International Scientific Conference *People, Buildings and Environment 2014* (PBE2014)
15-17 October, 2014, Kroměříž, Czech Republic, www.fce.vutbr.cz/ekr/PBE

MONTHLY STREAMFLOW TIME SERIES MODELLING OF CORUH RIVER

Meral Buyukyildiz[1]*

[1] Selcuk University, Engineering Faculty, Department of Civil Engineering, Konya 42075, Turkey

Abstract

In the present study, the application of Support Vector Machine (SVM), Artificial Neural Network (ANN) and Adaptive Network Based Fuzzy Inference System (ANFIS) models were investigated to model the monthly streamflows. The monthly streamflow data of Bayburt Station operated by EIE on the Çoruh River and located in Çoruh Basin of Turkey were used in the study. The data sample consisted of 59 years of streamflow records in the period of 1942-2000. In the first part of the study, the reliability of SVM, ANN and ANFIS models were determined based on the performance criteria such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Efficiency Coefficient (EC) and Determination Coefficient (R^2). The results verify that the improved SVM model is of better than the other models.

Key words

Streamflow; Çoruh River; Artificial Neural Networks; Support Vector Machine; ANFIS

To cite this paper: *Buyukyildiz, M. (2014). Monthly streamflow time series modelling of Coruh River, In conference proceedings of People, Buildings and Environment 2014, an international scientific conference, Kroměříž, Czech Republic, pp. 517-524, ISSN: 1805-6784.*

*Corresponding author: Tel.: +90-332-223-2139, Fax: +90-332-241-0635
E-mail address: meralbyildiz@selcuk.edu.tr

1 INTRODUCTION

Correct planning and operation of water resources systems is very important in terms of efficient use of decreasing water resources [1]. Flow models established to estimate river flows using available data are of great importance in terms of water resources project work. Recently, strategic works in water resources management as well as the existence of global climate change have led to a further increase in the interest to the said flow estimation models. Although flow models have a very important place in hydrological studies, it is one of the areas where work has continued today. Estimation of flow to be seen on a certain date in the future is important in terms of making flood warnings, operation of reservoirs for the purpose of flood control, determination of water potential of a river, planning hydroelectric power generation during dry periods, distribution of municipal water and irrigation water, and transport in rivers [2].

Several methods, including *Artificial Neural Networks (ANN), Support Vector Machines (SVM,) Adaptive Network-Based Fuzzy Inference System (ANFIS)* have been widely used by many researchers in estimation of many hydrological and meteorological parameters as well as river flow. The results obtained using these models were compared with the values measured and the results were found to have high degree of accuracy.

Kisi et al [3] used to compare the accuracy of adaptive neuro fuzzy inference system (ANFIS), artificial neural networks (ANNs) and support vector machine (SVM) for forecasting daily intermittent streamflow data from two stations, Uzunköpru and Babaeski in Turkey. They also compared the results of them with results of the local linear regression (LLR) and the dynamic local linear regression (DLLR).

Cigizoglu [4] investigated the availability of artificial neural networks (ANNs) to forecasting, estimation and extrapolation of the daily flow data belonging four river flow stations on the rivers Göksu, Lamas and Ermenek in the East Mediterranean region of Turkey. The ANN results are compared with regression models and AR models.

Lin et al [5] studied a SVM combined with the SCE-UA algorithm for predicting monthly river flow series in Manwan Hydropower Scheme. Their results are compared with the ARMA and ANN models.

Chang and Chen [6] investigated a fuzzy-neural network model to predict real time streamflow. They compared the accuracy of their results with the autoregressive moving average exogenous variables model (ARMAX).

Kalteh [7] studied the performance of ANN and SVM models coupled with wavelet transform to estimate monthly river flow. He compared the results of the ANN and SVM models are also compared to each other.

In this study, flow data of observation station no. 2304 located in Çoruh River Basin in Turkey was used to estimate flow by several methods, including Radial Basis Neural Network Algorithm (RBNN), Support Vector Regression (SVR), and Adaptive Network Based Fuzzy Inference System (ANFIS). Efforts were made to determine the most suitable model for estimating flow by comparing the results of the models with actual flow data.

2 METHODS

2.1 Radial Based Neural Network (RBNN)

The basic idea in radial basis neural networks (RBNN) is simply to collect a group of radial basis functions by weighting them so that they approximate to the desired f function [7]. Unlike conventional ANN structures, in the case of RBNNs, radial based activation functions and a non-linear cluster analysis are used in the transition from the input layer to the hidden layer.

