

## DALGACIK K-EN YAKIN KOMŞULUK YÖNTEMİ İLE HAVA KİRLİLİĞİ TAHMİNİ

*Abdüselam ALTUNKAYNAK* \*

*Eyyup Ensar BAŞAKIN* \*

*Elif KARTAL* †\*

Alınma: 14.10.2020; düzeltme: 19.11.2020; kabul: 17.12.2020

**Öz:** Son yıllarda artan insan nüfusu ile fosil yakıt kullanımı yaygınlaşmıştır. Enerji üretimi, ulaşım, ısınma gibi birçok kullanım alanına sahip fosil yakıtların yanması sonucunda atmosfere salınan zararlı maddelerin yoğunluğu hem kentsel hem de kırsal bölgelerde insan sağlığını tehdit edecek seviyelere ulaşabilmektedir. Lokal hava kalitesini muhafaza edecek önlemler almak ve kirleticilerin zararlarını en aza indirebilmek için ileriye yönelik emisyon tahminlerinde bulunmak büyük önem arz etmektedir. Çalışmamızda yanma sonucunda açığa çıkan önemli kirleticilerden PM<sub>10</sub> ve SO<sub>2</sub> maddelerinin mevcut günlük kayıtları kullanarak gelecekte olması muhtemel değerleri tahmin edilmeye çalışılmıştır. Erzincan ilinde 2016-2018 yılları arasında ölçülmüş toplam 651 adet veri kullanılarak bir model oluşturulmuştur. Model oluşturma aşamasında verilerin ilk 400 adeti eğitim, geriye kalan 251 adet veri doğrulama olmak üzere ikiye ayrılmıştır. Modeller K-En Yakın Komşuluk (KNN) algoritması kullanılarak kurulmuş ve modelleme başarısını arttırmak adına önişlem süreçlerinden biri olan dalgacık dönüşüm tekniği uygulanmıştır. Dalgacık dönüşümü ile oluşturulan modellerin, tahmin başarısını büyük derecede iyileştirdiği gözlemlenmiştir. Bu çalışma uygulaması basit makine öğrenmesi yöntemlerinden olan KNN'nin hava kirliliği tahmin modellerinde kullanılabileceğini kanıtlamıştır.

**Anahtar Kelimeler:** K-En Yakın Komşuluk, Hava kirliliği, Dalgacık dönüşümü, Tahmin

### Air Polution Prediction with Wavelet K-Nearest Neighbour Method

**Abstract:** In the last decades, the use of fossil fuels has become widespread with the increasing human population. The concentration of harmful substances released into the atmosphere as a result of the burning of fossil fuels, which have widely used for energy production, transportation, and heating. The burning of fossil fuels can reach levels that threaten human health in both urban and rural areas. It has great importance to estimate emission to take measures to protect local air quality and to minimize the damage of pollutants. The current study aims to predict the future concentration values of PM<sub>10</sub> and SO<sub>2</sub>, which are important pollutants, by using available daily records. A predictive model is implemented for Erzincan city by using a total 651 data points observed for period from 2016 through 2018. In the modeling process, data are divided into two groups; 400 the data points are utilized for training and the remaining 251 data points are used for verification. The wavelet transform technique is combined with the K-Nearest Neighbor (KNN) method to develop a predictive model called as Wavelet- KNN approach for increasing the modeling success. In the present study, the wavelet-KNN approach is provided better prediction results compared to stand-alone KNN method. It is noted that the combination of wavelet with KNN tool is enhanced the prediction performance of model. This study shows that the KNN method is one of the simplest machine learning methods and can be used for prediction of air pollution models.

**Keywords:** K-Nearest Neighbor, Air pollution, Wavelet transformation, Prediction

\* İstanbul Teknik Üniversitesi, İnşaat Fakültesi, Maslak, 34467 Sarıyer/İstanbul  
İletişim Yazarı: Elif Kartal (kartalel@itu.edu.tr)

## 1. GİRİŞ

Hızlı kentleşme ve sanayileşmenin negatif bir getirisi sayılabilecek hava kirliliği; insan sağlığı, ekosistem ve çevre üzerinde bir dizi olumsuz etkiye sahiptir. Günümüzde kirletici emisyonlarındaki artışın insan sağlığı ve çevre kalitesi üzerindeki doğrudan ilişkisi çeşitli çalışmalarda ortaya koyulmuştur (Gamble ve Lewis, 1996; Aalto ve diğ., 2005). Yerleşim alanlarında atmosferdeki öncelikli tehditler arasında, kükürt dioksit ( $SO_2$ ) ve 10 mikron ( $\mu$ ) ve daha küçük çaplı tanecikler ( $PM_{10}$ ) kirletici ve bu kirleticilerin etkileşimi sonucu oluşan maddeler sıralanabilir. Konut ısıtma amaçlı fosil yakıtların kullanıldığı ve endüstriyel tesislerin yoğunlukta olduğu kent merkezlerinde  $SO_2$  ve  $PM_{10}$  konsantrasyonlarının yüksek olması beklenen bir durumdur. Bununla birlikte atmosferik koşullardaki değişimler kirletici konsantrasyonlarında dalgalanmalara ve ani artışlara sebep olabilir (Isakov ve diğ., 2007). Kirletici yoğunluğundaki ani yükselişlerin önceden tahmin edilmesi ile kent sakinlerinin uyarılarak sağlık koruyucu önlemler almaları sağlanabilir. Başarılı hava kalitesi tahmin modelleri erken sağlık uyarı sistemlerinin geliştirilmesine önemli katkı sağlayacaktır.

