



Makine öğrenmesi yöntemleri ile kuraklık analizi Drought analysis with machine learning methods

Eyyup Ensar BAŞAKIN¹ , Ömer EKMEKÇİOĞLU¹ , Mehmet ÖZGER¹ 

¹İnşaat Mühendisliği Anabilim Dalı, İnşaat Fakültesi, İstanbul Teknik Üniversitesi, İstanbul, Türkiye.
basakin@itu.edu.tr; omer.ekmekcioglu@itu.edu.tr; ozgerme@itu.edu.tr

Geliş Tarihi/Received: 22.03.2019, Kabul Tarihi/Accepted: 31.07.2019

* Yazışılan yazar/Corresponding author

doi: 10.5505/pajes.2019.34392

Özel Sayı Makalesi/Special Issue Article

Öz

Çevresel etmenler canlı yaşamına doğrudan etki eden birçok doğal afeti tetiklemektedir. Bu afetlerin en önemlilerinden biri de kuraklıktır. Kuraklığın su kaynakları üzerindeki etkisi birçok şekilde canlıları etkilemektedir. Özellikle kuraklığın sebep olduğu, içme suyu ve tarımsal sulama amaçlı kullanılan su kaynaklarında görülen azalmalar, insan yaşamını önemli ölçüde tehdit edebilmektedir. Kuraklık diğer afetler gibi aniden ortaya çıkmadığı için, kuraklık oluşmadan önce tahmin edilip gerekli önlemlerin alınabilmesi imkânı bulunmaktadır. Kuraklık olayının belirlenebilmesi için çeşitli kuraklık indeksleri kullanılmakta ve bu da kuraklığı tahmin edebile imkânı sunmaktadır. Zaman içerisinde büyük değişiklikler gösteren kuraklık indekslerinin tahmini için birçok araştırma yapılmıştır. Bu çalışmada Kayseri iline ait 116 yıllık Palmer Kuraklık Şiddet İndeksi (PDSI - Palmer Drought Severity Index) değerleri, makine öğrenmesi yöntemleri kullanılarak modellenmiş olup, bir, üç ve altı ay sonraki kuraklık değerleri tahmin edilmiştir. Destek vektör makineleri (SVM) ve K-en yakın komşuluk (KNN) algoritmaları kullanılarak oluşturulan modeller ile yapılan tahminlerin başarı oranı istatistiksel olarak değerlendirilmiştir. Yapılan bu çalışma göstermiştir ki, makine öğrenmesi yöntemleri kuraklık problemlerin çözümüne önemli ölçüde katkı sağlamaktadır.

Anahtar kelimeler: Destek vektör makinesi, K-en yakın komşuluk, Kuraklık, Tahmin, PDSI

Abstract

Environmental factors, which directly affect the living beings cause the formation of many natural disasters. One of the most important of these disasters is drought. The effect of the drought on the water resources also affects many things in the way of living life. From the point of human life, diminution in water resources, may pose a significant threat. Drought does not appear suddenly, hence it is possible to predict and take necessary measures before it exists. In order to predict the drought, various drought indices are used to determine the drought phenomenon. A great deal of research has been made to estimate the drought values that have changed dramatically so far. In this study, the 116 - year Palmer Drought Severity Index (PDSI) values of Kayseri province were modeled using machine learning methods in order to predict future PDSI values. In this context, one, three and six months period of drought values were predicted. The success rate of the predictions constructed using support vector machines (SVM) and K-nearest neighbors (KNN) algorithms was evaluated statistically. This study indicates that machine learning methods provide a significant contribution to the solution of hydrological problems.

Keywords: Support vector machine, K- nearest neighborhood, Drought, Estimation, PDSI

1 Giriş

Kuraklık, azalan yağışlar sebebiyle su kaynaklarında yaşanan eksiklik ve bu eksiklik sebebiyle doğal yaşam sürecini sekteye uğratabilecek bir doğal afet olarak tanımlanabilir. Kuraklığı diğer doğal afetlerden ayıran en büyük fark ise ani olarak değil, yavaş şekilde gerçekleşiyor olmasıdır. Kuraklığın süresini ve şiddetini artırarak büyük bir doğal afet haline gelmesine neden olan etmenlerin ise küresel ısınma ve iklim değişikliği olduğu söylenebilir. Canlı yaşamında çok büyük öneme sahip olan su kaynakları, kuraklık neticesinde son derece olumsuz etkilenmektedir. Artan insan nüfusuna cevap verebilecek içme ve kullanma sularındaki noksanlık, bölgesel göç dalgalarına sebebiyet verebileceği gibi, diğer canlıları olumsuz etkileyerek doğal yaşam zincirinde büyük kopukluklara yol açabilecektir. Yaşanması muhtemel sorunları önceden tahmin edip, gereken önlemleri almak ise başta bilim insanlarının, daha sonra da tüm insanlığın ortak görevidir. Gelecekte gerçekleşmesi muhtemel kuraklık olayları tespit edilebilirse, su kaynakları yönetiminde hayati öneme sahip önlemler çok daha kolay şekilde alınabilecektir.

Literatürde üç çeşit kuraklık bulunmaktadır. Bunlar; yağış değerlerinin uzun süreler içerisinde olması gereken (normal) değerlerinin altında seyrettiği meteorolojik kuraklık, meteorolojik kuraklıkla çok yakından ilişki olarak tarımsal

ürünlerin yeterli su ihtiyacının karşılanmadığı tarımsal kuraklık ve son olarak meteorolojik kuraklığın süresinin uzaması neticesinde yer üstü ve yer altı sularının azaldığı hidrolojik kuraklık olarak sıralanabilir [1].