The structure between hidden layer and output layer continues its functioning as in the case of other types of ANN, and the actual training is performed there. The activation functions of the neurons in the hidden layer are determined by a C_j center and σ_j bandwidth.

Activation function is a Gaussian curve defined by the equation:

$$\varphi_j(X) = \exp\left(-\frac{\left\|X - C_j\right\|}{2\sigma_j^2}\right) \tag{1}$$

The general equation for the output of neuron j in the output layer is as follows :

$$s_j(X) = \sum_{i=1}^{k} w_{ij}\varphi_i(X) + b_j \tag{2}$$

where w_{ij} is the weight coefficient between hidden neuron i and the output neuron j, bj is the bias constant [8].

2.2 Support Vector Machines (SVM)

SVM was first developed as linear learning machines but most of the problems encountered in real life are not linear. This requires learning machines to be able to map non-linear data in the input space into linear data in a higher dimensional space. Although SVM method is mostly used in classification, it is also used to solve regression problems. In this study, ε-SVR model was used as SVR model. Detailed definition related with the mathematical expression of SVM is given in the literature [9, 10]. The three factors that affect prediction performance of SVM regression are ε, the insensitive error term, C, the regulatory factor (additional capacity control), type of kernel function and kernel parameter. Selection of kernel function as well as model parameters to be used plays an important role in determining SVR performance. The Radial Basis Kernel Function is used for SVR model in this study and denoted as

$$K(x_i, x) = \exp(-\gamma\left\|x_i - x_j\right\|^2) \tag{3}$$

where γ is the parameter of radial basis kernel function which gives the width of the kernel.

2.3 Adaptive Network-Based Fuzzy Inference System (ANFIS)

Adaptive-network based fuzzy inference system is a hybrid intelligence method that utilizes parallel calculation and learning ability of artificial neural networks and inference feature of fuzzy logic. Developed by Jang [11] in 1993, ANFIS model uses Sugeno type fuzzy inference system and hybrid learning algorithm. Learning algorithm of ANFIS is a hybrid learning algorithm consisting of combined use of the least squares method and backpropagation learning algorithm. ANFIS consists of six layers. The functioning of layers in ANFIS structure are the input layer, the fuzzification layer, the rule layer, the normalization layer, the defuzzification layer and overall layer in the order given.

3 CASE STUDY

Çoruh River Basin, located in northeastern Turkey, comprises 19,748 km^2 of drainage area and has an average annual flow of 6.3 billion m^3. The rivers of the basin consist of Çoruh River and its tributaries. The total length of Çoruh River is 431 km and it is the fastest flowing river in Turkey. In this study, flow data of Bayburt flow observation station no. 2304, located on Çoruh River in Çoruh River Basin was used (Figure 1). Flow data was 708 monthly average flow data between 1942 and 2000.

