

Hava Kirliliği Açısından Bölgelerin Kümeleme Analizi İle İncelenmesi: Türkiye Örneği

Dilek Altaş KARACA¹

Tuğçe ARIKAN²

¹Prof.Dr., Marmara Üniversitesi, İktisat Fakültesi, Ekonometri Bölümü, dilekaltas@marmara.edu.tr, ORCID: 0000-0001-5103-9018

²Marmara Üniversitesi, Sosyal Bilimler Enstitüsü, İstatistik Bilim Dalı, tugcearikan2@gmail.com, ORCID:0009-0009-7102-4177

Özet: Hava kirliliği insan sağlığını tehdit eden unsurlardan biri olduğundan tüm dünyada bu konuya büyük önem verilmektedir. Kümeleme analizi ise hava kirlleticilerinin mekansal ve zamansal kalıplarının tanımlanması, hava kirliticilerine maruz kalma, hava kalitesi yönetimi, hava kirlitici davranışının yerel sinoptik meteoroloji ile ilişkilendirilmesi, taşıma yolları ve kaynak paylaşırma gibi amaçlar için kullanılan etkin bir yöntem olarak öne çıkmıştır. Literatürde benzer amaçlara yönelik olarak farklı kümeleme yöntemlerinin kullanıldığı ancak yöntemler arasında sınırlı sayıda karşılaştırma yapıldığı gözlemlenmiştir. Bu yüzden gerek kümeleme yöntemlerine ilişkin gerekse hava kirliliğinde yapılan uygulamalarına ilişkin literatür detaylı şekilde incelenmiş ilgili verilerin yapısına uygun kümeleme yöntemleri belirlenmiştir. Bu çalışmada, Ocak 2021-Aralık 2023 arasında 86 istasyondan elde edilen aylık ortalama SO₂, NO₂, NO, NO_x, O₃ and PM₁₀ değerleri (i) çok değişkenli fonksiyonel temel bileşenler analizi sonucunda elde edilen skorlara k-ortalamlar kümeleme yöntemi uygulanan iki aşamada yapılan kümeleme yöntemi, (ii) çok değişkenli fonksiyonel temel bileşenler analizi sonucunda elde edilen skorların normal dağıldığı varsayımı altında geliştirilen model tabanlı kümeleme yöntemi, ve (iii) çok değişkenli karışık t dağılımı altında bir önceki yöntem olmak üzere üç çok değişkenli fonksiyonel kümeleme yöntemi ile gruplandırılmıştır. İlk iki yöntemle 5 küme, son yöntemle 4 küme oluştuğu ve yöntemlere göre kümelerdeki istasyonların farklılaştığı görülmüştür.

Anahtar Kelimeler: Kümeleme Analizi; Hava Kirliliği; Hava Kalitesi

Examining Regions In Terms Of Air Pollution With Cluster Analysis: Example Of Turkey

Abstract: Since air pollution is one of the factors that threaten human health, this issue is given great importance all over the world. Cluster analysis, on the other hand, became an effective method in this area for purposes such as defining spatial and temporal patterns of air pollutants, exposure to air pollutants, air quality management, associating air pollutant behavior with local synoptic meteorology, transport routes and resource allocation. It has been observed in the literature that different clustering methods are used for similar purposes, however, it is observed that a limited number of comparisons are made between the methods. Therefore, the literature on both clustering methods and their applications in air pollution has been examined in detail, and clustering methods suitable for the structure of the relevant data have been determined. In this study, monthly average SO₂, NO₂, NO, NO_x, O₃ and PM₁₀ values obtained from 86 stations between January 2021 and December 2023 are grouped with three multivariate functional clustering methods: (i) a two-stage clustering method in which the k-means clustering method is applied to the scores obtained from multivariate functional principal component analysis, (ii) a model-based clustering method developed under the assumption that the scores obtained from multivariate functional principal component analysis are normally distributed, and (iii) the previous method under the multivariate mixed t distribution. It is observed that 5 clusters are formed with the first two methods while 4 clusters are formed with the last method, and the stations in the clusters differ according to the method.

Keywords: Clustering; Air Pollution; Türkiye; Air Quality

1. GİRİŞ

Bu tez çalışması kapsamında, öncelikli olarak birinci bölümde hava kirliliği ve kümeleme analizine ilişkin Türkiye’de ve dünya genelinde yapılan k-ortalamlar, hiyerarşik kümeleme, iki aşamalı kümeleme, bulanık kümeleme, makine ve derin öğrenme algoritmalarını kullanan kümeleme ve diğer kümeleme yöntemi makaleleri incelenip derlenerek geniş kapsamlı bir literatür taraması yapılmıştır. İkinci bölümde, çalışmada kullanılan veri seti tanımlanmış, çalışmanın materyal ve yöntemleri, eksik verilerin tamamlanması, çok değişkenli fonksiyonel veriler için kümeleme yöntemlerinin tanımlanması, fonksiyonel verilerin düzleştirilmesi fonksiyonlarla temsili, iki

aşamada yapılan kümeleme yöntemi ve model tabanlı kümeleme yöntemi başlıkları altında teorik olarak ortaya konmuştur. Üçüncü bölümde ise grafik verileri üzerinden bulgular yorumlanmış, değişkenler arasındaki korelasyon yapısı tespit edilmiş ve iki aşamada yapılan kümeleme yöntemi, normal dağılımı altında model tabanlı kümeleme yöntemi ve t dağılımı altında model tabanlı kümeleme yöntemi sonuçları incelenmiştir.

Hava kirliliği, küresel olarak halk sağlığına yönelik en büyük çevresel tehdittir. İnsan faaliyetleri sonucu oluşan hava kirliliği temelde ısınma, sanayi ve trafik kaynaklıdır. İnsan sağlığı üzerinde olumsuz etkilere sahip başlıca kirlitici şunlardır: (<https://www.who.int/>)

Kükürt dioksit (SO₂), Partikül Madde (PM₁₀ ve PM_{2.5}), Azot Oksitler, Azot Monoksit ve Azot Dioksit (NO_x, NO, NO₂), Karbonmonoksit (CO), Ozon (O₃), Kurşun (Pb).

Hava kirliliği insan sağlığını bozmanın yanında küresel ısınma ve asit yağmurları gibi küresel problemlere de sebep olduğundan iklim değişikliği ile yakından bağlantılıdır. Türkiye de dahil olmak üzere birçok ülke ABD Çevre Koruma Ajansı (EPA) tarafından geliştirilen hava kalite indeksini kullanmaktadır. Bu indeks insan sağlığına etkileri nedeniyle 5 ana kirleniciye (SO₂, NO₂, CO, O₃ and PM₁₀) odaklanır. T.C. Çevre, Şehircilik ve İklim Bakanlığı da hava kirliliğinin daha etkin olarak izlenmesi için 8 bölgede (İstanbul merkez olmak üzere Marmara, Samsun merkezli Orta Karadeniz, Erzurum merkezli Doğu Anadolu, İzmir merkezli Ege, Adana merkezli Akdeniz, Konya merkezli Güney İç anadolu, Ankara merkezli Kuzey İç Anadolu ve Diyarbakır merkezli Güneydoğu Anadolu) temiz hava merkezleri kurmuş, Güneydoğu Anadolu hariç faaliyete geçirerek mevcut ölçüm istasyon sayılarını çoğaltmıştır. Ölçüm sonuçlarından elde edilen veriler değerlendirilerek aylık ve yıllık raporlar hazırlamakta ve izleme ağından elde edilen ham veriler de www.havaizleme.gov.tr adresinde eşzamanlı olarak yayınlamaktadır.

Konunun tüm dünyada gittikçe artan önemi nedeniyle bu konuya yönelerek bu alanda yaygın olarak kullanılan kümeleme analizine odaklanılmıştır. Kümeleme analizi, hava kirliliği araştırılmalarında 1980'lerden beri sıklıkla hava kirlenitçilerinin mekansal ve zamansal kalıplarının tanımlanması, hava kirlenitçilerine maruz kalma,

hava kalitesi yönetimi, hava kirlenitici davranışının yerel sinoptik meteoroloji ile ilişkilendirilmesi, taşıma yolları ve kaynak paylaşırma amacıyla uygulanan bir yöntemdir (Govender, Sivakumar, 2020). Kümeleme analizine ve uygulamalarına ait literatür tarama makaleleri bulunmakla birlikte hava kirliliği üzerine uygulamasına ilişkin tarama makalesinin az olduğu görülmektedir. Aghabozorgi vd. (2014) Malezya'da mekansal-zamansal ile zaman serileri kümeleme yöntemlerinin kullanıldığı hava kirliliği analizlerine ilişkin çalışmaların taramasını gerçekleştirirken; Bellinger vd. (2017) hava kirliliği epidemiyolojisinde veri madenciliği ve makine öğrenimini kullanan çalışmalara ilişkin literatür taramasını yapmışlardır. Son olarak Govender, Sivakumar (2020) hava kirliliği analizinde k-ortalamlar ve hiyerarşik kümeleme yöntemleri kullanan çalışmalara ait tarama gerçekleştirmiştir. Aghabozorgi vd. (2014) çalışmaları kullandıkları kümeleme yöntemlerine göre sınıflandırırken, Govender, Sivakumar (2020) çalışmaları yer bazlı hava kirliliği ölçümlerine ve hava kütlesi yörüngelerine ilişkin olmak üzere iki kategori altında incelemişlerdir. Bellinger vd. (2017) ise çalışmaları uygulama alanı (iç mekan, dış mekan), çalışmanın amacı ve kullanılan veri madenciliği yaklaşımı başlıkları altında özetlemişlerdir. Bu çalışmanın amacına yönelik olarak literatürdeki çalışmalar Aghabozorgi vd. (2014) gibi kullanılan kümeleme yöntemi bazında gruplandırılarak alt başlıklar halinde derlenmiş ve bahsi geçen tarama makalelerinde yer almayan çalışmalara değinilmiştir.

2. HAVA KİRLİLİĞİ VE KÜMELEME ANALİZİNE İLİŞKİN YAPILAN ÇALIŞMALAR

2.1. k-Ortalamlar Yöntemini Kullanan Çalışmalar

Kişi	Tarih	Amaç
Sahafizadeh ve Ahmadi	2009	İran'ın Buşehr şehrinin hava kirliliğinin göstergesi olan tozlu günlerini tahmin etmek
Austin ve ark.	2013	Hava kirliliği verilerindeki mekansal yapıları belirlemek
Carslaw ve Beevers	2013	Hava kirliliğinin kaynak tespiti ve karakterizasyonu iki değişkenli kutup grafiğini geliştirmişler ve benzer özellikleri belirlemek ve gruplandırmak
Baltacıbaşı	2014	Türkiye'deki hava kirliliği seviyesini belirleyebilmek için 2008-2010 yılları arasındaki PM ₁₀ konsantrasyonlarını mekansal ve zamansal olarak incelemiştir.
Kingsy ve ark.	2016	Hava kirliliği verilerini analiz etmek için geliştirilmiş k-ortalamlar kümeleme algoritması önermişler ve önerdikleri bu yöntemi değişik veri setleri ile olabilirlikli bulanık c-ortalamlar (PFCM) kümeleme algoritması ile

