The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM)
ISSN: 2602-3199

**The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM), 2020**

**Volume 11, Pages 125-130**

**IConTES 2020: International Conference on Technology, Engineering and Science**

# Evaluating Stream Flow Forecasting Performance Using Adaptive Network Based Fuzzy Logic Inference System, Artificial Neural Networks with Feature Selection

**Levent LATIFOGLU**
Erciyes University

**Abstract**: Water resources are needed to maintain the human life and the management the ecologic system for many areas. The most economical use, protection and development of water resources have a great importance for hydrological studies. Variable such as stream flow data are commonly used in hydrology. Accurate stream flow estimation is very important in terms of planning and management of water resources and minimizing the effects of natural disasters such as drought and flood. Monthly river flow data obtained from the Sakarya basin on Porsuk River between 1970-2000 years were used for the estimation study. For this purpose, forecasting performance has been analyzed using Adaptive Network Based Fuzzy Logic Inference System (ANFIS) and Artificial Neural Networks (ANN) models and performances of these two models were compared. In addition, the average monthly stream flow data, standard deviation values of these data were also used in the forecasting study and applied as an input to ANFIS and ANN models. For a one ahead estimation, models have been developed with different input combinations of 1-3 past value of stream flow data and standard deviation values. In this study, mean square error (mse), mean absolute error (mae) and correlation coefficient parameters were used to evaluate the performance of the models. According to the obtained results, it is seen that the ANN model has better forecasting performance for two inputs according to mse and mae parameters and for three inputs according to R and $R^2$ parameters. Also, it is seen that the ANFIS model has the best performance for two inputs according to mse, mae, R and $R^2$ parameters. There has been some improvement in the forecast performance if the monthly average river stream flow data as well as the standard deviation data has been applied as an input to the model.

**Keywords:** Stream flow forecasting, Adaptive Network Based Fuzzy Logic Inference System Method (ANFIS), Artificial Neural Networks (ANN)

## Introduction

In the planning of hydrological processes, river flow modeling and streamflow estimation are very important in terms of effective planning and use of water resources. Stream flow forecasting is essential for efficient operation and planning of reservoirs, sediment transport in rivers, hydropower generation, irrigation management decisions, scheduling of reservoir emissions and other hydrological applications. Accurate prediction of hydrological time series is a challenging task. This is because the dynamics of the stream flow data tends to have complex nonlinear dynamics and chaotic behaviors and nonstationary character. Increasing the predictions accuracy and reliability of non-stationary hydrological variables has been an important research topic. To date, no unique approach has been performed providing optimal stream flow estimation results. Artificial intelligence (or statistical) models are also classified as "black box" models. These models are very useful in modeling natural systems and these methods include mathematical optimization algorithms as they are used in logic, classification, statistical learning, and probability-based methods (Bayazit,, 1998; Chow et. al., 1988).

AI techniques such as Artificial Neural Networks (ANN), Support Vector Machines (SVM), Genetic programming (GP), Adaptive Network Based Fuzzy Inference System (ANFIS) have shown remarkable results



- Selection and peer-review under responsibility of the Organizing Committee of the Conference

in estimating nonlinear stream flow data (Adamowski et.al. , 2010; Cigizoglu, 2005; Dehghani et.al, 2019; Hadi et. al., 2018; He et. al, 2019; Jain et. Al, 2007, Mehr et al., 2015).

In this study, estimation of the average monthly stream flow data one of hydrological time series data has been performed. By investigating the appropriate methods for the forecasting of these data which are not linear and non-stationary by nature, a forecasting study has been made with ANN and ANFIS methods. In addition, the average monthly stream flow data, standard deviation values of these data were also used in the forecasting study as input parameter and applied as inputs to ANFIS and ANN models. The rest of the paper is organized as follows. Section 2 gives information about study area and the data. Section 3 provides a brief review of the ANN and ANFIS approaches for streamflow estimation. Section 4 describes the forecasting results obtained by ANN and ANFIS models using the proposed approach and concludes the paper.

