Prof. Dr. Abdüsselam ULUÇAM Armağanı

# Estimating the Long Term Average Flow Rates of Tigris Basin Using Machine Learning Methods

**Abdullah MURATOGLU[1] , Mehmet Ishak YUCE[2]**

[1]Batman University, Civil Engineering Department, 72100 Batman, Turkey
[2]University of Gaziantep, Civil Engineering Department, 27310 Gaziantep, Turkey
*Corresponding author: email: abdullah.muratoglu@batman.edu.tr

**ABSTRACT**

The discharge of a river is one of the most important parameters of the hydraulic and hydrological studies such as hydropower production, canal design, irrigation, basin management. Each basin has different climate and geological characteristics which influence the regional infiltration capacity and runoff. The aim of this study is to estimate the average annual flow rates of ungauged locations on the Tigris River Basin. In total, eleven machine learning methods were applied to the long-term average annual discharge and the drainage area data of 34 flow measurement stations (FMS). Among all methods employed here, the conventional regression analysis was found to be the most successful test with a correlation coefficient ($R^2$ value) of 0.96. The equation of the best fitted linear line represents the relationship between the drainage area and the discharge. The results of this study are expected to enable the prediction of the average annual flow rate of any sub-basin of the Tigris River.

**Key words:** Average annual discharge, the Tigris River, regression analysis, machine learning methods, ungauged basin.

**ÖZET**

Herhangi bir nehrin akım parametrelerinin bilinmesi, enerji üretimi, kanal tasarımı, sulama, havza planlama projeleri ve diğer hidrolik ve hidrolojik çalışmalarda hayati önem arz etmektedir. Her havzanın kendine has iklim durumu, yağış koşulları ve zemin yapısı olduğu için sızma (infiltrasyon) ve akış parametreleri de haliyle farklı olmaktadır. Bu çalışmanın amacı, Dicle havzası üzerinde, akım ölçüm istasyonu bulunmayan alt havzaların yıllık ortalama debi değerlerinin, yağış alanının bir fonksiyonu olarak belirlenmesidir. Bu kapsamda, uzun dönem yıllık ortalama akım ve yağış alanı verileri bulunan 34 adet akım ölçüm istasyonuna (FMS) 11 adet yapay öğrenme metodu uygulanmıştır. Klasik regresyon analizi, 0,96 korelasyon değeri ($R^2$) ile en başarılı test olarak elde edilmiş ve havza için debi ve yağış alanı arasındaki ilişkiyi gösteren lineer bir denklem türetilmiştir. Bu çalışma ile Dicle nehrinin alt havzalarında ve akım ölçüm istasyonu bulunmayan yerlerde yıllık ortalama debi tahmini yapılabilecektir.

**Anahtar kelimeler:** Ortalama debi, Dicle nehri, regresyon analizi, otomatik öğrenme, akım ölçüm istasyonu.

## Introduction

The Tigris River is the second biggest river of the Western Asia. Turkey, Iraq, Iran and Syria are the neighbor countries of the river [1]. The part of the basin located within the borders of Turkey has an area of 41058 km$^2$ which corresponds to 12 % of the whole basin [2]. Whereas, about 51 % of the total river discharge flows through Turkey with an average annual value of 594 m$^3$/s [3,4]. In this study, the part of the basin situated in Turkey is studied. The basin is located in the south-eastern part of Turkey with west (the Tigris) and east (Zab) tributaries. The western part of the Tigris Basin is illustrated in Figure 1. Measuring the flow rate of a river at every single point for long years is technically not applicable due to high costs, operation difficulties and other site-specific reasons and also it is not necessary. Those sites where flow measuring stations are not positioned are called as ungauged sites. Prediction of the surface runoff at ungauged sites is important for hydraulic and hydrologic works. The procedure of estimating stream flow at ungauged sites on the Tigris River are examined.

