Research Inventy: International Journal of Engineering And Science
Vol.11, Issue 10 (October 2021), PP 01-07
Issn (e): 2278-4721, Issn (p):2319-6483, www.researchinventy.com

Modeling of Urban Drinking Water with Genetic Expression Programming

Behzat Aslan[1], Fevzi Onen[2*]
[1]*Ilbank, Diyarbakir, Turkey,*
[2]*Department of Civil Engineering, Faculty of Engineering, Dicle University, Sur/Diyarbakır, Turkey,*
[*] *Corresponding: fonen@dicle.edu.tr;*

Abstract: *Total drinking water needs consist of domestic water needs, trade, industry, service sector, tourism, animal water needs, special needs and water losses. Although the estimation methods used in determining the need of drinking water used by the settlements or in the future will be used with the amount of population, it is a fact that there are many parameters affecting the water consumption today. In this study, parameters such as water amount, water loss, accrued, temperature, humidity, population, rainfall, development status and number of water subscribers were used for Diyarbakır City among 2005-2014 years. With these parameters affecting water demand and drinking water consumption, models based on Genetic Expression Programming, which is an artificial intelligence technique, have been developed for the estimation of future drinking water needs. Different models have been developed with Genetic Expression Programming in demand and estimation of drinking water and the effects of parameters affecting consumption have been examined. In addition, by looking at the Determination Coefficient (R^2) and the Mean Squared Error (RMSE), the predicted values obtained from the model were compared with the measured values.*

Keywords: *Drinking Water; Water Demand; Genetic Expression Programming; Water Loses*

Date of Submission: 18-10-2021 Date of Acceptance: 02-11-2021

I. INTRODUCTION

Mankind lives in the environment consisting of three main elements: water, land and air. Water, soil and air cannot live in an environment without living. Throughout history, water has been the most essential element of vital activities on earth. Before the BC In the 1760s Hammurabi Law some rules on the right to water, there were some sanctions related to violations. In Roman law, water was accepted as the common property of all humanity. For centuries, civilizations have been established near water resources. The surroundings of the Nile, Tigris and Euphrates Rivers were the first settlements in the world with the water they provided. In other words, large rivers and water resources in the world are the basis of the formation and development of civilizations. Since ancient times, people have worked hard to be close to water resources and to rule over these resources.

Climate, population crowds, social and economic status of water consumers and environmental factors affect water consumption. Population crowds have a direct impact on the use of water resources. The amount of water use per capita varies considerably due to differences in living standards at any settlement where the population is crowded. In many sectors established for production and development, water is absolutely needed. During the twentieth century, the world population increased threefold compared to the end of the 19th century and water use increased six fold [1]. The fact that the increase in population and water consumption did not follow the same rate shows that water consumption is not only dependent on the population.

When the current studies on water demand and estimation are examined, it is seen that the first studies are done by regression analysis. Howe and Linaweaver [2] developed housing water demand models with parameters arranged by regression analysis from cross-sectional data for many cities and found that the price variable was not relatively elastic in domestic water demand. Cassuto and Ryan [3] estimated housing elasticity in a long-term water conservation program in California's Oakland Region in a regression model where income and cost are used as independent variables. In this model, it is tried to predict the short term usage change in the precipitation and air temperature variables. Billings and Agthe [4] compared the regression analysis and time series space method with the basic average monthly approach to estimate short-term water demand in Arizona's Tucson city. Babel et al. [5] using the regression model, calculated the estimation of domestic water demand based on social and economic characteristics, climatic factors, multivariate econometric factors in which public water policies were defined.