## Materials and methods

Recently, artificial intelligence techniques have been used as an alternative to classical approaches for estimating hydrological data. Among them, ANN and ANFIS are common. In this paper the forecasting of stream flow models has been developed using ANN and ANFIS models.

### Study Area and Data

The mean monthly stream flow in ($m^3$/s) data has been continuously gauged over the 30 year period, between 1970 through 2000, hence consisting of 372 successive numbers are used as the material of this study. Region where data has been gauged has been from the Porsuk River at Beşdeğirmen Station lies between 30° 02' 12" East - 39° 31' 41" North, in the Sakarya basin in the South East Marmara Region in Turkey. Drainage basin from which the data has been recorded is shown in Figure 1. This data set has been obtained from Electrical Power Resources Survey Administration (EIEI), in Turkey (as seen in Figure 2) (www.dsi.gov.tr/faaliyetler/akim-gozlem-yilliklari). The drainage area at this site is 3838.4 $km^2$. 260 of these elements have been used for the training phase and the rest 112 values have been used for the testing stage. All of these data are normalized.

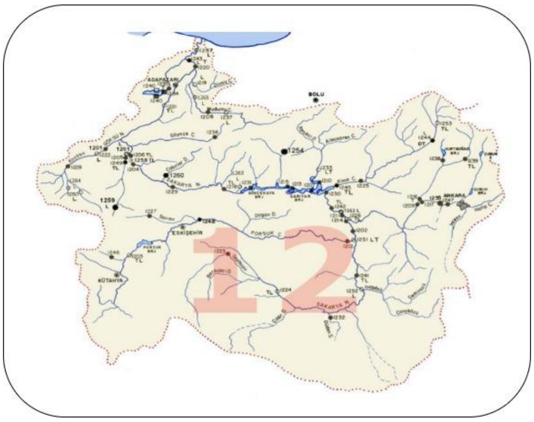


Figure 1. Basin of porsuk in turkey (www.dsi.gov.tr/faaliyetler/akim-gozlem-yilliklari)

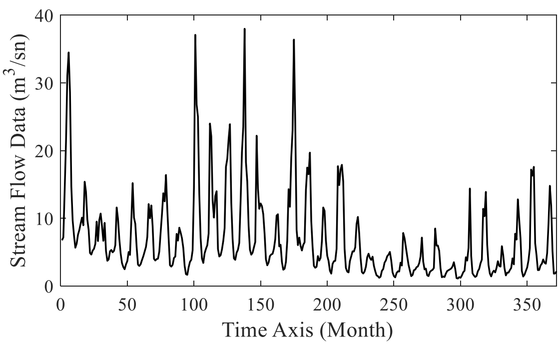



Figure 2. Streamflow data measured on porsuk river

**Artificial Neural Networks**

Recently, ANN has been widely used in forecasting study due to its advantages such as its ability to process nonlinear structures and its ability to process parallel and series. Just as there are nerve cells in biological neural networks, there are artificial nerve cells in ANN. Artificial nerve cells are called processor elements in engineering science. In an ANN structure, the sum and activation functions of the processor elements, the learning strategy and the learning rule are used. The topology resulting from the connection of the processor elements determines the model of the network. Artificial neurons come together to form ANNs. Neurons are not collected randomly. Generally, cells form a network of three layers called as the input, hidden and output layers, and these cells are positioned in parallel in each layer. In this study, the multi-layer perceptron (MLP) model has been used that has the input layer, one or more hidden layers, and output layer. The processor element in the input layer acts as a buffer that distributes the input signals to the processor element in the hidden layer and each one (layer) has at least one neuron. Each input is combined via the sum function of the multiplied input values (wij*ai) by the weight value (wij) that links it to the processing element.