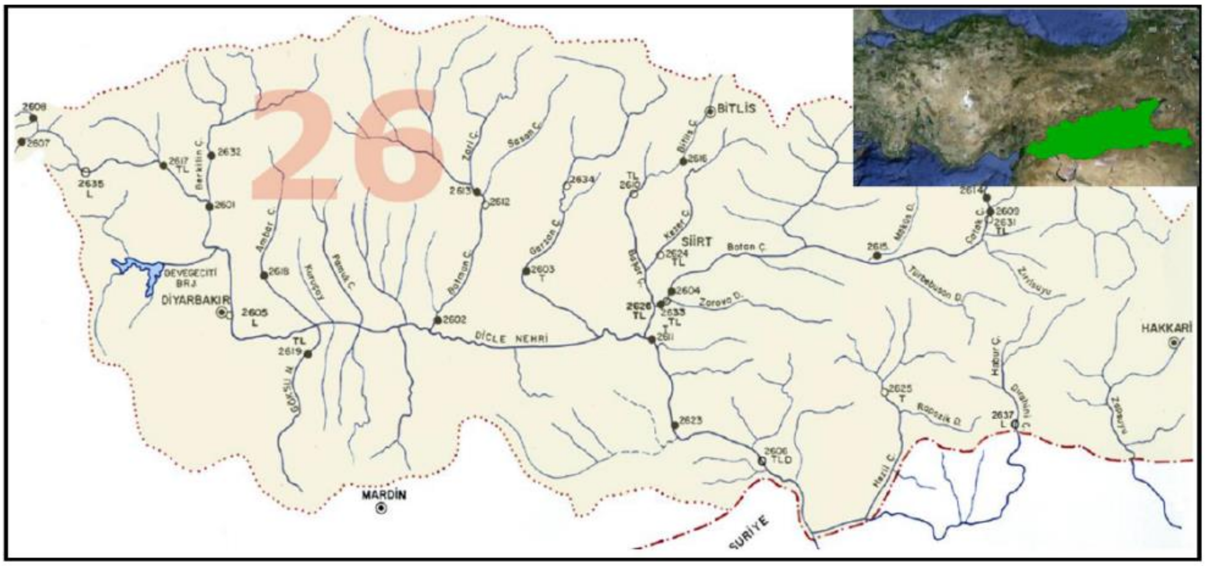



**Figure 1.** The western part of the Tigris basin [5], Turkey (Muratoglu and Yuce, 2012)

The fluctuation of the discharge of a river in time and space is a function of the geologic and climatic conditions of the river basin. Therefore, it can be said that, the long-term mean annual discharge is a function of mean annual precipitation and the hydrogeological characteristics of the basin. Because of several parameters which are not directly measurable should be involved in a complete prediction model, a simpler model is chosen. One of the oldest methods which delivers such a relationship was introduced by Kuichling (1889) [6]. Kuichling proposed an equation which shows the relationship between the stream flow and the drainage area.

Basically, the input and output of the hydrologic basins are analyzed with two approaches. These are parametric and non-parametric approaches. Parametric methods consider detailed basin parameters such as, precipitation, run off, topography, geological conditions, slopes, etc. and applied to smaller basins or sub basins.  On the other hand, non-parametric methods provide a black-box approach embedding the detailed parameters into the function as coefficient and exponents. Therefore, non-parametric techniques are very suitable to understand preliminary basin characteristics with a minimum amount of information and data. Latest studies show that the machine learning methods can be successfully applied to obtain basin-precipitation and discharge relationship with a black-box approach [7-9].

The main purpose of this study is to find a relation between the drainage area and the long-term annual average discharge (mean annual flow rate) of any sub-basin of the Tigris River. In total, eleven machine learning methods were applied to the mean annual discharge and the drainage areas of 34 flow measurement stations (FMS). The conventional regression analysis was found to be the most successful test with a correlation coefficient, $R^2$, value of 0.96.

## Material and Method

The stream flow and the drainage area data was obtained from the State Hydraulic Works (DSI) of Turkey. The mean annual stream flows were achieved by employing monthly average discharges. The flow measurement stations used in this study are presented in Table 1.