By using Artificial Neural Networks (ANN) and Fuzzy methods, estimations are made for different settlements with different parameters in drinking water demand estimation. Altunkaynak et al. [6] tried to make

Istanbul's Water Consumption Prediction by fuzzy logic method. In this study, it is tried to create the estimated water usage values in the following months by using the previous three month water consumption values. In the estimation of the need for drinking water is inadequate in population impact, besides this parameter, Gross Domestic Product (GDP), temperature, precipitation, humidity, average domestic water bill, households, such as the effect of parameters of the province of Izmir province by determining the effect of drinking water tried to do. Adamowski [7] tried to estimate the daily maximum water consumption of Ottawa, Canada using MLR, time-series analysis and ANN methods. Firat, et al. [8] evaluated artificial eural network techniques for municipal water consumption modeling. Ajbar and Ali [9] estimate the amount of water demand of the City of Mecca using the Artificial Neural Networks Method. Since the demand for drinking water is changing depending on the geographic, climatic and sudden changing population, the city of Mecca reveals a situation where water demand varies considerably both in terms of being dry and with a growing population at certain times of the year based on religious tourism. In this study, water forecasts were made for the city of Mecca monthly and yearly by using household income, settlement density, population density and highest temperature parameters. The parameters used in this model study with Artificial Neural Networks are expressed as follows: q=f(I,H,T,V). Where, q refers to the total annual water use, I annual income, H households, T monthly average temperature and V visitor numbers. Akdağ [10] presented a study on Estimation and Comparative Analysis of Urban Drinking Water Analysis using ANN, Support Vector Machines and Box-Jenkins Methods.

While drinking water projects are carried out in cities, districts and towns, estimation of drinking water is estimated by population estimation methods according to future populations of cities. However, there are many parameters that affect drinking water usage. Many factors ranging from geography to climate, from demographic structure to spatial type, from development to industry can affect drinking water consumption. Considering the effects of all these factors on drinking and using water, the estimation of the need for drinking water should not be evaluated according to the population. Therefore, in this study, data from the climate to the development level and from the water use in the past years to the amount of water losses have been used. Different models have been created with the GEP in the demand and estimation of Diyarbakır City Center and the effects of the parameters affecting the consumption have been examined.

II. MATERIAL AND METHOD

2.1 Properties of drinking water system of Diyarbakır

Nowadays, Diyarbakır drinking water is supplied from two points. One of them is Hamravat drinking water. The other one is the drinking water brought from the Dicle Dam with the transmission line. Population centers of Diyarbakir Metropolitan Municipality in Turkey exceed one million people. The total length of the water supply network is 1201 km. The network, which was built by Iller Bank, was transferred to DİSKİ (Diyarbakır Water and Sewerage Administration) in 2001. Water distribution network consists of 8 different regions. The raw water supplied from the Dicle dam comes to the treatment plant after approximately 35 km long pump line. After these processes are completed, the water transferred to the water tanks is distributed to the grid without pump. Table 1 shows the amount of water, water consumption and water loss occurring in Diyarbakır city center between 2007 and 2015. These losses occur in two ways, physical and administrative. In physical losses, leakages, faults and storage overflows in the network are effective. This type of loss can be exemplified by the fact that the network is old; operating pressure is high, and technically difficult to detect leakage. Administrative losses are referred to as illegal use and non-accrued losses. Examples of this type of loss are lost counters, losses due to fire hydrants, losses due to illegal connections and losses due to errors in data records.

Table1. Diyarbakır city annual water statistics report (DISKI)

