NĞÜ Müh. Bilim. Derg. / NGU J. Eng. Sci.

Niğde Üniversitesi Mühendislik Bilimleri Dergisi, Cilt 5, Sayı 2, (2016), 177-187
Nigde University Journal of Engineering Sciences, Volume 5, Number 2, (2016), 177-187

Araştırma / Research

METEOROLOJİK VERİLER KULLANILARAK YERALTI SU SEVİYESİNİN GENETİK PROGRAMLAMA İLE TAHMİNİ

Murat ÇOBANER[*1], Emin BABAYİĞİT[2], Bilal BABAYİĞİT[3]

[1]*İnşaat Mühendisliği Bölümü, Mühendislik Fakültesi, Erciyes Üniversitesi, Kayseri, Türkiye*
[2]*İnşaat Mühendisliği Anabilim Dalı, Fen Bilimleri Enstitüsü, Erciyes Üniversitesi, Kayseri, Türkiye*
[3]*Bilgisayar Mühendisliği Bölümü, Mühendislik Fakültesi, Erciyes Üniversitesi, Kayseri, Türkiye*

Geliş / Received: *20.04.2016*
Düzeltmelerin gelişi / Received in revised form: *24.05.2016*
Kabul / Accepted: *25.05.2016*

ÖZ

Tatlı su ihtiyacının önemli bir kısmı yeraltı suları tarafından karşılanmaktadır. Bu nedenle yeraltı sularının kalitesinin ve miktarının izlenmesi ve mevcut durumunun ortaya konması su yönetimi politikalarının geliştirilmesi bakımından önemlidir. Genetik programlama (GP) evrimsel tabanlı bir yaklaşımdır ve son yıllarda alternatif bir yapay zeka tekniği olarak kullanılmaktadır. GP, diğer yapay zeka tekniklerine göre değişkenler arasındaki ilişkilerden faydalanarak problemin bir matematiksel modelini ortaya koyabilme avantajına sahiptir. Bu çalışmada, meteorolojik veriler ve önceki günlere ait yeraltı su seviyeleri kullanarak yeraltı su seviyelerini tahmin etmek amacıyla çeşitli matematiksel modeller geliştirilmiştir. Matematiksel modellerin geliştirilmesine yeni bir GP yaklaşımı olan çok-genli genetik programlama (ÇGGP) uygulanmıştır. Çalışmada, Develi meteoroloji istasyonuna ve Karacaviran gözlem kuyusuna ait 2007-2009 yıllarına ait günlük veriler kullanılmıştır. Bir ay (30 gün) sonraki yeraltı su seviyelerini tahmin etmek için oluşturulan modellerin doğruluğu değerlendirilmiş ve çoklu doğrusal regresyon modelleri ile karşılaştırılmıştır. ÇGGP modelleri ile elde edilen sonuçların çoklu doğrusal regresyon modellerine göre dört farklı kritere göre daha iyi olduğu tespit edilmiştir. ÇGGP kullanarak oldukça basit ve kullanışlı modeller ortaya konmuştur.

Anahtar Kelimeler: Meteorolojik veriler, yeraltı su seviyesi tahmini, çoklu doğrusal regresyon, genetik programlama

ESTIMATION OF GROUNDWATER LEVEL WITH GENETIC PROGRAMMING USING METEOROLOGICAL DATA

ABSTRACT

The significant portion of freshwater needs is met by ground water. Therefore, paying attention to the monitoring of groundwater quality and quantity data and to reveal the current situation is necessary for the calculation of the potential for water development and water management policies. Genetic programming (GP) is an evolutionary-based approach and is used as an alternative to artificial intelligence techniques in recent years. GP has the advantage of revealing a mathematical model by using the relationships between variables. In this study, various mathematical models have been developed to predict the groundwater levels by using meteorological data and previous days of groundwater levels. To the development of mathematical models, a new GP approach, multi-gene genetic programming (MGGP), was implemented. The daily data obtained from Karacaviran observation wells and Develi meteorological station covering the years of 2007-2009 were used for the creation of the models. The accuracy of the generated models to predict the ensuing a month (30 days) groundwater levels were evaluated and compared with the multiple linear regression models. The results

[*]Corresponding author / Sorumlu yazar. Tel.: +90 352 207 66 66; e-mail/e-posta: mcobaner@erciyes.edu.tr

obtained with MGGP models were found to be better than multiple linear regression models based on four different criteria. Quite simple and useful models using MGGP have been revealed.

Keywords: Meteorological data, groundwater level estimation, multiple linear regression, genetic programming

1. GİRİŞ

Kendisini yenileyebilen bir kaynak olan su, canlı yaşamının devamı için hayati bir öneme sahiptir. Tarih boyunca yerleşim bölgeleri çoğunlukla su kaynaklarının yakınlarına kurulmuştur. Günümüzde ülkelerin nüfuslarının sürekli bir şekilde artmasına paralel olarak evlerde, tarımda ve sanayide kullanılan su sarfiyatı her geçen gün artmış bu sebeple de kişi başına düşen su miktarı azalmış ve suya olan ihtiyaç çok daha önemli hale gelmiştir.

Dünyada ve ülkemizde su kaynaklarının sürdürülebilir yönetimi, suların verimli kullanılması ve su kalitesinin korunması yönündeki çalışmalar giderek artmaktadır [1-3]. Bu çalışmalar kapsamında, artan nüfusa bağlı su kullanımındaki artışın yanı sıra iklim değişikliği, tarım ve sanayileşme gibi faktörler de dikkate alınarak gelecekte ihtiyaç duyulabilecek su miktarı belirlenerek modelleme çalışmaları yapılmaktadır. Modelleme çalışmalarının önemli parametrelerinden olan yağış-buharlaşma ilişkisi, yer altı ve yer üstü sularının etkileşimi ve miktar, depolama ve beslenme potansiyellerinin doğru tespit edilmesi gerekir. Bu parametrelerin tahmininde bölgedeki yeraltı sularının değişiminin belirlenmesi ve/veya önceden tahmin edilmesi hidrolojik döngünün diğer parametrelerinin belirlenmesinde önem arz etmektedir.

Yeraltı suyu seviyesi; yeraltı suyu yüzey suyu etkileşiminin, akifer besleniminin ve su kullanımının bir göstergesidir. Bu mekanizmaların belirlenmesinde önemli bir değişken olan yeraltı suyu seviyelerinin düzenli olarak ölçülmesi pahalı ve zahmetli bir iştir. Ancak meteorolojik veriler ve önceki günlere ait yeraltı su seviyeleri kullanılarak bir bölgedeki yeraltı su potansiyelinin belirlenmesi mümkündür. Yeraltı suyu seviyesinin düzenli olarak izlenebilmesi için ya izleme kuyuları vasıtasıyla doğrudan gözlem yapılması ya da gözlem olmayan veya eksik olan yerler için farklı metotlar kullanılarak tahmin edilmesi gerekmektedir [4].