Kuraklık değerlerinin tahmini fiziksel modeller kurularak elde edilebileceği gibi matematiksel modeller yardımı ile de yapılabilmektedir. Literatür incelendiğinde, kuraklık değerlerinin tahmini için kullanılan birçok matematik model yönteminin var olduğu görülmektedir. İtalya'nın Sicilya bölgesindeki Palmer Kuraklık İndeksi değerlerini tahmin etmek için, yapay sinir ağlarının kullanıldığı çalışmada geniş ölçekli iklim indeksi değerleri kullanılmıştır [2]. Kuzey Atlantik Salınımı ve Avrupa Bloğu İndekslerinin girdi olarak kullanıldığı modeller, yalnızca PDSI değerleri kullanılarak oluşturulan modellerden daha başarılı tahminlerde bulunmuşlardır. Yapay sinir ağları ve dalgacık dönüşümü tekniklerinin kullanıldığı çalışmada kuraklık tahmini yapılmıştır [3]. Ham zaman serisi değerleri dalgacık dönüşümü yöntemi ile alt bantlara ayrılmış ve bu bantlar kullanılarak model oluşturulmuştur. Doğrusal olmayan modeller için oldukça kullanışlı olan bu ayrıştırma yöntemi sayesinde başarılı tahminler gerçekleştirilmiştir. Makine öğrenmesi algoritmaları ve sinir ağları kullanılarak yapılan tahmin çalışmasında Belayneh ve diğ. [4], DVM algoritmasının tahmin başarısının, sinir ağları ile oluşturulan modellerin tahmin başarısından daha yüksek olduğunu

gözlemlemişlerdir. Ayrıca iki algoritmaya da uygulanan ayırıştırma teknikleri model başarısını daha da yükseltmiştir. Sonuçlar istatistiksel olarak incelendiğinde, DVM yönteminin daha iyi tahminlerde bulunduğu gösterilmiştir. Etiyopya'daki Awash Nehir Havzası'nda, kısa süreli kuraklık tahmini için Belayneh ve diğ. [5], Standartlaştırılmış Yağış İndeksi (SPI)'ni kullanmışlardır. Çalışmada; yapay sinir ağları (YSA), destek vektör regresyonu (DVR) ve dalgacık analizi (WA) kullanılarak modeller oluşturulmuştur. Her üç yöntemin tahmin sonuçları iki performans ölçütü (RMSE ve R²) kullanılarak karşılaştırılmıştır. Çalışmanın tahmin sonuçları, birleştirilmiş dalgacık sinir ağı (WA-ANN) modellerinin, 6 ay sonraki SPI değerleri tahmininde, 0.79 CE (coefficient of efficiency) değeri ile 0.67 CE değerine sahip normal modelden daha başarılı olduğunu göstermiştir. Araştırmacılar SPI 3 (3 aylık SPI) ve SPI 6 (6 aylık SPI) değerlerini tahmin etmek için en doğru modelin dalgacık sinir ağı (WA-ANN) olduğunu tespit etmişlerdir. El Nino-Güney Salınımı'nın (ENSO) bağımsız değişken olarak alındığı ve kuraklık özelliklerinin kalıcılığı da göz önünde bulundurulmuş çalışmada, uzun dönem kuraklık tahmini için dalgacık dönüşümü ve bulanık mantık (WFL) birleşik modeli geliştirilmiştir. WFL'nin ana fikri, her bir belirleyiciyi farklı frekansta bantlara ayırarak tahmin etmek ve tahmin edilen serileri, öngörülen bantlarını kullanarak yeniden oluşturmaktır. Çalışmalarında ayrıca, yapay sinir ağı (YSA) modeli ile birleştirilmiş dalgacık (WANN) model sonuçları karşılaştırılmış olup WFL'nin kuraklık tahmini için daha doğru olduğunu göstermişlerdir. Dalgacık dönüşümlü bulanık mantık modelinin R² değeri 0.90, normal bulanık mantık modelinin R² değerinin ise 0.12 olarak hesaplandığı görülmüştür. Ayrıca, ENSO değişkenliğinin, küresel bir kuraklık öncüsü olmadığını belirtmişlerdir. Bu nedenle, farklı bölgelerde böyle bir modelin uygulanmasından önce, uygun bağımsız belirleyicileri tanımlamak için önemli bir çalışma gerektirdiğini de vurgulamışlardır [6]. Eksik verilerin tamamlanması üzerine yapılan çalışmada Ma ve Zhong [7], k-en yakın komşu algoritmasını kullanmışlardır. Eksik verilerin tamamlanması için öncelikle ham değerler bulanık sayılar ile ifade edilmiş daha sonra k-en yakın komşu algoritması yardımıyla tahmin yapılmıştır. Öngörülen yöntemi kuraklık verileri üzerinde denemiş ve başarılı sonuçlar almışlardır. Yapılan çalışmada dalgacık bulanık mantık modeli (WFL) geliştirilmiştir [8]. Önerilen model, öngörücü olarak yaygın olarak bulunan çökeltme, sıcaklık ve büyük ölçekli iklim indeksleri ve PDSI kullanılmaktadır. WFL modelini, Teksas'taki on iklim bölümüne uygulamış ve performansı, geleneksel bulanık mantık (FL) modeli performansıyla karşılaştırmışlardır. WFL modelinin FL modelini geride bıraktığı gösterilmiştir. WFL model performansının, yağış zaman serilerinin ortalama dalgacık spektrumları ile değişimi değerlendirilmiştir. Sonuçlar WFL modelinin PDSI'yi tahmin edebildiğini göstermektedir. Mokhtarzad ve diğ. [9], İran'da bulunan Bojnourd bölgesinin meteorolojik verilerini kullanarak SPI değerlerini tahmin etmeye çalışmıştır. SPI değerlerinin tahmini için girdi olarak sıcaklık, nem ve yağış değerlerini kullanmışlardır. Tahmin modelleri oluşturulurken ise ANN, Nöro-bulanık adaptif çıkarım sistemleri (ANFIS) ve SVM yöntemlerini kullanmışlardır. Çalışma sonucunda SPI değerlerinin tahmin edilmesinde en yüksek başarıya sahip modelin SVM olduğunu kanıtlamışlardır. Ganguli ve Reddy [10], istatistiksel dağılım ölçütlerinden olan kapula yöntemlerini ve SVM yöntemini kullanarak SPI değerlerini tahmin etmişlerdir. Çalışmalarında 3 ay sonraki SPI değerlerini tahmin etmek üzerine yoğunlaşan araştırmacılar, kapula tabanlı SVM yönteminin daha güçlü

sonuçlar doğuracağını kanıtlamıştır. El Ibrahimi ve Baali, ANFIS, çok katmanlı algılayıcı yapay sinir ağı (ANN-MLP) ve DVR kullanarak kuraklığı tahmin etmeyi amaçlamışlardır. Bu amaçla, Ain Bittit istasyonundan elde ettikleri yağış verilerini kullanmışlardır. Ayrıca SPI değerlerini kısa vadeli (3 ve 6 ay) ve uzun vadeli (9 ve 12 ay) zaman ölçekleri için hesaplamışlardır. SPI'yi (t + 1) süre için tahmin etmek amacıyla, önceki aylardan itibaren yağışlar kullanılmıştır. Sonuçlar, zaman ölçeklerinin çoğunda, ANFIS modelinin SVM ve ANN modellerinden daha doğru sonuçlar verdiğini göstermiştir [11].