$$y_j = f\left(\sum_i w_{ij} a_i + \theta\right) \quad (1)$$

Where, j is the neuron number, i is the input number, $a_i$ input signal, $w_{ij}$ weight coefficient and θ is the bias term (or threshold). These operations are repeated for all processors in the hidden layers. The processor element in the output layer also acts as a hidden layer processor element and network output values are calculated. Thus, the output of neurons in the output layer has been calculated. The backpropagation algorithm is used in the training of MLP networks commonly, because it can be proved mathematically and is easy to apply. This strategy gets its name (backpropagation) by reducing errors from output to backward input. Back propagation strategy is the most common learning strategy with a supervised learning structure and used in many applications. In this study, for forecasting of stream flow data, MLP and Back Propagation Neural Network (BPNN) feed forward network structure have been used. In the ANN, different learning algorithms are used to train the network. The Levenberg Marquardt (LM) learning algorithm, which has a computational speed advantage in the ANN, is used here for the forecasting of the hydrological time series data (Hagan et. al. 1996: Haykin, 1994).

In order to be used in time series estimation of MLP networks, the structure of the network must be determined. Number of layers of the network, number of processing elements in each layer, and transfer function to be used for these process elements are important for determination of the ANN structure. The number of output neurons is determined depending on how many periods will be forecasted. Determination of the number of neurons to be used at the input is not as easy as determination of the number of output neurons, because it is necessary to determine the data value at time t is affected by how many past observation values.

In this study, different numbers of neurons in the hidden layer, different number of layers, inputs and transfer functions have been set in the architecture of the ANN, and the optimum architecture of the model has been determined based on the resulting minimum mean square error between forecasted and stream flow data during training performance (Marquardt, 1963).

**Adaptive Network Based Fuzzy Logic Inference System (ANFIS)**

Various methods have been used to regulate the relationship between many variables. Some of these methods are: ANNs and fuzzy inference systems (FIS). It is a combination of neural networks and FIS known as ANFIS. The adaptive neural fuzzy inference system (ANFIS) is the Tagaki-Sugeno-Kang fuzzy inference deep fuzzy map algorithm [16]. It was developed by Jang in the early 1990s and was used in modeling nonlinear functions and forecasting chaotic time series. The ANFIS model generally uses the hybrid learning algorithm. ANFIS combined with the hybrid learning algorithm can create an input-output structure based on forecasted input-output data pairs. ANFIS takes advantage of both structures as it integrates neural networks and fuzzy logic inference methods. In order to apply the ANFIS method, a data set based on input-output is generally required. The model established depending on the number and type of membership functions selected is created using a learning algorithm. The method uses the created fuzzy if-then set of rules.

The main structure of ANFIS consists of five layers. Layer 1 is the fuzzification layer, layer 2 is the rule layer, layer 3 is the normalization layer, layer 4 is the defuzzification layer and layer 5 is the output layer. For simplification, it is assumed that the framework of ANFIS has two inputs x; y and one output z. This framework has two rules which are the corresponding rule set with two fuzzy (if-then) rules. This set is used for a first order Sugeno fuzzy model expressed in below Equation.

**Rule 1:** if x is $A_1$ and y is $B_1$ then $z_1=p_1x+q_1y+r_1$
**Rule 2** : if x is A2 and y is B2 then $z_2=p_2x+q_2y+r_2$ (2)

where p1; q1; r1 and p2; q2; r2 are the consequent constants. A1; B1 and A2; B2 are the linguistic labels. Adaptive (compatible) networks consist of directly attached nodes. Each node represents a unit of processing. Connections between nodes indicate an uncertain weight between them. All or some of the nodes may be adaptive. Adaptation is created by determining the outputs of these nodes with variable parameters. Learning rules determine how variable parameters should be changed to make the difference between the output of the entire network and the target value, that has to minimize the error. The basic learning rule in adaptive networks is the steepest descent method. In this study, the ANFIS parameters; the number of rules, membership functions and learning rule were chosen by trial and error, also the initial set of values has been found as in Ref. (Jang, 1993).