Tahmin başarısı yüksek modeller oluşturabilmek için zamansal ve mekânsal olarak durağan karakter sergilemeyen  $PM_{10}$  ve  $SO_2$  değerlerinin hassas ve düzenli ölçümlerine ihtiyaç vardır. İlerleyen teknoloji ile hava kalitesi ölçüm ağı genişlemekte, toplanan veriler artmakta ve kirleticilerin sağlık üzerinde etkileri daha detaylı incelenmektedir. Yapılan çalışmalarda bir bölgede görülen  $PM_{10}$  konsantrasyondaki artışın, hem görüş mesafesini kısıtladığı hem de kısa vadeli maruziyetlerde dahi solunum yolları rahatsızlıklarına sebep olduğu ortaya koyulmuştur (Laden ve diğ., 2000; Chen ve diğ., 2007).  $SO_2$  ise nefes almayı kısıtlayan, akciğer ve boğazlarda tahrişe sebep olan bir gazdır. Uzun süreli maruziyetlerde ise kronik solunum hastalıklarında artışa sebep olur (Chen ve diğ., 2007). Ayrıca  $SO_2$ 'nin insan sağlığı üzerindeki tahrip edici etkisi havadaki PM konsantrasyonundaki artış ile artmaktadır. Partiküler maddelerin sadece kardiyovasküler ve solunum yolları gibi sağlık sorunları ile alakalı değil, aynı zamanda ruh sağlığı üzerinde de bir etkisi olduğu görülmektedir (Dominici ve diğ., 2006).

Hava kalitesini analiz etmek amacıyla kullanılan çeşitli kirletici parametrelerin modellenmesinde geçmiş yıllarda yürütülen çalışmalarda pek çok farklı tahmin yöntemi kullanılmıştır. Yıldırım ve Bayramoğlu (2006) kentsel bir alanda meteorolojik faktörlerin havadaki  $SO_2$  ve toplam askıya alınmış madde (TSP) miktarı etkisini araştırdıkları çalışmalarında, adaptif nöro-bulanık mantık yöntemi (ANFIS) ile  $SO_2$  ve TSP konsantrasyon seviyelerindeki eğilimleri sırasıyla %75-90 ve %69-80 arasında performansla tahmin edebilmiştir. Kentsel alanlarda yürütülen kirletici salınımlarının modellenmesinde kullanılan karar ağaçları (Kalapanidas ve Avouris, 2001), çok katmanlı algılayıcı (Gardner ve Dorling, 1999), yapay sinir ağları (YSA) (Özcan ve diğ., 2006; Yüksek ve diğ., 2007) gibi sezgisel yaklaşımlar, girdi çıktı kümeleri arasında zamansal ilişkilerin inşa edilmesinde klasik matematiksel analiz yöntemlerine göre daha yüksek başarı sağlamaktadırlar. Son yıllarda lineer olmayan doğal süreçlerin modellenmesinde, hızlı öğrenme ve iyi genelleme yetenekleri sayesinde öne çıkan aşırı öğrenme makineleri yaklaşımları hava kalitesi çalışmalarında yaygın kullanım alanı bulabilmiştir. Zhang ve Ding (2017) Hong Kong'da bulunan iki farklı hava kalitesi izleme istasyonundan altı yıl boyunca toplanmış sekiz kirleticiye ait zaman serilerinin modellenmesinde kullandığı farklı yöntemleri istatistiksel başarı kriterlerine göre kıyasladığı çalışmasında makine öğrenmesi yaklaşımının yüksek performansına işaret etmektedir. Atina havzasında bir sonraki günün en yüksek saatlik ozon konsantrasyonunu tahmin etmek için sinir ağları ve çoklu doğrusal regresyon modelleri ile bir karşılaştırma çalışması yapılmıştır (Chaloulakou ve diğ., 2003).

Yalnızca kirleticilerin zamansal değişkenliğini değil aynı zamanda iklimsel veya meteorolojik parametrelerin kirletici konsantrasyonları üzerindeki etkisinin incelenmesinde de farklı yaklaşımları birleştiren makine öğrenmesi tabanlı algoritmalar araştırmacılarca yol gösterici olarak değerlendirilmektedir. Örneğin; Lary ve diğ. (2015),  $PM_{2.5}$ 'in günlük dağılımlarını tahmin etmek amacıyla uzaktan algılama ve meteoroloji istasyonu gözlemlerini birlikte kullanarak geliştirdiği yenilikçi bir yaklaşımla hava kirliliğinin yüksek seyrettiği

mevsimlerde iklimsel modeller kurmuştur. Özcan ve diğ. (2006) İstanbul'da ozon konsantrasyonlarına iklimsel faktörlerin ve diğer kirletici parametrelerin etkilerini inceledikleri çalışmalarında, 8 iklimsel parametreye ve 6 kirletici parametreye bağlı olarak kurdukları YSA modeli ile başarılı sonuçlar elde etmişlerdir.

PM<sub>10</sub> ve SO<sub>2</sub> ölçüm kayıtları, durağan olmayan farklı frekans bileşenlerine sahip zaman serileri oluştururlar. Durağan olmayan zaman serilerinde, veriyi tahmin etmek için önışlem kademesinde dalgacık dönüşümü yaygın kullanılan bir sinyal analiz yöntemidir. Dalgacık sinyalleri eş enerjili alt-seri bileşenleri ayırarak verinin gürültüden kurtulmasını sağlar. Ayırıştırma işlemi sayesinde karmaşık zaman serilerinin frekans-zaman alanında daha homojen alt bileşenlerle temsil edilebilmesi, tahmin algoritmalarının girdi-çıkı kümeleri arasındaki ilişkileri yorumlama kabiliyetini artırır (Altunkaynak ve Kartal, 2019). Böylece dalgacık dönüşümü ile desteklen modellerde artan analiz kabiliyeti hem öğrenme hem de yeni değerler türetme sürecinde fayda sağlamaktadır.