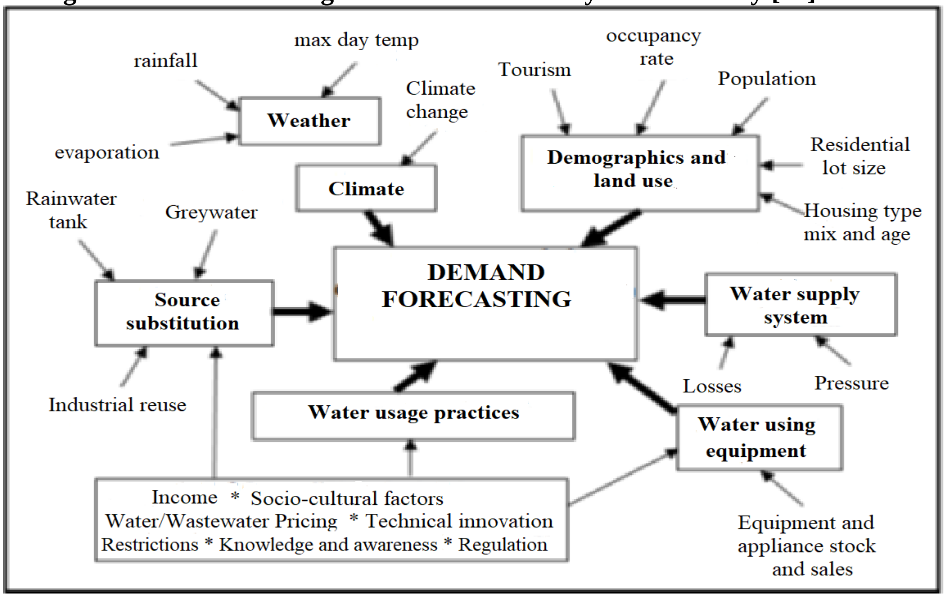
Years	Population	Water Production (m^3/year)	Water Consumption (m^3/year)	Water Loss (m^3/year)	Losses %
2007	717 986	59 625 000	26 721 105	32 903 895	55
2008	733 888	56 582 000	26 654 873	29 927 127	53
2009	768 450	60 175 971	28 781 083	31 394 888	52
2010	777 593	64 018 662	27 273 512	36 745 150	57
2011	806 667	64 635 899	28 795 827	35 840 072	55
2012	822 546	67 403 821	30 094 549	37 309 272	55
2013	841 762	72 224 265	30 586 858	41 637 407	57
2014	872 463	73 978 520	32 311 068	41 667 452	56
2015	887 348	76 381 010	34 269 666	42 111 344	55

2.2 Data used in the Study

Water demand is affected by various parameters such as population, employment, economic cycles, technology, weather, climate, price and so on. The effects of parameters such as local population growth, global warming, and change in the amount of urban green space, industrial development and change in living standards increase the consumption [11].

Numerous parameters affecting water demand directly and indirectly are shown in Fig. The factors include: water usage practices, water using equipment, demographics and land use, climate, water supply system, and source substitution.

Figure1. Factors affecting water demand directly and indirectly [12]



In this study, population, drinking water subscriber number, temperature, precipitation, humidity, water loss in the network, development status of the gross domestic product (GDP), accrual and amount of water supplied to the grid are used for estimation of drinking water requirement of Diyarbakır city. Between the years 2005-2014, population centers values of Diyarbakir in Turkey were obtained from the statistical institute. In this study, subscriber numbers, water usage, water losses and the amount of water supplied to the network were obtained from Diyarbakır Water and Sewerage Administration for Diyarbakır city center between the years 2005-2014. Average temperatures, average rainfall and average relative humidity values were obtained from Diyarbakır Meteorology Region Directorate between the years 2005-2014. In addition, Gross domestic product values were obtained from Turkey Statistical Institute between the years 2005-2014.

The main purpose of this study is to determine the parameters that make up and affect the drinking water consumption of Diyarbakır, to create different models by Genetic Expression Programming (GEP) method. The total estimated amount of drinking water (Q_T) was compared with the measured values R^2 and RMSE were used for the performance of the models.

2.3 Genetic Expression Programing (GEP)

The GEP approach is an algorithm based on genetic programming (GP) and genetic algorithms (GA). Computer program is developed which is encoded in linear chromosomes in fixed length. The GEP approach is a research model that includes computer program as mathematical expression, decision tree, and logical expression. There is a marked difference between GP and GEP approaches. In GP, individual is non-linear entities, introduced in different dimensions and shapes, as a parse trees; whereas the individuals in the GEP are also nonlinear entities in various dimensions and forms, also known as expression trees [13].