		karşılaştırmışlardır.
Kasparoğlu ve ark.	2018	Marmara Bölgesi'nin ozon ve ozon öncüllerinin seviyelerini zamansal ve mekansal özelliklerini ve sonuçlarını incelemek amacıyla, ozonun kırsal bölgelerde lokal kirleticilere katkısının olmadığını anlamak için regresyon analizini, taşıyım yörüngelerini belirlemek için HYSPLIT modelini ve hava kütlelerinin kaynağını belirlemek için elde edilen yörüngelere k-ortalama kümelemesini kullanmışlardır.
Koçak	2018	Aksaray İli'ne ait PM10 ve SO2 kirleticilerindeki zamansal değişimi incelemek amacıyla koşullu iki değişkenli olasılık fonksiyonu ile rüzgar hızı ve yönü bilgileri üzerinden kirleticili konsantrasyonunun nasıl değiştiğini belirlemiş ve elde ettiği iki değişkenli polar çizim sonuçlarını, potansiyel kaynak özelliklerinin daha iyi anlaşılması için k-ortalamalar yöntemi ile sınıflandırmıştır.
Chuang ve ark.	2018	PM2.5'in laboratuvar ortamındaki biyoreaktivitesi ile Tayvan'daki bir petrokimya kompleksi yakınındaki emisyon kaynakları arasındaki ilişkiyi belirlemek amacıyla önce kompleks yakınında toplanan PM2.5 kaynaklarını k-ortalama kümeleme yöntemi ile üç gruba ayırmışlar ve sonrasında neden olduğu sıralı toksiditeyi araştırmışlardır.
Paulose ve ark.	2018	Delhi'nin hava kirliliği yoğun olan bölgelerini ve buralardaki sağlık risklerinin nedenlerini belirlemek.
Zhang ve ark.	2016	Çin'in 74 şehrinde başlıca hava kirleticilerini ve bunların mekansal ve mevsimsel dağılımını araştırmak için faktör analizi (FA) ile veri boyutunu indirgedikten sonra k-ortalamalar küme yöntemini kullanmışlardır.
Shafi ve Waheed	2020	Southampton Şehri'nin 2 Aralık 2019 saat 17:00 ile 11:00 arasında gerçekleşen yangın boyunca toplanan PM2,5 değerlerini kullanarak yangın nedeniyle hava kalitesinde meydana gelen ani değişiklikleri k-ortalamalar yöntemi ile sınıflandırarak en toksik olduğu saatleri ve normale dönüş sürecini belirlemişler.
Jorquera ve Villalobos	2020	Çalışmalarında ortam PM2.5 ve PM10'un uzun vadeli, saatlik kaynak dağılımını tahmin etmek için, kirleticili kaynaklarını belirleyebilmekle birlikte kısa vadeli dönemler için sonuç veren alıcı modellerinin (RM), uzun vadeli sonuçlar için hava kirliliğine ilişkin k-ortalama kümeleme analizi ve meteorolojik ölçümler ile birleştirilmesini önermişler.
Annas ve ark.	2022	Makassar şehrinde gündüz ve gece ölçülen hava kalitesi parametrelerine dayalı bölgesel kümelemede k-ortalamalar yöntemini; kümeleme sonuçlarını görselleştirmek için de özdenleyici haritalar (SOM) yöntemini kullanmışlardır. Ayrıca Coğrafi Bilgi Sistemi (CBS) ile de bölgesel kümelenmenin sonuçlarını şehir haritası üzerinde görselleştirmişlerdir.
Omar ve ark.	2022	Putrajaya, Malezya için PM, NO2 ve bağıl nem verilerini kullanarak çok değişkenli zaman serisi analizinde k-ortalama kümeleme yönteminin uygunluğunu araştırmışlar ve oluşturulan kümeleri günlük kalıplarının daha kolay yorumlanabilmesi için takvim görünümünde incelemişlerdir.
Sarı	2023	Çanakkale'nin Biga ve Lapseki ilçelerindeki yüksek ozon konsantrasyonlarının oluşmasına neden olan atmosferik koşulları incelemek için HYSPLIT modeli ile geri yörüngeleri elde edilip k-ortalamalar kümeleme yöntemi ile gruplandırarak homojen hava kütlesi taşıma kalıplarını elde etmiş ve meteorolojik fotokimyasal modeller ile bölgedeki ozonun taşıyım, dağılım ve seviyelerini belirlemiştir.
Di Bernardino ve ark.	2023	Roma şehir merkezi ve çevresindeki kırsal kesimde sinoptik ve yerel ölçekte anemolojik koşullar ile ana atmosferik kirleticilerin (NO, NO2, NO2/NOx, O3, PM2.5 ve PM10) lokal konsantrasyonları arasındaki ilişkiyi araştırmak.
Sel	2023	PM10 ve SO2 kirleticilerini kullanarak Türkiye'deki illeri hava kalitesi açısından k-ortalamalar yöntemi ile sınıflandırmış ve elde edilen 5 sınıfa, özelliklerine özgü olarak hava kalitesini arttıracak öneriler sunmuştur.

2.2. Hiyerarşik Kümeleme Yöntemini Kullanan Çalışmalar

Kişi	Tarih	Amaç
Qian ve ark.	2001	Dört Çin şehiri Lanzhou, Chongqing, Wuhan ve Guangzhou'da biri kentsel biri banliyö bölgesi olmak üzere seçtikleri iki bölgede, hava kirliliği epidemiyolojik çalışması kapsamında, toplam askıdaki parçacıklar, PM _{2.5} , PM _{2.5} – 10 ve PM ₁₀ , SO ₂ ve NO _x miktarlarına ilişkin ölçümleri gerçekleştirmiş ve bu 8 bölgeyi hiyerarşik kümeleme yöntemi ile hava kirliliğine göre 4 gruba ayırmışlardır.
Anıl ve ark.	2009	10 istasyondan elde ettikleri veriler ile İstanbul'daki PM ₁₀ profilini incelemek için istasyon bazında belirlenen epizod sınır değerlerinin en küçüğünü ortak epizod eşiği olarak belirleyerek sadece uzun mesafeli PM ₁₀ taşımının meydana geldiği ortak epizod tarihlerini saptamışlardır. Ardından HYSPLIT modeli ile geri yörüngeleri belirlemişler ve son olarak PM ₁₀ derişimlerini incelemek üzere kümeleme analizi grafiği (dendogram) üzerinden uzman görüşlerine de başvurarak istasyonları gruplamışlardır.
Gao Yao ve ark.	2011	Çin'deki 81 şehrin Haziran 2004-2007 arası günlük hava kirliliği indeksini (API) kullanarak şehirleri hava kirliliği açısından sınıflandırmak.
Tian ve ark.	2020	Kuzey Çin'deki 9 şehri, aylık ortalama PM _{2.5} , SO ₂ , NO ₂ , CO, ve O ₃ kirlleticilerine göre sınıflandırarak hava kirliliğinin zamansal ve mekansal analizini gerçekleştirmişlerdir.
Güler	2011	Yüzey ozon değerlerindeki hafta içi-haftasonu ve mevsimsel olarak değişimleri incelemek ve bunların meteorolojik parametreler ile ilişkisini anlamak için HYSPLIT modelinden elde ettiği geri yörüngeleri Ward yöntemi ile kümelemiştir.
Yusuf ve ark.	2013	Hem temel bileşenler hem de hiyerarşik kümeleme analizi ile Nijerya'nın Lagos kentindeki Apapa'daki Tin-Can limanı sanayi bölgesinin ortam havasındaki SO ₂ , NO ₂ , PM ₁₀ , CO, H ₂ S, CH ₄ ve gürültünün dağılımını incelemişler.
Yaşa	2019	Kocaeli ilindeki hava kirliliğini belirlemek için, 50 farklı noktadan alınan liken örneklerinde ağır metal analizlerini gerçekleştirmiş, kirlilik bölgelerini belirlemek için tam bağlantı yöntemi ile liken elementlerini kümelemiş, korelasyon ve faktör analizleri gerçekleştirek kirlletici madde seviyelerini etkileyen faktörleri ve kirlletici kaynaklarını tespit etmiştir.
Arslan	2020	Ege Bölgesi'ne ait PM ₁₀ değerlerinin sinoptik klimatolojisini incelerken ikinci aşamasında Çanakkale'de belirlediği episodik günlerin solunum yolu hastalıkları üzerindeki etkisini araştırmak için önce HYSPLIT modeli ile geri yörünge analizi yapmış ve ardından Ward yöntemi ile kümeleme analizi gerçekleştirmiştir.
Hemavani ve Rao	2020	Hindistan'ın Haydarabad kentinin endüstriyel ve ticari bir bölgesine ait 10 yıllık veri ile hava kirleticileri ve meteorolojik parametreler arasındaki ilişkiyi mevsimsel bazda belirlemek için temel bileşenler analizi ile hiyerarşik kümeleme analizini kullanmışlar.
Sarı	2023	Tüm Türkiye'deki hava izleme istasyonlarını yıllık ortalama PM ₁₀ ve SO ₂ değerleri üzerinden Ward yöntemi ile yıl yıl kümelemiş ve her yıl 6 küme oluştuğunu ancak yıldan yıla oluşan kümelerin farklılık gösterdiğini belirtmiş ve bu farklılaşmada nüfus, sanayi tesisi, araç sayısı, inşaat alanları gibi etmenlerin etkisini irdelemiştir.

2.3. İki Aşamada Yapılan Kümeleme Yöntemini Kullanan Çalışmalar

Kişi	Tarih	Amaç
Kışlak	2009	Türkiye'deki 111 istasyondan elde edilen PM ₁₀ ve SO ₂ değerlerini kullanarak kış sezonu ortalama ölçüm sonuçları, yıllık ortalama ölçüm sonuçları ve günlük ortalama ölçüm sonuçları olmak üzere 3 farklı veri seti üzerinden önce küme sayısının belirlenmesi için Ward yönteminin uygulandığı ardından k-ortalamalar ile kümelemenin gerçekleştirildiği iki aşamalı kümeleme yöntemi uygulanmış.
Khedairia ve Khadir	2012	Cezayir'deki farklı meteorolojik sınıfları bulmak için önce SOM ardından k-ortalamaların kullanıldığı iki aşamalı kümeleme yöntemini önermişler ve gerek uygulama gerekse sonuçlar açısından iki aşamalı kümeleme prosedürünün yalnızca SOM veya yalnızca k-ortalamalar ile yapılacak doğrudan kümeleme yaklaşımından daha verimli olduğunu

		belirtmişlerdir.
Wu ve Kuo	2013	Tayvan'ın merkezindeki sekiz otomatik hava kalitesi izleme istasyonundan 1 Ocak 2010 ile 30 Eylül 2011 tarihleri arasında toplanan hava kalitesi verilerini kullanarak önce faktör analizi yöntemi ile hava kalitesini etkileyen ana faktörleri belirlemişler ve ardından hava kalitesinin farklı özelliklerini ve kirlilik derecelerini gözlemek için önce hiyerarşik kümeleme yöntemi ile yaklaşık kümelerin belirlendiği, sonrasında k-ortalamalar yöntemi ile farklı küme sayılarının test edildiği iki aşamalı kümeleme yöntemi ile Tayvan'ın merkezini beş kümeye ayırmışlardır.
Ljungman ve ark.	2016	Boston'a ait 2003-2008 yılları arasındaki veriler ile gerçekleştirdikleri çalışmada önce günleri, hava kirlleticileri ve meteorolojik değişkenleri Ward bağlantı ve k-ortalamaları içeren iki aşamalı kümeleme yöntemi ile beş farklı hava kirliliği karışımı elde edecek şekilde gruplamışlar ve ardından önceki 1 günlük hareketli PM _{2,5} ortalaması ile önceki gün hava kirliliği kümesi arasındaki etkileşim terimini içeren çok değişkenli doğrusal regresyon modelleri oluşturarak ortamdaki ince partikül maddeye maruz kalma ile mikrovasküler arasındaki ilişkinin hava kirliliğinin farklı karışımına göre değişip değişmediğini incelemişlerdir.
Zhang ve Yang	2022	Bölgesel hava kirliliği kontrolüne yardımcı olmak için ilk önce sık öge kümesi (frequent item set) kümeleme yaklaşımı ile şehirler arasındaki ortak kirlilik yapılarını belirledikten sonra ortak kontrol bölgelerinin kapsamını tanımlamak ve optimize etmek için hiyerarşik kümeleme yöntemini kullanmışlardır.
Ahn ve ark.	2013	Seul'un günlük kalıplarını dikkate alacak şekilde PM _{2,5} ve PM ₁₀ konsantrasyonlarının tipolojisini oluşturmak amacıyla Ward bağlantı ve k-ortalamalar yönteminden oluşan iki aşamalı kümeleme yöntemini kullanmışlardır.