Bu çalışmada Erzincan ili şehir merkezinde PM<sub>10</sub> ve SO<sub>2</sub> kirletici konsantrasyonlarını öngörebilen tahmin modelleri geliştirilmiştir. Kaydedilen günlük değerler arasındaki zamansal ilişkiyi temsil eden matematiksel ifadeler KNN algoritması inşa edilmiş, türetilen ifadeler yeni kirletici konsantrasyonlarının belirlenmesinde kullanılmıştır. KNN modellerinin dalgacık dönüşümü ile desteklenmesinin modellerin gelecekteki kirletici seviyelerini öngörebilme başarısına katkısı incelenmiştir. Hava kirliliği konusunda dalgacık dönüşümü tekniği ile birlikte KNN algoritmasının kullanılması adına bir ilk olma özelliği taşıyan çalışmamız, önışleme yöntemlerinin tahmin modelleri için önemine ışık tutmayı amaçlamaktadır.

## 2. MATERYAL YÖNTEM

### 2.1. Dalgacık Dönüşümü

Kronolojik sıra ile düzenli sıklıkta kaydedilen ölçümlerden oluşan zaman serilerinin analizinde frekans bilgisinin elde edilmesi büyük önem taşır. 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 çözümleyememesi ve sinüs dalgalarının geometrisinin ani sıçramaları yakalamadaki yetersizliği gibi eksikliklerin giderilmesi amacıyla zaman içerisinde dalgacık dönüşümü geliştirilmiştir. Dalgacıklar, prototip bir ana fonksiyondan türetilen sınırlı pencere fonksiyonlarıdır. Pencereleme işlemi ile sınırlı bir zaman alanında frekans analizi yapılır, bu işlem tüm zaman adımlarında gerçekleştirildiğinde verinin frekansının zamansal değişimi elde edilmiş olur. Dalgacık dönüşümünün (DD) temel prensibini pencereleme tekniğinde de görülen kaydırma ve dalgacık dönüşümü ile geliştirilen ölçeklendirme işlemleri oluşturur.

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

Dalgacık fonksiyonu  $\Psi_{a,b}(t)$  ana fonksiyonunun a kadar ölçeklendirilip b kadar zaman ekseninde ötelenmesi ile elde edilir. Ölçek parametresinin 1'den küçük olması durumunda yüksek frekanslı dalgacıklar ve geniş ölçekli analizler elde edilirken, ölçek parametresinin 1'den büyük olduğu durumlarda düşük frekanslı dalgacıklar ve ayrıntılı analizler elde edilir. Dönüşüm işlemi, dalgacık fonksiyonunun seçilen ölçekte her zaman adımında zaman serisi  $f(t)$  ile iç çarpımıdır.  $\Psi^*\left(\frac{t-a}{b}\right)$  dalgacık fonksiyonunun eşleniğidir.  $f(t)$  fonksiyonunun  $\Psi_{a,b}(t)$  pencereleme fonksiyonu ile dalgacık dönüşümü DD fonksiyonu ile ifade edilebilir.

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

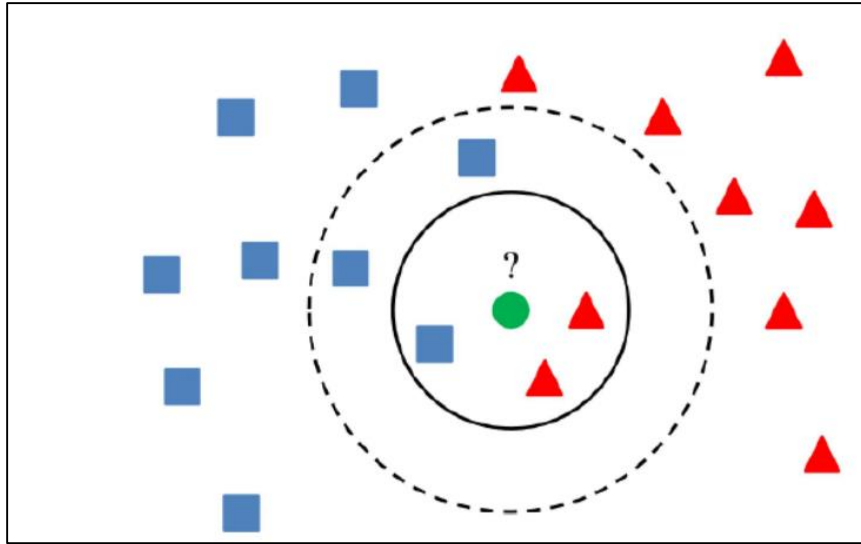
Skaler çarpım işlemi sonucunda elde edilen katsayılar dalgacık ve zaman serisinin korelasyonunu temsil eder. Yüksek katsayılar dalgacığın zaman serisini temsil etmedeki başarısına işaret eder. Çarpım adımının sonunda kullanılan ölçeğin karesine bölme işlemi enerji normalizasyonu için kullanılır. Yani eş ölçekli dalgacıklardan elde edilen alt-seri eş enerjili kabul edilir.