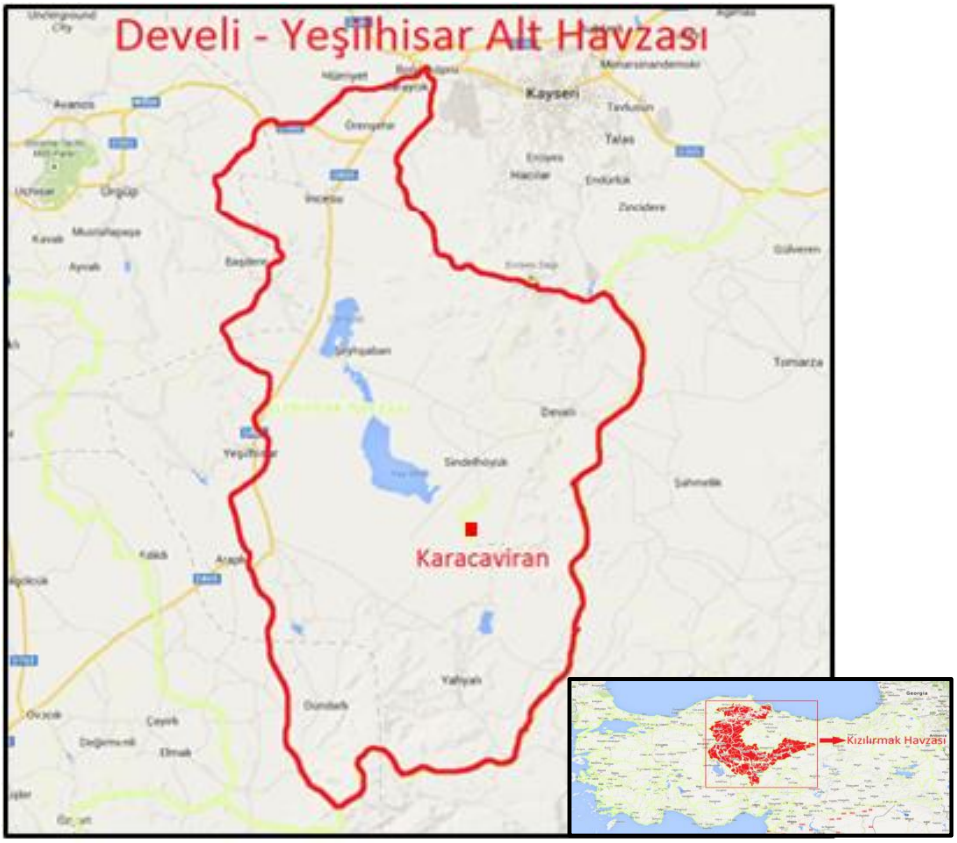
Hidrolojik veriler su mühendisliği çalışmalarında doğa karakterini ve davranışını analiz etmede ve öngörmede önemli yer tutmaktadır. Bu veriler doğası gereği lineer olmayan ve durağan olmayan karakterdedir. Bu verilerin tahmini ve simülasyonu için günümüze kadar pek çok çalışma yapılmıştır. Yeraltı sularının alansal ve zamansal değişiminin ölçümü veya belirlenmesi yeraltı suyu potansiyeli açısından önemlidir. Hemen hemen her bölgede hâlihazırda düzenli olarak ölçümü yapılan meteorolojik veriler ile yeraltı suyu verileri arasındaki ilişki yapay sinir ağları (YSA) [5], bulanık mantık (BM) [6] yöntemleriyle modellenmeye çalışılmıştır. Son yıllarda birçok inşaat mühendisliği problemine uygulanmış olan YSA ve BM modelleri oldukça başarılı sonuçlar vermesine rağmen, bu modeller giriş ve çıkış verileri arasındaki ilişkilerin belirlenmesini sağlayabilecek bir tahmin denklemi üretememektedir. Fakat genetik programlama (GP) yöntemi ile genelleştirilmiş genel tahmin denklemleri üretebilmek mümkündür. Bu çerçevede yeraltı su seviyesinin tahmin edilmesinde kullanılabilecek bir modelin formüle edilmesi oldukça önemlidir. Genetik programlama teknikleri birçok inşaat mühendisliği probleminde [7-10] kullanılmış olmasına rağmen GP'nin yeni bir versiyonu olan Çok Genli Genetik Programlama (ÇGGP)'nın [11] inşaat mühendisliği problemlerine uygulaması yok denecek kadar azdır [12, 13].

Bu çalışmada meteorolojik verilerden faydalanılarak yeraltı su seviyesinin tahmin edilmesi için ÇGGP yaklaşımı kullanılmıştır. Farklı meteorolojik veriler kullanılarak, gelecekteki yeraltı su seviyeleri tahmin edilmeye çalışılmıştır. ÇGGP modelinden elde edilen test sonuçları çoklu doğrusal regresyon (ÇDR) sonuçları ile karşılaştırılmıştır. Karşılaştırma kriterleri olarak, karekök ortalama karesel hata (KOKH), ortalama mutlak hata (OMH), ortalama mutlak rölatif hata (OMRH) ve determinasyon katsayısı (R^2) kullanılmıştır. Sonuçta, havzaya ait su bütçesinin önemli bir bileşeni olan yeraltı suyu potansiyeli *t* anındaki meteorolojik ve seviye ölçümleri kullanılarak sabit bir denklem yardımıyla tahmin edilebilecektir. Ayrıca ÇGGP ile elde edilecek regresyon denklemi ile yapılacak olan tahminler üç boyutlu yeraltı suyu modellerine göre daha az bilgi kullanarak yeterince sağlıklı sonuçlar vermektedir. Bu açıdan ölçümü kolay olan meteorolojik ve yeraltı su seviyesi (YASS) verileri kullanılarak yeraltı su potansiyelinin belirlenmesi su kaynaklarının temininde ve planlanmasında karar vericilere büyük kolaylık sağlayacaktır.

2. MATERYAL VE METOT

2.1. Develi-Yeşilhisar Alt Havzası

Develi-Yeşilhisar alt havzası Kızılırmak havzasında yer almaktadır (Şekil 1). Bu alt havza Kayseri ilinin güneybatısında L33-L34-L35 paftasındadır. Kuzeyinde Erciyes dağı, doğusunda Develi dağları, güneyinde Aladağlar, batısında ise Derinkuyu havzası bulunmaktadır. Develi-Yeşilhisar havzası kapalı bir havzadır ve toplam drenaj alanı 3080 km^2'dir. Develi ovası, ortadaki göl çöküntüleri ile birlikte 908 km^2 yüzölçümüne sahiptir ve ovayı oluşturan düzlük 1070 m'den 1150 m kotuna kadar yayılır. Bu havzada karasal iklim görülür. Yazlar kurak ve sıcak, kışlar soğuk, gece-gündüz, yaz-kış ısı farkı yüksektir. En sıcak aylar Temmuz-Ağustos ayları, en soğuk ayları ise Ocak ve Şubat aylarıdır. Yıllık Class A Pan buharlaşma değerleri ortalama 1660 mm, yıllık yağış ise 334,5 mm'dir [14].