2.4. Bulanık Kümeleme Yöntemini Kullanan Çalışmalar

Kişi	Tarih	Amaç
Ojeda-Magaña ve ark.	2010	Beklenmedik durumların daha iyi yönetilmesi için Meksika'nın Salamanca şehrindeki üç istasyondan birleşik bir ölçüm elde etmeyi hedeflemişler ve bunun için PFCM kümeleme yöntemini kullanmışlardır.
D'Urso ve ark.	2013	Zaman serilerini sınıflandırmak için en uygun otoregresif hareketli ortalama (ARMA) modeline ait tahmin edilen katsayıların bulanık c-medoids kümeleme yöntemi ile sınıflandırılmasına dayalı AR tabanlı bulanık c-medoids kümeleme modelini geliştirmişlerdir.
Dogruparmak ve ark.	2014	Marmara bölgesindeki 11 ilde bulunan 22 istasyondan elde edilen günlük SO ₂ ve PM ₁₀ değerlerini temel bileşenler analizini (PCA) ve FCM yöntemini kullanarak gruplayıp iki yöntemin performansını kıyaslamışlar.
D'Urso ve ark.	2015	AR tabanlı bulanık c-medoid yöntemini üstel mesafe ölçütü ile entegre ederek yöntemi aykırı değerlere karşı sağlam hale getirmişlerdir. Simüle edilen senaryolar altında önerdikleri bu sağlam kümeleme yöntemini sağlam olmayan AR tabanlı bulanık c-medoid (D'Urso vd, 2013a), Gürültü Kümeleme ile AR tabanlı bulanık c-medoid (D'Urso ve ark. 2013b), bunların crisp halleri ve hiyerarşik kümeleme (tekli, tam ve ortalama bağlantı ile Ward's yöntemi) yöntemleri ile karşılaştırmışlar.
Bekar	2017	Bulanıklaştırma adımı aykırı ve gürültülü değerlere karşı daha dayanıklı olan bulanık küme algoritmasının kullanımına dayanan yeni bir mevsimsel bulanık zaman serisi yöntemi önermiş ve bu yöntemi Ocak 2013-Nisan 2016 arası Türkiye'deki 112 istasyondan elde edilen haftalık PM ₁₀ ve SO ₂ verilerine uygulayarak FCM ve Gustafson-Kessel yöntemleri ile karşılaştırarak önerilen yöntemin öngöründe daha başarılı olduğu sonucuna varmıştır.
Dincer ve Akkuş	2018	Hava kirliliğini tahmin etmek amacıyla bulanık zaman serisinin bulanıklaştırma adımı bulanık k-medoid kümeleme algoritmasını kullanarak aykırı değerlerin ve anormal gözlemlerin tahminlerdeki olumsuz etkisini azaltan yeni bir bulanık zaman serisi modeli önermişler ve Türkiye'deki 65 izleme istasyonundan elde edilen haftalık SO ₂ konsantrasyon verilerini kullanarak bu yöntemin performansını bulanık zaman serileri tabanlı FCM ve Gustafson-Kessel tahmin modelleri ile karşılaştırmışlardır.

2.5. Makine ve Derin Öğrenme Algoritmalarını Kullanan Çalışmalar

Kişi	Tarih	Amaç
Li ve Shue	2004	Tayvan Devleti'nin kirlilik kontrol kurumuna referans olabilecek mevcut kirlilik dağılımını ortaya çıkarmak amacıyla ülke genelindeki 71 istasyondan elde edilen PM10 verilerini kullanarak dalgacık dönüşümü (WT) ve SOM sinir ağlarını ardışık olarak uygulamışlardır.
Karaca ve Camci	2010	İstanbul'daki 10 istasyona ait 2008 yılı ortalama saatlik PM10 değerlerini kullanarak hava kütlelerinin uzun menzilli ulaşım yapılarının bölgesel PM profiline etkilerini değerlendirmişlerdir. Bunun için öncelikle hibrit tek parçacıklı langrange entegre yörünge (HYSPLIT) modeliyle beş-gün saatlik geriye doğru hava kütlesi yörüngelerini elde etmiş, ardından bu yörüngeleri SOM ile kümelemişlerdir.
Bougoudis ve ark.	2016	Hava kirlleticilerin konsantrasyonlarının öngörülerek şehir merkezindeki insanların korunmasına yönelik doğru kararların alınmasına yardımcı olmak üzere veri kümesinin yeniden örneklemesini gerçekleştirmek için SOM uygulanması, ardından elde edilen kümelenmiş veriler üzerinden ileri beslemeli yapay sinir ağları (FFANN) ve RF'den oluşan toplulukların eğitilmesi ve son olarak elde edilen modellerin Mamdani kural tabanlı bulanık çıkarım sistemi kullanılarak optimize edilmesi aşamalarından oluşan HISYCOL adını verdikleri bir yaklaşım önermişler.
Wu ve ark.	2017	Veriye dayalı olarak kentsel hava kirliliği bölgelerini belirlemek için öncelikle AQI ile meteorolojik veriler arasındaki ilişkiyi sık örüntü büyüme (FP-Growth) algoritmasını kullanarak belirleyen ve ardından şehirleri coğrafi olarak sınırlı bir aralıkta kalacak şekilde kümeleyebilen Voronoï'ye dayalı yoğunluk kümelemesi (DCBV) ile benzer yapıya sahip bölgelere ayırtıran bir yöntem önermişlerdir.
Maciag ve ark.	2019	Hava kirliliğini öngörmek amacıyla zaman serilerinin eğitim setlerini belirlemek için kümeleme algoritmasının kullanıldığı ve elde edilen her küme için tek bir gelişen uyarımlı sinir ağı (eSNN) kurulduğu yeni bir kümeleme tabanlı topluluk modeli (CEE-SNN) önermişler ve bu yöntemi 2016-2017 yılları arasında Londra'daki 3 istasyondan elde edilen veriler üzerinden yalnızca tek bir uyarımlı sinir ağı (SNN) kullanıldığı düz (plain) NeuCube modeli, MLP ve ARIMA modelleri ile karşılaştırarak geliştirilen yöntemin diğerlerine göre çok daha iyi öngörü değerleri verdiğini belirtmişlerdir.
Althuwaynee ve ark.	2020	Malezya'nın başkenti Kuala Lumpur ve çevresindeki alanlarda PM10 yüzdelik değerleri ile diğer hava kirleticileri arasındaki ilişkiyi test etmek için Ocak 2016-Aralık 2016 tarihleri arasındaki saatlik verileri değişik makine öğrenme algoritmaları ile incelemişlerdir. Sınıflandırma analizi için R programlama ortamındaki Karar Ağacı (DT) algoritmasının doğrusal olmayan yinelemeli bölümlenme ve regresyon ağacı (RPart) modelini kullanmışlar, kirleticilerin PM10 oluşumu açısından önemini ise DT algoritmalarının Rastgele Orman (RF) algoritmasını kullanarak değerlendirmişler, K-ortalama kutupsal kümeleme fonksiyonu ile benzer özellikleri tanımlayarak gruplandırmışlar ve endüstriyel alanlar çevresindeki ortam izleme verilerinde hayati kümeleri tespit etmişlerdir.
Yan ve ark.	2021	Pekin'in 6 saatlik AQI değerini öngören zamansal-mekansal kümelemeye dayalı üç derin öğrenme yöntemi geliştirerek performanslarını geri beslemeli sinir ağları (BPNN) kontrol modeli üzerinden karşılaştırmışlardır. Önerilen yöntemde, AQI'nın zamansal ve mekansal analizleri yapıldıktan sonra tüm veriler zamansal veya mekansal boyutta birkaç kümeye bölünmüş ve uzun kısa-süreli hafıza (LSTM), evrimsel sinir ağları (CNN), CNN-LSTM öngörü modellerini ya toplam verilere ya da farklı mekansal-zamansal kümelerdeki verilere uygulamışlardır.
Harishkumar ve ark.	2021	Batı Tayvan'ın 76 istasyonun bulunduğu, değişik kirleticili kaynaklarına ve meteorolojik yapıya sahip bölgesindeki PM2,5 kirliliğinin analizi ve etiketlenmemiş bilgi koleksiyonları ile kümeleme incelemesi gerçekleştirecek bir model oluşturmak için kümeleme analizinde ilki kümeleme modelinin eğitilmiş ağırlıklarının yüklendiği, ikincisi evrimsel otokodlayıcı ile modelin kurulduğu, üçüncüsü gruplamayı ve otomatik kodlayıcıyı evrimsel hazırlayan ve son olarak gruplamayı ve otomatik kodlayıcıyı aynı anda (tamamen ilişkili) hazırlayan olmak üzere dört farklı derin öğrenme modeli geliştirmişler.
Özel	2022	Çerkezköy istasyonundan elde edilen veriler ile sadece kirleticili verilerinin, sadece meteorolojik verilerinin ve hem kirleticili hem meteorolojik verilerin kullanıldığı 3 veri seti üzerinden PM10 ve PM2,5 konsantrasyonlarını tahmin etmek için RF, CART, SVR, KNN, ELM ve ANFIS makine öğrenme algoritmalarını kullanmış.
Hulkkonen ve ark.	2022	Helsinki şehrine ait saatlik verileri kullanarak değişen antropojenik aktivitenin, partikül madde ve diğer hava kirleticilerin atmosferik konsantrasyonları üzerindeki etkilerini belirlemek amacıyla, RF ve SOM algoritmalarını, geçmiş meteorolojik veriler, kirleticili konsantrasyonları ve trafik göstergeleri ile eğitmişlerdir. Yerel PM2,5 için farklı değişkenlerin önemi RF ile belirlenirken SOM ile boyut azaltma ve kümeleme gerçekleştirilmiş ve eğitilen SOM, değişen antropojenik aktiviteye sahip koşulları temsil eden pandemi dönemindeki yeni verilerin ait olduğu kümenin

		öngörülmesinde kullanılmıştır.
Wang ve ark.	2023	Meteorolojik verilerin analizinde kullanılmak üzere fonksiyonel veri analizi ile kümeleme regresyonunu birleştirerek elde ettikleri fonksiyonel kümeleme regresyon heterojenliği öğrenme modeli (FCR-HL) ile küme sayısını otomatik olarak seçecek şekilde algoritmasını önermişlerdir.
Ren ve Ji	2023	Hava kalitesini öngörmek için çok değişkenli hava kalitesi zaman serisilerinin dinamik, mekansal-zamansal ve doğrusal olmayan özelliklerini ele alan Takagi-Sugeno bulanık modelini kullanarak kümeleme tabanlı bir yöntem önermişlerdir.

2.6. Diğer Çalışmalar

Kişi	Tarih	Amaç
Bougoudis ve ark.	2014	Attika bölgesinin hava kirliliği yapısını belirlemek amacıyla 2000-2012 arası verileri kullanarak SOM'un performansını çok katmanlı yapay sinir ağıları (ML-ANN) ile karşılaştırarak değerlendirmişler.
Bao ve ark.	2015	Çin anakarasındaki 27 başkent ve 4 belediyeye ait Ocak 2001-Aralık 2012 ve Ocak-Aralık 2014 yıllarındaki API ve AQI'daki trendi inceledikten sonra hem Moran's I global mekansal otokorelasyon hem de lokal Moran's I otokorelasyon değerlerine göre her iki indeksi mekansal olarak kümelemişlerdir. Son olarak her iki indeksin ağırlık merkezinde kayma olup olmadığını incelemişlerdir.
Chen ve ark.	2017	Hava kirliliğine ilişkin kümeleme çalışmalarında kullanılmak üzere alandan bağımsız, doğru ve ölçeklenebilir bir özellik çıkarma yöntemi geliştirmişlerdir. Bu yöntem üç aşamalı olup ilk adımda veri setindeki tüm zaman serilerinin toplam alınarak birleştirilmiş haline topluluk ampirik mod ayrışımı (EEMD) algoritması uygulanır. Daha sonra PCA ile boyut küçültülür ve son olarak da en küçük kareler projeksiyon yöntemi ile tüm zaman serileri elde edilen vektörlere yansıtılır.
Ye ve ark.	2018	Çin'in 338 şehrine ait 2016 yılı verilerini hem Moran's I global mekansal otokorelasyon değerleri üzerinden hem de lokal Gi* mekansal otokorelasyon değerlerine göre şehirleri yüksek ve düşük değerlere sahip olacak şekilde kümeleyerek PM2,5 yoğunluğunu aylık ve yıllık olarak mekansal analizlerle incelemişler ve sıcak noktalar ile bu yoğunluğu etkileyen meteorolojik faktörleri belirlemişlerdir.
Ye ve ark.	2018	Çin'in 338 şehrine ait 2016 yılı verilerini kullanarak altı kirliletiçi üzerinden kapsamlı hava kalitesi endeksini geliştirmişler ve bu endeks üzerinden şehirlerin hava kirliliğini hem Moran's I global mekansal otokorelasyon değerleri üzerinden hem de lokal Moran's I otokorelasyon değerlerine göre şehirleri yüksek ve düşük değerlere sahip olacak şekilde kümeleyerek aylık ve yıllık bazda incelemişlerdir.
Chen ve ark.	2020	Hava kirliliği durumunu ve bölgesel farklılıkları kesitsel verileri yıl bazında Ward yöntemi ile sistematik ve panel verileri gri korelasyon kümelemesi ile incelemişler
Sunori ve ark.	2021	Hindistan'ın Uttarakhand eyaletine ait kapanma dönemini de içine alan 2020 pandemi yılı AQI'yı FCM ve olasılıksal sinir ağıları (PNN) ile "temiz" ve "zararlı" olacak şekilde 2 kümeye ayırarak karşılaştırmışlar.
Zareba ve ark.	2023	Krakow'da bulunan yaklaşık 100 sensörden Mart 2021-Şubat 2022 arası saatlik PM10 verilerini kullanarak dinamik zaman bükülmesine (DTW) sahip k-ortalamlar kümeleme ile ağaç kenarının kaldırılmasıyla uzamsal 'K'luster analizi (SKATER) algortimalarının performansını karşılaştırmışlar ve SKATER algoritmasının kendine özgü çalışma modu nedeniyle hızlı ve mekansal olarak değişen verileri analiz etmede optimal olmadığı; bunun yerine DTW'li k-ortalamlar algoritmasının bu tür verileri analiz etmede daha etkili olduğu sonucuna varmışlardır.