### Training and Test Sets

The stream flow data has been splitted into two subsets before performing the model fitting. The subset of training was composed of 70% of original length. Thus, during the training, the 70% of the hydrological time series has been applied to the ANN and ANFIS model and the remaining 30% has been applied as test set to compare the quality of the k-steps-ahead forecasted values. The performance parameters to measure the trained for the k-step-ahead forecasting model are the Mean Squared Error (mse), Mean Absolute Error (mae), the correlation coefficient (R) and Coefficient of Determination ($R^2$). One to three numbers of inputs obtained from monthly mean of stream flow data (x(t)) and standard deviation of stream flow data (st(t)) in each month have been performed to show forecasting performance. The forecasting process has been summarized in below table.

Table 1. One to three ahead forecasting process

| One Ahead Forecasting | | |
|---|---|---|
| One input | Two inputs with std | Two inputs |
| $\hat{x}(t) = f(x(t-1))$ | $\hat{x}(t) = f(x(t-1), st(t-1))$ | $\hat{x}(t) = f(x(t-2), x(t-1))$ |
| Three inputs | Four inputs | |
| $\hat{x}(t) = f(x(t-2), x(t-1), st(t-1))$ | $\hat{x}(t) = f(x(t-3), x(t-2), x(t-1), st(t-1))$ | |

## Results and Discussions

In this study, to evaluate the forecasting performance of stream flow data ANN and ANFIS models have been performed with different number of inputs and features. The obtained results have been summarized in Table 2 and 3 for one to four input ANN and ANFIS models for one step ahead forecasting. Also forecasted data using ANN and ANFIS model vs observed data can be seen in Figure 3.

Table 2. Testing performance of ANN model for one to four input

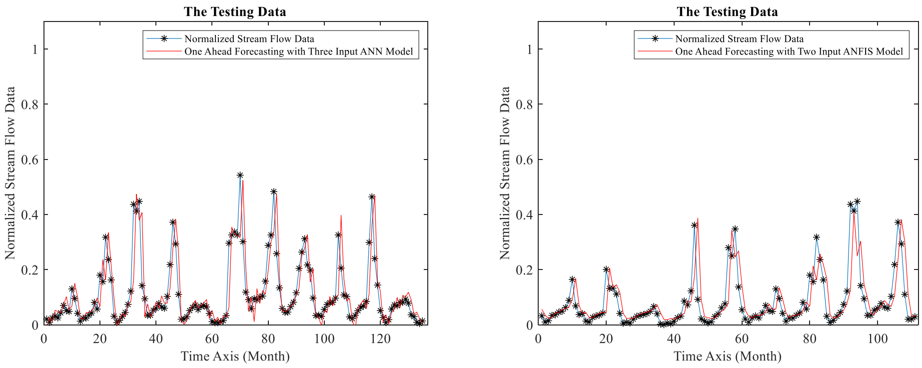
| | One input | Two inputs with std | Two inputs | Three inputs | Four inputs |
|---|---|---|---|---|---|
| MSE | 0.0057 | 0.0054 | 0.0065 | 0.0061 | 0.0062 |
| MAE | 0.0476 | 0.0466 | 0.0500 | 0.0485 | 0.0516 |
| R | 0.6896 | 0.6951 | 0.7051 | 0.7466 | 0.7140 |
| $R^2$ | 0.4707 | 0.4784 | 0.4725 | 0.5534 | 0.5052 |

Table3. Testing performance of ANFIS model for one to four input

| | One input | Two inputs with std | Two inputs | Three inputs | Four inputs |
|---|---|---|---|---|---|
| MSE | 0.0069 | 0.0055 | 0.0066 | 0.0067 | 0.0065 |
| MAE | 0.0562 | 0.0439 | 0.0559 | 0.0565 | 0.0551 |
| R | 0.7267 | 0.7371 | 0.7379 | 0.7331 | 0.7377 |
| $R^2$ | 0.4992 | 0.5303 | 0.5045 | 0.5196 | 0.5142 |

