

Evaluation of Gene Expression Method for Rainfall-Runoff Relationship: Case study of the E21A057 Station

¹İbrahim Halil İFŞAAT, *¹Veysel GÜMÜŞ and ¹Oğuz ŞİMŞEK
¹Faculty of Engineering, Department of Civil Engineering Harran University, Turkey

Abstract

A reliable model for the prediction of rainfall-runoff process is important for water resource planning and management. The accurate estimation of monthly rainfall and runoff also can be considered a key element for predicting drought in the near future. Rainfall is the most important factor for flow. Therefore, knowing the amount of rainfall in a basin can help to improve the accuracy of forecasting models for river flow. Precipitation and flow have a complex and nonlinear relationship. Additionally, the effect of precipitation on the flow depends on basin characteristics such as topography, shape, slope, type of soil, soil moisture, temperature and evaporation. These parameters are generally considered in conceptual models. However, since the basin characteristics are difficult to determine, in recent years the relationship between precipitation and flow can be made with limited data by artificial intelligence methods. In this study, monthly mean flow values of E21A057 Stream Gauging Station (SGS) are estimated by Genexpression Programming (GEP) and Multiple Linear Regression (MLR) methods by using monthly total precipitation values of 17204 meteorological observation station. Models which are prepared using different input parameters, the GEP method is found to be more successful in almost all cases compared to the MLR method. The most successful model among all models is determined as DM04 using four input parameters in GEP method. These parameters are Precipitation (P), 1 month lagged flow rate (Q-1), 1 month lagged precipitation (P-1) and 2 month lagged precipitation (P-2).

Key words: Euphrates Basin, Gene expression programming, rainfall-runoff relationship, Multiple Linear regression

Yağış-Akış İlişkisinde Genekspressyon Yönteminin Değerlendirilmesi: E21A057 İstasyonu Örneği

Özet

Yağış-akış sürecinin tahmini için oluşturulacak güvenilir bir model, su kaynakları planlaması ve yönetimi için önem arz etmektedir. Ayrıca, aylık yağış ve akımların doğru tahmin edilmesi, yakın gelecekte kuraklığın önceden tahmin edilmesi için kilit bir unsur olarak kabul edilebilir. Yağış ise akım için en önemli etki unsuru olarak karşımıza çıkmaktadır. Bu nedenle, bir havzadaki yağış miktarını bilmek, nehir akışı için uygulanan tahmin modellerinin doğruluğunu arttırmaya yardımcı olabilir. Yağış ile akış karmaşık ve doğrusal olmayan bir ilişkiye sahiptir. Ayrıca, yağış miktarının akışa etkisi havzanın topografyası, şekli, eğimi, zemin türü, zemin nemi, sıcaklığı ve buharlaşma miktarı gibi birçok havza karakteristiğine bağlıdır. Bu parametreler genellikle kavramsal modellerde göz önünde bulundurulur. Ancak havza karakteristiklerini belirlemek oldukça güç olduğundan son yıllarda yağış ile akış arasındaki ilişki yapay zekâ yöntemleri ile daha sınırlı bir veri ile yapılabilmektedir. Bu çalışmada, E21A057 numaralı Akım Gözlem İstasyonuna (AGİ) ait aylık ortalama akım değerleri, 17204 numaralı meteoroloji gözlem istasyonuna ait aylık toplam yağış değerleri kullanılarak, Genekspressyon Programlama (GEP) ve Çoklu Doğrusal Regresyon (ÇDR) yöntemleri ile tahmin edilmeye çalışılmıştır. Farklı girdi parametreleri kullanılarak hazırlanan modellerde GEP yönteminin neredeyse ele alınan tüm durumlarda ÇDR yöntemine göre daha başarılı olduğu görülmüştür. Oluşturulan modeller arasında en başarılı model GEP yönteminde Yağış (P), 1 aylık gecikmeli akış hızı (Q-1), 1 aylık gecikmeli yağış (P-1) ve 2 aylık gecikmeli yağış (P-2) girdi parametresini kullanan DM04 olarak belirlenmiştir.