2. MATERYAL VE YÖNTEM

www.havaizleme.gov.tr sitesinden elde edilen veriler incelendiğinde, 2021 yılına kadar eklenen istasyonlar olduğu görülmektedir. Ayrıca Güneydoğu Anadolu Temiz Hava Bölgesine ait

istasyonlarda halen sadece PM10 ve SO2 değerlerinin ölçüldüğü gözlemlenmiştir. Yine genel olarak PM2.5 ve CO ölçümlerinin yapıldığı istasyon sayısının daha az olduğu ve eksik veri yüzdesi yüksek olan ve/veya tek bir değişkende fazla eksik veriye sahip olan istasyonların olduğu tespit

edilmiştir. Tüm bu nedenle çalışma, en fazla veriye sahip olunacak şekilde Ocak 2021-Aralık 2023 arasında Tablo 3.1’de listelenen 86 istasyondan elde edilen aylık ortalama SO₂, NO₂, NO, NO_x, O₃ and PM₁₀ değerleri ile yürütülmüştür. Buna rağmen veride %2,9’luk eksik veri kalmıştır. Bunun

için öncelikle eksik veriler tamamlanmış ve sonrasında çok değişkenli fonksiyonel veriler için geliştirilen kümeleme yöntemleri uygulanmıştır. Tüm analizler R version 4.4.2 ve RStudio kullanılarak gerçekleştirilmiştir.

Tablo 3.1. Çalışmada Yer Alan İstasyonlar

İstasyon Adı	İstasyon Adı	İstasyon Adı
Afyon Selçuklu Cami	Erzurum	Konya Karatay Sunaypark
Ağrı	Erzurum Palandöken	Kütahya Heymeana Cad
Ağrı Doğubeyazıt	Giresun Gemilerçekeği	Kütahya Kentpark
Ağrı Patnos	İğdır	Manisa Akhisar
Ankara Keçiören Sanatoryum	İğdır Aralık	Manisa Kırkağaç
Ankara Ostim	İstanbul Alibeyköy	Manisa Salihli
Ankara Polatlı	İstanbul Arnavutköy	Manisa Soma
Ankara Sıhhiye	İstanbul Avcılar	Muğla Fethiye
Ankara Ulus	İstanbul Başakşehir MTHM	Muğla Milas
Ankara Yaşamkent	İstanbul Beşiktaş	Muğla Yatağan
Antalya Gazipaşa	İstanbul Esenyurt MTHM	Nevşehir
Antalya Manavgat	İstanbul Kadıköy	Ordu Ünye
Ardahan	İstanbul Kağıthane	Rize
Artvin	İstanbul Kartal	Rize Ardeşen
Aydın Efeler	İstanbul Maslak	Sakarya Ozanlar MTHM
Aydın Germencik	İstanbul Sarıyer	Samsun Atakum
Aydın Nazilli	İstanbul Şile MTHM	Tekirdağ
Bayburt	İstanbul Silivri MTHM	Tekirdağ Çerkezköy MTHM
Bilecik Bozuyuk MTHM	İstanbul Tuzla	Trabzon Akçaabat
Bolu Karacayır Parkı	İstanbul Ümraniye	Trabzon Uzungöl
Burdur	İzmir Eğitim İstasyonu	Trabzon Valilik
Bursa Kestel Hilal Parkı	İzmir Karabağlar	Uşak
Bursa Kestel MTHM	İzmir Kemalpaşa	Yalova Armutlu MTHM
Bursa Nilüfer	İzmir Ödemiş	Zonguldak Çatalağzı Cumayanı
Çanakkale Biga İcdas	İzmir Torbalı	Zonguldak Çatalağzı Kuzyaka
Çankırı	Kars İstasyon	Zonguldak MusluTepekoy
Denizli Civril	Kastamonu	
Edirne	Kırklareli	
Edirne Keşan MTHM	Kırşehir	
Erzincan	Kocaeli Körfez MTHM	

3.1. Eksik Verilerin Tamamlanması

Josse, Husson (2012) Temel Bileşenler Analizi (TBA) özelinde tekli veya çoklu tamamlamaya izin veren düzenlenmiş iteratif TBA yöntemini geliştirmişlerdir. Her ne kadar çalışmalarında eksik verilerin en iyi tahminine odaklanmak yerine eksik veri altında temel bileşen katsayılarının ve bunların varyanslarının en iyi tahminini amaçlamış olsalar da yaptıkları karşılaştırmalar sonucu geliştirdikleri yöntemin bu amaçla kullanımının da avantajlı olduğunu göstermişlerdir. Ayrıca yöntemin uygulanabilmesi için R'da missMDA kütüphanesini geliştirmişlerdir.

3.2. Çok Değişkenli Fonksiyonel Veriler için Kümeleme Yöntemleri

Fonksiyonel veri analizinde (FDA) verilerin düzgün eğrilerden üretildiği yani türevlenebilir olduğu varsayılmakta ve analizler klasik istatistikteki nokta veriler yerine hepsinin tek bir gözlem sayıldığı eğriler üzerinden gerçekleştirilmektedir (Ramsay, Silverman 2005). Bu nedenle FDA'da ilk adım verinin temel özelliklerini taşıyan bu fonksiyonların oluşturulmasıdır.

3.2.1. Fonksiyonel Verilerin Düzgünleştirilmiş Fonksiyonlarla Temsili

En sık kullanılan baz fonksiyonları:

i. *Fourier fonksiyonları*: Periyodik ve herhangi bir aralıkta diğer aralıklara göre çok daha hızlı dalgalanmalar göstermeyen verilerde sıklıkla tercih edilen baz fonksiyonlardır. Ayrıca n ikinin katı ve aralıklar eşit olduğunda tüm katsayılar işlemsel olarak etkin bir şekilde hesaplanabildiğinden önemli bir işlemsel avantaj sağlar ve uzun zaman serileri için tercih edilmelerine sebep olur (Ramsay, Silverman 2005).

ii. *B-Splayn fonksiyonları*: düğümlerin sağladığı esnekliğin yanısıra hesaplama avantajına da sahip olur. Her türlü veriye uygulanabilir.

Baz fonksiyonlara karar verildikten sonra c_r katsayılarını elde etmek ve R değerine karar vermek için en küçük kareler, ağırlıklı en küçük kareler, yerleştirilmiş en küçük kareler, kantil regresyon, pürüzlü ceza veya TBA yöntemleri uygulanabilmektedir (Ramsay, Silverman 2005). Bu tez çalışması kapsamında seçilen yöntemler TBA tabanlı olduğundan diğer yöntemler ayrıca tartışılmamıştır.

3.2.2. İki Aşamada Yapılan Kümeleme Yöntemi

Bu yöntemde önce çok değişkenli fonksiyonel TBA (MFPCA) uygulanarak skorlar elde edilmiştir. Happ, Greven (2018) değişik boyuttaki ve/veya değişik aralıklardaki verilerin bir arada kullanılmasına olanak sağlayan ve verinin yoğun veya seyrek olmasına bakılmaksızın uygulanabilecek çok değişkenli fonksiyonel Karhunen–Loève temsili geliştirmiş ve bunun tek değişkenli TBA ile ilişkisini göstererek etkin bir algoritma sunmuşlar ve 4 adımlı MFPCA algoritması önermişlerdir:

S1. Her bir $X^{(1)}$ elemanı $x_{1(i)}, x_{2(i)}, \dots, x_{N(i)}$ gözlemlerine dayanan tek değişkenli FPCA tahmin edilir. Böylece $\hat{\phi}_m^{(j)}$ tahmini özfonksiyonlar ve $\hat{\xi}_{i,m}^{(j)}$ skorlar, $i = 1, \dots, N$, $m = 1, \dots, M_j$ elde edilir. Burada tek değişkenli FPCA için literatürdeki herhangi bir yöntem kullanılabilir.

S2. Her bir satırı tek bir gözleme ait tahmin edilen

$$(\hat{\xi}_{i,1}^{(1)}, \dots, \hat{\xi}_{i,M_1}^{(1)}, \dots, \hat{\xi}_{i,1}^{(p)}, \dots, \hat{\xi}_{i,M_p}^{(p)})$$

skorlarından oluşan $\Xi \in \mathbb{R}^{N \times M^+}$ matrisi

oluşturulur ve buna bağlı olarak

$$\hat{Z} = (N - 1)^{-1} \Xi^T \Xi, \hat{Z} \in \mathbb{R}^{M^+ \times M^+}$$

tahmini matrisi elde edilir.

S3. $\hat{Z}$ için matris özelemleri gerçekleştirilerek $\hat{v}_m$ özdeğerleri ve $\hat{e}_m$ ortonormal özvektörleri elde edilir.

S4. Çok değişkenli özfonksiyonlar için tahminler,

$$\hat{\psi}_m^{(j)}(t_j) = \sum_{n=1}^{M_j} [\hat{e}_m]_n^{(j)} \hat{\phi}_n^{(j)}(t_j) \quad t_j \in [0, T_j], \quad m = 1, \dots, M^+$$

elemanları tarafından verilirken çok değişkenli skorlar

$$\hat{\rho}_{i,m} = \sum_{j=1}^p \sum_{n=1}^{M_j} [\hat{e}_m]_n^{(j)} \hat{\xi}_{i,n}^{(j)} = \Xi_{i,*} \hat{e}_m$$

şeklinde hesaplanır.

Bu ilk aşama için yine Happ, Greven (2018) tarafından R'da yazılan MFPCA kütüphanesi kullanılmış olup ikinci aşamada elde edilen bu

skorlara k-ortalamlar yöntemi uygulanarak istasyonlar gruplandırılmışlardır. Bu tez çalışmasında, iki aşamada yapılan kümeleme yöntemleri incelendiğinde literatür taraması yapılırken bu yöntemin iki aşamalı kümeleme yöntemleri olarak adlandırılabilirdiği de tespit edilmiştir.

3.2.3. Model Tabanlı Kümeleme Yöntemi

Jacques, Preda (2014), Bölüm 2.2.2’de detayları anlatılmış olan MFPCA yönteminin özel hali olan bir MFPCA ile oluşturulan temel bileşenlerin belli bir oranını kullanarak elde ettikleri skorların normal dağıldığı varsayımı altında, verileri, grup

spesifik fonksiyonel alt uzaylara yerleştiren fonksiyonel gizli karışım modeli tanımlayarak

teorik garantisi olmayan sezgisel EM benzeri bir algoritma ile tahmin etmişlerdir. Daha sonra Schmutz vd. (2020) bu yaklaşımı tüm temel bileşenleri kullanacak şekilde genişleterek sezgisel algoritma yerine teorik garantisi olan EM algoritması ile tahmin etmişlerdir.

Uygulamada latent değişken değerleri bilinmediğinden Schmutz vd. (2020) EM algoritmasını önermişler ve R’da funHDDC kütüphanesini geliştirmişlerdir. Küme sayısının (K) ise Akaike bilgi kriteri (AIC) veya Bayesçi bilgi kriteri (BIC) gibi kriterleri maksimize eden değer olarak alınabileceği belirtilmiştir.

Daha sonra bu yöntem Anton, Smith (2024) tarafından normallik varsayımı esnetilerek çok değişkenli t dağılımlarının karışımı altında revize edilerek aykırı değerlere karşı daha sağlam hale getirilmiş ve R’da TFunHDDC kütüphanesi oluşturulmuştur.

4. BULGULAR VE TARTIŞMA

Öncelikle toplanan istasyon verileri istasyonların bulunduğu coğrafi bölgelere göre gruplandırılarak her bir kirlenici açısından incelenmiştir. **Şekil 4.1’**de Akdeniz Bölgesi’ndeki istasyonlara ait hava kirlenicilerinin zaman grafikleri verilmiştir. Manavgat ve Burdur’un hava kirlenicilerinin zaman içinde SO₂ hariç benzer seyrettiği, SO₂ itibarıyla ise Gazipaşa ve Burdur’un daha benzer olduğu görülmektedir. Gazipaşa PM₁₀ ve O₃ hariç en düşük değerlere sahipken, O₃ itibarıyla en yüksek değerlere sahip olup PM₁₀ açısından diğerlerine yakın seyretilmektedir. Genel olarak kış aylarında O₃ azalırken diğer kirlenicilerin arttığı görülmektedir. O₃’ün ise yaz aylarında arttığı görülmektedir.