Bu çalışmada 1, 3 ve 6 ay sonrası PDSI değerlerini tahmin etmek için En Yakın Komşuluk ve DVM yöntemleri kullanılmıştır. Daha sonra model sonuçlarını iyileştirmek için Dalgacık-En Yakın Komşuluk (DKNN) ve Dalgacık-Destek Vektör Makinesi (DDVM) hibrid teknikleri kullanılmıştır.

2 Materyal ve yöntem

2.1 Palmer kuraklık şiddeti indeksi (PDSI)

Belirli bir bölgedeki toprak-su dengesine dayanan kuraklık indeksi olan PDSI çok yaygın biçimde kullanılan meteorolojik bir kuraklık indeksidir. PDSI faydalı toprak-su içeriğinin yanı sıra yağış ve sıcaklık verileri temel alınarak hesaplanmaktadır. Palmer indeksi genellikle aylık olarak hesaplanmaktadır. PDSI değeri hesaplarında, toplam yağış, ortalama sıcaklık, faydalı su içeriği ve potansiyel evapotranspirasyon değerleri kullanılmaktadır [12].

Palmer kuraklık indeksi aşağıdaki şekilde hesaplanmaktadır.

1. İlk olarak "d" ile gösterilen nemlilik anomalisi hesaplanır:

$$d = P_j - \hat{P}_j \quad (1)$$

Denklem (1)'de bulunan P_j, j'nci aydaki toplam yağış miktarı, \hat{P}_j , j'nci aydaki klimatolojik koşulların oluşturduğu uygun yağış miktarını ifade etmektedir.

2. Denklem (1)'deki \hat{P}_j ifadesi ise;

$$\hat{P}_j = \alpha PET_j + \beta PR_j + \gamma PRO_j + \delta PL_j \quad (2)$$

şeklinde hesaplanmaktadır. Denklem (2)'de bulunan ifadeler ise sırasıyla;

PET= Potansiyel Evapotranspirasyon

PR= Potansiyel Toprak Nemi

PRO= Potansiyel Akış

PL= Potansiyel Toprak Nem Kaybı

Denklem (2)'de bulunan α , β , γ , δ ifadelerinin hesaplanması ise kısaca Denklem (3)'de verilmiştir.

$$\alpha_j = \frac{\overline{ET}_j}{\overline{PET}_j}, \quad \beta_j = \frac{\overline{R}_j}{\overline{PR}_j}, \quad \gamma_j = \frac{\overline{RO}_j}{\overline{PRO}_j}, \quad \delta_j = \frac{\overline{L}_j}{\overline{PL}_j} \quad (3)$$

ET; evapotranspirasyon, PET; potansiyel evapotranspirasyon, R; toprak su dolumu, PR; potansiyel toprak su dolumu, RO; akış, PRO; potansiyel akış, L; kayıp su, PL; potansiyel kayıp su.

3. Palmer Nem Anomali İndeksi hesaplanır.

$$Z_i = K_i \times d_i \quad (4)$$

Denklem (4)'de bulunan K_i; i'nci ayın ağırlık faktörüdür. Ağırlık faktörünün hesabı ise;

$$K_i = 17.67 \times \frac{K'_i}{\sum_{i=1}^{12} \bar{D}_i \times k'_i} \quad (5)$$

şeklinde hesaplanmaktadır. Denklem (5)' de bulunan K'_i ; bölgesel iklim düzeltme katsayısını ifade etmektedir.

4. Bölgesel iklim düzeltme katsayısı ise denklem (6) yardımıyla aşağıdaki şekilde hesaplanır;

$$K'_i = 1.5 \times \log_{10} \left[\left(\frac{M_i + 2.8}{\bar{D}_i} \right) \right] + 0.5 \quad (6)$$

$$M'_i = \left(\frac{\bar{P}E + \bar{R} + \bar{R}O}{\bar{P} + \bar{L}} \right) \quad (7)$$

5. Son olarak PDSI değeri;
- $$X_i = 0.897X_{i-1} + \frac{Z_i}{3} \quad (8)$$

şeklinde hesaplanır.

Tablo 1'de PDSI'ya göre kuraklık değerlerinin sınıflandırılması verilmiştir. Değerler -4 ve 4 arasında değişmekle birlikte -4' ten küçük 4' ten büyük değer de alabilmektedir.

Tablo 1: PDSI sınıflandırma tablosu.

PDSI Değeri	Kuraklık Sınıfı
≥ 4	Aşırı nemli
3 3.99	Çok nemli
2 2.99	Orta nemli
1 1.99	Hafif nemli
0.5 0.99	Yeni başlayan nemli dönem
0.49 -0.40	Normal
-0.5 -0.99	Yeni başlayan kurak dönem
-1 -1.99	Hafif kurak
-2 -2.99	Orta kurak
-3 -3.99	Şiddetli kurak
≤ -4	Aşırı kurak

2.2 Destek vektör makinesi (DVM)

DVM, yapısal risklerin minimum hale getirilmesi kuralı ile çalışan optimizasyona dayalı makine öğrenmesi algoritmasıdır. DVM sınıflandırma ve regresyon problemlerinin çözümü için kullanılabilir [13]. Amaç, verilerin sınıflandırılması sırasında doğru şekilde ayrılabilmesini sağlamaktır. Bu ayırma işleminde doğrusal bir sınır çizgisi kullanır ve sonsuz sayıda doğrusal sınır çizgisi çizebilmek mümkündür. En iyi ayırmayı sağlayan çizgi için marjin adı verilen aralık kullanır. Marjin aralığı çizilirken en geniş aralığın bulunması hedeflenir. Maksimum marjin bulunabilmesi için aşağıdaki denklem çözülmelidir. Minimizasyon:

$$\min 1/2 \|w\|^2 \quad (9)$$

Kısıt:

$$y_i(w \cdot x_i + b) \geq +1, \quad \forall i \quad (10)$$

Denklem (10)'da; w ; ağırlık vektörü, x ; hiper düzlem üzerinde bir nokta ve b ; bias olarak adlandırılmaktadır. Denklem (9) çözülmek istenirse doğrusal olmayan bir optimizasyon problemi ile karşılaşılır ve bu problem Lagrange fonksiyonu kullanılarak çözülür. Problem kuadratik olduğundan, optimizasyon süreci her zaman global bir maksimum ile sonuçlanmaktadır. O halde maksimal marj sınıflandırıcı;

$$f(x) = \text{sig} \left(\sum_{i=1}^l y_i \alpha_i \langle x_i, x \rangle + b \right) \quad (11)$$

kullanılarak türetilir.