Şekil 1. Kızılırmak havzası ve Develi-Yeşilhisar alt havzası

Çalışmada, Karacaviran gözlem kuyusunun 2007-2009 yıllarına ait günlük veriler ile Develi meteoroloji istasyonuna ait günlük meteorolojik veriler kullanılmıştır. Modelleme çalışmasında kullanılacak verilere ait istatistiki bilgiler Tablo 1'de verilmiştir. Tablo 1'den görüleceği üzere değişkenlerin standart sapma değerleri, dolayısıyla varyansları yüksektir. Bu durum verilerin geniş bir aralıkta değiştiğini göstermektedir. Çarpıklık katsayılarına göre sıcaklık haricindeki değişkenler sağa çarpık dağılım göstermektedir. Sağa çarpık dağılım değişkene ait verilerin modunun medyandan, medyanının da aritmetik ortalamadan küçük olduğu yani küçük verilerin fazla olduğu veri setini temsil etmektedir. Yeraltı su seviye değişimlerine bakıldığında yaklaşık 5 m değiştiği görülmektedir. Yeraltı su seviyesi ve diğer değişkenler arasındaki çapraz korelasyon değerleri incelendiğinde yağış haricindeki diğer tüm verilerin yeraltı su seviyeleri ile ilişkili olduğu görülmektedir.

2.2. Genetik Programlama (GP)

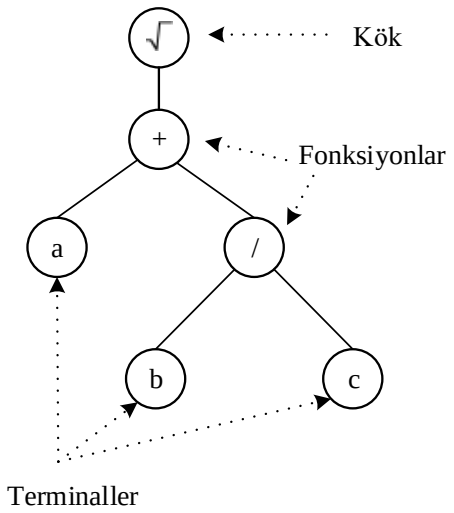
Genetik Programlama (GP) bir problemi çözmek için bilgisayar programları oluşturabilen bir sembolik optimizasyon yöntemidir. GP, genetik algoritma (GA)'nın bir dalıdır ve geleneksel GP ağaç-tabanlı GP olarak da

anılır [15]. Genetik programlama problemin yapı taşını oluşturan ilkel çözüm tarzlarının belli bir uyum kriterine göre evrilerek mükemmelleşmesini amaçlayan bir evrimsel programlama tekniğidir [16]. Genetik programlama, genetik algoritmaların bir uzantısı olmakla birlikte GP ve GA arasında çeşitli farklılıklar bulunmaktadır. GP ve GA arasındaki en önemli fark çözümün gösterimidir. GA'da çözüm ikili veya gerçek sayılardan oluşan bir dizi şeklindeyken, GP'de genellikle ağaç yapıları ile temsil edilen bilgisayar programları şeklindedir. GP ağaçlarının sayısı veya uzunluğu oluşum veya koşma sırasında değişebilirken, GA'da sayı dizileri sabittir [17]. GP, GA'ya göre oldukça yeni bir alandır. GP üç temel aşamada matematiksel model oluşturmaktadır. Bunlar; başlangıç popülasyonunun oluşturulması, yeni popülasyon oluşturulurken mutasyon ve çaprazlama gibi genetik işlemler ile sonlandırma kriteridir.

Tablo 1. Çalışmada kullanılan değişkenlerin istatistikî bilgileri

Değişkenler	Minimum	Maksimum	Ortalama	Standart Sapma	Çarpıklık	YASS ile Korelâsyon
Ortalama Sıcaklık, T_t (°C)	-16,40	30,10	12,10	9,46	-0,31	0,57
Ortalama Nispi Nem, NN_t (%)	22,00	94,70	56,87	13,13	0,20	-0,36
Ortalama Buhar Basıncı, BB_t (hPa)	1,10	18,30	8,50	3,74	0,29	0,57
Referans evapotranspirasyon, $ET_{0,t}$ (mm)	0,31	8,68	3,64	2,21	0,29	0,41
Yerlatı su seviyesi, $YASS_t$ (cm)	535,00	1038,00	722,81	130,08	0,71	1,00

GP'de bireylerin (bilgisayar programları) popülasyonu rastgele bir şekilde oluşturulur. Başlangıç popülasyonu, mümkün olan geniş bir çözüm uzayı içerisinden oluşturulduğundan yüksek çeşitlilik içerir. GP'de bireylerin (bilgisayar programları) popülasyonu rastgele bir şekilde oluşturulur. Başlangıç popülasyonu, mümkün olan geniş bir çözüm uzayı içerisinden oluşturulduğundan yüksek çeşitlilik içerir. Bir popülasyon bireyi, fonksiyonlar ve terminaller içeren hiyerarşik bir ağaçtır. Fonksiyonlar ve terminaller fonksiyon ve terminal kümelerinden seçilir. Fonksiyon kümesi aritmetik operatör (+, -, /, x), mantıksal operatör (and, or, not) veya herhangi matematiksel fonksiyonu (sin, cos, tan, exp, log, tanh) içerebilir. Terminal kümesi kullanıcı tarafından belirlenen fonksiyon parametrelerinden (sayısal sabitler, mantıksal sabitler ve değişkenler) oluşur. Rastgele bir şekilde seçilen fonksiyonlar ve terminaller, beraber bir kök noktası ve uzanan dalları olan ağaç benzeri bir yapıda bir model oluştururlar. Örnek olarak, bir GP modelinin ($\sqrt{a+\frac{b}{c}}$ sembolik ifadesinin) ağaç yapısı gösterimi Şekil 2'deki gibidir.