Şekil 4.2 incelendiğinde, Ege Bölgesi’ndeki istasyonlara ait hava kirlenicilerinin zaman grafiklerinde kirlenicilerdeki mevsimselliğin Akdeniz bölgesindeki istasyonlarla benzerlik gösterdiği görülmektedir. Bununla birlikte bazı istasyonların Antalya-Gazipaşa istasyonundaki gibi NO, NO_x, NO₂ ve SO₂ açısından diğerlerine görece daha düzgün olup mevsimsellik göstermediği ya da görece düşük mevsimsellik gösterdiği söylenebilir.

Şekil 4.3 incelendiğinde, Marmara Bölgesi’ndeki istasyonlara ait hava kirlenicilerinin de benzer mevsimsellik gösterdikleri ve yine bazı istasyonlardaki ölçümlerin diğerlerine göre oldukça yüksek olduğu görülmektedir. Özellikle İstanbul Alibeyköy’de NO_x değerinin 2023 yılının Ağustos ayında tüm istasyonlar arasında en yüksek değere ulaştığı (grafik bu değerden dolayı görüntüsünü kaybettiği için y-ekseni kesilmiştir) ve bu değer kendi içindeki diğer değerlerle uyumsuz olması ve sonrasında da bu istasyonda NO_x ölçümünün olmaması burada bir cihaz problemi ile karşılaşıldığını düşündürdüğünden bundan sonraki analizlerde bu değer analiz dışı bırakılmıştır.

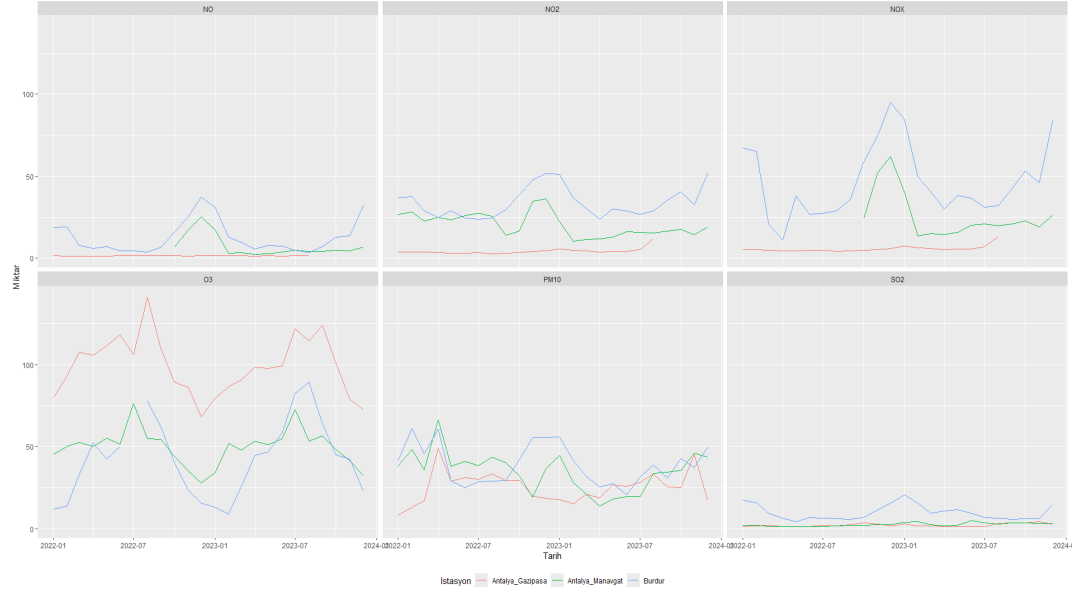
Şekil 4.4’de İç Anadolu Bölgesi’ndeki istasyonlara ait ilgili grafikler verilmiş olup benzer mevsimsel etki burada da gözlemlenmiştir. Yine diğerlerine görece daha yüksek değerlere sahip istasyonların varlığı ve bunların miktarsal olarak İstanbul Alibeyköy istasyonundan elde edilen değerlere yakınlık gösterdiği görülmektedir.

Şekil 4.5’te görüldüğü gibi Karadeniz Bölgesi’ndeki istasyon verilerinin de diğer bölgelerdekine benzer mevsimsellik içerdiği görülürken, özellikle Bayburt istasyonunda NO ve NO_x değerlerinde hem diğer kirlenicilere göre hem de bölge geneline göre kış aylarında aşırı bir yükseliş olduğu görülmektedir.

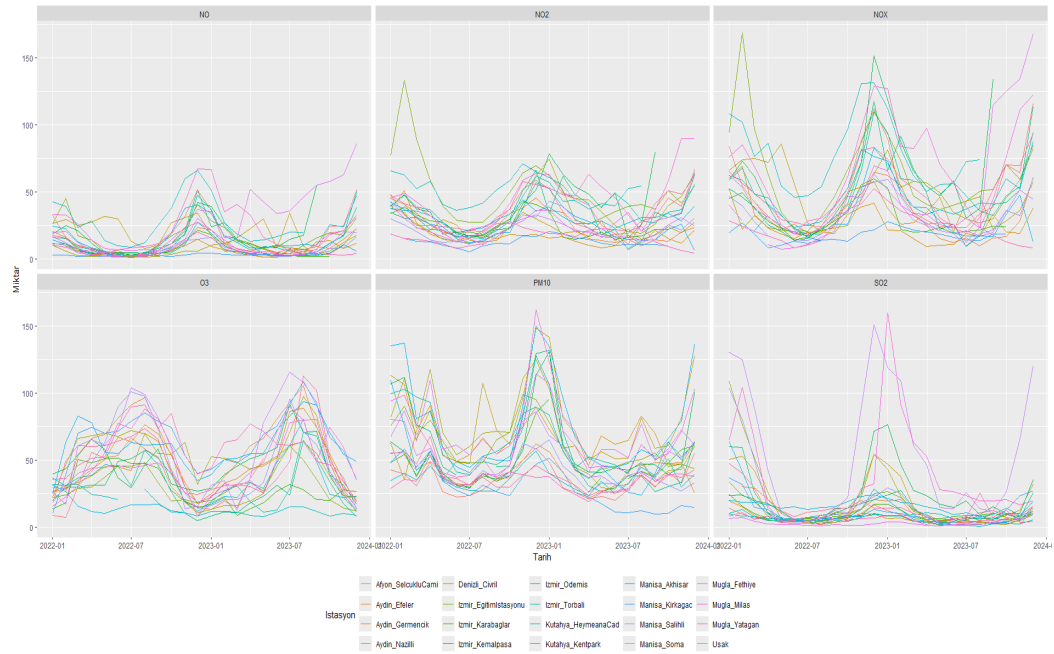
Şekil 4.6’da Benzer mevsimsellik yapısının Doğu Anadolu Bölgesi’nde de korunduğu ancak yine özellikle NO_x açısından İstanbul Alibeyköy değerlerine ulaşan istasyonların olduğu gözlemlenmektedir.

Şekil 4.7’de Türkiye geneli olarak mevsimsel yapının benzer olduğu görülmekle birlikte birkaç istasyonun yapıca ayrıştığı, bazı istasyon verilerinin ise zaman içinde daha düzgün seyrettiği görülmektedir. NO₂, NO ve NO_x kirlenicileri itibarıyla diğerlerinden daha yüksek değerlere sahip bir grup istasyonun olduğu, farklı istasyon grubunun ise PM₁₀ açısından daha yüksek değerler göstererek ayrıştığı gözlemlenebilmektedir.

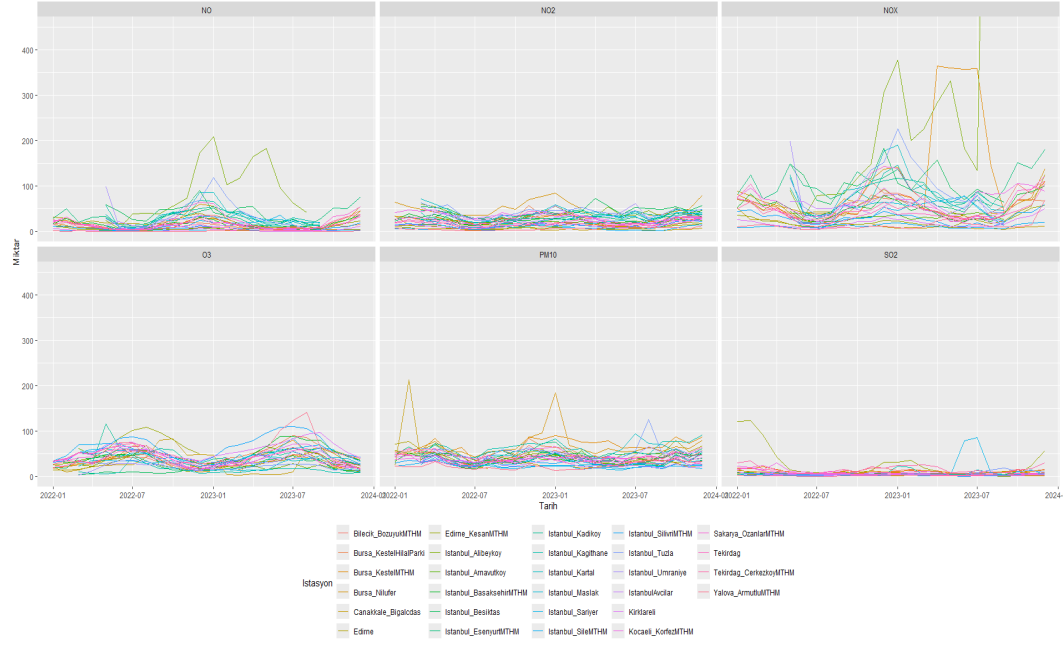
Şekil 4.1. Akdeniz Bölgesi'ndeki İstasyonlara Ait Zaman Grafikleri



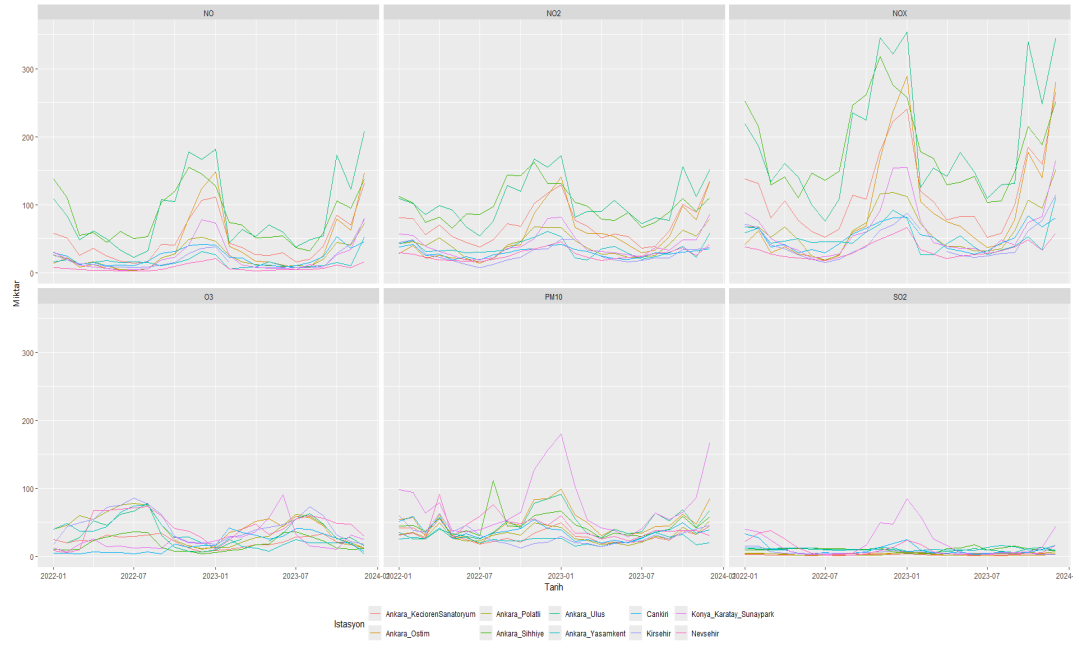
Şekil 4.2. Ege Bölgesi'ndeki İstasyonlara Ait Zaman Grafikleri