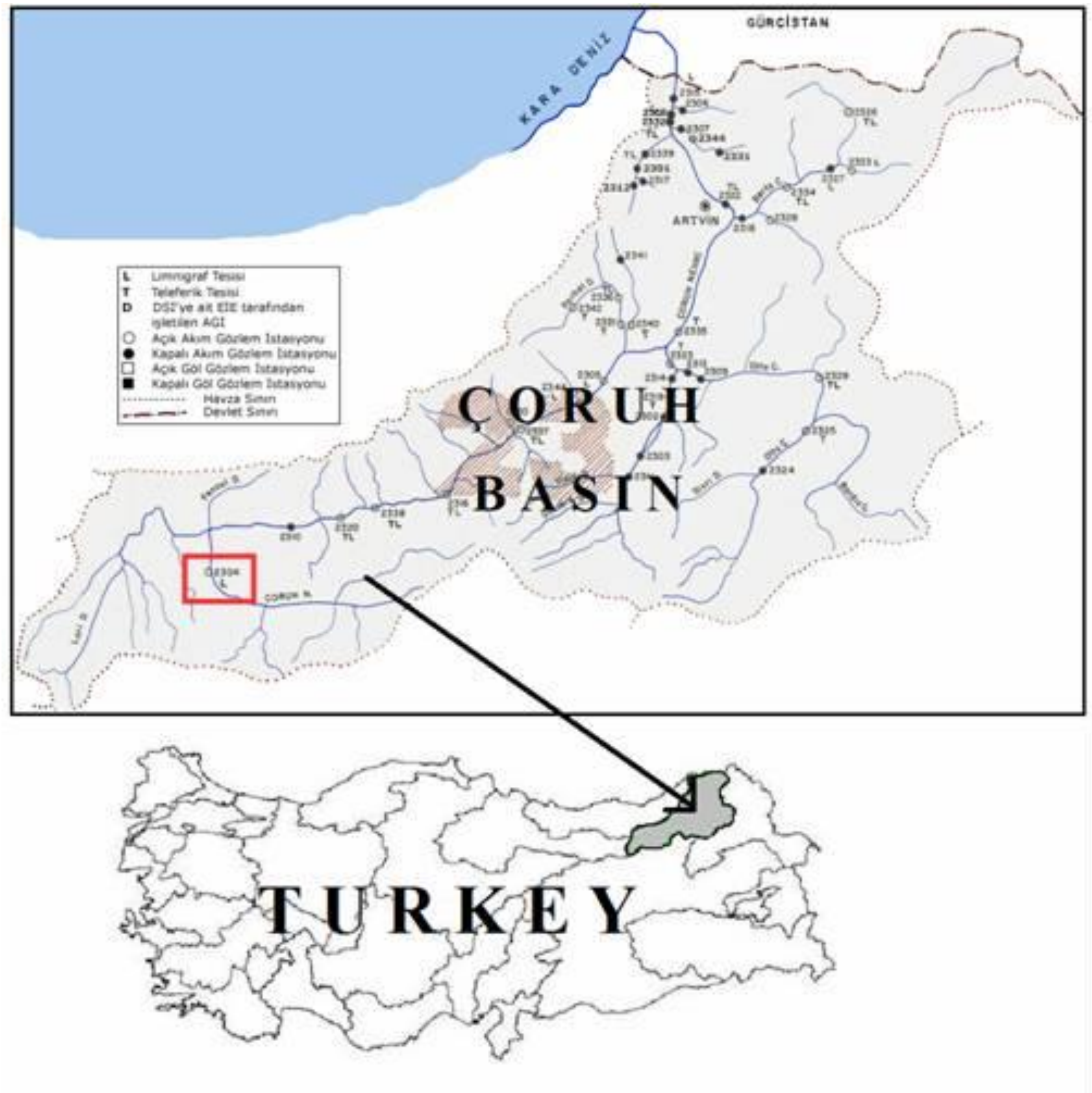

Fig. 1: Location of the study area and Bayburt Station (No:2304)

4 APPLICATION AND RESULTS

In this study, 708 monthly average flow data of Bayburt observation station no. 2304 on Çoruh River between 1942 and 2000 was used to estimate flow by RBNN, SVR, and ANFIS methods. In order to estimate flow, Q_t was obtained by the models comprised of 4 different inputs created by time backpropagation.

528 of 708 monthly average flow data were used as training data and the rest as testing data. Root mean square error (RMSE), mean absolute error (MAE), efficiency coefficient (EC) and determination coefficient (R^2) were used to evaluate the success of the models. Before applying RBNN and SVR methods to these flow data, the observed data were normalized between 0 and 1 by the following equation.

$$x_{norm} = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \tag{4}$$

where X_{norm}, $X_{i,}$ X_{min} and X_{max} represent normalized, observed, minimum and maximum values, respectively.

In order to find the most successful RBNN model, the number of neurons in the hidden layer was tested with increments of 0.01 between 2 and 20, while the spread parameter (σ) with increments of 0.01 between 0.01 and 5. Performance criteria of test data for the most successful RBNN models obtained for 4 different inputs are given in Table 1.

Tab. 1. The MAE, RMSE, R^2 and EC statistics of RBNN in testing period

Model Inputs		Model Structure	MAE	RMSE	R^2	EC
M1	Q(t-1)	RBNN (1-20-0.13-1)	8.802	14.166	0.362	0.360
M2	Q(t-1) Q(t-2)	RBNN (2-20-1.58-1)	5.948	9.960	0.684	0.684
M3	Q(t-1) Q(t-2) Q(t-3)	RBNN (3-20-0.41-1)	6.003	9.979	0.684	0.682
M4	**Q(t-1) Q(t-2) Q(t-3) Q(t-4)**	**RBNN (4-20-0.22-1)**	**5.929**	**9.923**	**0.687**	**0.686**

In view of Table 2, the most successful RBNN model in terms of all performance criteria was RBNN-M4 (4-20-0.22-1) model in M4 model, in which 4 variables were used. In this model, numbers in parenthesis represent the number of inputs, the number of neurons in the hidden layer, spread parameter and the number of neurons in the output layer, respectively.

While creating models using ε-SVR method, Radial Basis Kernel Function was used as kernel function. Model parameters for ε-SVR method are error term (ε), regulatory factor (C) and γ parameter of Radial Basis Kernel function. The most successful ε-SVR models were created by an increment of 0.01 in the range of [0.01 0.5] for ε, an increment of 1 in the range of [1 100] for C, and an increment of 0.1 in the range of [0.1 8] for γ parameter (Table 2). The most successful model among the ε-SVR models created was ε-SVR (56, 0.05, 3.5) model obtained with inputs in model M4.