Anahtar kelimeler: Fırat Havzası, Genekspressyon programlama, yağış-akış ilişkisi, çoklu doğrusal regresyon

*Corresponding author: Address: Faculty of Engineering, Department of Civil Engineering Harran University, 63050, Şanlıurfa TURKEY. E-mail address: gumus@harran.edu.tr, Phone: +904143183000/1274

1. Giriş (Introduction)

Yağış ile akış arasındaki ilişkinin belirlenmek istenmesinin temel nedeni hidrolojik ölçüm tekniklerinin sınırlı olmasıdır. Hidrolojik sistemler hakkında bilmek istenen her şey ölçülemede ve buna bağlı olarak uzay ve zamanda yalnızca sınırlı ölçüm teknikleri ile ve sınırlı ölçüm yelpazesine sahip olunmaktadır. Gelecekteki hidrolojik değişimin muhtemel etkisini değerlendirmek için hem mekânsal hem de zamansal ölçümlerden yararlanılarak, akım ölçümü yapılmamış bölgelerde veya geleceğe dönük tahmin yöntemlerine ihtiyaç duyulmaktadır. Bu nedenle, yağış akış arasındaki ilişkiyi veren farklı tipteki modeller karar verme noktasında yardımcı olmaktadır.

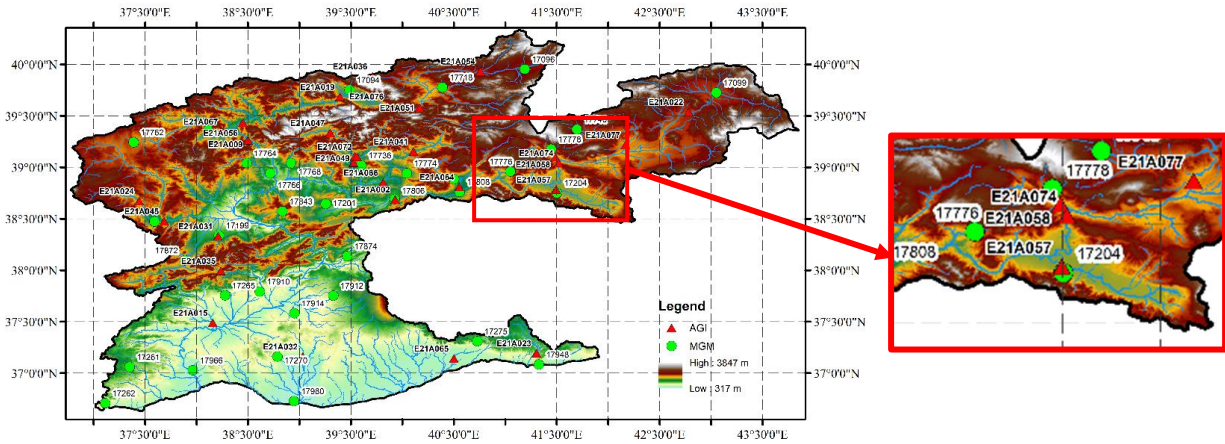
Akarsu akımları, havzadaki yağış, sıcaklık, nem gibi birçok meteorolojik parametre ve havza yapısına bağlıdır. Bunun yanında, akarsu akımlarının ana kaynağı olan yağış ile, yağışın akışa dönmesi olayı arasında belli bir gecikme bulunmakta ve akım tahminlerinde bu gecikmeden yararlanılmaktadır [1, 2]. Yağış-Akış ilişkisinin belirlenmesinde parametrik modellerin yanında Yapay Sinir Ağları, Genetik Algoritma ve Bulanık Mantık gibi Yapay Zeka Yöntemlerinin kullanılmaktadır [3-6]. Bu yapay zeka yöntemleri genellikle kara kutu yöntemler olarak anılmakta ve yağış ile akış arasındaki ilişkiyi açık denklem şeklinde vermemektedir [1, 7, 8]. Bunun yanında, son yıllarda su mühendisliğinde farklı alanlarda kullanılan bir yapay zeka yöntemi olan Gen ekspresyon programla (GEP) ile ilişki açık denklemler ile ifade edilmekte ve herhangi ekstra bir yazılım ya da donanım gerektirmeden kullanılabilir [9, 10]. Gümüş, Soydan, Şimşek, Aköz, Kırkgöz [3] çalışmalarında, Orta Fırat Havzası'nda bulunan E21A057 numaralı Karasu-Karaköprü istasyonuna ait aylık ortalama akım değerlerini, 17204 numaralı Muş meteoroloji gözlem istasyonu aylık toplam yağış değerlerine göre tahmin etmeye çalışmışlardır. Çalışmalarında, üç farklı yapay sinir ağı yöntemini kullanmışlar ve çalışma sonucunda, Radyal Tabanlı Yapay Sinir Ağı yönteminin akımı tahmin etmede en başarılı yöntem olduğunu bildirmişlerdir.