Doğrusal ayrılabilir marj, mühendislikte karşılaşılan problemlerin çoğu için uygun çözümü üretmekte yetersiz kalabilir. Bu durumu ortadan kaldırmak için yumuşak marj adı verilen hataya müsaade eden bir sistem geliştirilmiştir. Marj içerisindeki tüm veriler, ister ayırıcı doğrunun yanlış tarafında isterse doğru tarafında olsun ihmal edilerek, hatayla işbirliği yapmak mümkündür. Yumuşak marjlar, hatayı minimize ederek ayrılmayan sınıfları çözmeye yardımcı olmaktadır. Verilerin doğrusal olarak ayrılmadığı durumlarda sınıflandırma işlemi, çekirdek fonksiyonları kullanılarak verilerin yüksek boyutlu bir uzaya taşınması ile gerçekleşir.

En çok kullanılan çekirdek fonksiyonları şunlardır:

- Doğrusal Çekirdek Fonksiyonu:

$$K(x_i, x_j) = x_i^T \cdot x_j \quad (12)$$

- Polinom Çekirdek Fonksiyonu:

$$K(x_i, x_j) = (\gamma x_i^T \cdot x_j + r)^d \quad \gamma > 0 \quad (13)$$

- Sigmoid Çekirdek Fonksiyonu

$$K(x_i, x_j) = \tanh(\gamma x_i^T \cdot x_j + r)^d \quad (14)$$

- Radyal Tabanlı Çekirdek Fonksiyonu:

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad \gamma > 0 \quad (15)$$

Burada γ , r ve d çekirdek parametreleridir.

Bu çalışmada radyal tabanlı çekirdek fonksiyonu kullanılmıştır.

2.3 K-en yakın komşuluk

K-En Yakın Komşuluk tahmin yöntemi parametrik olmayan sınıflandırma ve regresyon algoritmasıdır. Uygulanabilme kolaylığı, basit bir matematiksel temele sahip olması pek çok farklı alanda tahmin modelleri kullanımını yaygınlaştırmıştır. Bir vakanın sonucu kendisine en yakın komşu vakalarının sonucu ile aynı olur fikrini esas alır. Geçmiş gözlemlere dayalı olan eğitim seti vasıtasıyla, veri setinin her elemanının (vakanın) sonucu olan bağımlı değişkenler belirlenir. İleriye dönük tahminler, mevcut vakaların sonuçlarının eğitim veri setindeki en yakın elemanların sonuçlarının ortalamasına eşit olacaktır. İsminde geçen "k" harfi komşu sayısını ifade etmektedir. En iyi sınıflamanın veya tahminin yapılması için k sayısının önemi büyüktür. Doğru k sayısının seçimi için deneme yanılma ya da çapraz doğrulama tekniklerinden yararlanılabilir [14]. Verilerin hangi sınıfa ait olduğu, eğitim kümesinde en yakın mesafede olduğu hesaplanan k adet verinin ortalaması alınarak bulunabilmektedir. Bulunan değer yorumlanması için eşik değer belirlenmelidir. Regresyon problemlerinde eşik değere gerek olmamaktadır. Bağımsız parametreleri birbirine olan yakınlıkları ile tahminledikleri değerlerin ortalaması alınarak istenen değer hesaplanabilir.

Literatürde en çok kullanılan 3 uzaklık ölçüsü bulunmaktadır. Bu uzaklık ölçüleri; Öklid uzaklığı, Manhattan uzaklığı ve Chebyshev uzaklığıdır. Öklid uzaklığı genelde en çok kullanılan uzaklık ölçüsüdür. Öklid uzaklığı denklem (16)'de tanımlanmıştır.

$$d_{i,j} = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{ip} - x_{jp})^2} \quad (16)$$

Manhattan uzaklık ölçüsü ise denklem 17'de tanımlanmıştır.

$$d_{i,j} = (|x_{i1} - x_{j1}| + |x_{i2} - x_{j2}| + \dots + |x_{ip} - x_{jp}|) \quad (17)$$

Chebyshev uzaklığı iki nesne arasındaki farkın mutlak büyüklüğünü dikkate alır. Maksimum değer uzaklığı da denir ve denklem (18)'de tanımlanmıştır.

Denklemlerde bulunan $d_{i,j}$ uzaklıkları, x_{i1} ve x_{j1} bağımsız değişkenlerin konumunu ifade etmektedir.

$$d_{i,j} = \max_m |x_{1m} - x_{2m}| \quad (18)$$

2.4 Dalgacık dönüşümü

Zaman serilerinin analizinde frekans bilgisinin eldesi büyük önem taşımaktadır. Dalgacık dönüşümünün temelleri klasik bir frekans analiz yöntemi olan Fourier dönüşümüne dayanır. Fourier dönüşümü, zaman serilerinin sürekli sinüs dalgaları ile temsil edilerek frekans boyutuna taşınmasını sağlar. Fourier dönüşümünün durağan olmayan serileri temsil edememesi, zaman ve frekans bilgisinin eşzamanlı çözümlememesi ve sinüs dalgalarının geometrisinin ani sıçramaları yakalamadaki yetersizliği gibi eksikliklerin giderilmesi amacıyla zaman içerisinde dalgacık dönüşümü geliştirilmiştir. Dalgacık dönüşümünün temel prensibini pencereleme tekniğinde de görülen kaydırma ve dalgacık dönüşümü ile geliştirilen ölçeklendirme işlemleri oluşturur.