Chromosome and expression tree (ET) are the two main elements of GEP. Chromosomes can be composed of one or more germs that represent a mathematical expression. These genetic mathematical codes are specified in two different lines, the language of gene and expression tree (ET), which is called the Karva Language. GEP genes are consisting of two parts, the head and tail name. The head contains some mathematical operators, variables and constants.

In this model the monthly period between 2005-2014, drinking water data (d0) A, population data (d1) P, drinking water subscriber numbers (d2) S, temperature data (d3) T, rainfall data (d4) R, humidity data (d5) H,

water loss data (d6) Q_L, GDP per capita for Diyarbakır (income) (d7) D are shown as input parameters. In the period of 2005-2014, total amount of drinking water supplied to the grid (D.V.) and Q_T is the output parameter.
In the creation of the GEP Model, the equation showing the independent variables affecting drinking water consumption is given below.

$$Q_T = fonk(A, P, S, T, R, H, Q_L, D) \quad (1)$$

where, A Drinking water accrue, P Population, S House subscriber number, T Temperature, R Rainfall, H Moisture, Q_L Water losses, D per capita GDP and Q_T Total drinking water amounts for Diyarbakır.
The development of approaches GEP involves five steps, and the first step is the selection of the fitness function, f_i. For this problem, an individual program i suitability measured by the following expression:

$$f_i = \sum_{j=1}^{c_i} (M - |C_{(i,j)} - T_j|) \quad (2)$$

where M=selection interval; $C_{(i,\ j)}$ =value returned by the individual chromosome i for fitness case j; and T_j=target value for fitness case j. In the second stage, the terminals T_j set and function f group are selected to produce chromosomes. In the problem, the terminal set explicitly accounts for the independent variables, Q_T=f (A, P, S, T, R, H, Q_L, D). Although the selection of the fitness function set is not so clear, a good prediction can be made to include all necessary functions. In this case, the four main operators (+, -, *, /) and some basic mathematical expressions (1/x, $x^{1/2}$, $x^{1/3}$, x^2, x^3) are utilized. Third stage is to structure the chromosome architecture, the head size and the number of genes. The fourth main step is to select the link function. Finally, the fifth big step is to choose the set of genetic operators that cause variations and proportions. Other details of the GEP modeling structure are expressed in the literature [14, 15]. This is a big step, chromosomal architecture, ie the head size and to choose the number of genes. After several attempts, the head size is 8 and the number of genes is 3 to get the best result for GEP models. The Linking function is selected as the multiplication function. Finally, a set of genetic operators have been used as a set of genetic operators.

2.4 Creation of GEP Model

In this study, ten years (120) data of eight independent parameters were used to construct GEP model. A total of 120 data, 80% (96) was used for training purposes and 20% (24) was used for testing purposes. Training data (96) was first used to construct the model. Other 24 data models were used to test [16]. In the model studies, firstly three different models were formed for different genes considering all parameters. Thus, it was tried to determine the suitability of the proposed GEP models in the case of different genes.

The model 1 (one Gen) was developed for the 8 input GEP formulations and the Expression Tree (ET) was obtained. For the training set, R^2 =0.93, RMSE =2.64E+0.5and for the test set R^2 =0.905, RMSE=3.26E+0.5 was obtained. The equation obtained for Model 1:

((((d[3]+d[3])*d[7])+(d[6]+d[1]))+((d[7]*d[3])+(d[2]+d[1]))) (3)

where: the constants in the formulation are G1C0=9.90 and G1C1=-6.31. The real variables in the GEP formulation are d[1]=P, d[2]=S, d[3]=T, d[6]=Q_L, d[7]=D. The final values of the equation after the corresponding values are left:

Q_T = (2T*D+QL+P) + (D*T+S+P) (4)

In this model Q_T=Total drinking water amounts for Diyarbakır was effected by parameters such as T=Temperature, D=per capita GDP, Q_L=Water losses, P=Population, S=House subscriber number. This equation contains five independent variables.
The equation obtained by simplifying the equation (4) above:

Q_T = 3T*D+Q_L+2P+S (5)

This equation contains five independent variables in total. The program eliminated these variables because the effect of other variables was low (Accrued / A, Rainfall / R, Humidity / H).