Bu çalışma kapsamında ise, Gümüş, Soydan, Şimşek, Aköz, Kırkgöz [3] tarafından farklı yapay sinir ağı yöntemleri ile 17204 numaralı istasyona ait yağış verileri ile E21A057 numaralı Karasu-Karaköprü istasyonuna ait akım verileri tahmin edilmeye çalışılmıştır. Ancak iki çalışma arasındaki temel fark, 2000 ile 2007 yılları aralığında ölçülen yedi yıllık yeni veri hesaba katılmış ve güncel bir tahmin yöntemi olan GEP yöntemi ile akım değerleri tahmin edilmeye çalışılmıştır. Elde edilen sonuçlar ÇDR yöntemi ve Gümüş, Soydan, Şimşek, Aköz, Kırkgöz [3]'te verilen farklı tipteki yapay sinir ağı yöntemleri ile karşılaştırılmıştır.

2. Materyal ve Yöntem (Materials and Method)

2.1 Çalışma Alanı (Study Area)

Çalışma kapsamında, Türkiye'nin toplam akış hacminin %16.2'sine denk gelen, 27.65 km³ akış hacmine sahip Türkiye'nin en büyük su havzası olan Fırat Havzası'nda bulunan ve Şekil 1'de verilen E21A057 numaralı Karaköprü istasyonuna ait aylık akım değerleri 17204 numaralı MUŞ istasyonun aylık toplam yağış değerleri kullanılarak tahmin edilmeye çalışılmıştır. Kullanılan istasyonlar ve onlara ait coğrafi bilgiler Tablo 1'de verilmiştir.



Şekil 1. Çalışma Alanı

Tablo 1'e göre, ele alınan istasyonların yükseklik ve koordinatlarının birbirine yakın olduğu görülmektedir. İstasyonlara ait veri aralıkları incelendiğinde ise 17204 istasyonuna ait ölçüm aralığı E21A057 numaralı istasyona göre fazla olmasına rağmen, akım değerlerinin ölçüm aralığı ile örtüşen 1969-2007 yılları arasındaki veriler tahmin için kurulacak modellerde kullanılmıştır.

Tablo 1. Kullanılan istasyonlar ve coğrafi bilgileri

Türü	İstasyon No	Adı	Koordinatı	Yükseklik (m)	Ölçüm Aralığı	Kullanılan Veri Aralığı
AGİ	E21A057	KARAKÖPRÜ	38.4702(K),41.2943(D)	1250	1969-2007	1969-2007
MGİ	17204	MUŞ	38.7509(K),41,5023(D)	1322	1964-2017	1969-2007

Beş farklı durum için yapılan akım tahmin çalışmasında kullanılan girdi ve çıktılar Tablo 2'de verilmiştir. Tabloda bulunan, P ele alınan aydaki yağış verisini, P-1 ele alınan aydan bir ay önceki yağış verisini yani bir ay gecikmeli yağışı, P-2 iki ay önceki yağışı, P-3 üç ay önceki yağışı ve Q-1 ise bir önceki aya ait ortalama akımı temsil etmektedir. Buna göre, Durum 1'de (DM01) girdi olarak sadece yağış verileri tanımlanırken, diğer durumlarda akım değerini tahmin etmek için bir önceki aya ait ortalama akım ve yine bir önceki aya ait toplam yağış verileri girdi parametresi olarak kullanılmıştır.

Tablo 2. Yağış-Akış ilişkisindeki kullanılan model girdi ve çıktıları

Durum	Girdi	Çıktı
DM01	P	Q
DM02	P, Q-1	Q
DM03	P, P-1, Q-1	Q
DM04	P, P-1, P-2, Q-1	Q
DM05	P, P-1, P-2, P-3, Q-1	Q

2.2.Yöntem (Method)

Çalışma kapsamında, veri setinde bulunan bağımsız değişkenlerle bağımlı değişkenin değerinin tahmin edilmesi ve hangi bağımsız değişkenin bağımlı değişkeni etkilediğini belirlemek amacıyla kullanılan klasik bir yöntem olan Çoklu Doğrusal Regresyon (ÇDR) yöntemi [6] ve genetik bireylerin nüfusunu kullanarak onları uygunluklarına göre seçen ve bir veya daha fazla genetik operatör ile genetik varyasyonu tanıtan genetik algoritma ve genetik programlamaya benzer bir yöntem olan Gen ekspresyon programlama (GEP) yöntemi [11] kullanılmıştır.