Dalgacıklar, sinyalin frekans bilgisini, kendi ölçeği ile eşleştirilmiş çözümlülüklerde işleyen, fonksiyonlardır. Dalgacık dönüşümde arkasındaki ana fikir, sinyalleri belirlenen bir ölçeğe göre analiz etmektir [15]. Dalgacık fonksiyonları keskin geometri, sıkışabilirlik, sonlu enerjisi özellikleri ile zaman serilerindeki değişimleri Fourier fonksiyonlarından daha iyi tespit edebilmektedir. Dalgacık analizi; zaman-ölçek alanını kullanarak analiz yapmaktadır. Dalgacık analizinde en önemli özellik yerel analiz yapabilmesidir. Bunlarla birlikte dalgacık analizi geleneksel yöntemlerle karşılaştırıldığında, bir sinyali sıkıştırma veya arındırma işlemini, sinyalin orijinalini bozmadan yapılabilir. Dalgacık dönüşümleri; Sürekli Dalgacık Dönüşümü (SDD) ve Ayrık Dalgacık Dönüşümü (ADD) olmak üzere 2'ye ayrılır. Matematiksel kuram olarak 2 metodunda birbirinden farkı yoktur. Bu çalışmada sürekli dalgacık dönüşümü kullanılmıştır.

Dalgacık dönüşümünde kullanılan ana dalgacık fonksiyonu ölçek parametresi ve konum parametresine bağlı olarak denklem 19 yardımıyla hesaplanabilmektedir. İfadede bulunan a ve b ölçek parametresi, Ψ^* ana dalgacık fonksiyonunun eşleniğini ifade etmektedir. Ölçek parametresi olan a, Fourier dönüşümündeki 1/frekans ifadesi ile eşleşmektedir. Dolayısıyla dar yani düşük ölçekli dalgacıklar yüksek frekans bilgisine sahip detay bileşenlerin tespitinde, büyük ölçekli yüksek zaman bilgisine sahip dalgacıklar ise serinin eğilim karakterini tespitinde fayda sağlar.

$$\Psi_{a,b}(t) = |a|^{-\frac{1}{2}} \Psi\left(\frac{t-a}{b}\right) \quad (19)$$

Dalgacık fonksiyonu $\Psi(t)$ ana fonksiyonunun "a" kadar ölçeklendirilip "b" kadar zaman ekseninde ötelenmesi ile elde edilir. Dönüşüm işlemi, dalgacık fonksiyonunun seçilen ölçekte her zaman adımında zaman serisi $f(t)$ ile iç çarpımınıdır.

Sürekli dalgacık dönüşümü (SDD) prototip bir dalgacığın farklı ölçeklerde zaman serisi boyunca çarpılmasıyla elde edilir. Sinyal dalgacık çarpımı sonucu elde edilen dönüşüm

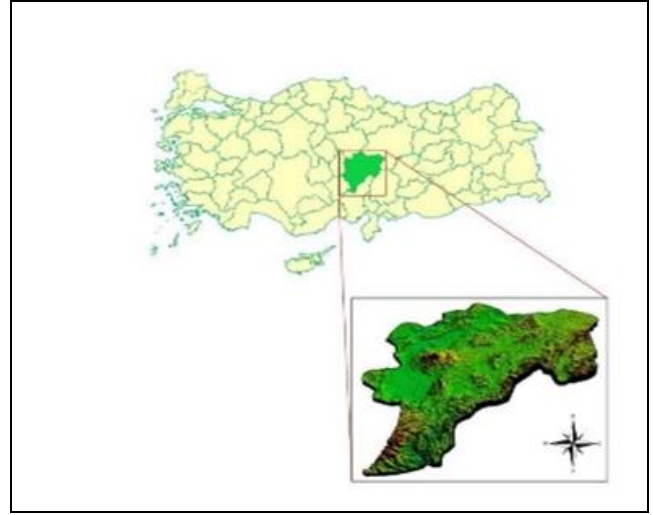
katsayıları, dalgacık ve zaman serisi uyumuna işaret eder. Sürekli dalgacık dönüşümü tüm ölçek ve öteleme adımlarında elde edilen dönüşüm katsayıları, detaylı sinyal bilgisini içerir ancak bu bilgilerin tekrar eden değerleri içermesi yüksek kapasitede hafıza gerektirir. Bu durum işlem hacmini arttırmaktadır. Sürekli dalgacık dönüşümü katsayıları denklem (20) yardımıyla hesaplanmaktadır.

$$DD\{f(t), \Psi_{a,b}(t)\} = |a|^{-\frac{1}{2}} \int f(t) \Psi^*\left(\frac{t-a}{b}\right) dt \quad (20)$$

Skaler çarpım işlemi sonucunda elde edilen katsayılar dalgacık ve zaman serisinin korelasyonunu temsil eder. Yüksek katsayılar dalgacığın zaman serisini temsil etmedeki başarısına işaret eder.

2.5 Çalışma alanı ve veriler

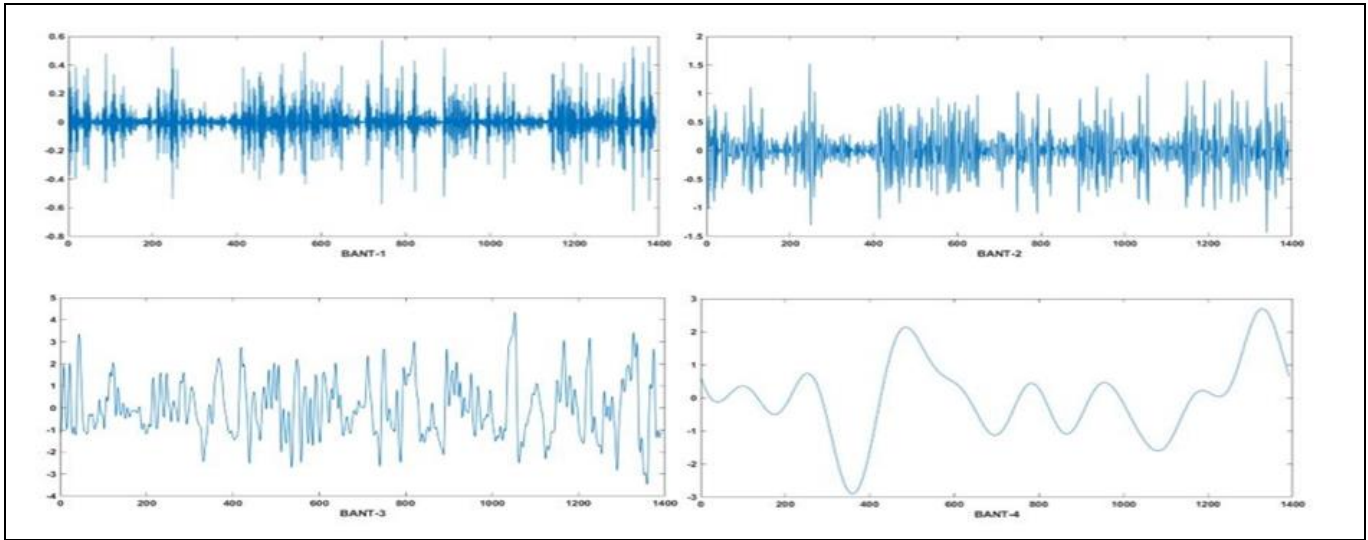
Bu çalışmada kullanılan veriler Kayseri iline ait 116 yıllık PDSI değerleridir. PDSI değerleri zaman serisi grafiği Şekil 2'de verilmiştir. Kayseri ili Türkiye'nin Orta Anadolu Bölgesi'nin güneyinde bulunmaktadır. Kayseri Kızılırmak havzasına dâhil olup, yaklaşık 17 bin km² yüzölçümüne sahiptir (Şekil 1).