**Table 1.** The list of flow measuring stations and related information

| FMS | Stream Name | Station Name | Elevation (m) | Drainage area (km$^2$) | Mean annual discharge (m$^3$/s) |
|---|---|---|---|---|---|
| 2602 | Batman Stream | Sinan | 518 | 4988.4 | 146 |
| 2603 | Garzan Stream | Beşiri | 545 | 2450.4 | 48.7 |
| 2604 | Botan Stream | Billoris | 465 | 8747.3 | 143 |
| 2605 | Tigris River | Diyarbakır | 570 | 5655.2 | 70.3 |
| 2606 | Tigris River | Cizre | 370 | 38280.7 | 534 |
| 2607 | Behramaz Stream | Hatunköy | 1075 | 108.4 | 2.55 |
| 2609 | Çatak Creek | Çatak | 1625 | 2339.5 | 27.5 |
| 2610 | Bitlis Creek | Baykan | 698 | 636.5 | 18.7 |
| 2611 | Tigris River | Rezuk | 427 | 34493.1 | 420 |
| 2612 | Batman Creek | Malabadi | 597 | 4105.2 | 123 |
| 2613 | Batman Creek | Hüseyinkan | 650 | 3427.6 | 60.7 |
| 2614 | Sortkin Stream | Çatak | 1615 | 426.4 | 4.67 |
| 2615 | Müküs Creek | Beğendik | 1250 | 505.6 | 19.1 |
| 2616 | Bitlis Stream | Karınca | 1145 | 346.4 | 12.1 |
| 2617 | Tigris River | Çayönü | 695 | 1186 | 24.3 |
| 2618 | Ambar Steam | Köprübaşı | 595 | 976 | 7.68 |
| 2619 | Göksu Stream | Çınarköprü | 657 | 667.8 | 2.59 |
| 2620 | Zap Creek | Üzümcü | 1072 | 5270.3 | 57.3 |
| 2621 | Zap Creek | Musahan | 1725 | 2504.4 | 13 |
| 2622 | Nehil Creek | Konak | 1694 | 1136 | 19.1 |
| 2624 | Kezer Creek | Pınarca | 530 | 1169.6 | 20.2 |
| 2625 | Nezil Creek | Girikhan | 780 | 1127.2 | 17.7 |

| | | | | | |
|---|---|---|---|---|---|
| 2626 | Botan Creek | Billoris | 457 | 8761.2 | 157 |
| 2627 | Zap Creek | Narlı | 775 | 6771.9 | 107 |
| 2628 | Cemilkatli Creek | Kamışlı | 1620 | 290 | 8.1 |
| 2629 | Semdinli Creek | Yeşilöz | 1627 | 290 | 9.45 |
| 2630 | Zap Creek | Teknisyenler | 1412 | 4172 | 37.5 |
| 2631 | Catak Creek | Tüliran | 1482 | 2455 | 26.6 |
| 2632 | Berkilin Creek | Çayüstü | 689 | 1503.6 | 28.2 |
| 2633 | Botan Creek | Billoris | 465 | 8747.3 | 133 |
| 2634 | Garzan Creek | Köprübaşı | 630 | 1407.7 | 38.1 |
| 2635 | Tigris River | Kalender Köp. | 843 | 388 | 6.53 |
| 2636 | Semdinli Creek | Şemdinli | 1290 | 312.5 | 10.6 |
| 2637 | Habur Creek | Habur II | 935 | 1217.1 | 33.5 |