## 2.2. K-En Yakın Komşuluk

K-En Yakın Komşu (KNN) tahmin yöntemi parametrik olmayan sınıflandırma ve regresyon algoritmasıdır. Uygulanabilme kolaylığı, basit bir matematiksel temele sahip olması pek çok farklı alanda tahmin modelleri kullanımını yaygınlaştırmıştır (Kermani ve diğ. 2018, Ni ve diğ., 2012). 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. Genellikle, en yakın gözlemler, göz önünde bulundurulmuş veri noktasına en küçük Öklid mesafesine sahip olanlar olarak tanımlanır. Gözlemler arasındaki öklid mesafesi 2 boyutlu çözüm kümesi örneğinde x düzlemindeki doğrusal uzaklık  $x_i$ 'ye ve y düzlemindeki doğrusal uzaklık  $y_i$ 'ye bağlı olarak bulunabilir.

$$\text{Öklid mesafesi} = \sqrt{\sum_{i=1}^k (x_i - y_i)^2} \quad (3)$$

Dikkate alınacak komşu sayısını gösteren k'nın belirlenmesi için model optimizasyonu gerekir. Komşuluk sayısı k değeri arttıkça hesaplama adımı artsa da eğitim setindeki gürültüye karşı kırılabilirliği azalacak ve test verisine uyum artacaktır.



**Şekil 1:**

*KNN model temsili*

KNN yönteminin çalışma prensibi şekil 1 üzerinden izah edilmeye çalışılmıştır. Bir sınıflandırma problemi karşısında, iki farklı sınıfa ayrılmış olan verilerden komşu sayısına (k) bağlı olarak bir eğitim işlemi gerçekleştirilmektedir. Eğitim işleminin amacı optimum komşu sayısını belirlemektir. Belirlenen optimum komşu sayısı eğer 3 (k=3) ise, hangi sınıfa dahil olacağı tahmin edilmek istenen yeşil nokta kırmızı üçgenlerin bulunduğu sınıfa dahil edilecektir. Komşu sayısı 3 değil de 5 alınacak olursa yeşil noktanın dahil olduğu sınıf, mavi karelerden oluşan sınıf olacaktır. Regresyon problemlerinde ise; tahmin edilmek istenen bağımlı değişken

değeri, tahminleyicisi olan bağımsız değişkenin optimum k sayıdaki komşu bağımsız değerlerinin bağımlı değerlerinin aritmetik ortalaması olarak hesaplanabilmektedir.

### 2.3. Model Başarısı Ölçütleri

Literatürde, tahmin modellerinin doğruluğunu belirlemek için farklı performans ölçütleri önerilmiştir. Bu çalışmada model başarıları, boyut hassasiyetine sahip karesi ortalaması (HKO) ve değişkenlerin boyutlarından bağımsız verimlilik katsayısı (VK) (Nash ve Sutcliffe, 1970) olmak üzere iki farklı ölçüt ile değerlendirilmiştir.

$$HKO = \frac{1}{n} \sum_{i=1}^n (T_i - G_i)^2 \quad (4)$$

HKO, n uzunluğunda bir veri seti için her bir tahmin değerleri  $T_i$ 'nin aynı zamandaki gözlemlenen  $G_i$  değerinden ortalama sapmasını temsil eder.

$$VK = 1 - \left[ \frac{\sum_{i=1}^n (T_i - G_i)^2}{\sum_{i=1}^n (\bar{G} - G_i)^2} \right] \quad (5)$$

VK ölçütü, HKO'da temsil edilen hata büyüklüğüne ilave olarak gözlem değerlerinin varyansını da başarı değerlendirilmesinde dikkate alınmasında fayda sağlar. VK denkleminde paydada verilen ifade,  $G_i$  gözlem değerlerinin  $\bar{G}$  gözlem ortalamasından sapmasını gösterir. Verimlilik katsayısı değerinin 1'e yaklaşması tahmin başarısının yüksek olduğunu gösterir. Negatif VK değerleri ise tüm değerlerin serinin ortalaması ile temsil edildiği bir modelin mevcut tahminlerden daha başarılı sonuçlar vereceğine işaret eder.

Bir diğer değerlendirme ölçütü Ortalama mutlak hata (OMH) ile model başarısı ölçülmüştür.

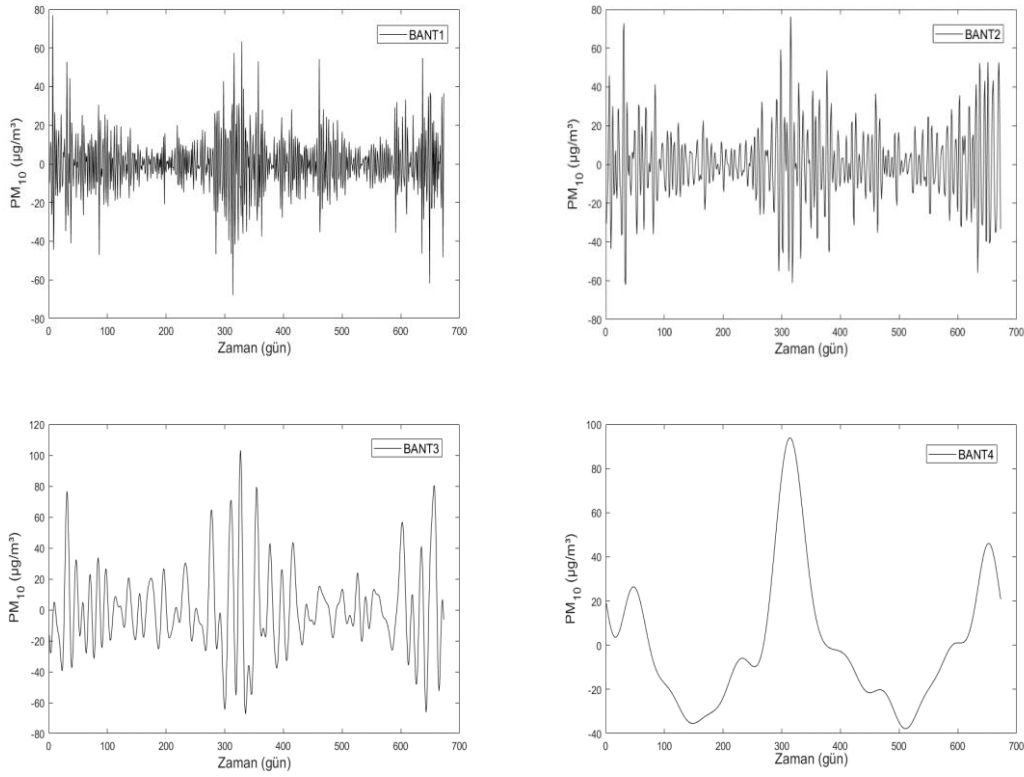
$$OMH = \frac{\sum_{i=1}^n |T_i - G_i|}{n} \quad (5)$$