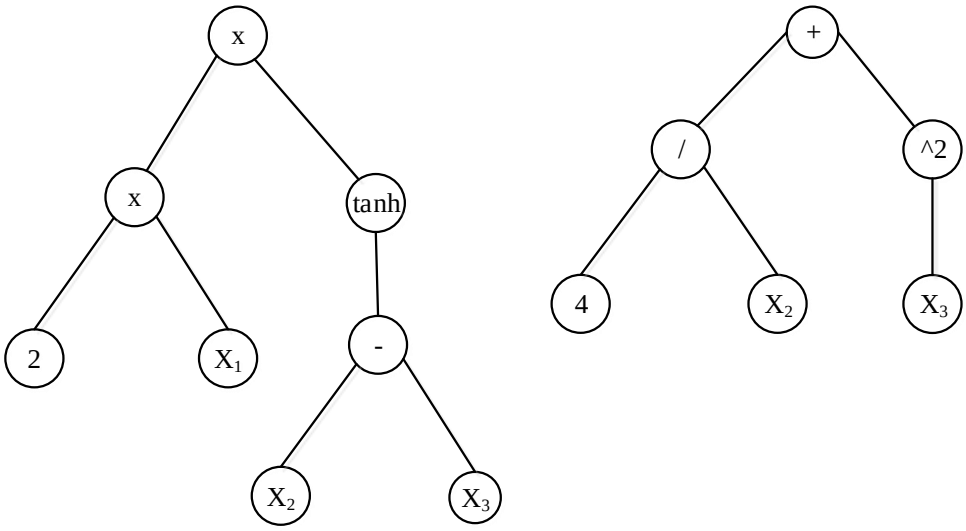


Şekil 2. Bir GP modelinin ağaç yapısı gösterimi

Genetik programlama doğal seleksiyon ve biyolojik olarak esinlenilmiş operatörler prensibi ile genetik olarak üreyen popülasyonları kullanır [17]. Genetik programlama, bir problemi çözmek ya da çözümüne yaklaşmak için boyut ve şekli belirlenmemiş bilgisayar programları bulmak için bir yol sağlar [15]. Genetik programlama genetik algoritmalara göre çok daha güçlüdür. Genetik algoritmada sonuç bir nicelik iken, genetik programlamada sonuç bir programdır. Yani bir bilgisayar programının çözüm için yeni bir program oluşturması anlamına gelir. Kesin sonucu olmayan problemlerin optimum değerlerde modellenmesinde rahatlıkla kullanılabilir. Mesela araba sürme programını ele alırsak, bir tane çözüm yoktur. Bazı çözümler çok zaman harcayıp güvenli sürüşü seçerken, bazıları ise hızlı ama güvensiz bir sürüşü benimseyebilir. Bu durumda araba sürmek basitçe güvenlik ve hız arasındaki ilişki olarak düşünülebilir. Ayrıca diğer bir avantajı da sürekli değişen değişkenler içinde çözüm kümesi oluşturabilmesidir. Mesela araba örneğinde düz yol için sonuç bulan program yer yer asfaltlanmamış bir yol için birden çok çözüm bulabilir [18].

2.3. Çok Genli Genetik Programlama (ÇGGP)

Sembolik regresyon geleneksel GP ile ağaçlardan oluşan bir popülasyonu geliştirmek amacıyla gerçekleştirilir. Her bir ağaç bir matematiksel ifadeye karşılık gelmektedir. Geliştirilen ifadelerle (*Nx1*) çıkış vektörü (*y*) (*N-M*) giriş (*x*) kullanılarak tahmin edilmeye çalışılır. *N* çıkış değişkeni gözlem sayısı *M* ise giriş değişkenlerinin sayısıdır. Geleneksel GP'nin aksine, Çok Genli Genetik Programlama (ÇGGP)'daki her bir sembolik model (ve GP popülasyonunun her bir üyesi) GP ağaçlarındaki çıktıların ağırlıklandırılmış bir doğrusal kombinasyonudur. Bu ağaçların her biri "gen" olarak isimlendirilir. Daha açık bir şekilde belirtilirse, gelenekse GP'de bir model tek bir ağaç ile ÇGGP'de ise her bir geni ağırlıklandırılmış doğrusal birleşimine bias terimi eklenmiş birçok ağaçla ($y = d_0 + d_1\, x\, gen_1 + d_2\, x\, gen_2 + .. + d_n\, x\ \ gen_n$) ifade edilmektedir. Örnek birçok genli model Şekil 3'te görülmektedir. Şekil 3'te görülen modelin denklemi $y = d_0 + d_1\left(2x_1 tanh(x_2 - x_3)\right) + d_2\left(\frac{4}{x_2} + x_3^2\right)$ şeklindedir. Bu model, üç giriş değişkenini (x_1, x_2, x_3) kullanarak çıkış değişkenini tahmin etmektedir. Bu model yapısı doğrusal olmayan terimleri ($tanh$) içerdiği gibi d_0, d_1, d_2 katsayıları ile ilgili parametreleri doğrusaldır. d_1 1. genin d_2 2. genin ağırlık katsayısı, d_0 ise modelin bias terimidir. Uygulamada model için maksimum gen sayısı (G_{maks}) ve herhangi bir genin maksimum ağaç derinliği (D_{maks}) kullanıcı tarafından belirlenmektedir. Bu şekilde modellerin karmaşıklığı kontrol edilebilmektedir. Özellikle 4 veya 5 gibi sabit derinlik kısıtları konulduğunda oldukça uygun modeller elde edilebilmektedir. Ancak problemin karmaşıklığı ve eğitim veri kümesinin büyüklüğüne göre bu değerler değişebilmektedir. Her bir model için, en küçük kareler yöntemi kullanarak eğitim verilerinden doğrusal katsayılar türetilir. Çok genli sembolik regresyon ile standart GP olan sembolik regresyona göre daha doğru ve daha etkili sonuçlar elde edilmiştir [11].

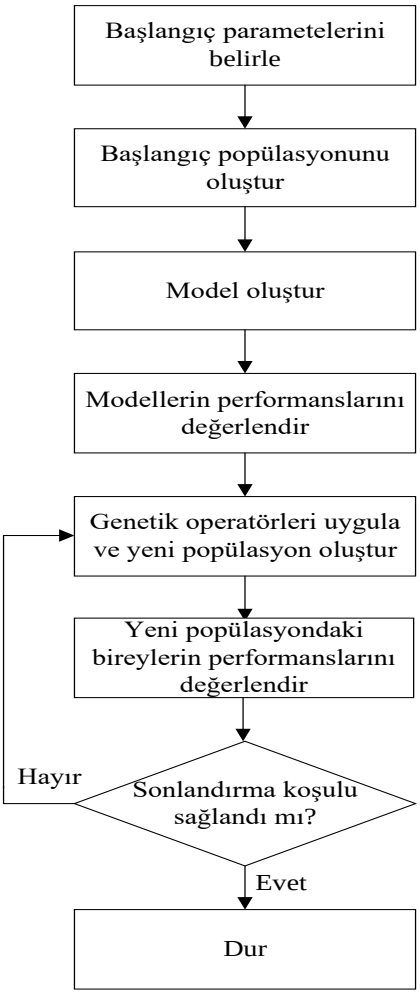


Şekil 3. Bir ÇGGP modeli

ÇGGP'de başlangıç popülasyonu rastgele bir şekilde farklı genlerden $(1 - G_{maks})$ oluşan GP ağaçlarını içeren bireylerden meydana gelir. Bir ÇGGP koşması boyunca, iki noktalı yüksek seviyeli çaprazlama olarak adlandırılan bir ağaç çaprazlama operatörü kullanılarak genler eklenir veya silinir. Bu operatör geleneksel