Obtained performance parameters showed in the Table 2 and 3 indicates that the forecasting performance of the models. As an example, ANN model performance has been better than ANFIS model for one step ahead forecasting according to mse and mae performance parameters. But the forecasting performance of the ANFIS model better than ANN model for one step ahead forecasting according to R and $R^2$ performance parameters. While the mse, mae, R and $R^2$ values have been obtained as 0.0057, 0.0476, 0.6897 and 0.8641 in the ANN method for one input one ahead forecasting, it has been obtained as 0.0069, 0.0562, 0.7267 and 0.4992 respectively in the ANFIS method. When Table 3 is examined, it is seen that the performance of the ANFIS model increases by adding the standard deviation values as input parameters of ANFIS model. It is same as also for the ANN model for three input (two of past values of the time serie with standard deviation value), as can be seen from Table 2. Also, the optimal performance has been obtained from ANFIS model using two inputs (one of time series past values, one of standard deviation values) and the optimal performance has been obtained from ANN model using three inputs.

In this study monthly streamflow data forecasting performance of ANN and ANFIS model has been analyzed and the effect of input features to the forecasting performance has been tested using the past value of the data and standard deviation values of the data. As a result, when the standard deviation value is used for the input parameters, it has been observed that the forecasting performance increases even more.




a b

Figure 3. One ahead forecasted stream flow data a- obtained from three input ANN model, b- obtained from two input ANFIS model

## References

Adamowski, J., & Sun, K. (2010). Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds. *Journal of Hydrology, 390,* 85-91.

Bayazıt M., (1998). *Hidrolojik Modeller*, İTÜ İnşaat Fakültesi Matbaası, İstanbul.

Chow V.T., Maidment D.R. & Mays L.W., (1988). *Applied Hydrology*, McGraw-Hill, NY.

Cigizoglu, H. K. (2005). Application of generalized regression neural networks to intermittent flow forecasting and estimation. *Journal of Hydrologic Engineering*, *10*(4), 336-341.

Jain, A., & Kumar, A. M. (2007). Hybrid neural network models for hydrologic time series forecasting. *Applied Soft Computing, 7*(2), 585-592.

Jang, J.S.R., (1993). ANFIS: adaptive-network-based fuzzy inference system, *IEEE Trans. Syst. Man Cybern, 23*, 665-685.

Dehghani, M., Seifi A., & Riahi-Madvar H., (2019). Novel forecasting models for immediate-short-term to long-term influent flow prediction by combining ANFIS and grey wolf optimization, *Journal of Hydrology, 576*, 698-725.

Hadi, S.J., & Tombul M., (2018), Monthly streamflow forecasting using continuous wavelet and multi-gene genetic programming combination, *Journal of Hydrology, 561*, 674-687.

Hagan, M.T., Demuth H.B., & Beale M.H., (1996). *Neural Network Design*. Boston. MA: PWS Publishing.

Haykin, S., (1994), *Neural Networks: A Comprehensive Foundation*, Macmillan College Publishing Company Inc., New.

He, Y., Yan Y., Wang X., & Wang C., (2019). Uncertainty Forecasting for Streamflow based on Support Vector Regression Method with Fuzzy Information Granulation. *Energy Procedia, 158*, 6189-6194.

Marquardt, DW., (1963). An algorithm for least-squares estimation of nonlinear parameters, *Journal of the Society for Industrial and Applied Mathematics, 11*, 431-441.

Mehr, A. D., Kahya, E., Şahin, A., & Nazemosadat, M. J. (2015). Successive-station monthly streamflow prediction using different artificial neural network algorithms. *International Journal of Environmental Science and Technology*, *12*, 2191-2200.

www.dsi.gov.tr/faaliyetler/akim-gozlem-yilliklari , (accession date February, 2020)

## Author Information

**Levent Latifoğlu**
Erciyes University
Civil Engineering Dept.
Contact e-mail: *latifoglu@erciyes.edu.tr*