Şekil 1: Çalışma alanı.

Bu bölgede karasal iklim hâkimdir. Meyve, sebze, şeker pancarı, hububat gibi tarım ürünleri üretimine, büyükbaş ve küçükbaş hayvan üretimine uygun iklim özelliklerine sahip olması çalışma alanı olarak Kayseri ilinin seçiminde önemli rol oynamıştır [16]. PDSI değerleri 1900-2016 yılları arasında modellenmiş aylık, 1392 adet veriden oluşmaktadır. Şekil 2'deki grafik incelendiğinde PDSI değerleri "orta kurak" sınıfı içerisinde salınım göstermektedir. Zaman serisi analizi yapıldığı için tarihsel sırası bozulmadan ilk 900 veri ile model eğitilmiş geri kalan 492 adet veri ile model test edilmiştir.

PDSI değerleri <https://crudata.uea.ac.uk/cru/data/drought/> isimli bağlantıdan temin edilmiştir. Kullanılan değerler Wells ve diğ. [17]'nin çalışmaları kapsamında interpolasyon yöntemi ile üretilen değerlerdir. PDSI değerlerinin istatistiksel özellikleri incelendiğinde, en yüksek değer 5.3, en düşük değer -5.3, ortalama değer -0.923, standart sapma değerinin 1.73 ve çarpıklık katsayısının 0.361 olduğu görülmüştür. Çarpıklık katsayısı göz önüne alındığında verilerin simetrik dağılmadığı dolayısıyla normal dağılıma sahip olmadığı anlaşılmaktadır. Ayrıca yapılan Ki-Kare dağılım testine göre de verilerin normal dağılmadığı sonucuna varılmıştır.



Şekil 2: Dalgacık dönüşümü ile ayrıştırılan yeni zaman serileri.

2.6 Model performans başarı ölçütleri

Bu çalışma kapsamında kurulan tahmin modellerinin başarısının sınanması için hata kareleri ortalaması (HKO) ve Nash-Sutcliffe, (NS) katsayısı ölçütleri kullanılmıştır. HKO değeri 0 ile sonsuz arasında değişmekte olup en iyi değer 0 olarak kabul edilir. Sıfır (0) değeri tahmin işleminin hatasız yapıldığını göstermektedir. NS katsayısı ise hata kareleri ortalamaları ile gerçek değerlerin varyansının bir oranıdır. Bu oran 1'den çıkarılarak bir katsayı elde edilir. Elde edilen katsayı 1 ile eksi sonsuz arasında değişmektedir ve en iyi tahmini temsil eden değer 1'dir. Literatürde kabul gören oranlar ise, 0.5-0.7 arası "kabul edilebilir tahmin", 0.7-0.9 arası "iyi tahmin" ve 0.9-1 arası ise "mükemmel" tahmin olarak adlandırılabilir [18]. HKO ve NS değerlerinin denklemleri denklem (21) ve denklem (22)'de verilmiştir.

$$HKO = \sqrt{\frac{1}{n} \sum_{i=1}^n (P_{pi} - P_{oi})^2} \quad (21)$$

$$NS = 1 - \frac{\sum_{i=1}^n (P_{pi} - P_{oi})^2}{\sum_{i=1}^n (P_{oi} - \bar{P}_o)^2} \quad (22)$$

Denklemlerdeki ifadeler; n=toplam gözlem ve tahmin değeri sayısı, P_{pi} =i'nci PDSI tahmin değeri, P_{oi} =i'nci PDSI gözlenen değeri, \bar{P}_o =Gözlenen PDSI değerleri ortalamasıdır.

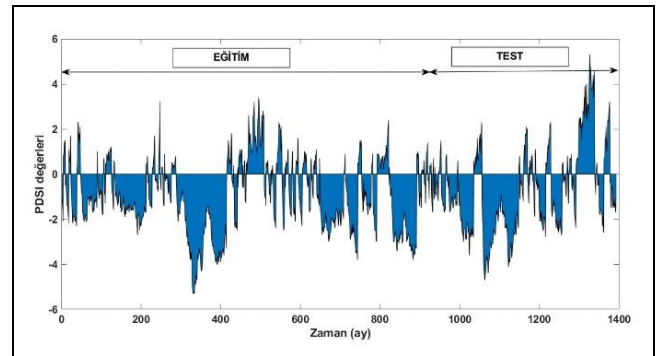
3 Sonuç ve tartışma

Uygulamada modeller, makine öğrenmesi yöntemlerinden K-en yakın komşuluk algoritması ve DVM, STATISTICA paket programı yardımıyla oluşturulmuştur. Her iki yöntemde de farklı girdi değişkenleri ile birçok model oluşturulmuştur. Bir, iki ve üç ay önceki değerleri içeren modellerden en düşük hatayı veren model olan, bir ve iki ay önceki (T-1, T-2) girdi değerlerinin beraber modele dahil edildiği kombinasyon, seçilmiştir. Ayrıca çalışmada bir, üç ve altı ay sonraki PDSI değerleri tahmin edilmeye çalışılmıştır.

Tahminler istatistik değerlendirme kriterleri olan hata kareleri ortalaması (HKO) ve Nash-Sutcliffe (NS) [19] denklemleri yardımıyla değerlendirilmiştir. K-en yakın komşuluk

algoritmasında komşu sayısı, DVM'de ise çekirdek fonksiyonu parametreleri deneme yanılma yoluyla tespit edilmiştir.

