). The procedure of developing a FIS using the framework of adaptive neural networks is called adaptive neuro fuzzy inference system (or ANFIS for short), which was for the first time introduced by Jang (1993). This type of modeling rectifies the main problem of FIS to determine If-Then rules by using the ANFIS. The general structure of the ANFIS is presented in Figure 2. Selection of the FIS is the major concern when designing an ANFIS to model a specific target system. Various types of FIS are reported (Takagi and Sugeno, 1985). In this study Sugeno's Fuzzy Inference System, Gaussian membership functions, hybrid learning algorithm and subtractive fuzzy clustering were used to construct ANFIS models. For more information on ANFIS models and its structure, one may refer to Jang (1993).

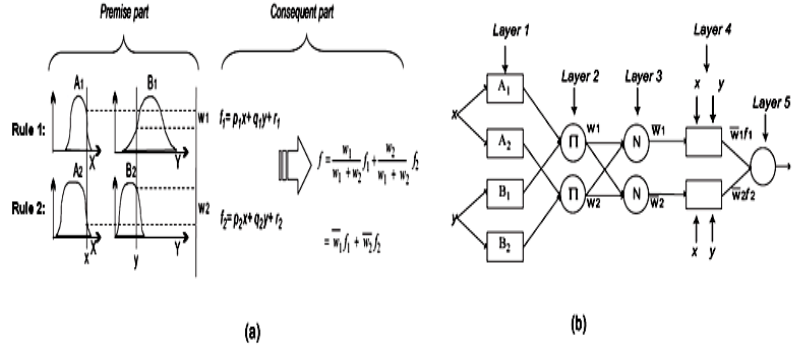


Figure 2. (a) FIS, (b) Equivalent ANFIS architecture (after Jang, 1993)

6 Evaluation indices

In order to evaluate the performance of the forecast models, some evaluation indices including Determination Coefficient (R^2), Root of Mean Squared Error (RMSE) and Mean Absolute Error (MAE) has been used in this study.

7 Results and Discussions

7.1 ANN models

7.1.1 One month ahead stream flow forecasts

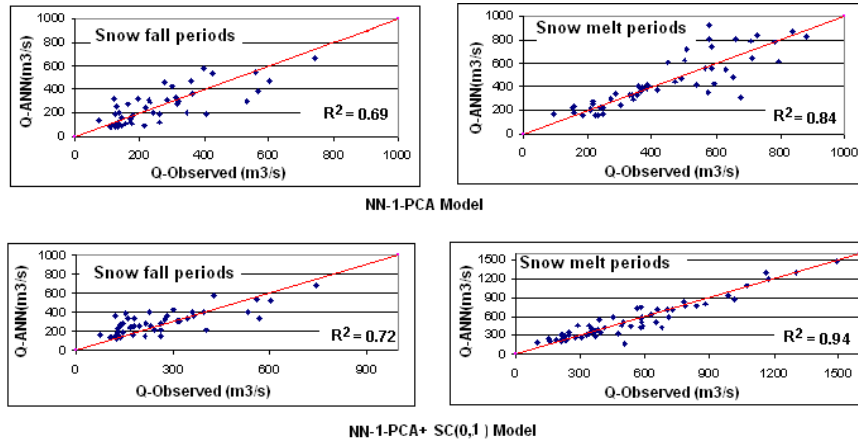
In this stage, ANN with three input patterns was performed and using evaluation indicis, the best pattern was known as the selected pattern. Then the effect of Snow Cover (SC) data at t , $t-1$ and $t-2$ times, labeled as SC(0), SC(1), SC (2) respectively, was investigated on it. Also, the single and compound role of SC data at $(t, t-1)$, $(t, t-2)$, $(t-1, t-2)$, $(t, t-1, t-2)$ which labeled as SC(0,1), SC(0,2), SC (1,2), SC (0,1,2) were investigated on the selected pattern. Some of the more accurate results were presented in Table 1.