The model 2 (two Genes) was developed for the 8 input GEP formulations and the Expression Tree (ET) was obtained. For the training set, R^2 =0.952, RMSE =2.13E+0.5 and for the test set R^2 =0.907, RMSE=3.44E+0.5 was obtained. The equation obtained for Model 2:

sqrt(sqrt(sqrt(((d[0]/sqrt(G1C1))+d[0]))))*((((d[5]*d[4])*d[3])+(d[1]+d[6]))+((d[5]/G2C1)*pow(d[3],3))); (6)

where: the constants in the formulation are G1C0 = 6.08, G1C1 = 5.78, G2C0 = 9.96, G2C1=1.82. The real variables in the GEP formulation are d[0]=A, d[1]=P, d[3]=T, d[4]=R, d[5]=H, d[6]=Q_L. The final values of the equation after the corresponding values are left:

Q_T= $(A/(5.78)^{1/2} + A)^{1/8}$∗(H∗R∗T+P+$Q_L$ + (H/1.82∗T^3)) (7)

The equation obtained by simplifying the equation (7) above:

Q_T= $(1.417A)^{1/8}$∗(H∗R∗T+P+Q_L + (H/1.82∗T^3)) (8)

This equation contains five independent variables in total. The program eliminated these variables because the effect of other variables was low (Development / D, House subscriber number / S).

The model 3 (three Genes) was developed for the 8 input GEP formulations and the Expression Tree (ET) was obtained. For the training set, R^2 =0.982, RMSE =1.23E+0.5 and for the test set R^2 =0.947, RMSE =2.52E+0.5 was obtained. Fig. 2 shows the ET of the formulation which actually is:

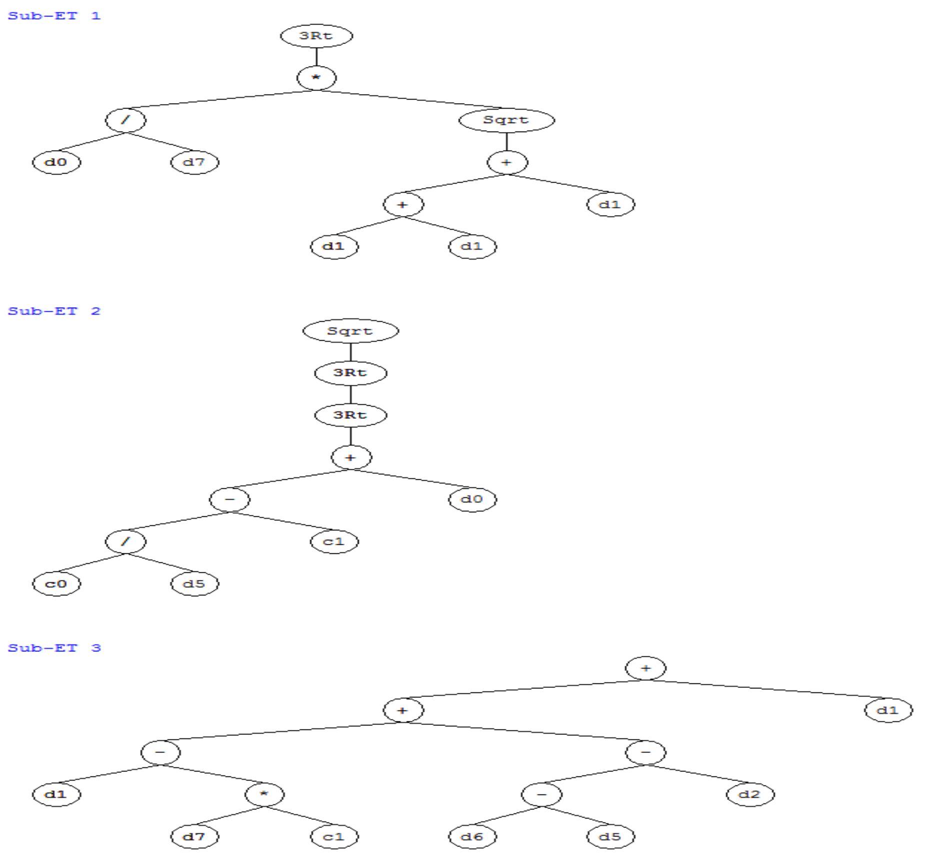
pow(((d[0]/d[7])*sqrt(((d[1]+d[1])+d[1]))),(1/3))*sqrt(pow(pow((((G2C0/d[5])-G2C1)+d[0]),(1/3)),(1/3)))
*(((d[1]-(d[7]*G3C1))+((d[6]-d[5])-d[2]))+d[1]), (9)