GP'nin yeniden oluşturma operatörüne ek olarak kullanılır ve bireyler arasındaki gen alışverişine izin verir. Bir bireyin i'inci geni G_i olsun. Bu gene iki noktalı yüksek seviyeli çaprazlama gerçekleştirilme durumunu bir örnek üzerinde açıklayalım. İlk ebeveyn birey dört gen $(G_1G_2G_3G_4)$, ikinci ebeveyn birey ise yine farklı dört gen $(G_5G_6G_7G_8)$ içersin. İlk önce iki adet rastgele seçilmiş çaprazlama noktası her birey için oluşturulur. Genlerdeki çaprazlama noktaları […] ile gösterildiğini varsayalım örneğin $(G_1[\boldsymbol{G_2G_3}]G_4)(G_5G_6[\boldsymbol{G_7G_8}])$ şeklinde olsun. Sonra çaprazlama noktaları değiş tokuş edilir ve iki yeni birey $(G_1G_7G_8G_4)(G_5G_6G_2G_3)$ gibi olur. Değiş tokuş sonucu bir bireydeki gen sayısı G_{maks}'dan daha fazla ise, bireydeki gen sayısı G_{maks} oluncaya kadar rastgele genler seçilir ve silinir.

Geleneksel GP'nin alt ağaç çaprazlama operatörü ÇGGP'de düşük seviye çaprazlama operatörü olarak adlandırılır. Bu durumda, bir gen her bir ebeveynden rastgele bir şekilde seçilir. Sonra, standart alt ağaç çaprazlama uygulanır ve oluşturulan ağaçlar ebeveyn ağaçlar ile değiştirilir aksi takdirde sonraki jenerasyonda değiştirilmemiş birey olurlar. ÇGGP'de çok çeşitli alt ağaç mutasyon ve çaprazlama operatörleri bulunmaktadır. Çaprazlama, yeniden oluşturma ve mutasyon olaylarının olasılık değerleri kullanıcı tarafından seçilmektedir. Bu üç olayın olasılık değer toplamı bir olmalıdır. ÇGGP algoritmasının operatörleri hakkında detaylı bilgi literatürde bulunmaktadır [11, 19, 20]. ÇGGP'nin temel akış diyagramı Şekil 4'te verilmiştir. Şekil 4'teki akış diyagramında ilk önce terminal ve fonksiyon seti, popülasyon büyüklüğü, jenerasyon sayısı, genetik operatörlerin başlangıç değerleri ile maksimum gen sayısı gibi parametrelerin değerleri atanır. Başlangıç popülasyonu oluşturulduktan sonra en küçük kareler yöntemi kullanarak gen kümelerini birleştirerek modeller oluşturulur. Oluşturulan modellerin performans değerleri eğitim verileri üzerinden belirlenir. Performans kriteri kullanıcı tarafından belirlenebilir. Daha sonra modellere genetik operatörler uygulanır ve yeni popülasyon oluşturularak performans değerlerine bakılır. Sonlandırma kriterinin sağlanıp sağlanmadığına göre iterasyon devam eder. Sonlandırma kriteri sağlanmış ise performans kriterine göre en küçük eğitim hatasını veren en iyi birey (model) seçilir.



Şekil 4. ÇGGP temel akış diyagramı

2.4. Çoklu Doğrusal Regresyon (ÇDR)

Çoklu doğrusal regresyonda amaç, bağımlı değişkeni etkilediği belirlenen bağımsız değişkenler yardımıyla bağımlı değişkenin değerinin kestirilmesi ve bağımlı değişkeni etkilediği düşünülen bağımsız değişkenlerden hangisi ya da hangilerinin bağımlı değişkeni daha çok etkilediğini bulmaktır. Çoklu regresyon çözümlemesinde, bağımlı değişken y, bağımsız değişkenler $x_1, x_2,....,x_p$ ile gösterildiğinde aralarındaki ilişki

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_j x_j + ... + \beta_p x_p + \varepsilon \tag{1}$$

olarak yazılabilir. Burada; $\beta_0, \beta_1, \beta_2, ..., \beta_j, ... \beta_p$ regresyon katsayılarıdır. Herhangi bir β_j regresyon katsayısı, diğer değişkenler sabit tutulduğunda (diğer değişkenlerin etkisi ortadan kaldırıldığında) x_j değişkeninde ortaya çıkan bir birimlik değişmeye karşılık y değişkenindeki beklenen değişiklik miktarını vermektedir. Diğer bir ifadeyle; $\beta_1, \beta_2, ..., \beta_j, ... \beta_p$; bağımsız değişkenlerin y'nin saptanmasına yaptıkları göreceli katkıya ilişkin ağırlıklardır. Bu nedenle, β_j (j=1, 2, ..., p) parametreleri genellikle kısmi regresyon katsayıları olarak adlandırılmaktadır. β_0'a ise kesim noktası veya sabit denir ve tüm x_j değişken değerleri sıfır olduğunda bağımlı değişkenin aldığı değeri gösterir. Denklemde yer alan ε ise hata terimini ifade etmektedir [21].

3. BULGULAR VE TARTIŞMA

Çalışmada Develi Yeşilhisar alt havzasında bulunan Karacaviran yeraltı gözlem kuyusunun 2007-2009 yıllarına ait YASS verileri ile Develi meteoroloji istasyonuna ait bazı günlük meteorolojik veriler kullanılmıştır. Bu tarihler arasında günlük birimlerdeki, t anındaki ortalama sıcaklık (T_t), ortalama nispi nem (NN_t), ortalama buhar basıncı (BB_t) ve referans evapotranspirasyon ($ET_{0,t}$) verileri Karacaviran kuyusundaki yeraltı su seviyesini (K_t) etkileyecek bağımsız değişkenler olarak seçilmiştir. Anlamlı değişkenler belirlendikten sonra elde edilen ÇDR sonuçları ile ÇGGP teknikleri karşılaştırılmıştır. Veriler ilk olarak eğitme (2007 ve 2008 yılı verileri) ve test (2009 yılı verileri) olmak üzere iki gruba ayrılmıştır. Eğitme verisi öğrenme amaçlı (genetik evrim ve ÇDR denkleminin elde edilmesi) kullanılmıştır. Test verisi ise elde edilen denklemlerin genelleme kapasitesini görmek için kullanılmıştır.