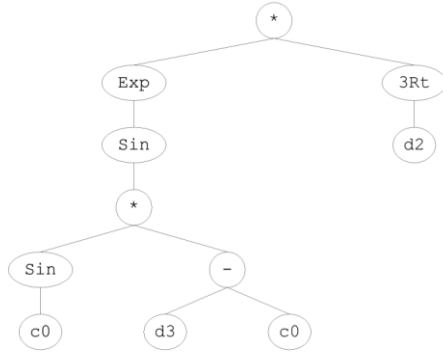
3. Bulgular (Results)

E21A057'e ait aylık akım verilerinin tahmin edilmesi için ÇDR ve GEP yöntemleri ile analizler gerçekleştirilmiş, eğitim ve test süreçleri için sonuçlar Tablo 3'te verilmiştir. Buna göre, ele alınan tüm modellerin test süreci dikkate alındığında akımı tahmin etmede GEP yönteminin ÇDR yöntemine kıyasla daha başarılı olduğu görülmüştür. Gözlenen ve tahmin edilen veriler arasındaki korelasyon katsayısı ise test sürecinde en yüksek Durum 4'te 0.808 olarak, aynı modelde OKH değerinin ise 331.73 m²/s² olarak GEP yöntemi ile belirlenmiştir. Durum 4'e GEP yöntemi ile elde edilen "ifade ağacı" ve formüller Şekil 2'de verilmiştir. Bu denklemlerin birbirleri ile olan bağlantı fonksiyonu (linking function) toplama olarak seçilmiştir. Buna göre en başarılı modelde d0, d1, d2 ve d3 sırasıyla 1, 4, 2 ve 1 kez kullanılmış olup, akımın tahmin edilmesinde en fazla d2'nin kullanıldığı belirlenmiştir. Buna göre, akımı tahmin etmede en etken değişkenin yağışın kendinden önceki değeri olan P-1'i temsil eden d2 değişkeni olduğu söylenebilir.

Tablo 3. Tüm durumlar için eğitim ve test sonuçları

	Durum 1						Durum 4					
	Eğitim			Test			Eğitim			Test		
	R	OKH	OMGH	R	OKH	OMGH	R	OKH	OMGH	R	OKH	OMGH
ÇDR	0.503	1101.707	20.833	0.45	771.44	18.84	0.698	769.249	16.6	0.689	467.439	15.232
GEP	0.529	1060.391	20.881	0.459	804.764	19.314	0.809	506.72	12.676	0.808	331.733	11.471
	Durum 2						Durum 5					
	Eğitim			Test			Eğitim			Test		
	R	OKH	OMGH	R	OKH	OMGH	R	OKH	OMGH	R	OKH	OMGH
ÇDR	0.692	772.919	16.364	0.679	471.249	14.838	0.7	768.323	16.698	0.69	469.396	15.426
GEP	0.791	549.012	12.554	0.787	341.13	11.436	0.761	653.297	12.497	0.529	1060.391	20.881
	Durum 3											
	Eğitim			Test								
	R	OKH	OMGH	R	OKH	OMGH						
ÇDR	0.694	771.672	16.4	0.691	461.009	14.818						
GEP	0.799	530.282	12.931	0.79	340.833	11.921						