where: the constants in the formulation are G1C0=2.61, G1C1=2.547, G2C0=9.92, G2C1=0.273, G3C0=8.55 and G3C1=4.83. The real variables in the GEP formulation are d[0]=A, d[1]=P, d[2] =S, d[3]=T, d[4]=R, d[5]=H, d[6]=Q_L, d[7]=D. The final values of the equation after the corresponding values are left:

$$Q_T=\left(\sqrt[3]{\frac{A}{D}}+\sqrt{3P}\right)*\left(\sqrt[18]{\frac{9.92}{H}}-0.273+A\right)*(2\mathrm{P}-4.83\mathrm{D}+Q_L-\mathrm{H}-\mathrm{S}) \quad (10)$$

In this model Q_T=Total drinking water amounts for Diyarbakır was effected by parameters such as A=Drinking Water Accrued, P=Population, S=House subscriber number, T=Temperature, R=Rainfall, H=Moisture, Q_L=Water losses, D=per capita GDP. This equation contains seven independent variables. It is seen that seven variables are effective in this equation.

Figure 2. Expression tree (ET) for the proposed GEP formulation



It is calculated as R^2=0.982 for the training set in Fig. 3. In Fig. 4, the test (validation) set is calculated as R^2=0.947. When the models for different amplitudes are examined, it is seen that the three genes model gives better results than the one and two genes models. As shown in Fig. 3 and Fig. 4, the graphs of the estimated values obtained and measured in the model are close to each other. This indicates that the GEP Model contains values close to the measured data.

Figure3. Gep estimations of Q_t versus measured ones for train set (Model 3)