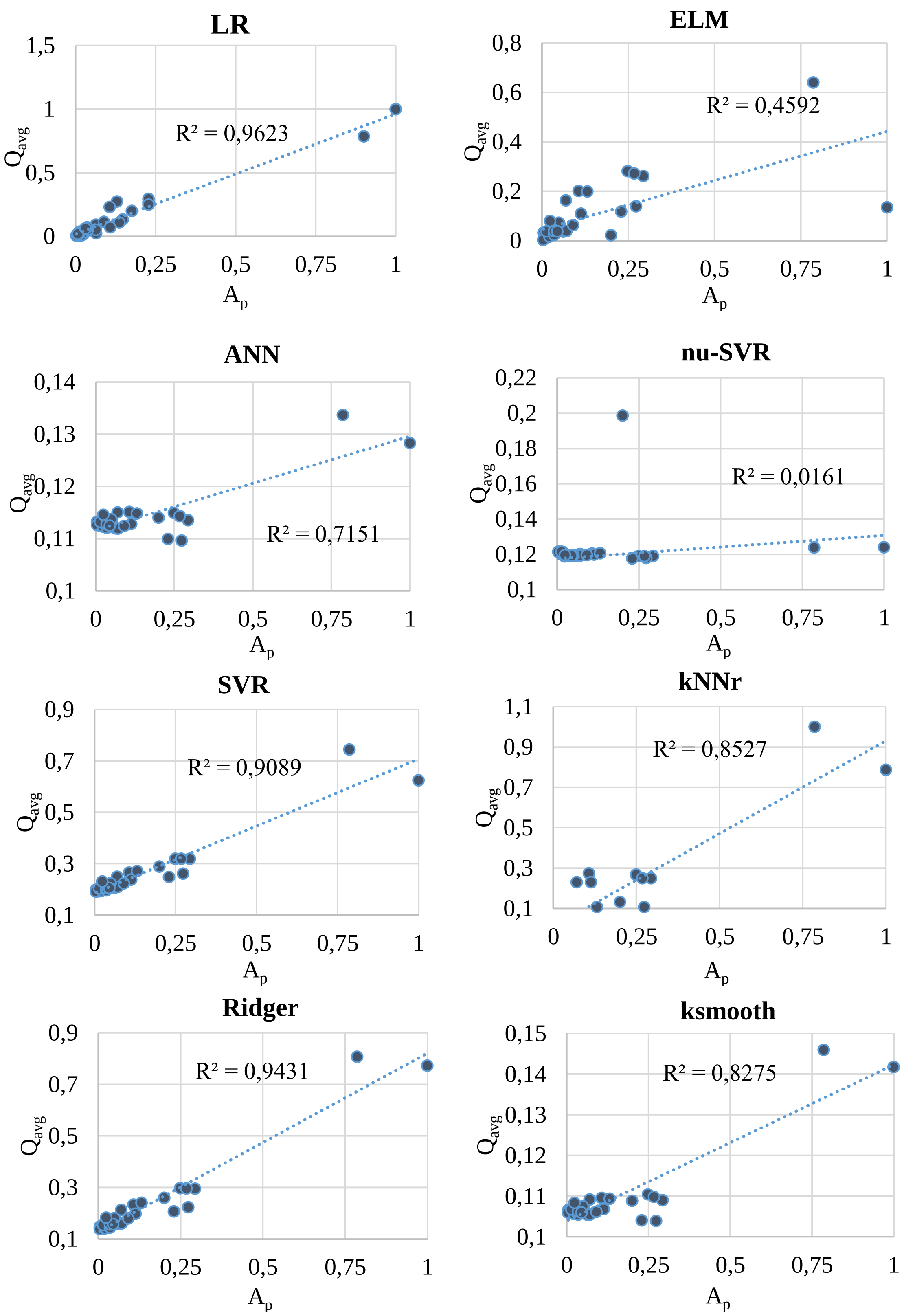
Machine learning methods are defined as the set of procedures in which the pattern of the data is automatically detected and employed to predict the future data or the other decision making procedures [10]. Various machine learning methods have been proposed for different purposes. In this study, the extreme learning machine (ELM), artificial neural networks (ANN), support vector regression (SVR), nu-support vector regression (nu-SVR), k nearest neighbor regression (kNNr), Ridge regression (Ridger), Kernel smoother (ksmooth), Pseudo-inverse regression (PINVR), partial least squares regression (PLSR) and Gaussian process regression (Gaussian-R) and the simple linear regression (LR) methods have been used.