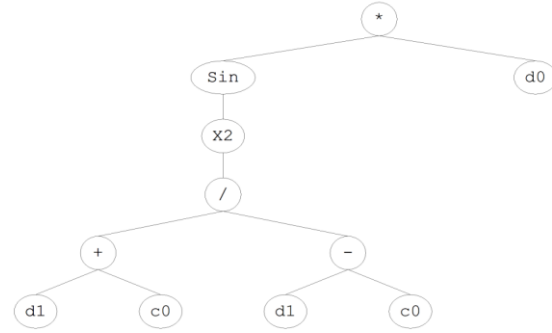
Sub-ET 1



$$c_0 = -4.48652607806635$$

$$e^{\sin(\sin(-4.48652)(d_3 + 4.4865))} \sqrt[3]{d_2}$$

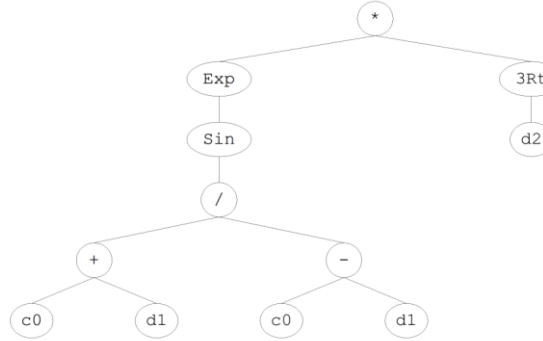
Sub-ET 2



$$c_0 = -4.03787347025971;$$

$$\sin\left(\frac{d_1 - 4.03787}{d_1 + 4.03787}\right)^2 d_0$$

Sub-ET 3

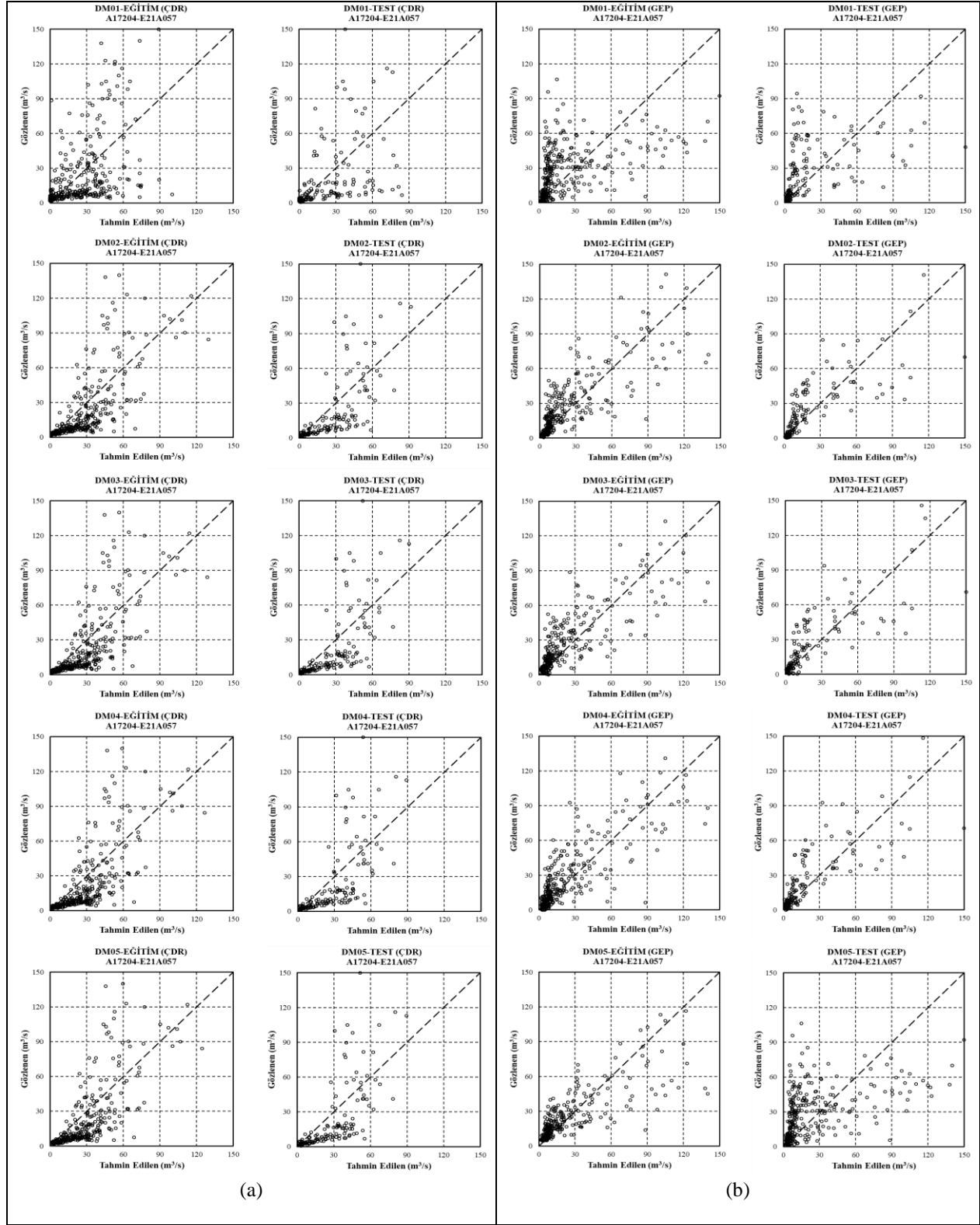


$$c_0 = -4.48652607806635;$$

$$e^{\sin\left(\frac{-(d_1 - 4.48652)}{d_1 + 4.48652}\right)} \sqrt[3]{d_2}$$

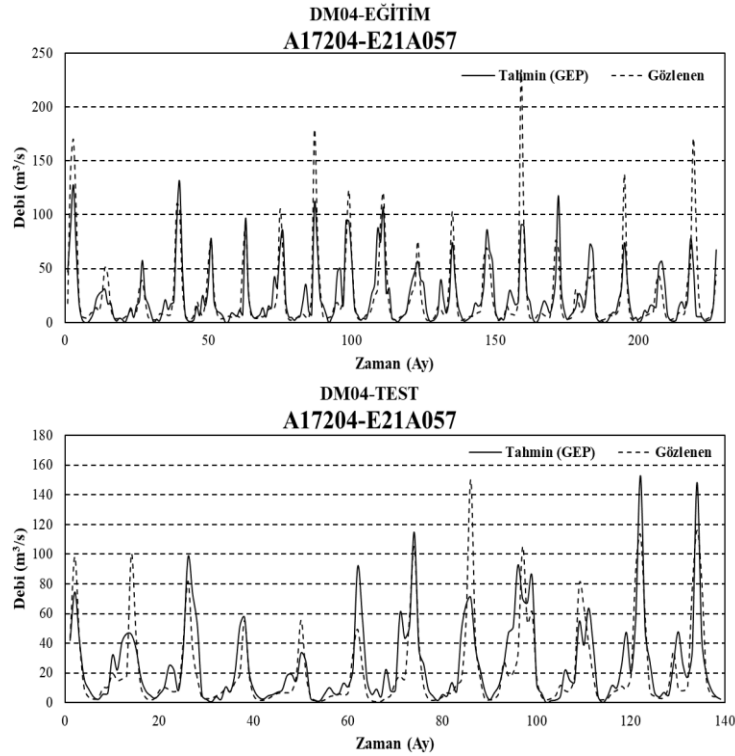
Şekil 3. GEP yöntemi ile Durum 4 için elde edilen “İfade Ağacı”

Tüm durumlar için E21A057 istasyonunu tahmin edilen ve gözlenen akım değerlerinin saçılım grafikleri ÇDR için Şekil 4-a’da, GEP için Şekil 4-b’de verilmiştir. Buna göre, eğitim sürecinde ÇDR ve GEP modelleri ile elde edilen sonuçlar ile test sürecinde elde edilen sonuçlara benzer olup, akımı tahmin etmede birbirleri ile uyumlu bir grafik sergilemektedir.



Şekil 4. E21A057 istasyonuna ait tahmin edilen ve gözlenen akım değerlerinin saçılım grafiği ((a) ÇDR, (b) GEP)

OKH ve R değerlerine göre en başarılı tahmini yapan GEP yöntemi ile DM04'te elde edilen değerlerin gözlenen değerler ile zamansal dağılımı Şekil 5'te verilmiştir. Buna göre eğitim sürecinde tahmin edilen değerler gözlenen değerler yakın olmuş ve eğitim sürecinde tahmin edilen değerler 0.1 ile 150 aralığında olup, özellikle 0.1 m³/s değerlerinden yüksek değerleri tahmin etme başarısı yüksek olmuştur, test sürecinde ise düşük ve yüksek debi değerlerine yakınsamada genel olarak başarılı olmuştur. GEP yöntemi ile elde edilen bulgular Gümüş, Soydan, Şimşek, Aköz, Kırkgöz [3] ile yapılan çalışma ile karşılaştırıldığında, RTYSA yönteminin GEP yönteminden daha başarılı olduğu görülmektedir. Ancak, bu çalışmada GEP yöntemi ile elde edilen formülün kullanılabilirlik açısından kapalı kutu yöntemlerden daha uygun olduğu değerlendirilmektedir.