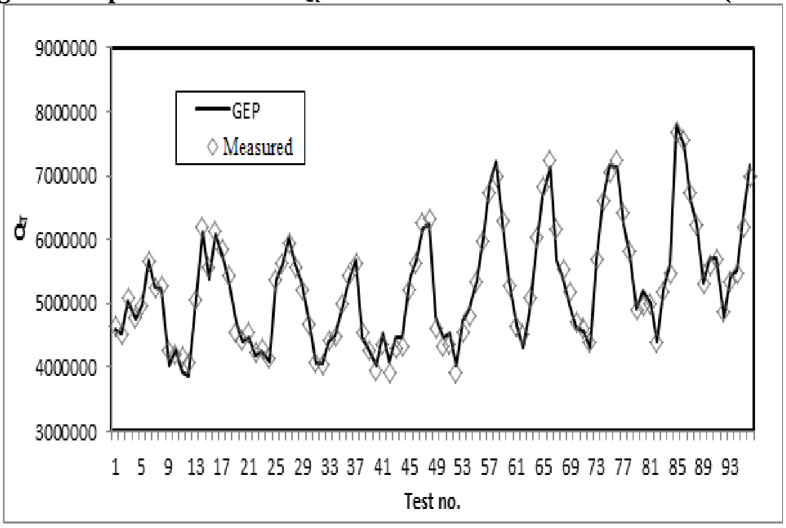
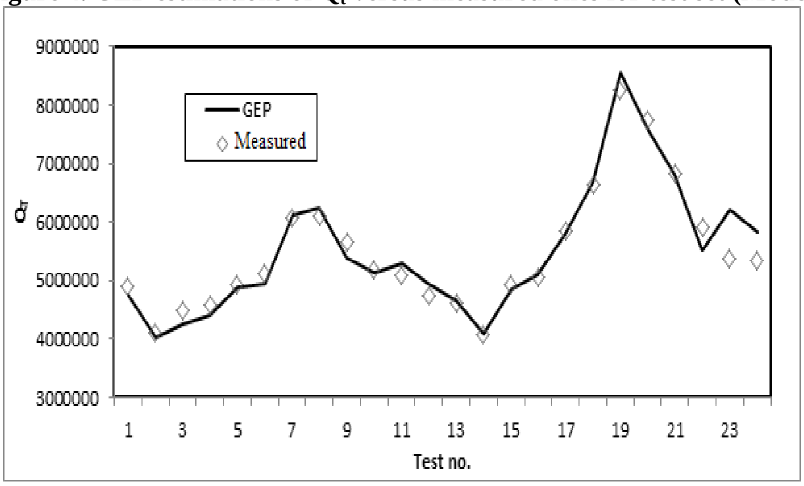


Figure 4. GEP estimations of Q_t versus measured ones for test set (Model 3)



III. RESULTS AND DISCUSSION

As seen in Table 2, 3 models with different gens of 8 input parameters have been developed. The training set, R^2=0.93, RMSE=2.64E+0.5 and the test set R^2=0.905, RMSE=3.26E+0.5 was obtained for Model 1. The training set, R^2=0.952, RMSE =2.13E+0.5 and the test set R^2=0.907, RMSE=3.44E+0.5 was obtained for Model 2. The training set, R^2=0.982, RMSE=1.23E+0.5 and the test set R^2=0.947, RMSE =2.52E+0.5 was obtained for Model 3. As it can be seen from the models the performance of the model increases with the increase in the number of genes with the constant input parameters. According to Model 1 and Model 2, in the model 3 (three genes), both high R^2 and small RMSE values were obtained in both training and test (0.982, 1.23E+0.5 and 0.947, 2.52E+0.5).

Table2. Statistical performance of GEP with different genes

			TRAINING		TEST	
MODEL	**MODEL TYPE**	Gen	**R^2**	**RMSE**	**R^2**	**RMSE**
1	Q_T=F (A, P, S, T, R, H, Q_L, D)	1	0.930	2.64E+0.5	0.905	3.26E+0.5
2	Q_T=F (A, P, S, T, R, H, Q_L ,D)	2	0.952	2.13E+0.5	0.907	3.44E+0.5
3	Q_T=F (A, P, S, T, R, H, Q_L ,D)	3	0.982	1.23E+0.5	0.947	2.52E+0.5

IV. CONCLUSIONS

In this study, an Artificial Intelligence (AI) technique, Genetic Expression Programming (GEP) was used to mathematically model the parameters that may affect drinking water consumption. In the models formed within the scope of data that may affect drinking water consumption of Diyarbakır city, ten years data have been used between 2005-2014 years.

Various models with different parameters were developed in the study. In models 1, 2 and 3; Drinking water accrued, Population, House subscriber number, Temperature, Rainfall, Moisture, Water losses, per capita GDP (A, P, S, T, R, H, Q_L and D) used for different genes (1, 2 and 3 genes). In the other models, the inputs were randomly subtracted and the performance of the models were compared with the R^2 and RMSE values in order to better examine the effect of the inputs on the measured and generated models. Model 3's performance was better and higher R^2 values were achieved (R^2=0.982).