ELM is used for the single hidden layer feedforward neural networks. The method overcomes the slow training speed and over-fitting problems, compared with the conventional methods [11]. ANN methods are non-linear mapping structures which mimics the human brain [12]. SVR is a regression based machine learning method. It is very effective in high dimensional spaces. k-NN is a non-parametric method which is used for regression and classification. The method is described as lazy learning methods and it is the simplest machine learning method. The close neighbors provide more contribution than the other data. Ridger method is used to analyze multiple regression data which suffer from the multicollinearity [13]. Ksmooth technique is generally used to estimate a real valued function using noisy observations [14]. PINVR method solves some of the problems which encountered at linear regression techniques. Gaussian-R method is depend on Bayesian probability model which assumes that the random variables are normally distributed [15]. Consequently, regression is the technique based on the relationship between two quantitative variables in which the value of the dependent variable can be predicted via independent variable [16]. If the relationship can be stated as a straight line and if there is a unique independent variable, then, the regression is called simple linear regression [17]. Regression is the technique based on the relationship between two quantitative variables in which the value of the dependent variable can be predicted via independent variable [16]. If the relationship can be stated as a straight line and if there is a unique independent variable, then, the regression is called simple regression [17].

**Results and Discussion**

The machine learning methods were applied to the mean annual discharge and the drainage area data of 34 flow measurement stations (FMS). The classical regression analysis was found to be the most successful test with a coefficient of determination, $R^2$ value of 0.96, among the methods utilized here. Although, there should be more parameters which will affect the predictions, the $R^2$ is still consistent and the theoretical flow rates were estimated by approximating to the measured discharges with a consistency of 96 %. For computational simplicity, the machine learning tests were performed on the normalized data. The relationship between the annual average discharge values and the drainage areas of the sub-basins of Tigris River Basin are presented in Figure 2. The coefficient of determination, $R^2$ indicates the fitness of the data to the model, the methods utilized and the $R^2$ values achieved are given in Table 2. The simple linear regression analysis was found to generate the highest performance. ELM and nu-SVR methods were failed with extremely low coefficients of determination. Remaining tests except from ANN represented reasonable results. Equation (1) represents the linear relationship between the drainage area and the annual average stream flow, which was obtained from the linear regression.

$$Q_{avg} = 0.0132\,(A_p) + 9.4647 \qquad (1)$$


LR
R² = 0,9623
ELM
R² = 0,4592
ANN
R² = 0,7151
nu-SVR
R² = 0,0161
SVR
R² = 0,9089
kNNr
R² = 0,8527
Ridger
R² = 0,9431
ksmooth
R² = 0,8275
Qavg
Ap