### 3. SONUÇLAR

Bu çalışma kapsamında incelenen  $PM_{10}$  ve  $SO_2$  kirleticilerinin gelecek konsantrasyon değerlerinin tahmin edilmesi amacıyla KNN algoritması kullanılmıştır. Mevcut istasyondan elde edilen 651 verinin yaklaşık ilk %60'lık kısmıyla modelin öğrenme sürecinde değerlendirilmiş, geri kalan %40'lık kısımda ise model doğrulaması yapılmıştır. Eğitim setine bağlı olarak temsil kabiliyeti yüksek en yakın komşu sayısının belirlenmesinde çapraz-doğrulama yöntemi kullanılmıştır. Tahmin değerleri, -k sayısı optimizasyonu ve tahmin setlerinin oluşturulmasını eş zamanlı yürüten Statistica programının veri madenciliği aracı ile gerçekleştirilmiştir.

Matematiksel analiz yöntemleri, pek çok farklı frekanslı yürütücü mekanizma etkisi ile oluşan zaman serilerinde, geçmiş ile gelecek arasındaki ilişkileri inşa etmede yetersiz kalabildiğinden, verilerin sadeleştirilerek modele sunulması tahmin başarısına katkı sağlayabilmektedir. Bu çalışmada, dalgacık dönüşümü zaman serilerinin farklı enerji seviyelerinde ayrıştırılarak homojen alt-seriler oluşturulmasında kullanıldı. Ayrıştırma sonrası orijinal seri dört alt seriye ayrıştırılmıştır.  $PM_{10}$  serinin alt-serileri Şekil 2'de gösterilmiştir. İlk enerji seviyesinde ayrıştırılan BANT-1 yüksek frekanslı bilgiye işaret ederken, BANT-4 düşük frekanslı değişimleri gösterir. DD destekli KNN modellerinde (D-KNN) her bir alt-seri tahmin edildi ve alt-seri model çıktılarının ters dalgacık işlemine tabi tutulması ile zaman serilerinin bütüncül tahminleri elde edildi. Pencere fonksiyonu olarak farklı alternatifler arasından başarısı ile öne çıkan Morlet dalgacık fonksiyonu tercih edildi.

Dalgacık dönüştürme tekniği, Fourier tekniğinden farklı olarak hem zaman temelinde hem de frekans temelinde ayarlanabilir bir temsiliyet sağlayabilmektedir. Seçilen ana dalgacık fonksiyonunun nonlinear bir fonksiyon olması ve doğal sinyaller ile uyumunun yüksek olması ile pik değerler rahatlıkla yakalanabilmektedir. Pik değerlerin temsiliyeti ölçekleme parametresi değerinin sürekli olarak güncellenmesi ve orijinal seri ile korelasyonuna bakılması ile sağlanabilmektedir. Pencereleme işlemi sırasında pik değer ile en yüksek korelasyona sahip ölçek parametresi tespit edilerek ayrıştırma işlemine devam edilmektedir. Sürekli dalgacık dönüştürme tekniğinin bir dezavantajı olan bu işlem yüksek hafıza gereksinimine ihtiyaç duymaktadır (Wu ve Liui 2008). Fakat çok uzun olmayan sinyal boyutları ve düşük seviyede ayrıştırma işlemi ile hafıza probleminin bir miktar önüne geçilebilmektedir. Ana dalgacık fonksiyonu ile temsil edilen alt sinyaller kendi içerisinde tahmin edilip tüm alt sinyal tahminleri toplanırsa nihai tahmin değerine ulaşılmış olmaktadır. Pik değerlerin tüm alt serilerde temsil edilmiş olması yalnız modellere nazaran daha isabetli tahminler gerçekleştirilmesini sağlamaktadır.