Seçilen bu meteorolojik veriler birçok meteorolojik veri arasından 30 gün sonraki yeraltı su seviyesi (K_{t+30}) ile arasındaki korelasyon katsayısına göre belirlenmiştir (Tablo 2). Bir ay (30 gün) sonraki YASS'ye etki eden anlamlı bağımsız değişkenler belirlendikten sonra elde edilen ÇDR ve ÇGGP analiz sonuçları karşılaştırılmıştır. Çalışmadaki ET_0 değerleri Citakoğlu ve ark. [22, 23] tarafından modifiye edilen Hargreaves-Samani eşitliği ile tahmin edilmiştir. Bu eşitlik aşağıdaki gibidir:

$$ET_0 = 0.0012 \times R_a \times (T_{max} - T_{min})^{0.386} \times \left(\frac{T_{max} + T_{min}}{2} + 82.019 \right) \tag{2}$$

Bu denklemde, ET_0, Hargreaves eşitliğinin tahmin ettiği referans evapotraspirasyon değerini (mm/gün); R_a, extraterrestrial radyasyonu (MJ/m^2/gün); T_{max}, maksimum sıcaklığı (°C); T_{min}, minimum sıcaklığı (°C) ifade etmektedir.

Tablo 2. Modelde kullanılan verilerin t anındaki değerli ile 30 gün sonraki YASS arasındaki korelasyonlar

Kuyu	**T_t (°C)**	**$ET_{0,t}$ (mm)**	**NN_t (%)**	**BB_t (hPa)**	**K_t (cm)**
Karacaviran, K_{t+30}	0,77	0,73	-0,47	0,79	0,81

ÇGGP ile meteorolojik veriler kullanarak yeraltı su seviyesinin tahminde Tablo 3'te verilen parametreler kullanılarak modeller oluşturulmuştur.

Tablo 3. ÇGGP parametre değerleri

Parametre	Değer
Fonksiyon kümesi	+, -, x, √, sin, abs, tanh
Popülasyon büyüklüğü	300
Maksimum jenerasyon sayısı	100
Çoğalma (reproduction) oranı	0,05
Mutasyon oranı	0,1
Çaprazlama oranı	0,85
Seçkincilik (elitism)	0,02
Maksimum gen derinliği	4
Maksimum gen sayısı	4
Turnuva oranı	4
Ağaç oluşturma yöntemi	Ramped half and half
Turnuva için seçim yöntemi	Lexicographic

Parametre seçimi ÇGGP ile oluşturulacak modellerin genelleme yeteneğini etkilemektedir. Bu sebeple, Tablo 3'te verilen parametreler daha önceki çalışmalarda [12, 13, 19, 20] tavsiye edilen değerlere ve yapılan ön çalışmalar göre belirlenmiştir. Özellikle popülasyon büyüklüğü ve jenerasyon sayıları problemin karmaşıklığına göre seçilmektedir. Bu iki değerin büyük olması çalışma zamanının artmasına sebep olmaktadır. Maksimum gen sayısı (G_{maks}) ve maksimum gen derinliği (D_{maks}) arama uzayı büyüklüğünü ve bu arama uzayındaki keşfedilecek çözümleri etkilemektedir. Bu parametrelerin artırılması ile daha başarılı sonuçlar elde edilebilmektedir ancak bu durumda da problemin karmaşıklığı ve çalışma zamanı artmaktadır. Daha doğru sonuç veren model belirlemek için G_{maks} ve D_{maks} değerleri için farklı değerler denenmiş ve en iyi model her iki parametre için 4 alındığında elde edilmiştir. Uygunluk (fitness) değerleri olarak OKH değerleri kullanılmıştır ve en düşük OKH değerini veren model en iyi model olarak belirlenmiştir. Durdurma ölçütü olarak maksimum jenerasyon sayısı kullanılmıştır. Modellerin değerlendirilmesinde OKH, OMH, OMRH ve determinasyon katsayısı R^2 kriterleri kullanılmıştır. N toplam veri sayısı ve $\overline{YASS}_{t+30}$ ortalama YASS değeri olmak üzere OKH, OMH ve OMRH ifadeleri aşağıdaki gibidir.

$$KOKH = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left((YASS_{t+30})_{i_{Gözlenmiş}} - (YASS_{t+30})_{i_{Tah\min}}\right)^2} \tag{3}$$

$$OMH = \frac{1}{N}\sum_{i=1}^{N}\left|(YASS_{t+30})_{i_{Gözlenmiş}} - (YASS_{t+30})_{i_{Tah\min}}\right| \tag{4}$$

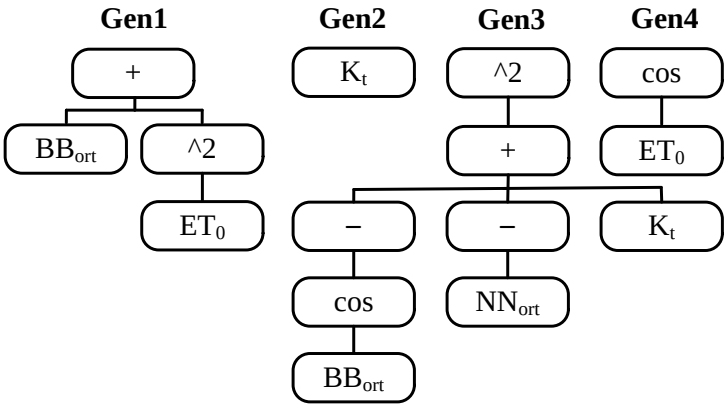
$$OMRH = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{(YASS_{t+30})_{i_{Gözlenmiş}} - (YASS_{t+30})_{i_{Tah\min}}}{(YASS_{t+30})_{i_{Gözlenmiş}}}\right| \times 100 \tag{5a}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}\left((YASS_{t+30})_{i_{Gözlenmiş}} - (YASS_{t+30})_{i_{Tah\min}}\right)^2}{\sum_{i=1}^{N}\left((YASS_{t+30})_{i_{Gözlenmiş}} - (\overline{YASS}_{t+30})_{i_{Gözlenmiş}}\right)^2} \tag{5b}$$

Otuz gün sonraki YASS (K_{t+30}) tahmini için beş veri girişi (T_t, NN_t, BB_t, $ET_{0,t}$ ve K_t) ile elde edilen regresyon denklemi Denklem 6'da, ÇGGP ile oluşturulan model ise Denklem 7'de verilmiştir. ÇGGP ile elde formülleri ifade eden açıklama ağaç diyagramı Şekil 5'de verilmiştir.

$$K_{t+30} = 123 - 2{,}80 \times T_t - 0{,}185 * NN_t + 5{,}84 * BB_t + 29{,}2 * ET_{0,t} + 0{,}640 * K_t \quad (6)$$

$$K_{t+30} = 3{,}11 \times (BB_{ort} + (ET_0)^2) + 1{,}383 \times K_t - 0{,}00056 \times (K_t - NN_{ort} - cos(BB_{ort}))^2 + 9{,}35 \times cos(ET_0) - 96{,}65 \quad (7)$$