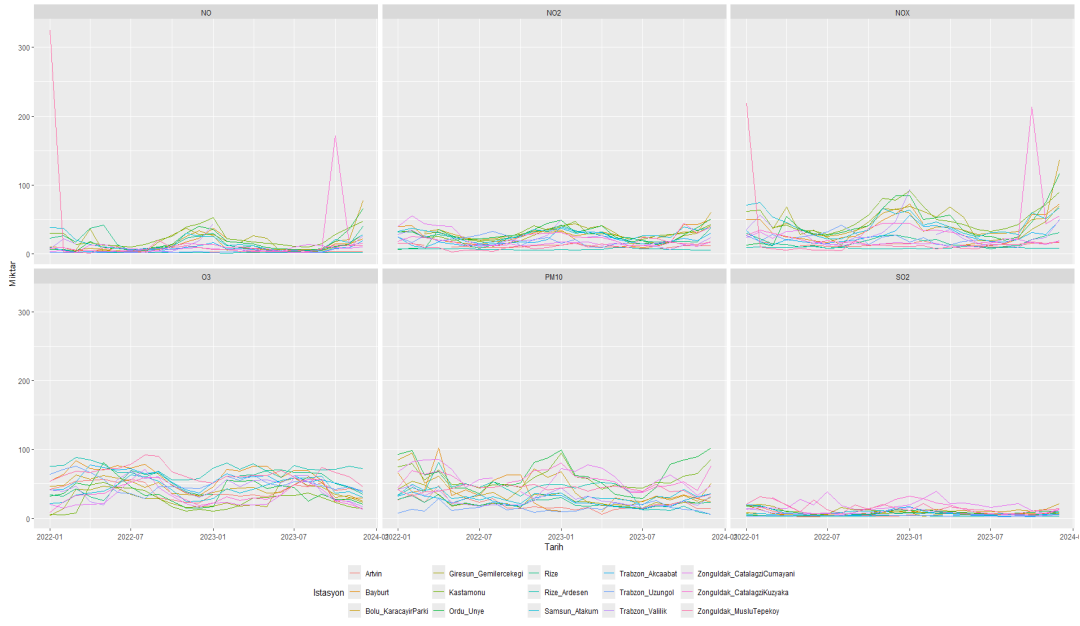
Şekil 4.3. Marmara Bölgesi'ndeki İstasyonlara Ait Zaman Grafikleri



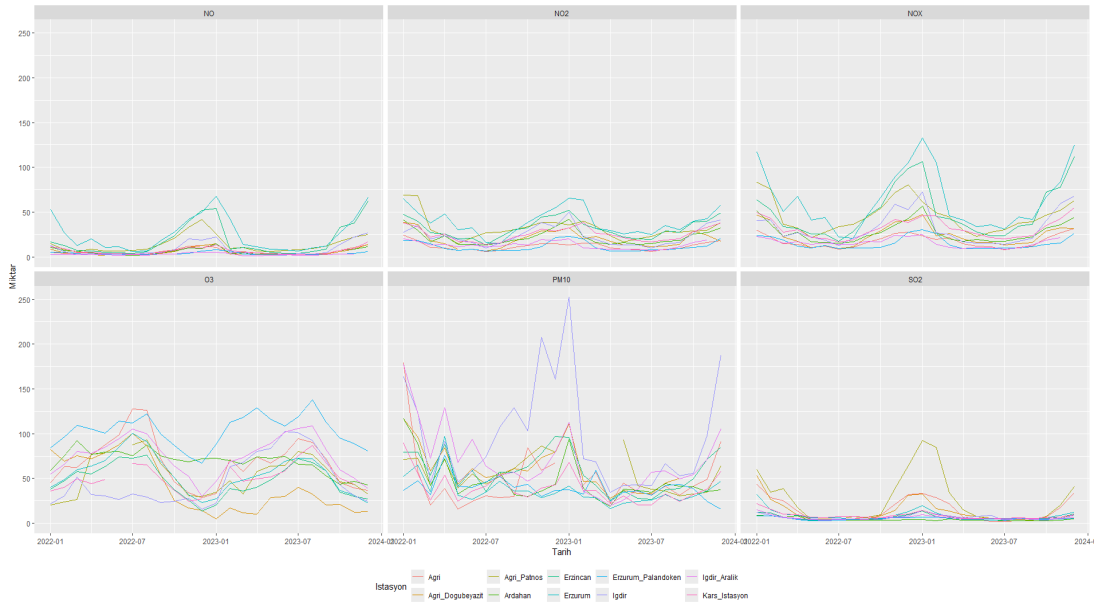
Şekil 4.4. İç Anadolu Bölgesi'ndeki İstasyonlara Ait Zaman Grafikleri



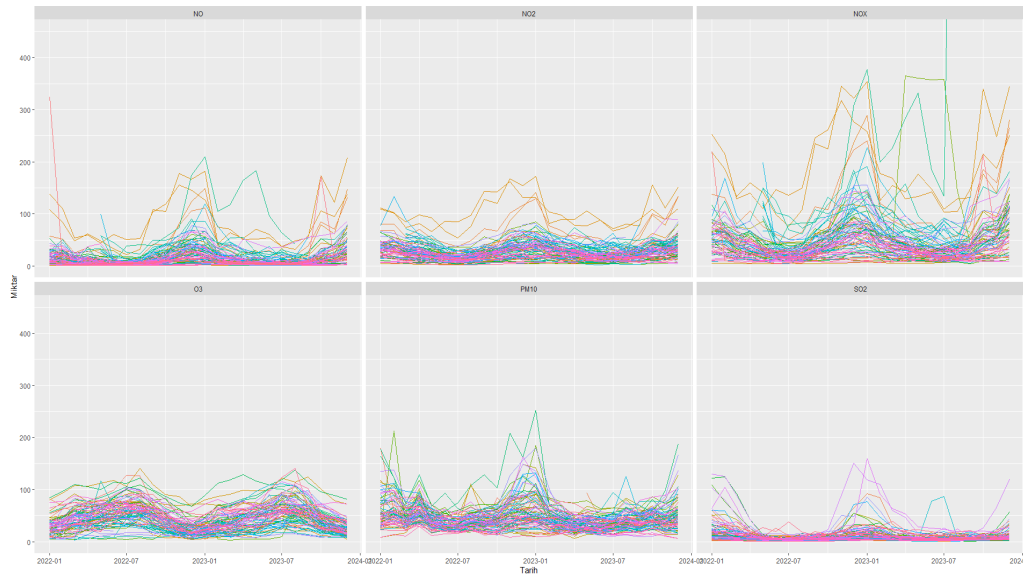
Şekil 4.5. Karadeniz Bölgesi'ndeki İstasyonlara Ait Zaman Grafikleri



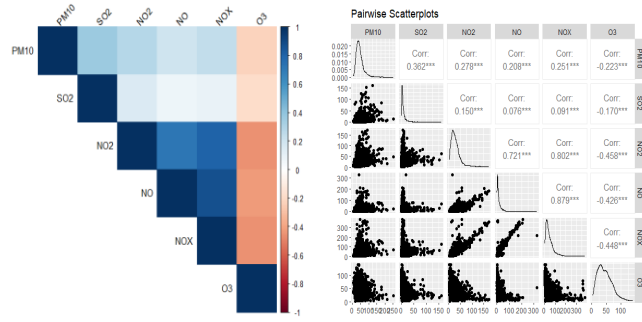
Şekil 4.6. Doğu Anadolu Bölgesi'ndeki İstasyonlara Ait Zaman Grafikleri



Şekil 4.7. Bütün İstasyonlara Ait Zaman Grafikleri

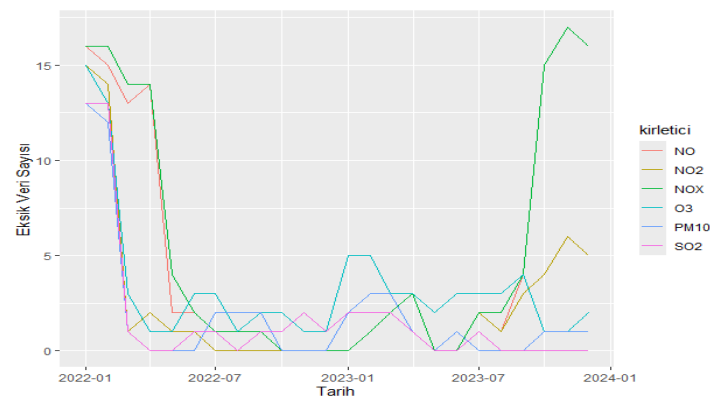


Şekil 4.8. Hava Kirleticileri Arasındaki Korelasyon Grafikleri



İstanbul Alibeköy NOX son değeri veri setinden çıkartıldıktan sonra kirleticiler arasındaki korelasyon değerleri incelenmiştir. Şekil 4.8'den görüleceği gibi NO, NO2 ve NOX arasındaki kuvvetli bir ilişki vardır. SO2'nin PM10 hariç diğer kirleticiler ile en düşük ilişkiye sahip kirleticisi olduğu görülmektedir. Bu ilişki istatistiksel olarak anlamlı olmakla birlikte korelasyon değerlerinin çok düşük olması nedeniyle analizler açısından anlamlı bir ilişki olmadığı görülmektedir. O3 kirleticisinin ise diğerleri ile negatif ilişkiye sahip olduğu görülürken genel olarak tüm kirleticiler arasında ilişki olduğu söylenebilir. Dolayısı ile analizlerde bu ilişkiyi dikkate almaları bakımından MFPCA tabanlı kümeleme yöntemleri tercih edilmiştir.

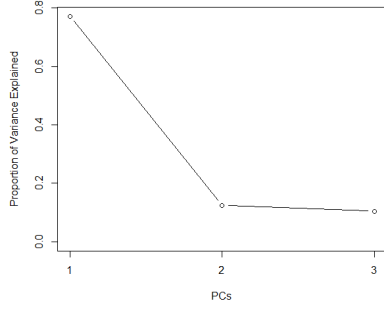
Şekil 4.9. Her Kirleticisiye Ait Toplam Eksik Veri Sayılarının Zaman Grafiği



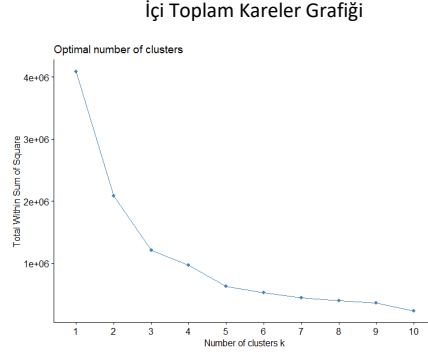
Tüm bunlara ek olarak veri setinde %2,9 oranında eksik veri olduğu ve Şekil 4.9 incelendiğinde bunların genellikle ilk veya son veriler olduğu görülmektedir. Bu eksik veriler, değişkenler arasında yüksek korelasyon olması nedeniyle Josse, Husson (2012) tarafından önerilen çoklu atama şeklindeki iteratif TBA yöntemiyle her bir ay için ayrı ayrı tamamlanmıştır. Eksik veri oranı düşük olduğundan ve tezin ana amacına yönelik olmadığından bu atamanın farklı yöntemler denenmesi yolu ile sonuçlara etkisi tartışılmayarak kapsam dışı bırakılmıştır.