Table 1 represents the NN-1-PCA model with 5 principal components as the input has the best performance and know as the selected pattern. SC data as the input, generally has improved the accuracy of this model. Regarding to SC effectiveness on selected model the most influence belongs to the combined form of SC at t and $t-1$ times i.e. NN-1- PCA+ SC (0, 1). In the training phase, this model has improved the performance of selected model about 13.26%, 6.1% and 13.71% in RMSE, R^2 and MAE values respectively. In the testing phase, this model could also improve the its performance about 7.73%, 3.61% and 1.17% in RMSE, R^2 and MAE values, respectively.

Table 1. Results of three input patterns of ANN and investigation of SC data effectiveness

Input Model	Training Phase			Test Phase			Total		
	RMSE	R ²	MAE	RMSE	R ²	MAE	RMSE	R ²	MAE
NN-1- PCA	117.56	0.82	76.59	108.54	0.83	69.42	114.59	0.82	74.17
NN-1-Var PCA	124.61	0.80	79.39	119.17	0.79	79.85	122.80	0.80	79.55
NN-1-CC	95.31	0.88	58.58	152.05	0.76	100.50	117.58	0.82	72.74
NN-1- PCA+ SC(0)	117.55	0.82	76.73	108.60	0.83	69.89	114.60	0.82	74.42
NN-1- PCA+ SC(1)	126.75	0.79	80.30	118.83	0.80	83.75	124.13	0.79	81.47
NN-1- PCA+ SC(2)	111.24	0.84	76.15	107.55	0.85	78.55	110.01	0.84	76.96
NN-1- PCA+ SC(0, 1)	101.97	0.87	66.09	100.15	0.86	70.23	101.36	0.86	67.49
Degree of improvement (%)	13.26	-6.10	13.71	7.73	-3.61	1.17	11.55	-4.88	9.01
NN-1- PCA+ SC(0, 2)	115.18	0.83	77.42	109.58	0.83	73.27	113.32	0.83	76.02
NN-1- PCA+ SC(0, 1, 2)	121.22	0.81	82.75	141.29	0.81	89.26	128.35	0.79	84.95

According to Table 1, for all data series the degrees of improvement are about 11.55%, 4.88% and 9.01% for RMSE, R² and MAE values respectively. Also, to determine the efficiency of SC data during snowfall (November to February) and snowmelt periods (March to July), the performance of two best models i.e. NN-1-PCA and NN-1-PCA + SC (0, 1), was investigated separately for snowfall and snow melt periods (Figure 3) .

**Figure 3.** Separation of NN-1-PCA and NN-1-PCA+ SC(0,1) model performance during snowfall and snowmelt periods

As Figure 3 shows, SC data at t and $t-1$ times improves the performance of NN-1-PCA model for both snowfall and snowmelt periods, but the degree of improvement in the snowmelt periods is more considerable. The reason could be referred to physiographic characteristics of the region. Since Karoon river has located in a

mountainous basin with a lot of snow in the winter, these snow budgeting lead to occur the runoff during snow melt periods. Therefore applying SC data in the forecasts models directly, one can map the relation between inputs and outputs, which is stream flow here, more appropriately.

7.1.2 Two- and three- months ahead stream flow forecasts

For 2- and 3- months ahead forecasts of Karoon river flow, NN-2-PCA and NN-3-PCA models also known as the selected models, respectively. SC data at different times caused to improve NN-2-PCA model's performance averagely about 7.16%, 4.74% and 5.82% in RMSE, R² and MAE values respectively. Also for NN-3-PCA the degree of improvement is about 4.97%, 5.63% and 8.92% in RMSE, R² and MAE values respectively.

7.2 ANFIS models

Performance of ANFIS models by means of trial and error procedure on cluster radius was evaluated in this section. Result of ANFIS models for 1- to 3- months ahead stream flow forecasts were presented as follow. The most accurate model for 1- month ahead flowforecast was NF-1-PC that named selected pattern. Therefore the effect of SC data on its performance was investigated through Figure 4. According to this figure the effect of snow cover data with different lags was positive.