Şekil 5. E21A057 istasyonu için en uygun modelin tahmin edilen ve gözlenen akım değerlerinin zamansal dağılımı

4. Sonuçlar (Conclusions)

Bu çalışmada, E21A057 numaralı akım gözlem istasyonuna ait aylık ortalama akım değerleri, 17204 numaralı meteoroloji istasyonuna ait aylık toplam yağış değerleri kullanılarak tahmin edilmeye çalışılmıştır. Çalışma kapsamında iki farklı yöntem kullanılmış ve elde edilen bulgular literatür ile karşılaştırılmıştır. Buna göre,

- GEP yönteminin neredeyse ele alınan tüm durumlarda ÇDR yöntemine göre daha başarılı olduğu görülmüştür.
- Tüm modeller arasında en başarılı model GEP yöntemine ait DM04 olmuştur. Buna göre, bu modelde d0 (P), d1 (Q-1), d2 (P-1) ve d3 (P-2) sırasıyla 1, 4, 2 ve 1 kez kullanılmış olup, akımın tahmin edilmesinde en fazla d2'nin kullanıldığı ve akımı tahmin etmede en etken

değişkenin yağışın kendinden önceki değeri olan P-1'i temsil eden d2 değişkeni olduğu belirlenmiştir.

Bu çalışma kapsamında kullanılan GEP yönteminin açık formül vermesinin literatürde kullanılan ve modeli kullanmak için yazılım desteği gereken “kapalı kutu” yöntemlerden daha kullanışlı olduğu değerlendirilmektedir.

Teşekkür (Acknowledgements)

Bu çalışma Harran Üniversitesi Bilimsel Araştırma Projeleri Koordinasyon Birimi (HÜBAP) tarafından desteklenmiştir (Proje No:19108).

Kaynaklar (References)

- [1] Wang W-C, Chau K-W, Cheng C-T, Qiu L. A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. *Journal of Hydrology* 2009;374(3-4):294-306.
- [2] Wu CL, Chau KW, Li YS. Predicting monthly streamflow using data-driven models coupled with data-preprocessing techniques. *Water Resources Research* 2009;45(8).
- [3] Gümüş V, Soydan NG, Şimşek O, Aköz MS, Kırkgöz MS. Yağış-Akış İlişkisinin Belirlenmesinde Farklı Yapay Sinir Ağı Yöntemlerinin Karşılaştırılması. *Çukurova Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi* 2013;28(1):37-50.
- [4] El-Shafie A, Mukhlisin M, Najah AA, Taha MR. Performance of artificial neural network and regression techniques for rainfall-runoff prediction. *International Journal of Physical Sciences* 2011;6(8):1997-2003.
- [5] Adamowski J, Chan HF, Prasher SO, Sharda VN. Comparison of multivariate adaptive regression splines with coupled wavelet transform artificial neural networks for runoff forecasting in Himalayan micro-watersheds with limited data. *Journal of Hydroinformatics* 2012;14(3):731-44.
- [6] Aksoy H, Dahamsheh A. Artificial neural network models for forecasting monthly precipitation in Jordan. *Stochastic Environmental Research and Risk Assessment* 2008;23(7):917-31.
- [7] Machado F, Mine M, Kaviski E, Fill H. Monthly rainfall-runoff modelling using artificial neural networks. *Hydrological Sciences Journal* 2011;56(3):349-61.
- [8] Dahamsheh A, Aksoy H. Artificial neural network models for forecasting intermittent monthly precipitation in arid regions. *Meteorological Applications* 2009;16(3):325-37.
- [9] Danandeh Mehr A. An improved gene expression programming model for streamflow forecasting in intermittent streams. *Journal of Hydrology* 2018;563:669-78.
- [10] Hashmi MZ, Shamseldin AY, Melville BW. Statistical downscaling of watershed precipitation using Gene Expression Programming (GEP). *Environmental Modelling & Software* 2011;26(12):1639-46.
- [11] Ferreira C. *Gene expression programming: mathematical modeling by an artificial intelligence*. Springer; 2006.