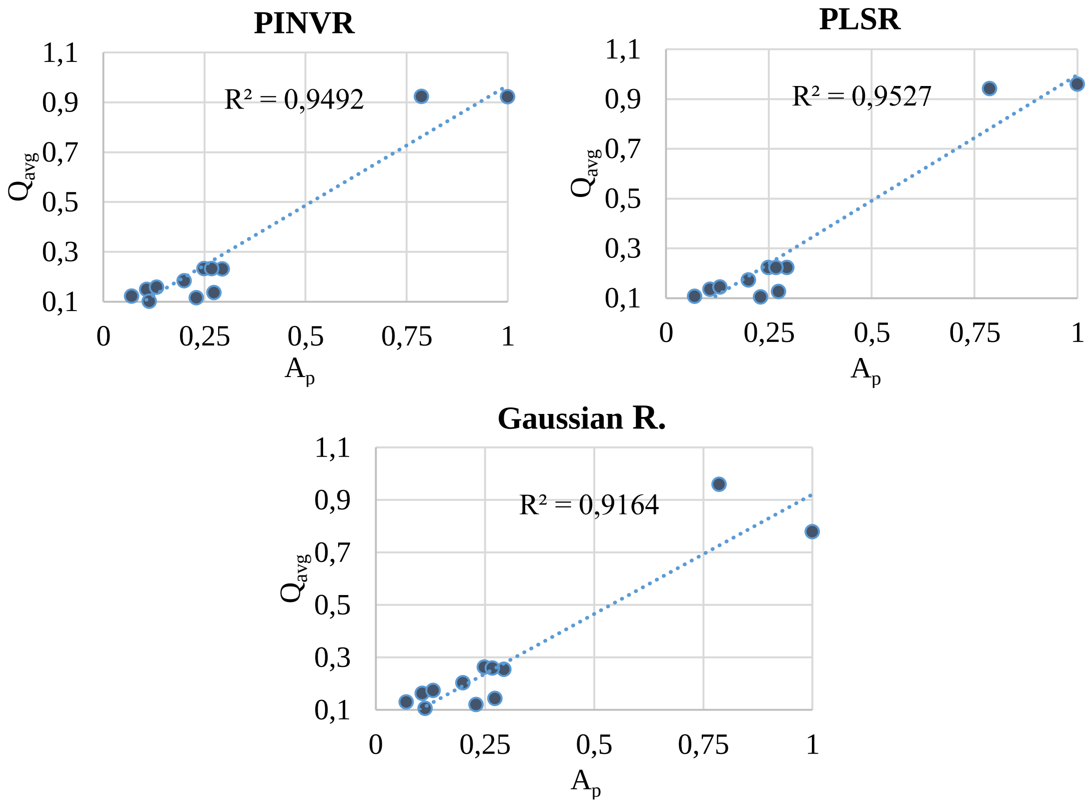



**Figure 2.** The linear relationships between the drainage area and the annual average stream flow data of Tigris River Basin for various machine learning methods

**Table 2.** Machine learning methods and coefficients of determination

| Abbrev. | Method name | $R^2$ |
|---|---|---|
| LR | Linear regression | 0.96 |
| ELM | Extreme learning machine | 0.17 |
| ANN | Artifical neural networks | 0.71 |
| nu-SVR | nu-Support Vector Regression | 0.02 |
| SVR | Support Vector Regression | 0.91 |
| kNNr | k nearest neighbor regression | 0.85 |
| Ridger | Ridge regression | 0.94 |
| ksmooth | Kernel smoother | 0.83 |
| PINVR | Pseudo-inverse regression | 0.95 |
| PLSR | Partial least squares regression | 0.95 |
| Gaussian R. | Gaussian Process regression | 0.92 |

The model introduced in this work (Eq. 1), is a function of the drainage area. Therefore, the model offers a very rough approximation of the annual averaged discharge values. In real, the amount of the stream flow depends on the meteorological, geomorphological and most importantly, on the hydrological characteristics of a basin [18]. These parameters are tabulated in Table 3. Consequently, for a more accurate model, all these parameters should be included into the model as independent variables.

**Table 3.** The characteristics of the basin effecting the streamflow

| Meteorological | Geomorphological | Hydrological |
|---|---|---|
| Temperature | Area | Stream shape |
| Pressure | Shape | Infiltration capacity |
| Humidity | Slope | Soil condition |
| Wind | | Vegetal cover |
| Solar radiation | | Groundwater |
| | | Storage and seepage |

Increasing the number of parameters in the model requires intensive amount of data and study. Lack of data is one of the most important problems encountered in the basin related studies in most countries. Even though this study reflects a very rough approximation, where the only independent variable is the drainage area, the stream flow at any ungauged site of any tributary of Tigris River can roughly be estimated by only considering the area of the sub-basin. It should be stated that, this model can be used only for the Tigris River Catchment or the neighboring basins having similar hydrological characteristics. It is obvious that, the model will deliver impractical results for other catchments and specific models should be developed for different basins.