Conflict of interest
There is no conflict to disclose.

ACKNOWLEDGEMENT
This study was supported by Dicle University Scientific Research Projects (DÜBAP) Engineering 16.005 project.

REFERENCES

[1]. State Planning Organization Annual Report; DPT (2007) Ankara

[2]. Howe C.W, Linaweaver Jr,F.P (1967) The impact of price on residential water demand and its relation to system design and price structure. Water Resour. Res. 3:13–32, DOİ.org/10.1029/WR003i001p00013.

[3]. Cassuto AE, Ryan S (1979) Effect of price on the residential demand for water within an acency. JAWRA, Journal of the American Water Resources Association, 15:345–353, DOİ.org/10.1111/j.1752-1688.1979.tb00337.x.

[4]. Billings RB, Agthe DE (1998) State-space versus multiple regressions for forecasting urban water demand. Journal of Water Resources Planning and Management, 124:113-117, DOİ: 10.1061/(ASCE)0733-9496(1999)125:4(240).

[5]. Babel MS, Gupta AD, Pradhan P (2007) A multivariate econometric approach for domestic water demand modeling: an application to Kathmandu, Nepal. Water Resources Management 21:573-589, DOI:10.1007/s11269-006-9030-6.

[6]. Altunkaynak A, Ozger M, Cakmakcı M (2005) Water consumption prediction of Istanbul city by using fuzzy logic approach. Water Research Management 19:641-654, DOI: 10.1007/s11269-005-7371-1.

[7]. Adamowski F (2008) Peak daily water demand forecast modeling using artificial neural networks. J. Water Resour. Plann. Manage, 134:119-128, DOI:10.1061/(ASCE)0733-9496(2008)134:2(119).

[8]. Fırat M, Yurdusev MA, Turan ME (2009) Evaluation of artificial neural network techniques for municipal water consumption modeling. Water resources management *23*:617-632 DOI 10.1007/s11269-008-9291-3

[9]. Ajbar A, Ali E (2012) Water demand prediction for touristic mecca city in Saudi Arabia using neural networks. International Scholarly and Scientific Research 6:342-346

[10]. Akdag R (2016) Estimation and comparative analysis of urban drinking water analysis using artificial neural networks, support vector machines and box-Jenkins methods. Business and Economics Research Journal, 7:123, DOI:10.20409/berj.2016116808

[11]. Wu L, Zhou H (2010) Urban Water Demand Forecasting Based on HP Filter and Fuzzy Neural Network. Journal of Hydroinformatics, 12:172-184, DOİ: 10.2166/hydro.2009.082.

[12]. White S, Robinson J, Cordel D, Jho M, Milne G (2003) Urban Water Demand Forecasting and Demand Management: Research Needs Review and Recommendations. Astralia:Water Services Association of Australia.

[13]. Ferreira C (2006) Gene expression programming: mathematical modeling by an artificial intelligence. 2nd Ed., Springer-Verlag, Germany

[14]. Bagatur T, Onen F (2014) A predictive model on air entrainment by plunging water jets using GEP and ANN. *KSCE* Journal of Civil Engineering, 18:304–314, DOI 10.1007/s12205-013-0210-7.

[15]. Onen F, Bagatur T (2017) Prediction of flood frequency factor for Gumbel distribution using regression and GEP model. Arabian Journal for Science and Engineering, 42:3895-3906, DOI: 10.1007/s13369-017-2507-1

[16]. Aslan B (2017) Diyarbakır kenti içmesuyu ihtiyacının genetik ifadeli programlama ile modellenmesi**,** Yüksek Lisans Tezi, Dicle üniversitesi Fen Bilimleri Enstitüsü, Diyarbakır, Turkey (Master Thesis)

Behzat Aslan. "Modeling of Urban Drinking Water with Genetic Expression Programming." *International Journal of Engineering and Science*, vol. 11, no. 10, 2021, pp. 01-07.