**Şekil 2:**

*Dalgacık dönüşümü ile ham zaman serisinin ayrılmış olduğu bantlar*

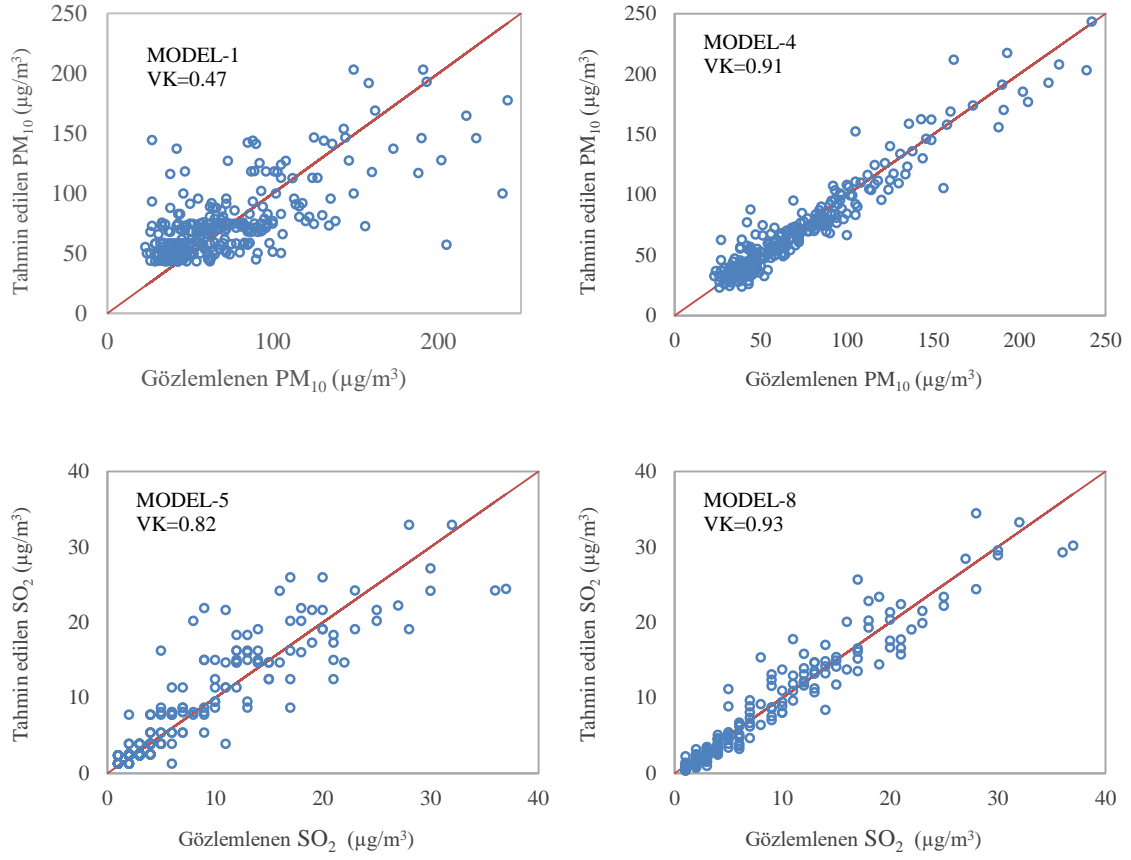
Model yapısı optimizasyonu sırasında farklı girdi kombinasyonlarının tahmin başarısına etkisi de incelenmiş, alternatif modelleme prosedürleri Tablo 1’de verilen başarı kriterlerine göre değerlendirilmiştir. Bir girdili modellerde tahmin edilecek  $PM_{10}(t)$  ve  $SO_2(t)$  konsantrasyonları, sırasıyla bir önceki  $PM_{10}(t-1)$  ve  $SO_2(t-1)$  değerlerinden elde edilmiştir. İki girdili modellerde, iki gün önceki kirletici değerleri ( $t-2$ ) girdi setine dahil edilmiştir. Tablo 1’de verilen HKO ve VK kriterlerine göre iki gün önceki değerlerden oluşan zaman serisinin girdi kümesine ilavesi, ham zaman serilerini kullanan KNN modellerine önemli bir katkı sağlamazken, dalgacık analizi sonucundan elde edilen alt serilerin kullanıldığı D-KNN modellerinde tahmin başarısını iyileştirmiştir.  $PM_{10}$  tahmin eden Model-1 ve MODEL-2 sonuçları iki girdili model yapısının sonuçları olumsuz etkilediğine işaret etmektedir. Bununla birlikte zaman serilerinin dalgacık dönüşümüne tabi tutulduğu  $PM_{10}$  ve  $SO_2$  tahmin modellerinde iki günlük verinin kullanılması VK

değerlerini 0,9'un üzerine çıkartmıştır. Özellikle dalgacık destekli PM<sub>10</sub> modellerinde (t-1) zaman serisi ile kurulan MODEL-4 ile VK=0,7 yakalanabilirken, (t-2) ve (t-1) zaman serilerinin kullanıldığı MODEL-5'te ulaşılan VK=0,91 değeri girdi sayısının D-KNN modellerine yüksek katkısına işaret etmektedir.

**Tablo 1. Performans kriterlerine göre modellerin başarısı**

MODEL	GİRDİ	YÖNTEM	ÇIKTI	HKO	VK	OMH
MODEL-1	PM <sub>10</sub> (t-1)	KNN	PM <sub>10</sub> (t)	896	0,47	21,5
MODEL-2	PM <sub>10</sub> (t-2),(t-1)	KNN	PM <sub>10</sub> (t)	947	0,44	22,1
MODEL-3	PM <sub>10</sub> (t-1)	D-KNN	PM <sub>10</sub> (t)	510	0,70	16,4
MODEL-4	PM <sub>10</sub> (t-2),(t-1)	D-KNN	PM <sub>10</sub> (t)	144	0,91	8,6
MODEL-5	SO <sub>2</sub> (t-1)	KNN	SO <sub>2</sub> (t)	8,42	0,82	1,7
MODEL-6	SO <sub>2</sub> (t-2),(t-1)	KNN	SO <sub>2</sub> (t)	8,16	0,83	1,6
MODEL-7	SO <sub>2</sub> (t-1)	D-KNN	SO <sub>2</sub> (t)	5,89	0,88	1,4
MODEL-8	SO <sub>2</sub> (t-2),(t-1)	D-KNN	SO <sub>2</sub> (t)	3,12	0,93	1,1