Çalışma kapsamında farklı frekansta 4 bantta ayrılan dalgacık dönüşümlü modeller, klasik dalgacık modellerinden farklı olarak, tüm bantların girdi olarak alındığı bir model oluşturulmuştur. Dalgacık dönüşümü ile ayrıştırılan bantlar Şekil 2'de gösterilmiştir. İlk bantın orijinal zaman serisi ile arasında çok düşük korelasyon bulunması, bu bantın beyaz gürültü olarak adlandırılmasının nedeni olmaktadır. Bu bant istatistiksel ve grafiksel olarak incelendiğinde otokorelasyonun sıfıra çok yakın olduğu da görülmektedir. Bu özelliği ile de rastgele bir davranış sergilediği düşünülmektedir. Üçüncü ve dördüncü bant ise orijinal zaman serisinin karakterine en yakın bantlardır. Şekil 3'te ise eğitim ve test serilerinin ayrımı ifade edilmektedir.



Şekil 3: PDSI değerleri zaman serisi grafiği.

Klasik çalışmalarda her bant kendi içerisinde tahmin edilip toplandıktan sonra tahmin gerçekleştirilir. Bu çalışmada ise işlem sayısını azaltmak adına bantlar 4 girdi şeklinde sisteme tanıtılmış ve istenilen çıktıyla doğrudan bağlantı kurması sağlanmıştır. Bu da çalışmamızın özgün değere sahip olmasını sağlayan en önemli aşama olarak görülmektedir.

Ayrıca k-en yakın komşuluk algoritması kullanılarak yapılan tahminlerin son derece karmaşık olan DVM'ye yakın sonuç verdiği görülmüş, önışleme tabi tutulan zaman serilerinin basit algoritmalar kullanılarak iyi tahminler gerçekleştirildiği göz önüne serilmiştir. Modellere ait tahmin değerleri Tablo 2 ve Tablo 3'te gösterilmiştir.

Tablo 2: K-en yakın komşuluk tahmin modelleri ve sonuç tablosu.

Model	Girdi	Çıktı	HKO	NS
MODEL1	P(T-1)	P(T)	0.507	0.859
MODEL2	P(T-1),P(T-2)	P(T)	0.519	0.857
MODEL3	P(T-1),P(T-2),P(T-3)	P(T)	0.522	0.856
MODEL4	P(T-1),P	P(T+3)	1.297	0.644
MODEL5	P(T-1),P	P(T+6)	2.079	0.432
MODEL6(Dalgacık Dönüşümü)	P(T) _{B1} ,P(T) _{B2} ,P(T) _{B3} ,P(T) _{B4}	P(T+3)	0.867	0.761
MODEL7(Dalgacık Dönüşümü)	P(T) _{B1} ,P(T) _{B2} ,P(T) _{B3} ,P(T) _{B4}	P(T+6)	1.595	0.563

P(T-1); bir ay önceki PDSI değeri, P(T-2); iki ay önceki PDSI değeri, P(T-3); üç ay önceki PDSI değeri, P; içinde bulunan ay PDSI değeri P(T+3); üç ay sonraki PDSI değeri, P(t+6); altı ay sonraki PDSI değeri, B1;bant 1, B2;bant 2, B3; bant3, B4; bant4 ü ifade etmektedir.

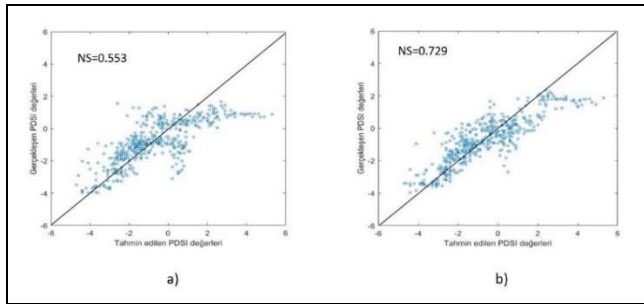
Tablo 3: Destek vektör makinası tahmin modelleri ve sonuç tablosu.

MODEL	GİRDİ	ÇIKTI	KOMŞU SAYISI	HKO	NS
MODEL1	P(T-1)	P(T)	19	0.639	0.823
MODEL2	P(T-1),P(T-2)	P(T)	10	0.609	0.832
MODEL3	P(T-1),P(T-2),P(T-3)	P(T)	19	0.643	0.823
MODEL4	P(T-1),P	P(T+3)	10	1.628	0.553
MODEL5	P(T-1),P	P(T+6)	20	2.316	0.368
MODEL6(Dalgacık Dönüşümü)	P(T) _{B1} ,P(T) _{B2} ,P(T) _{B3} ,P(T) _{B4}	P(T+3)	17	0.986	0.729
MODEL7(Dalgacık Dönüşümü)	P(T) _{B1} ,P(T) _{B2} ,P(T) _{B3} ,P(T) _{B4}	P(T+6)	15	1.761	0.518

P(T-1); bir ay önceki PDSI değeri, P(T-2); iki ay önceki PDSI değeri, P(T-3); üç ay önceki PDSI değeri, P; içinde bulunan ay PDSI değeri P(T+3); üç ay sonraki PDSI değeri, P(t+6); altı ay sonraki PDSI değeri, B1;bant 1, B2;bant 2, B3; bant3, B4; bant4.

Modellerin karakterine ışık tutması açısından gerçekleşen tahmin edilen saçılma grafikleri de aşağıda Şekil 5'te verilmiştir. Kuraklık tahmini çalışmalarında 1 ay ilerisi tahmininden, 3 ay sonraki değerlerin tahmin edilmesi daha önemli olduğu için çalışmada 3 ay sonrasına ait saçılma grafikleri verilmiştir.

Şekil 4'te görüldüğü üzere dalgacık dönüşümü sonrası 3 ay sonraki K-En Yakın Komşuluk (KNN) tahminleri için 0.553 olan NS değeri 0.729 değerine ve 6 ay sonrası için de 0.368' den 0.518'e yükselmiştir.

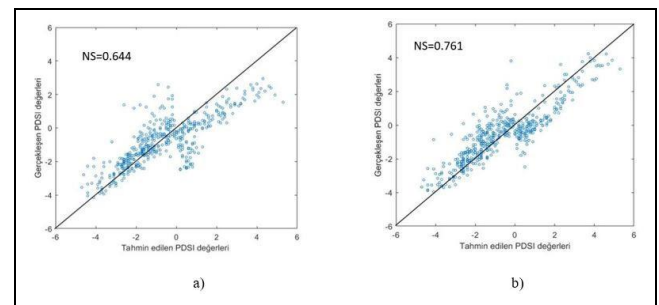


Şekil 4: 3 ay sonrası PDSI tahmin değerleri, a) KNN yöntemi tahmin değerleri, b) D-KNN tahmin değerleri.