Tab.2: The C, ε, γ and performance criteria values for ε-SVR models in testing period

Model Inputs		Model Structure ε-SVR (C, ε, γ)	MAE	RMSE	R^2	EC
M1	Q(t-1)	ε-SVR (100, 0.05, 5)	8.638	14.622	0.349	0.347
M2	Q(t-1) Q(t-2)	ε-SVR (73, 0.05, 4.4)	6.284	10.316	0.661	0.660
M3	Q(t-1) Q(t-2) Q(t-3)	ε-SVR (15, 0.05, 7)	6.448	10.416	0.654	0.652
M4	**Q(t-1) Q(t-2) Q(t-3) Q(t-4)**	**ε-SVR (56, 0.05, 3.5)**	**6.242**	**10.316**	**0.662**	**0.660**

For the purpose of estimating mean amount of monthly flow, hybrid learning algorithm was used to create ANFIS models, while Grid Partition (ANFIS-GP) techniques were used to set the rules. Gaussmf was selected as input membership function type, while linear was selected as output membership function type. Values in the range of 2 to 4 were tried as the number of membership functions. The number of epochs was taken as 150 in ANFIS models established. The results obtained for ANFIS models are given in Table 3. Among the ANFIS models created, unlike RBNN and ε-SVR models, the most successful ANFIS model was obtained in the case of M3 model, in which 3 input variables were used. In this model, the number of membership functions was obtained as 2 among 3 input functions.

Tab. 3: The characteristics and performance criteria of ANFIS models in testing period

Model	**Model Inputs**		**Number of MFs**	**MAE**	**RMSE**	**R^2**	**EC**
ANFIS-GP	M1	Q(t-1)	3	8.752	14.447	**0.356**	**0.354**
	M2	Q(t-1) Q(t-2)	2 2	6.928	10.824	0.613	0.610
	M3	**Q(t-1) Q(t-2) Q(t-3)**	**2 2 2**	**6.652**	**10.648**	**0.638**	**0.637**
	M4	Q(t-1) Q(t-2) Q(t-3) Q(t-4)	2 2 2	7.274	11.128	0.588	0.587

Among the models created using RBNN, SVR and ANFIS methods used to estimate flow, when the most successful models were compared by test data, RBNN (4-20-0.22-1) model with a determination coefficient of R^2=0.687 was the most successful model in estimating flow. Scatter diagram and time series of test data of this model are given in Figure 2 and Figure 3, respectively.

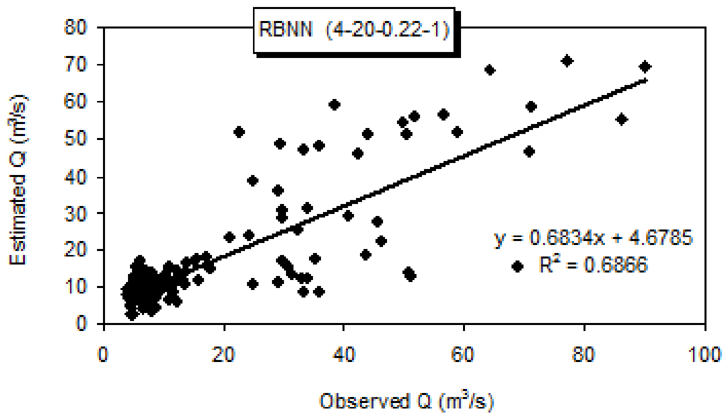

Fig. 2: Scatter diagram of estimated and observed monthly mean flow for RBNN (4-20-0.22-1) in test period

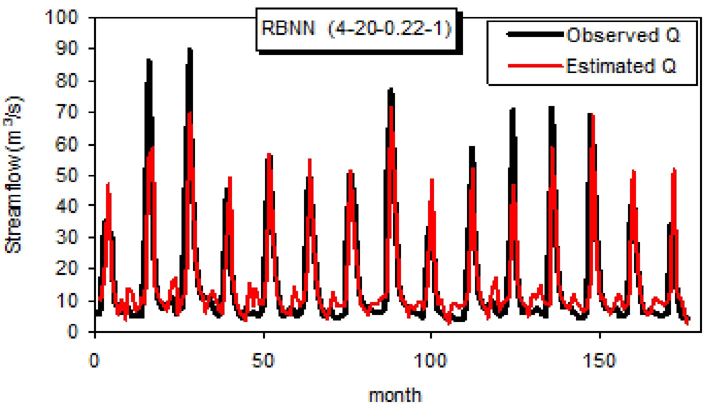

Fig.3: Time series of estimated and observed monthly mean flow for RBNN (4-20-0.22-1) in test period