Her bir kirleniciye ait tahmin değerlerinin aynı tarihlerdeki gözlem değerleri ile uyumunu gösteren saçılma grafikleri, artan girdi sayısının ve ilave ayrıştırma işleminin model başarısına katkısını değerlendirmek amacıyla Şekil 3'te paylaşılmıştır. Grafiklerde mükemmel model uyumunu temsil eden 45°'lik eğri üzerinden sapmalar tahmin hatasına işaret ederken, çizginin altında gözlenen yığılma ise modellerin yüksek değerlerin olduğundan daha düşük tahmin edebildiğini gösterir. PM<sub>10</sub> tahmininde kullanılan MODEL-1'e ve MODEL-4'e ait grafiklerdeki saçılma düzeyleri kıyaslandığında, dalgacık dönüşümü ile desteklenen KNN modellerinin kirliliğin hem ortalama hem de aşırı değerlerde seyrettiği koşullarda, gözlemlenen değerleri yakalamakta daha başarılı olduğu söylenebilir. Optimal model yapısı tercihinin model performansına katkısını SO<sub>2</sub> tahmininde kullanılan MODEL-5 ve MODEL-8'de verilen tahmin-gözlem eğrileri üzerinden değerlendirmek de mümkündür. MODEL-8'de girdi sayısının artırılması ve ham zaman serilerinin dalgacık dönüşümü ile ayrıştırılması modellerin geçmiş ve gelecek arasındaki bağlantıları daha iyi analiz etmesine ve kurulan zamansal ilişkiler ile daha başarılı tahminler üretilmesi sağlamıştır. Tüm modellerde KNN yönteminin aşırı değerlere özel bir hassasiyeti gözlemlenmemiştir.



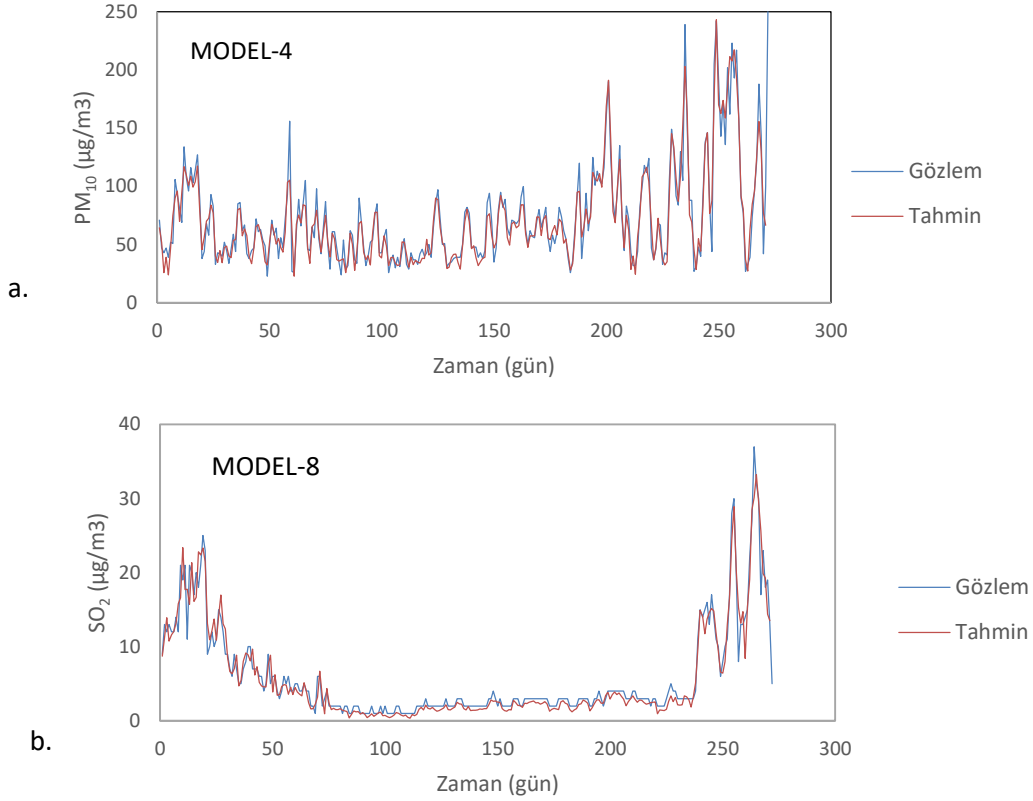
**Şekil 3:**  
*PM<sub>10</sub> ve SO<sub>2</sub> modelleri saçılma grafikleri*

Yapılan ön çalışmalar sonucunda PM<sub>10</sub>(t) ve SO<sub>2</sub>(t) konsantrasyonlarının tahmininde en yüksek başarı sağlayan iki gecikme zamanlı D-KNN modellerinin kullanılmasının uygun olduğuna karar verilmiştir. Modellerin test kısımlarına ait çıktıları Şekil 4’te paylaşılmıştır.

#### 4. TARTIŞMA

Kirletici parametreler ile ilgili tahmin çalışmaları, uzun vadede hava kalitesinin istenilen seviyede muhafaza edilmesi ve kısa vadede oluşabilecek ani yükselişlere karşı acil önlem planı oluşturulması alanında karar vericilere yol gösterici araçlar olarak hizmet etmektedir. Bu çalışma ile PM<sub>10</sub> ve SO<sub>2</sub> maddelerinin günlük konsantrasyonları KNN algoritması ile modellenmiş ve gelecekte olması muhtemel değerleri tahmin edilmiştir. İstatistiki başarı kriterlerine göre KNN modellerinin dalgacık dönüşümü tekniğinin ile desteklenmesi tahmin başarısına yüksek katkı sağlamıştır. Ayrıca gelecek tahminlerini oluşturmada kullanılacak geçmiş bilgisini temsil eden girdi sayısının artırılması algoritmaların analiz kabiliyetine olumlu katkı sağlamıştır. KNN modellerinin performanslarında kirletici konsantrasyonlarının ortalama seviyede seyrettiği durumlarda veya aşırı değerler gözlemlenen günlerde önemli bir farklılık gözlemlenmemiştir. Uygun modelleme yapısı oluşturulduğu takdirde iki durumu da temsil edebilen modeller inşa etmek mümkündür. Araştırmacılara hibrit yöntemlerin önemi için yol gösterici olan bu çalışma gelecekte daha basit önileme yöntemleri kullanılarak daha kısa sürede ve daha güvenilir sonuçların alınabileceğine de birer kanıt niteliği taşımaktadır.