## Conclusions

A simple and general model was developed to predict the annual average discharge at ungauged sites on the Tigris River and its tributaries by employing the corresponding area of the sub-basin. In total, eleven machine learning methods were applied to the long-term average annual discharge and the drainage areas of 34 flow measurement stations (FMS). Among all methods employed here, the classical regression analysis was found to be the most successful test with an $R^2$ value of 0.96. The equation of the best fitted linear line represents the relationship between the drainage area and the discharge. The results of this study expected to enable the prediction of the average annual flow rate of any sub-basin of the Tigris River. It is expected to guide engineers in planning and designing of hydropower plants, canal design, irrigation structures, basin management and other hydraulic and hydrological studies.

**Acknowledgements**

The authors would like to thank to Dr. Ömer Faruk Ertuğrul and Dr. Şahnaz Tiğrek due to their contributions.

## References

[1] UN-ESCWA and BGR (United Nations Economic and Social Commission for Western Asia; Bundesanstalt für Geowissenschaften und Rohstoffe) (2013). Inventory of Shared Water Resources in Western Asia. Beirut.

[2] Zehir C., 2003. Ortadoğuda su medeniyetlerinden su savaşlarına.Su Vakfı Yayınları. İstanbul.

[3] Sehsuvaroğlu, L., 1997. Su barışı-Türkiye ve Ortadoğu su politikaları. Gümüş Motif Yayınları. İstanbul

[4] Muratoglu A., 2011. Assessment of Tigris River Hydropower Potential. MSc thesis, Graduate School of Natural and Applied Sciences, Gaziantep University, Gaziantep.

[5] Muratoglu A, Yuce MI., 2012. Trend analysis of Tigris River basin stream flow data. In: Proceedings of 10$^{th}$ International Congress on Advances in Civil Engineering, 17–19 October 2012, Middle East Technical University, Ankara, Turkey.

[6] Kuichling, E, 1889. The relation between the rainfall and the discharge of sewers in populous areas. *American Society of Civil Engineering*, 20, 1-56.

[7] Gümüş V, Soydan NG, Şimşek O, Aköz MS, Kırkgöz, MS., (2013). Yağış-akış ilişkisinin belirlenmesinde farklı yapay sinir ağı yöntemlerinin karşılaştırılması. *Çukurova Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi*, 28(1), 37-49.

[8] Okan U, Mollamahmutoğlu A., (2010). Yiğitler çayı günlük akımlarının yapay sinir ağları ve regresyon analizi ile modellenmesi. Dumlupınar Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 1302-3055.

[9] Kişi Ö, Shiri J, Tombul M., (2013). Modeling rainfall-runoff process using soft computing techniques. *Computers and Geosciences*, 51, 108-117.

[10] Murphy KP., 2012. Machine learning: a probabilistic perspective. MIT Press. USA.

[11] Ding S, Zhao H, Zhang Y and Ru N., 2013. Extreme learning machine: algorithm, theory and applications. *Artificial Intelligence Review*. 44(1).

[12] Lek S. Guegan JF., 1999. Artificial neural networks as a tool in ecological modelling, an introduction. *Ecological modelling*. 120 (2-3).

[13] NCSS Statistical Software, Chapter 335, Ridge Regression. Available at: http://www.ncss.com. Accesed at: March 2016.

[14] Wikipedia, 2016. Kernel smoother. Available at: https://en.wikipedia.org/. Accessed at: March 2016.

[15] Pasolli, E., Melgani F., 2011. Gaussian Process Regression within an Active learning Scheme, *IGARSS*, pp. 3574-3577

[16] Paulson DS., 2007. Handbook of Regression and Modelling: Applications for the Clinical and Pharmaceutical Industries. Chapman & Hall/CRC. USA.

[17] Voelker D., Orton P. and Adams S., 2001. CliffsQuickReview Statistics. Hungary Minds. New York.

[18] Usul N., 2005. Engineering Hydrology. Metu Press, Ankara.