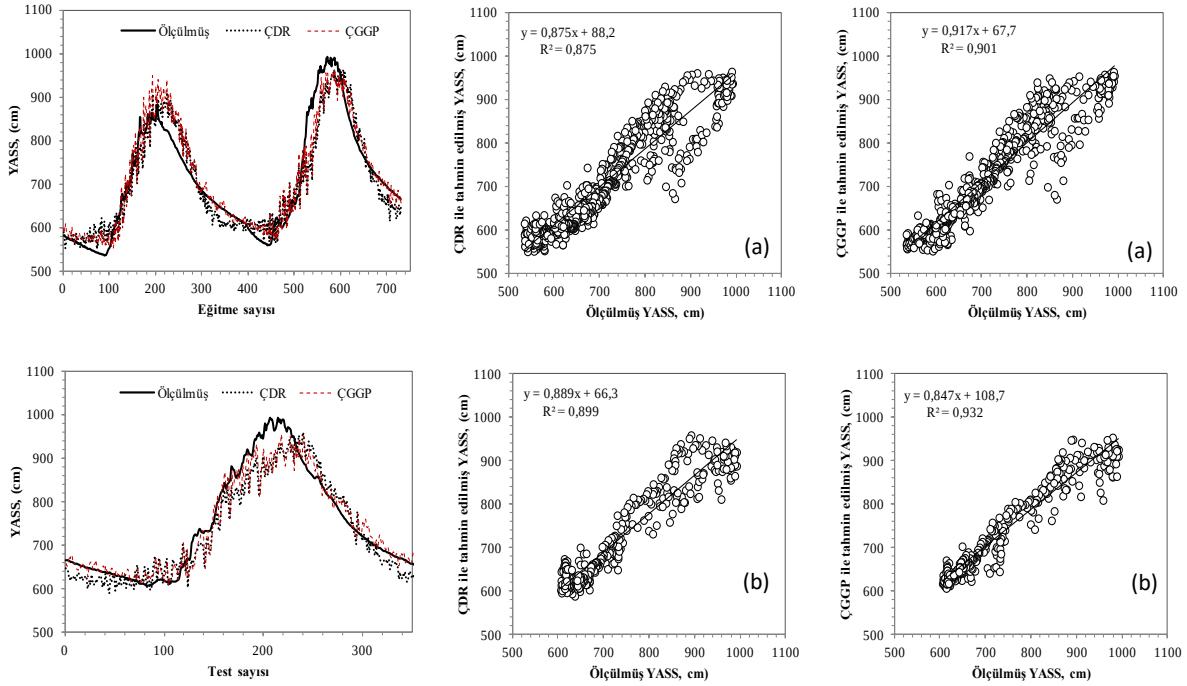
Şekil 5. ÇGGP ile oluşturulan model yapısı

Denklem 6 ve 7 ile elde edilen sonuçlar ile ölçülmüş değerlerin eğitme, doğrulama ve test aşamasındaki OKH, OMH, OMRH ve R^2 değerlerinin karşılaştırılması Tablo 4'te görülmektedir. Tablo 4'e göre ÇGGP'nin KOKH ve OMH hata istatiklerini eğitme için yaklaşık % 20, test için ise % 30 oranında, röletif hata değerini ise sadece % 1 oranında azalttığı görülmektedir. Bu durum küçük seviye değerlerindeki meydana gelen sapmaların rölatif değerlerinin yüksek olmasından kaynaklanmaktadır. Dolayısıysla OMRH değeri diğer istatistiklere göre yeterince azalmamıştır. ÇDR denkleminde 5 değişken kullanılmasına karşın eğitme sonrası ÇGGP modelinde T_t haricindeki 4 girdi kullanılmıştır. Eğitme aşamasındaki mutasyonlar esnasında model T_t değişkenini denklemden çıkarmıştır. ÇGGP modeli 4 girdi ile de ÇDR'den daha doğru sonuç vermiştir. ÇGGP modelinde sıcaklık değişkenininden daha çok evapotranspirasyon ($ET_{0,t}$) değerinin daha etkili olduğu görülmektedir. Evapotranspirasyon değeri bitki su ihtiyacını göstermekte olup aynı zamanda alandan çekilebilecek su miktarını etkilemektedir. Bu nedenle ÇGGP ile Denklem 7 oluşturulurken ET_0 değişkeni öne çıkmış, modelleme esnasındaki evrim gereği *Tt* değişkeni denklemden çıkarılmıştır. Giriş verileri ile çıkış verileri arasındaki ilişkinin doğrusala yakın olmasından (Tablo 2'deki korelasyon katsayı değerlerine göre) dolayı doğrusal olmayan problemlerde daha doğru sonuçlar veren GP modeli ÇDR'den çok da iyi tahminde bulunamamıştır.

Tablo 4. ÇDR ve ÇGGP modellerinin hata istatistikleri

MODEL		KOKH	OMH	OMRH	R^2
ÇDR	Eğitme	43,88	33,92	4,70	0,875
	Doğrulama	41,30	33,63	4,51	0,899
ÇGGP	Eğitme	39,66	29,17	3,95	0,901
	Doğrulama	33,60	22,43	2,90	0,932

Şekil 6'da verilen saçılma diyagramı incelendiğinde ÇGGP ile elde edilen denklemin sonuçlarının eğilim çizgisinin etrafında daha az saçıldığını, ÇDR ile elde edilen denklemin ise saçılmaların fazla olduğu görülmektedir. Tablo 4 ve Şekil 6 birlikte değerlendirildiğinde, 30 gün sonraki YASS değerlerini ÇGGP modelinin (Denklem 7) daha iyi tahmin ettiği görülmektedir. ÇGGP modelinin en önemli avantajlarından biri ÇDR gibi uygulaması basit bir denklem vermesidir.



Şekil 6. Otuz gün sonraki Karacaviran kuyusuna ait ÇDR ve ÇGGP modelleri ile tahmin edilen YASS'lerin eğitme (a) ve test (b) aşamasındaki saçılma diyagramları

4. SONUÇLAR

Bu çalışmada Develi havzasındaki meteorolojik veriler ile yeraltı su seviyesi arasındaki ilişki incelenmiştir. Modellemede bağımsız değişkenler olarak *t* anındaki Karacaviran kuyusundaki YASS değerleri ve bu verilere ek olarak meteorolojik verilerden *t* anındaki ortalama sıcaklık (T_t), ortalama nispi nem (NN_t), ortalama buhar basıncı (BB_t) ve referans evapotranspirasyon ($ET_{0,t}$) verileri eklenerek bir ay (30 gün) sonraki (K_{t+30}) YASS değerleri tahmin edilmiştir. ÇGGP ile YASS değerlerini tahmin etmede meteorolojik verilerden $ET_{0,t}$, BB_t, ve NN_t değişkenlerinin etkili olduğu tespit edilmiştir. ÇDR ve ÇGGP ile elde edilen denklemler OKH, OMH, OMRH ve R^2 istatistikleri ile karşılaştırılmıştır. Modellerin eğitme ve test aşamasındaki tahmin performansları karşılaştırıldığında ÇGGP denklemin ÇDR'ye göre eğitme için yaklaşık % 20, test için ise % 30 daha düşük KOKH ve OMH değerine sahip olduğu görülmektedir. Çalışma serbest akiferlerdeki yeraltı su seviye değerleri ile meteorolojik veriler arasında güçlü bir ilişki olduğunu göstermiştir. Özellikle ET_0 değerinin YASS üzerinde etkili olduğu bunun sebebi olarak da ET_0 değerinin direk olarak sulama suyu ihtiyacını dolayısıyla yeraltından çekilecek su miktarını etkilemesinden kaynakladığı kanaatine varılmıştır. Bu çalışmaya ek olarak yüzeysel su kaynaklarının seviye değerleri de modele eklenerek modelin tahmin gücü arttırılabileceği düşünülmektedir. Yapay zekâ tekniklerinin belirli bir problemi çözmek için teorik bilgiye ihtiyaç duymadan örnekleri kullanarak sonuca vardığı için su kaynaklarının planlanmasında rahatlıkla ve kolaylıkla kullanılabileceği görülmüştür.