4.1 İki Aşamada Yapılan Kümeleme Yönteminin Sonuçları

Şekil 4.10. Temel Bileşenlere Ait Yamaç Grafiği

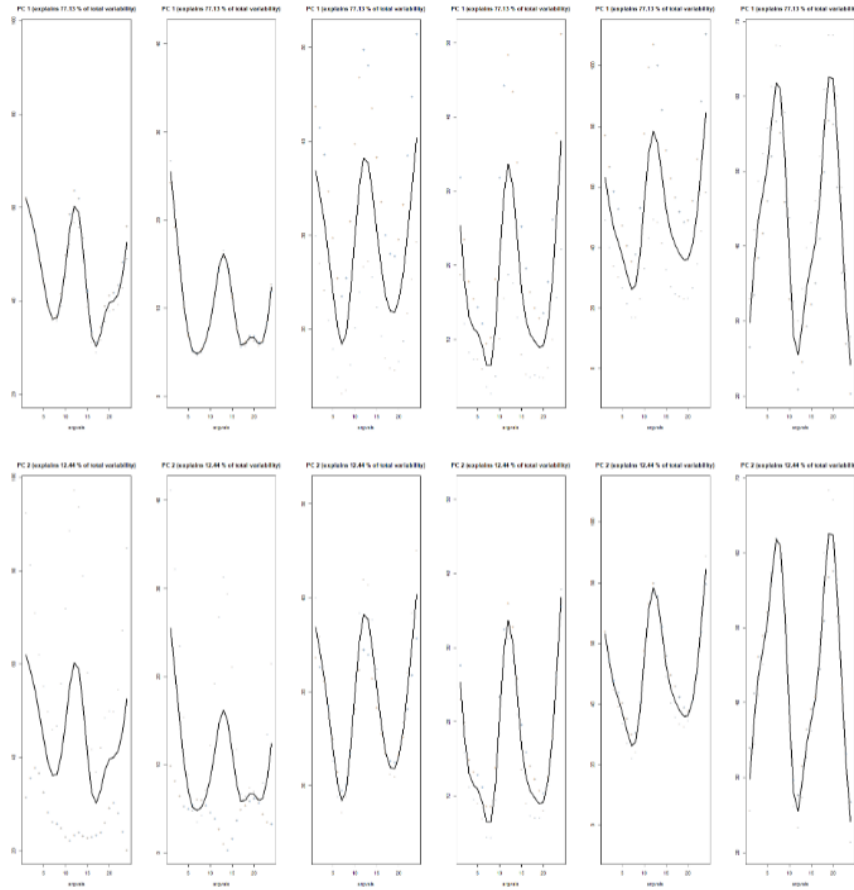


Şekil 4.11. Küme Sayısına Karşın Kümeler İçi Toplam Kareler Grafiği



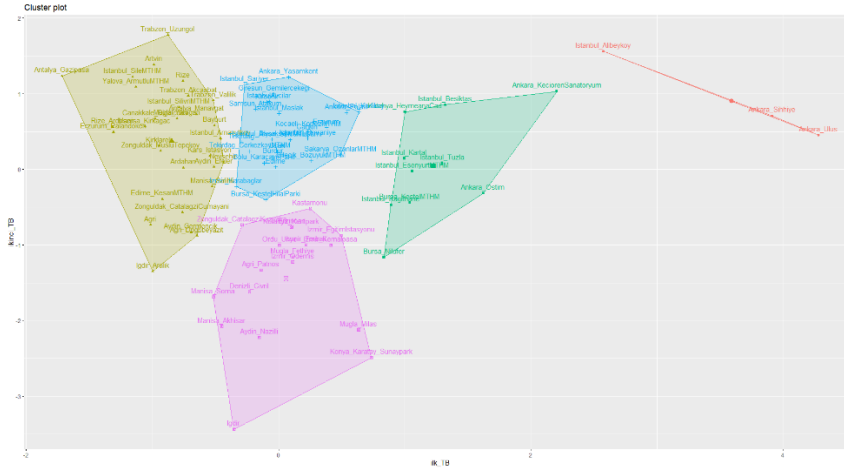
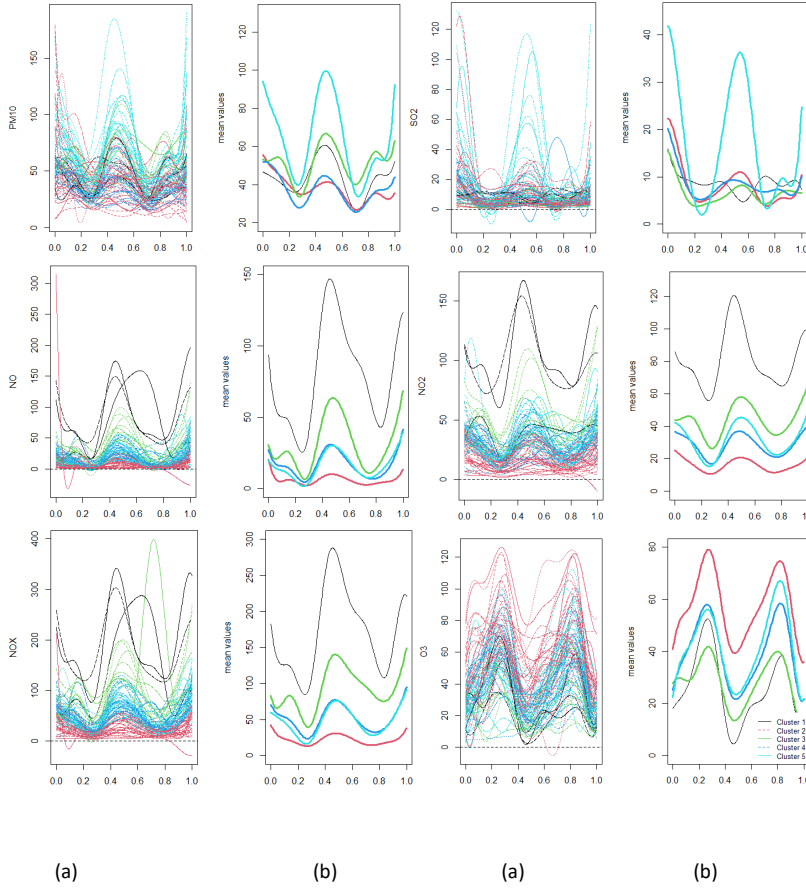
Öncelikle verilere MFPCA uygulanmıştır. Elde edilen özvektörlere ait **Şekil 4.10'**da verilen yamaç grafiği incelendiğinde ilk temel bileşenin varyanstaki değişimin %77'sini açıklarken ilk iki bileşenin toplamda değişimin %89,6'sını açıkladığı görülebilir. Bu nedenle k-ortalamalar kümelemesinin ilk iki temel bileşene ait skorlar üzerinden gerçekleştirilmesine karar verilmiştir. Küme sayısına karar vermek için **Şekil 4.11'**de çizilen küme sayısına karşın kümeler içi toplam kareler grafiğine bakıldığında uygun küme sayısının 5 olduğu görülmektedir.

Şekil 4.12. Değişkenlere Ait Birinci ve İkinci Tahmini Çok Değişkenli Fonksiyonel Temel Bileşen Grafiği



Değişkenlere ait birinci ve ikinci tahmini çok değişkenli fonksiyonel temel bileşenin çizdirildiği **Şekil 4.12'**de (üst satırdaki grafikler birinci, alt satırdakiler ikinci bileşeni verirken, grafikler sırası ile PM10, SO2, NO2, NO, NOX ve O3 için çizilmiştir), ilk temel bileşen PM10 ve SO2 ile uyumlu olduğu, ikinci bileşenin ise diğer değişkenlerle daha uyumlu olduğu görülmektedir.

Şekil 4.13. İstasyonların İlk 2 Temel Bileşene Göre Oluşturdukları Kümeler

Şekil 4.14. İki Aşamada Yapılan Kümeleme Yöntemine Göre Kirleticilerin
Grup (a) Eğrileri (b) Ortalama Eğrileri

5 küme sayısı ile elde edilen kümelerin temel bileşen grafiğindeki konumları **Şekil 4.13**'de verilmiştir. 1. kümedeki istasyonların her iki temel bileşene göre yüksek değerler aldığı, 3. kümedekilerin ise yine her iki temel bileşene göre yüksek değer almakla birlikte görece birinci kümedekilerden daha düşük olduğu görülmektedir. 4 ile 5. kümelerin birinci temel bileşene göre ortalamada olmalarına rağmen 4. kümenin ikinci temel bileşene göre yüksek puanlar alırken 5. kümenin düşük puanlar aldığı görülmektedir. 2. kümedeki istasyonların ise birinci temel bileşene göre en düşük puanları alırken ikinci temel bileşene göre puanlarının yüksek olduğu hatta görece 1. kümedeki istasyonlardan daha yüksek değerlere sahip istasyonları içerdiği görülmektedir.

Her grup için grup eğrileri ile grup ortalama eğrileri de **Şekil 4.14**'te, her kümedeki istasyonların listesi ise **Tablo 4.1**'de verilmiştir. 1. Gruptaki istasyonların NO, NO₂ ve NO_x açısından en yüksek grup ortalama değerlerine sahip iken O₃ açısından düşük, PM₁₀ açısından ortalamada oldukları gözükmemektedir. 2. gruptaki istasyonların ise NO, NO₂, NO_x ve PM₁₀ açısından en düşük grup ortalamasına sahip oldukları görülürken O₃ açısından en yüksek grup ortalamasına sahip oldukları gözlemlenmektedir. 3, 4 ve 5. Gruplardaki istasyonların NO, NO₂ ve NO_x açısından ortalamada oldukları gözlemlenirken birbirlerinden özellikle PM₁₀ miktarları açısından ayrışmakta oldukları görülmektedir. SO₂ kirleticisinde ise grupların hem miktarsal hem de yapısal anlamda farklılaşarak gruplandıkları dikkat çekmektedir.

Tablo 4.1. İki Aşamada Yapılan Kümeleme Yöntemi İle Oluşturulan İstasyon Grupları

Küme	İstasyonlar	İstasyon sayısı
1	İstanbul Alibeyköy, Ankara Sıhhiye, Ankara Ulus	3
2	Antalya Gazipaşa, Antalya Manavgat, Aydın Efeler, Aydın_Germencik, Manisa Kırkağac, Manisa Salihli, Muğla Yatağan, Çanakkale Biga İcdas, Edirne Keşan MTHM, İstanbul Arnavutköy, İstanbul Şile MTHM, İstanbul Silivri MTHM, Kırklareli, Yalova Armutlu MTHM, Nevşehir, Artvin, Bayburt, Rize, Rize Ardeşen, Trabzon Akçaabat, Trabzon Uzungöl, Trabzon Valilik, Zonguldak Çatalağzı Cuma yanı, Zonguldak Muslu Tepeköy, Ağrı, Ağrı Doğubeyazıt, Ardahan, Erzurum Palandöken, Iğdır Aralık, Kars İstasyon	30
3	Kütahya Heymeana Cad, Bursa Kestel MTHM, Bursa Nilüfer, İstanbul Beşiktaş, İstanbul Esenyurt MTHM, İstanbul Kağıthane, İstanbul Kartal, İstanbul Tuzla, Ankara Keçiören Sanatoryum, Ankara Ostim	10
4	Burdur, Afyon Selçuklu Cami, İzmir Karabağlar, Uşak, Bilecik Bozuyuk MTHM, Bursa Kestel Hilal Parkı, Edirne, İstanbul Avcılar, İstanbul Başakşehir MTHM, İstanbul Kadıköy, İstanbul Maslak, İstanbul Sarıyer, İstanbul Ümraniye, Kocaeli Körfez MTHM, Sakarya Ozanlar MTHM, Tekirdağ, Tekirdağ Çerkezköy MTHM, Ankara Polatlı, Ankara Yaşamkent, Çankırı, Kırşehir, Bolu Karaçayır Parkı, Giresun Gemilerçekeği, Samsun Atakum, Erzurum	25
5	Aydın Nazilli, Denizli Civrili, İzmir Eğitim İstasyonu, İzmir Kemalpaşa, İzmir Ödemiş, İzmir Torbalı, Kütahya Kentpark, Manisa Akhisar, Manisa Soma, Muğla Fethiye, Muğla Milas, Konya Karatay Sunaypark, Kastamonu, Ordu Ünye, Zonguldak Çatalağzı Kuzyaka, Ağrı Patnos, Erzincan, Iğdır	18

4.2 Model Tabanlı Kümeleme Yönteminin Sonuçları

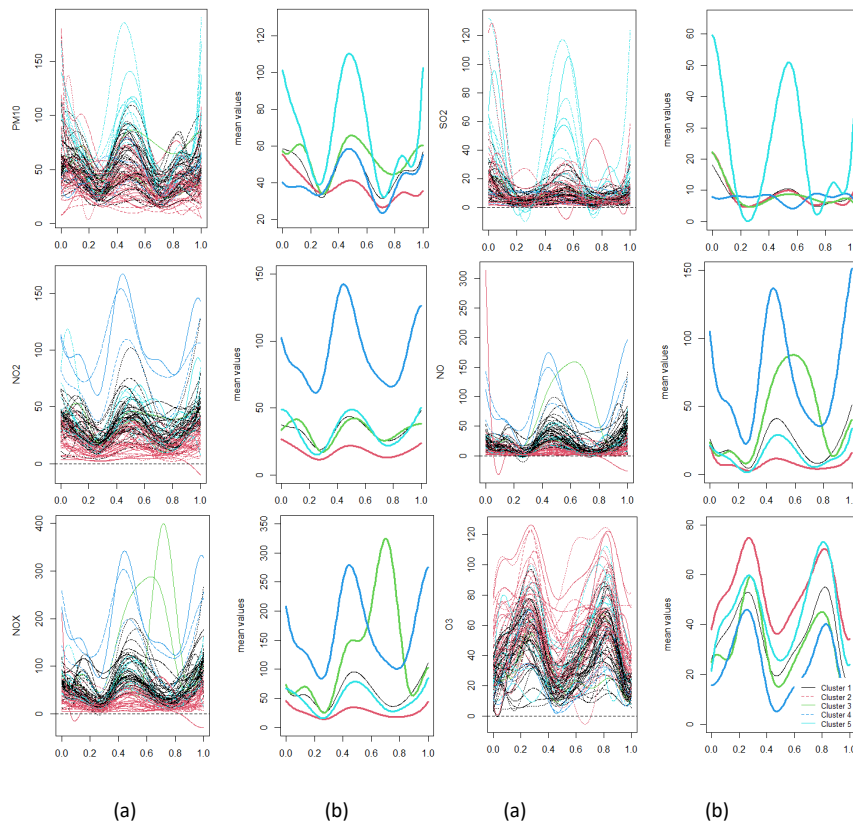
Tablo 4.2. Küme sayısı ve modele göre en iyi ilk 15 sonucun BIC değerleri