5 CONCLUSION

In this study, RBNN, SVR and ANFIS models were developed to estimate flow of Bayburt observation station on Çoruh River. With Q_t representing monthly average flow value of this month, 4 different input combinations, including 1) Q_{t-1}; 2) Q_{t-1}, Q_{t-2}; 3) Q_{t-1}, Q_{t-2}, Q_{t-3}; 4) Q_{t-1}, Q_{t-2}, $Q_{t-3,}$ Q_{t-4} were tested. In the case of ANFIS method, the most successful model was the model which included the inputs created by flow values at time Q_{t-1}, Q_{t-2}, Q_{t-3} , while in the case of SVR and RBNN methods, the most successful model was the model which included the inputs created by flow values at time Q_{t-1}, Q_{t-2}, $Q_{t-3,}$ Q_{t-4} . When the most successful models of the three methods used were compared, RBNN (4-20-0.22-1) model with R^2=0.687 was the most successful model among other models. The most successful model in RBNN method was the one in the case of which flow values at time (t-1), (t-2), (t-3), (t-4) were taken as input, and flow at time t was estimated. Therefore, in RBNN (4-20-0.22-1) model, 4 represents the number of cells in the input layer; 20, the number of neurons in the hidden layer; 0.22, the spread parameter and 1, the number of cells in the output layer. In conclusion, the value of R^2=0.686 obtained in the case of the most successful model can be evaluated as usable for the estimation of flow, even though it is not very successful. Possibility of creating more successful models for the estimation of flow using flow values in combination with precipitation values can be investigated as the subject of another study. By this study, it was demonstrated that RBNN method can be used to estimate flow from flow on Çoruh River and to complete missing data or in cases where a measurement cannot be made.

ACKNOWLEDGEMENT

This study has been funded By Selcuk University Scientific Research Projects Department.

REFERENCES

[1] Önöz, B. and Albostan, A. (2007). Hidroelektrik Santral Planlama ve İşletmesinde Yüksek Akımların Mevsimselliğinin Belirlenmesi. *TMMOB Türkiye VI. Enerji Sempozyumu – Küresel Enerji Politikaları ve Türkiye Gerçeği*, Ankara, pp. 196-206 (in Turkish).

[2] Bayazıt, M. (1998). *Hidrolojik Modeller*. İTÜ İnşaat Fakültesi Matbaası. 228 p. İstanbul (in Turkish).

[3] Kisi, O. Nia, A.M. Gosheh, M.G. Tajabadi, M.R.J. Ahmadi, A. (2012). Intermittent streamflow forecasting by using several data driven techniques. *Water Resour Manag*, **26**(2), pp. 457–47.

[4] Cigizoglu, H.K. (2003). Estimation, forecasting and extrapolation of flow data by artificial neural networks. *Hydrological Science Journal*, **48**(3), pp. 349–361.

[5] Lin, J.Y. Cheng, C.T. Chau, K.W. (2006). Using support vector machines for long-term discharge prediction. *Hydrological Sciences Journal*, **51**(4), pp. 599–612.

[6] Chang, F-J. and Chen Y.C. (2001). A counterpropagation fuzzy-neural network modeling approach to real time streamflow prediction. *J Hydrol*, **245**, pp.153–164.

[7] Kalteh, A.M. (2013) Monthly river flow forecasting using artificial neural network and support vector regression models coupled with wavelet transform. *Computers & Geosciences*, **54**, pp. 1–8

[8] Verleysen, M. Hlavackova, K. (1994). *An optimized rbf network for approximation of functions, roceedings european symposium on artificial neural networks*, Brussels, Belgium, pp. 175-180.

[9] Paredes, V. and Vidal, E. (2000). A class-dependent weighted dissimilarity measure for nearest neighbor classification problems. *Pattern Recognıtıon Letters*, **21**, pp. 1027-1036.

[9] Vapnik, V. (1995) The Nature of Statistical Learning Theory. *Springer Verlag*. New York, USA.

[10] Cristianini, N. Shaw, J. Taylor, J. (2000) *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge University Press: New York.

[11] Jang, J.S.R. (1993). ANFIS Adaptive-Network-Based-Fuzzy Inference Systems. *IEEE Transactions on Systems, Man, and Cybernetics*, **23**(3), pp. 665-685.