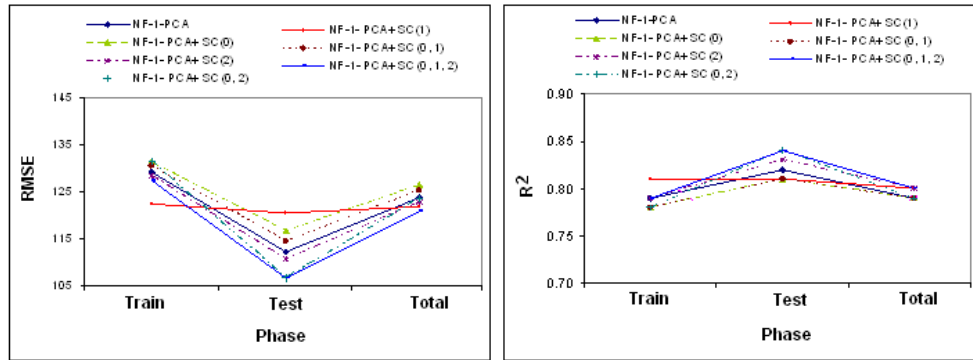


Figure 4. Graph of NF-1-PCA model and the effect of snow cover input

The most accurate model for 2- month ahead flow forecasts was NF-2-PC. Therefore the effect of Snow Cover on its performance was averagely about 4.8%, 5.3% and 5.06% in RMSE, R² and MAE values respectively. The effect of snow cover data on 3- month ahead flows was also positive.

7.3 Results Comparison

In this section, the results of two best models extracted from ANN and ANFIS for 1- to 3- months ahead of Karoon flow forecasts were compared using RMSE and R² evaluation indices. Results comparison showed ANNs outperform ANFIS models for both training and test phases. SC data as inputs has also improved their performance.

Obtained results for 2- and 3- months ahead flow forecasts of Karoon river also confirmed this issue.

8. CONCLUSIONS

In this study, the potential of ANN and ANFIS models to forecast 1- to 3- month ahead flows of Karoon river was investigated. Also the effect of Snow Cover (SC) data extracted from NOAA satellite images as the input variable through ANN and ANFIS models was assessed. Results demonstrated that use of PCA technique to decrease the number of input variables was effective. The influence of SC data on the performance of ANN and ANFIS models was positive. For example for ANN models, the degree of improvements for all lead times were about 7.8%, 5.1% and 7.9% for RMSE, R2 and MAE values respectively. Also, in order to investigate the effectiveness of SC data, the performance of selected patterns from ANN and ANFIS models in combination with SC data were analyzed separately during snowfall and snowmelt periods. Results revealed SC data can improve the performance of selected models for both periods but the degree of improvements for melt periods was more considerable. Finally, the results comparison between ANN and ANFIS models exhibited the ANN models outperformed.