KAYNAKLAR

[1] AKKAYA, C., EFEOGLU, A., YEŞİL, N., "Avrupa Birliği Su Çerçeve Direktifi ve Türkiye'de Uygulanabilirliği", TMMOB Su Politikaları Kongresi, 195–204, Ankara, Türkiye, 2006.

[2] ÇOBANER, M., Kıyı Akiferlerinde Tuzlu Su Girişiminin Üç Boyutlu Simülasyonu Göksu Deltası Örneği, Doktora Tezi, Çukurova Üniversitesi, Fen Bilimleri Enstitüsü, Adana, 2009.

[3] YUKSEL, I., "Water management for sustainable and clean energy in Turkey", Energy Reports, 1, 129–133, 2015.

[4] BAŞÇİFTÇİ, F., DURDURAN, S.S., İNAL, C., "Konya Kapalı Havzasında Yeraltı Su Seviyelerinin Coğrafi Bilgi Sistemi (CBS) ile Haritalanması", Harita Teknolojileri Elektronik Dergisi, 5, 1–15, 2013.

[5] KARABORAN, O., ÇOBANER, M., "Serbest Akiferlerdeki Yeraltı Su Seviyeleri ile Meteorolojik Veriler Arasındaki İlişkinin Yapay Sinir Ağları ile Belirlenmesi", Erciyes Üniversitesi, Fen Bilimleri Enstitüsü Dergisi, 30, 376–384, 2014.
[6] YURTÇU, Ş., UYGUNOĞLU, T., İÇAĞA, Y., "Yeraltı Suyu Akımı ile Diğer Meteorolojik Değişkenler Arasındaki İlişkinin Bulanık Mantıkla Modellenmesi", Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi, 12, 285–292, 2006.
[7] AYTEK, A., KISI, O., "A Genetic Programming Approach to Suspended Sediment Modelling", Journal of Hydrology, 351, 288–298, 2008.
[8] KISI, O., GUVEN, A., "A Machine Code-Based Genetic Programming for Suspended Sediment Concentration Estimation", Advances in Engineering Software, 41, 939–945, 2010.
[9] GUVEN, A., KISI, O., "Monthly Pan Evaporation Modeling Using Linear Genetic Programming", Journal of Hydrology, 503, 178–185, 2013.
[10] MEHR, A.D., KAHYA, E., OLYAIE, E., "Streamflow Prediction Using Linear Genetic Programming in Comparison with a Neuro-Wavelet Technique", Journal of Hydrology, 505, 240–249, 2013.
[11] SEARSON, D.P., LEAHY, D.E., WILLIS, M.J., "GPTIPS: An Open Source Genetic Programming Toolbox for Multigene Symbolic Regression", Proceedings of the International Multi Conference of Engineers and Computer Scientist (IMECS'10), 17–19, Hong Kong, China, 2010.
[12] KUMAR, B., JHA, A., DESHPANDE, V., SREENVINASULU, G., "Regression Model for Sediment Transport Problems Using Multi-Gene Symbolic Genetic Programming", Computers and Electronics in Agriculture, 103, 82–90, 2014.
[13] MUDULI, K., DAS, S.K., "CPT-Based Seismic Liquefaction Potential Evaluation Using Multi-Gene Genetic Programming Approach", Indian Geotecnical Journal, 1, 86-93, 2013.
[14] KARABORAN, O., Serbest akiferlerdeki Yeraltı Su Seviyeleri ile Meteorolojik Veriler Arasındaki İlişkinin Yapay Sinir Ağları ile Belirlenmesi, Yükesk Lisans Tezi, Erciyes Üniversitesi, Fen Bilimleri Enstitüsü, Kayseri, 2013.
[15] KOZA, J.R., Genetic Programming. Massachusetts Institute of Technology, MIT Press, London, UK, 1992.
[16] KOZA, J.R., Genetic Programming III: Darwinian Invention and Problem Solving, Morgan Kaufmann, San Francisco, USA, 1999.
[17] TELCİOĞLU, M.B., Veri Madenciliğinde Genetik Programlama Temelli Yeni Bir Sınıflandırma Yaklaşımı ve Uygulanması, Yüksek Lisans Tezi, Erciyes Üniversitesi, Fen Bilimleri Enstitüsü, Kayseri, 2007.
[18] AYDINER, Z., Genetik Programlama ile Akışkan Kontrolü Tahmini, Yüksek Lisans Tezi, TOBB Ekonomi ve Teknoloji Üniversitesi, Fen Bilimleri Enstitüsü, Ankara, 2009.
[19] GANDOMI, A.H., ALAVI, A.H., "A New Multi-Gene Genetic Programming Approach to Non-Linear System Modeling. Part I: Materials and Structural Engineering Problems", Neural Computing and Applications, 21, 171–187, 2012.
[20] GANDOMI, A.H., ALAVI, A.H., "A New Multi-Gene Genetic Programming Approach to Non-Linear System Modeling. Part II: Geotechnical and Earthquake Engineering Problems", Neural Computing and Applications, 21, 189–201, 2012.
[21] CHAPRA, S.C., CANALE, R.P., Numerical Methods for Engineers (4th ed.), McGraw-Hill, New York, 2002.
[22] ÇİTAKOĞLU, H., ÇOBANER, M., HAKTANIR, T., YELKARA, F., "Regional Calibration of Hargreaves-Samani Equaton Based on Temperature and Humudity Data for Aegean and Inland Anatolia Region of Turkey", Internatıonal Congress on Natural and Engineering Sciences, 23–23, Sarayova, Bosna Hersek, 2015.
[23] ÇİTAKOĞLU, H., KİŞİ, Ö., HAKTANIR, T., ÇOBANER, M., "Hargreaves Denkleminin İç Anadolu Bölgesinde Referans Evapotransyonun Hesabı için Modifiye Edilmesi", İnşaat Mühendisliği'nde 100. Yıl, "Teknik Kongre", 6-7, İstanbul, Türkiye, 2012.