**Şekil 4:**  
Gözlemlenen ve optimum modellerle tahmin edilen  
(a)  $PM_{10}$  ve (b)  $SO_2$  zaman serileri

## KAYNAKLAR

1. Aalto, P., Hämeri, K., Paatero, P., Kulmala, M., Bellander, T., Berglind, N., ... ve Marconi, A. (2005) Aerosol particle number concentration measurements in five European cities using TSI-3022 condensation particle counter over a three-year period during health effects of air pollution on susceptible subpopulations. *Journal of the Air & Waste Management Association*, 55(8), 1064-1076. doi:10.1080/10473289.2005.10464702
2. Altunkaynak, A. ve Kartal, E. (2019) Performance comparison of continuous wavelet-fuzzy and discrete wavelet-fuzzy models for water level predictions at northern and southern boundary of Bosphorus, *Ocean Engineering*, 186, 106097. doi:10.1016/j.oceaneng.2019.06.002.
3. Chaloulakou, A., Saisana, M. ve Spyrellis, N. (2003) Comparative assessment of neural networks and regression models for forecasting summertime ozone in Athens, *Science of the Total Environment*, 313(1-3), 1-13. doi:10.1016/S0048-9697(03)00335-8.
4. Chen, T. M., Kuschner, W. G., Gokhale, J. Ve Shofer, S. (2007) Outdoor air pollution: nitrogen dioxide, sulfur dioxide, and carbon monoxide health effects, *The American journal of the medical sciences*, 333(4), 249-256. doi:10.1097/MAJ.0b013e31803b900f.
5. Dominici, F., Peng, R. D., Bell, M. L., Pham, L., McDermott, A., Zeger, S. L., & Samet, J. M. (2006). Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases. *Jama*, 295(10), 1127-1134. doi:10.1001/jama.295.10.1127.

6. Gamble, J. F. ve Lewis, R. J. (1996). Health and respirable particulate (PM<sub>10</sub>) air pollution: a causal or statistical association?, *Environmental health perspectives*, 104(8), 838-850. doi:10.1289/ehp.96104838
7. Gardner, M. W. ve Dorling, S. R. (1999) Neural network modelling and prediction of hourly NO<sub>x</sub> and NO<sub>2</sub> concentrations in urban air in London, *Atmospheric Environment*, 33(5), 709-719. doi:10.1016/S1352-2310(98)00230-1
8. Isakov, V., Touma, J. S. ve Khlystov, A. (2007) A method of assessing air toxics concentrations in urban areas using mobile platform measurements, *Journal of the Air & Waste Management Association*, 57(11), 1286-1295. doi:10.3155/1047-3289.57.11.1286
9. Kalapanidas, E. ve Avouris, N. (2001) Short-term air quality prediction using a case-based classifier, *Environmental Modelling & Software*, 16(3), 263-272. doi:10.1016/S1364-8152(00)00072-4
10. Kermani, E. F., Barani, G. A. ve Hesarroeyeh, M. G. (2018) Cavitation damage prediction on dam spillways using Fuzzy-KNN modeling. *J Appl Fluid Mech*, 11(2), 323-329. doi:10.29252/JAFM.11.02.28356
11. Laden, F., Neas, L. M., Dockery, D. W., ve Schwartz, J. (2000) Association of fine particulate matter from different sources with daily mortality in six US cities, *Environmental health perspectives*, 108(10), 941-947. doi:10.1289/ehp.00108941
12. Lary, D. J., Lary, T. ve Sattler, B. (2015) Using machine learning to estimate global PM<sub>2.5</sub> for environmental health studies, *Environmental health insights*, 9, EHI-S15664. doi:10.1177/EHI.S15664
13. Nash, J. E. ve Sutcliffe, J. V. (1970) River flow forecasting through conceptual models part I—A discussion of principles, *Journal of hydrology*, 10(3), 282-290. doi:10.1016/0022-1694(70)90255-6
14. Ni, J., Qiao, F., Li, L. ve Di Wu, Q. (2012, July). A memetic PSO based KNN regression method for cycle time prediction in a wafer fab. *In Proceedings of the 10th World Congress on Intelligent Control and Automation*, 474-478. IEEE. doi:10.1109/WCICA.2012.6357922
15. Özcan, H. K., Şahin, Ü., Bayat, C. ve Uçan, O. N. (2006) İstanbul İli Tropsoferik Ozon (O<sub>3</sub>) Konsantrasyonlarının Hücreyel Yapay Sinir Ağ Yöntemiyle Modellenmesi, *Gazi Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi*, 21(2).
16. Yıldırım, Y. ve Bayramoğlu, M. (2006) Adaptive neuro-fuzzy based modelling for prediction of air pollution daily levels in city of Zonguldak, *Chemosphere*, 63(9), 1575-1582. doi:10.1016/j.chemosphere.2005.08.070
17. Yüksek, A. G., Bircan, H., Zontul, M. ve Kaynar, O. (2007) Sivas İlinde Yapay Sinir Ağları İle Hava Kalitesi Modelinin Oluşturulması Üzerine Bir Uygulama, *Cumhuriyet Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 8(1), 97-112.
18. Zhang, J. ve Ding, W. (2017) Prediction of air pollutants concentration based on an extreme learning machine: the case of Hong Kong, *International journal of environmental research and public health*, 14(2), 114. doi:10.3390/ijerph14020114
19. Wu, J.D ve Liu, C.H. (2008) Investigation of engine faultdiagnosis using discrete wavelet transform and neural network. *Expert Systems with Applications*, 35(3), 1200-1213. doi:10.1016/j.eswa.2007.08.021