REFERENCES

- Aqil, M., I. Kita, A. Yano and S. Nishiyama (2006) Analysis and prediction of flow from local source in a river basin using a Neuro-fuzzy modeling tool, *Journal of Environmental Management*, 85, 215-223.
- ASCE Task Committee on Application of Artificial Neural Networks in Hydrology (2000a) Artificial neural networks in hydrology-I. Preliminary concepts, *Journal of Hydrologic Engineering*, 5(2), 115-123.
- ASCE Task Committee on Application of Artificial Neural Networks in Hydrology, 2000b. Artificial neural networks in hydrology-II, Hydrologic applications, *Journal of Hydrologic Engineering*, 5(2), 124-137.
- Bazartseren, B., G.Hildebrandt, and K.P. Holz (2003) Short-term water level prediction using neural networks and neuro-fuzzy approach, *Neurocomputing*, 55, 439-450.
- Besaw, L.E., D.M. Rizzo, P.R. Bierman and W.R. Hackett (2010) Advances in ungauged streamflow prediction using artificial neural networks, *Journal of Hydrology*, 386, 27-37.
- Coulibaly, P., F. Anctil and B. Bobée (2000) Daily reservoir inflow forecasting using artificial neural networks with stopped training approach, *Journal of Hydrology*, 230, 244-257.
- Demirei, M. C., A. Venancio and E. Kahya (2009) Flow forecast by SWAT model and ANN in Pracana basin, Portugal, *Advances in Engineering Software*, 40, 467-473.
- Efron, B. and R.J. Tibshirani (1993) *An Introduction to the Bootstrap*, Chapman and Hall, London.
- Firat, M. and M. Gungor (2007) River flow estimation using adaptive neuro fuzzy inference system, *Mathematics and Computers in Simulation*, 75, 87-96.
- Hall, D. K., Riggs, G. A., Salomonson, V. V., Digirolamo, N. E., and Bayr, K. J., 2002, MODIS snow-cover products. *Remote Sensing of Environment*, 83, 181-194.
- Jain, A. and A.M. Kumar (2007) Hybrid neural network models for hydrologic time series forecasting, *Applied Soft Computing*, 7, 585-592.
- Jang, J. (1993) Anfis: Adaptive-network-based fuzzy inference system, *IEEE transactions on systems, man, and cybernetics*, 23, 665-685.
- Kisi, Ö. (2005) Daily river flow forecasting using artificial neural networks and auto-regressive

- models, *Turkish Journal Engineering Environmental Science*, 29, 9-20.
- Lee S, Klein AG, Over TM. 2003. In Review, An Assessment of the Suitability of MODIS Snow Products for Simulating Streamflow in the Upper Rio Grande River Basin using the Snowmelt Runoff, *Hydrological Processes*, http://geog.tamu.edu/klein/publications/in_review/hp_2003b.pdf.
- Morrison, D.F. (1986) *Multivariate Statistical Methods*. 2nd Fed. Auckland, Mcgraw-Hill Book Company, Pp:266-302.
- Nayak, P.C., K.P.Sudheer, D.M. Rangan and K.S. Ramasastri (2004) A neuro-fuzzy computing technique for modeling hydrological time series, *Journal of Hydrology*, 291, 52-66.
- Saghafian, B. and Davtalab, R. (2007). Mapping snow characteristics based on snow observation probability. *International Journal Of Climatology*, 27(10): 1277–1286.
- Sedki, A., D. Ouazar, and E. El Mazoudi (2009) Evolving neural network using real coded genetic algorithm for daily rainfall–runoff forecasting, *Expert Systems with Applications*, 36, 4523-4527.
- Shiri, J. and O. Kisi (2010) Short-term and long-term streamflow forecasting using a wavelet and neuro-fuzzy conjunction model, *Journal Of Hydrology*, **394**(3-4), 486-493.
- Simpson JJ, Stitt JR, Sienko M. 1998. Improved estimates of the areal extent of snow cover from AVHRR data. *Journal of Hydrology* 204: 1–23.
- Takagi, T., Sugeno, M., 1985. Fuzzy identification of systems and its application to modeling and control. *IEEE Transactions on Systems, Man and Cybernetics* 15 (1), 116–132.
- Tekeli, E. A., Akyurek, Z., Sorman, A. A., Sensoy, A., & Sorman, Ü. A. (2005). Using MODIS snow cover maps in the modeling snowmelt runoff process in the eastern part of Turkey. *Remote Sensing of Environment*, 97,216–230.
- Wang, W., P.H.A.J.M .Van Gelder, J.K. Vrijling and J. Ma (2006) Forecasting daily streamflow using hybrid ANN models, *Journal of Hydrology*, **324**, 383-399.
- Zealand, C., D. Burn and S. Simonovic (1999) Short term streamflow forecasting using artificial neural networks, *Journal of Hydrology*, 214, 32-48.
- Zounemat-Kermani, M. and M. Teshnehlal (2007) Using adaptive neuro-fuzzy inference system for hydrological time series prediction. *Applied Soft Computing*, **8**, 928-936.