Şekil 5'te görüldüğü üzere ise yine dalgacık dönüşümü sonrası 3 ay sonraki (DVM) tahminleri için 0.644 olan NS değeri 0.761 değerine ve 6 ay sonrası için de 0.432' den 0.563'e çıkmıştır. Bu da tek başına bir yöntem kullanmaktan dalgacık dönüşümü ile beraber hibrit modeller kullanmanın model performanslarındaki iyileşmeye yaptığı katkıyı göstermektedir. Genel olarak sonuçlar incelendiğinde, DVM tahmin sonuçları hem dalgacık dönüşümsüz modellerde hem de dalgacık dönüşümü ile birlikte kurulan modellerde, KNN yönteminden daha iyi sonuçlar vermektedir.

Sonuç olarak, dalgacık-en yakın komşuluk ve DDVM yaklaşımları 3-ay sonrası kuraklık tahminleri için tatmin edici sonuçlar (NS>0.70) verdiği tespit edilmiştir. Gelecekteki

çalışmalarda araştırmacıların önışlem yöntemlerini kullanarak kuraklık analizi yapmalarının faydalı sonuçlar doğuracağı kanaatine varılmaktadır. Karmaşık yapıya sahip modeller görece iyi sonuçlar verse de, basit modellerin de önışlem yöntemleri ile birlikte kullanımının göz ardı edilemeyeceği de bu çalışma ile varılan önemli bir diğer sonuçtur. İfade edildiği üzere kuraklık belirsizlikler içeren rasgele bir olaydır ve önceden tahmin edilmesi oldukça zordur. Kuraklık yönetiminin etkili bir şekilde planlanması ve işletilmesi için ileriye dönük uzun dönemli kuraklık tahminlerinin önemli bir girdi parametresi olarak değerlendirilmesi gerekir. Karar vericiler gelecekteki kurak devreleri ne kadar önceden bilebilirlerse zararları da o kadar iyi bertaraf etme imkanı bulacaklardır. Ayrıca kuraklık karar destek sisteminin vazgeçilmez bir parçası olan kuraklık tahminlerinin başarısı, karar vericilerin de başarısında önemli rol oynayacaktır.



Şekil 5: 3 ay sonrası PDSI tahmin değerleri, a) DVM yöntemi tahmin değerleri, b) D-DVM tahmin değerleri.

4 Kaynaklar

- [1] Wilhite DA, Glantz MH. "Understanding the Drought Phenomenon: the role of definitions". *Water International*, 10(3), 111-120, 1985.
- [2] Cutore P. "Forecasting Palmer index using neural networks and climatic indexes". *Journal of Hydrologic Engineering*, 6(14), 585-595, 2009.

- [3] Kim TW, Valdes JB. "Nonlinear model for drought forecasting based on a conjunction of wavelet transforms and neural networks". *Journal of Hydrologic Engineering*, 8(6), 319-328, 2003.
- [4] Belayneh A, Adamowski C, Khalil B, Quilty J. "Coupling machine learning methods with wavelet transforms and the bootstrap and boosting ensemble approaches for drought prediction". *Atmospheric Research*, 172, 37-47, 2016.
- [5] Belayneh A, Adamowski J. "Drought forecasting using new machine learning methods". *Journal Of Water And Land Development*, 18, 3-12, 2013.
- [6] Özger M, Mishra AK, Singh VP. "Long lead time drought forecasting using a wavelet and fuzzy logic combination model: a case study in Texas". *Journal Of Hydrometeorology*, 13, 284-297, 2012.
- [7] Ma X, Zhong Q. "Missing value imputation method for disaster decision-making using k nearest neighbor". *Journal of Applied Statistics*, 43(4), 767-781, 2016.
- [8] Özger M, Mishra AK, Singh VP. "Estimating palmer drought severity index using a wavelet fuzzy logic model based on meteorological variables". *International Journal of Climatology*, 31, 2021-2032, 2011.
- [9] Mokhtarzad M, Eskandari F, Jamshidi Vanjani N. "Drought forecasting by ANN, ANFIS, and SVM and comparison of the models". *Environmental Earth Science* 76, 729, 2017.
- [10] Ganguli P, Reddy M. "Ensemble prediction of regional droughts using climate inputs and the SVM-copula approach". *Hydrological Processes*, 28, 4989- 5009, 2014.
- [11] El Ibrahimy A, Baali A. "Application of several artificial intelligence models for forecasting meteorological drought using the standardized precipitation index in the saïss plain (Northern Morocco)". *International Journal of Intelligent Engineering and Systems*, 11, 267-275, 2018.
- [12] Palmer WC. *Meteorological Drought*. Research Paper No. 45, U.S. Weather Bureau, Washington, D.C. 1995.
- [13] Cortes C, Vapnik V. "Support-Vector networks". *Machine Learning*, 20(3), 273-297, 1995.
- [14] Yiğiter ŞY, Sarı SS, Karabulut T, Başakın EE. "Kira sertifikası fiyat değerlerinin makine öğrenmesi metodu ile tahmini ". *International Journal of Islamic Economics and Finance Studies*, 4, 74-82, 2018.
- [15] Serroukh, A, Walden AT, Percival CB. "Statistical properties and uses of the wavelet variance estimator for the scale analysis of time series". *Journal of the American Statistical Association*, 95(449), 184-196, 2000.
- [16] Kayseri İli Tarımsal Yatırım Rehberi, T.C. Gıda Tarım ve Hayvancılık Bakanlığı Strateji Geliştirme Başkanlığı, 2018.
- [17] Wells N, Goddard S, Hayes MJ. "A self-calibrating Palmer Drought Severity Index". *Journal of Climate*, 17, 2335-2351, 2004.
- [18] Altunkaynak A, Başakın EE. "Zaman serileri kullanılarak nehir akım tahmini ve farklı yöntemlerle karşılaştırılması". *Erzincan Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 11, 92-101, 2018.
- [19] Nash JE, Sutcliffe JV. "River flow forecasting through conceptual models: Part 1. A Discussion Of Principle". *Journal of Hydrology*, 10, 282-290, 1970.