Model	Küme Sayısı	BIC değeri
$\alpha_k b Q_k d_k$	5	-381,072.92
$\alpha_{kj} b Q_k d_k$	5	-384,092.84
$\alpha_{kj} b Q_k d_k$	2	-425,135.97
$\alpha_k b Q_k d_k$	2	-425,190.01
$\alpha_k b Q_k d_k$	3	-439,292.63
$\alpha b Q_k d_k$	3	-439,387.08
$\alpha b Q_k d_k$	2	-447,214.10
$\alpha_k b Q_k d_k$	4	-455,467.87
$\alpha_k b Q_k d_k$	2	-464,037.16
$\alpha_{kj} b Q_k d_k$	3	-488,252.58
$\alpha_{kj} b Q_k d_k$	4	-500,338.85
$\alpha b Q_k d_k$	4	-568,005.56
$\alpha b Q_k d_k$	5	-591,343.89
$\alpha_k b Q_k d_k$	6	-668,798.82
$\alpha b Q_k d_k$	6	-952,260.20

Önerilen 5 model için 2'den 10'a kadar tüm küme sayıları denenerek BIC kriterine göre en iyi sonucu veren modele karar verilmiştir. Ancak pek çok modelin yakınsama problemi yaşadığı gözlemlenmiş ve sonuçlar elde edilememiştir. Sonuç alınabilenler içinde en iyisinin 5 kümeli α_k parametrelerinin her grup içinde, b_k parametrelerinin ise gruplar arasında ortak olacak şekilde sabitlendiği $\alpha_k b Q_k d_k$ modeli olduğu görülmüştür.

(Tablo 4.2).

Şekil 4.15. Model Tabanlı Kümeleme Yöntemine Göre Kirleticilerin Grup (a) Eğrileri (b) Ortalama Eğrileri



Belirlenen bu model altında elde edilen her grup için grup eğrileri ile grup ortalama eğrileri de Şekil 4.15'te verilirken, gruplardaki istasyonların listesi de Tablo 4.3'de verilmiştir. 1. gruptaki istasyonların genel olarak her kirletici için ortalama değerlere sahip olduğu görülebilir. 2. gruptaki istasyonların ise O3'de yüksek değerlere sahipken diğer kirleticiler açısından genel olarak en düşük seviyelerde seyrettiği söylenebilir. 3. gruptakilerde ise NO ve NOX yüksek değerlerde seyrederken diğerlerinin ortalamada olduğu görülmektedir. 4. gruptakilerin NO, NO2 ve NOX açısından daha yüksek değerlere sahipken O3 değerlerinin en düşük, diğer kirleticilerin ise ortalama olduğunu görülmektedir. 5. gruptaki istasyonlarda PM10 ve SO2 değerlerinin diğerlerinden yüksek iken diğer kirleticiler açısından diğerlerine göre ortalamada oldukları görülmektedir. Genel olarak küme ayrışımının bazen miktarsal bazen ise yapısal farklılıklardan kaynaklandığı gözlemlenebilir.

Tablo 4.3. Model Tabanlı Kümeleme Yöntemi İle Oluşturulan İstasyon Grupları

Küme	İstasyonlar	İstasyon Sayısı
1	Burdur, Afyon Selçuklu Cami, İzmir Torbalı, Kütahya Heymeana Cad, Kütahya Kentpark, Muğla Fethiye, Uşak, Bilecik Bozuyuk MTHM, Bursa Kestel Hilal Parkı, Bursa Nilüfer, Edirne, İstanbul Başakşehir MTHM, İstanbul Beşiktaş, İstanbul Esenyurt MTHM, İstanbul Kadıköy, İstanbul Kağıthane, İstanbul Kartal, İstanbul Tuzla, İstanbul Ümraniye, Kocaeli Körfez MTHM, Sakarya Ozanlar MTHM, Tekirdağ Çerkezköy MTHM, Ankara Ostim, Ankara Polatlı, Ankara Yaşamkent, Çankırı, Kırşehir, Bolu Karaçayır Parkı, Giresun Gemilercekeği, Kastamonu, Ordu Ünye, Zonguldak Çatalağzı, Kızvaka, Erzincan, Erzurum	34
2	Antalya Gazipaşa, Antalya Manavgat, Aydın Efeler, Aydın_Germencik, İzmir Karabağlar, Manisa Kırkağaç, Manisa Salihli, Muğla Yatağan, Çanakkale Biga Icdas, Edirne Keşan MTHM, İstanbul Arnavutköy, İstanbul Avcılar, İstanbul Maslak, İstanbul Sarıyer, İstanbul Şile MTHM, İstanbul Silivri MTHM, Kırklareli, Tekirdağ, Yalova Armutlu MTHM, Nevşehir, Artvin, Bayburt, Rize, Rize Ardeşen, Samsun Atakum, Trabzon Akçaabat, Trabzon Uzungöl, Trabzon Valilik, Zonguldak Çatalağzı Cuma yanı, Zonguldak Muslu Tepeköy, Ağrı, Ağrı Doğubeyazıt,	36
3	Bursa Kestel MTHM, İstanbul Alibeyköy	2
4	Ankara Keçiören Sanatoryum, Ankara Sıhhiye, Ankara Ulus	3
5	Aydın Nazilli, Denizli Civril, İzmir Eğitim İstasyonu, İzmir Kemalpaşa, İzmir Ödemiş, Manisa Akhisar, Manisa Soma, Muğla Milas, Konya Karatay Sunaypark, Ağrı Patnos, Iğdır	11

Tablo 4.4. Temel Bileşenlerin Normallik Testine Ait W Test İstatistikleri (p-değerleri)

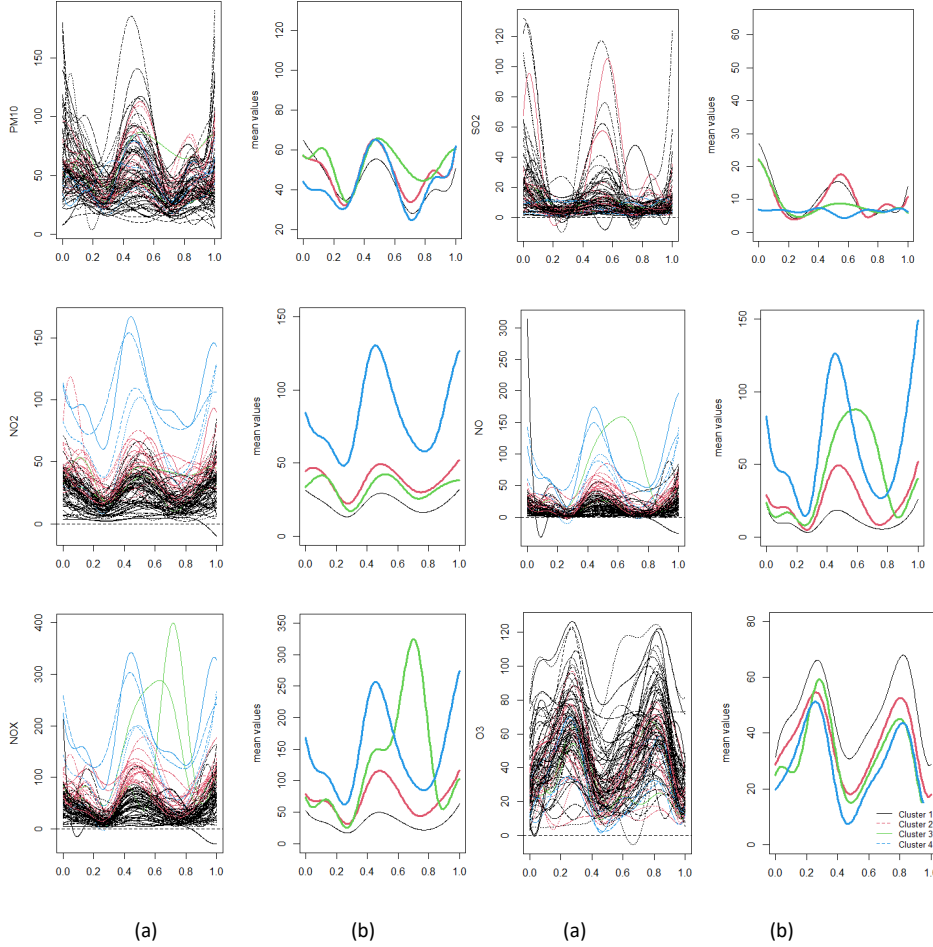
Küme	Tb 1	Tb 2	Tb 3	Tb 4	Tb 5	Tb 6
1	0.87268 (0.0009)	0.94549 (0.0899)	0.96491 (0.3359)	0.9297 (0.0307)	0.939 (0.0576)	0.96011 (0.245)
2	0.95491 (0.1492)	0.98559 (0.9107)	0.93263 (0.0301)	0.91218 (0.0075)	0.67243 (1.07e-07)	0.97295 (0.5113)
3						
4	0.86634 (0.2852)	0.97787 (0.7148)	0.99586 (0.8771)	0.77089 (0.04675)	0.82422 (0.1737)	0.94035 (0.5288)
5	0.95921 (0.7618)	0.96158 (0.7912)	0.93205 (0.432)	0.96999 (0.8864)	0.93266 (0.4383)	0.86605 (0.0690)

Normallik varsayımının sağlanıp sağlanmadığını test etmek için skorlara uygulanan Shapiro-Wilk test sonuçları sadece ilk 6 temel bileşen için **Tablo 4.4**'de raporlanmıştır. Diğer temel bileşenler için p-değerleri 0.10'dan büyük olduğundan raporlanmamıştır. Ayrıca 3. kümede istasyon sayısı az olduğundan test uygulanmamaktadır. Test sonuçları incelendiğinde genel olarak normallik varsayımının sağlandığı görülmekle birlikte kırmızı ile işaretlenenlerin normal dağılım varsayımını sağlamadığı görülmektedir. Az miktardaki bu sonuçların analiz sonuçlarını geçersiz hale getireceğini düşünülmemekle birlikte analizler, bir de t dağılımı altında benzer şekilde yinelenmiştir.

Normal dağılım varsayımı altında en iyi modelin 5 kümeli α_k parametrelerinin her grup içinde, bk parametrelerinin ise gruplar arasında ortak olacak şekilde sabitlendiği α_k b Qk dk modeli olduğu bulunmuşken t dağılım altında en iyi modelin verilerin ortak bir edinim süreciyle toplandığı durumlar için geçerli olan bk parametrelerinin gruplar arasında ortak olacak şekilde sabitlendiği 4 kümeli α_{kj} b Qk dk modelin daha uygun olduğu

bulunmuştur (BIC = -69737.15). Verilerin değişik istasyonlardan benzer ölçüm aletleri ile toplandığını düşünecek olursak, seçilen bu modelin veri için daha uygun olduğu öne sürülebilirse de BIC değerleri karşılaştırıldığında normal dağılım altındaki modellerin tercih edilebileceği görülmektedir. Bu modelin normallik varsayımı altında 11. sırada yer aldığı ve BIC değerinin (-500,338.85) daha yüksek olduğu yani bu modele göre daha iyi olduğu görülmektedir.

Şekil 4.16. t Dağılım Altında Kirleticilerin Grup (a) Eğrileri (b) Ortalama Eğrileri



Elde edilen yeni fonksiyonlar ve gruplar sırası ile **Şekil 4.16'**da ve **Tablo 4.5'te** verilmiştir.

Eğriler incelendiğinde genel olarak ayrışmada NO, NO2 ve NOX değerlerinin miktarsal ve/veya yapısal farklılıklarının etkili olduğu, PM10 ve O3 açısından pek ayrışım olmazken SO2 açısından yapısal farklılaşma gözlemlenmektedir.

Tablo incelendiğinde ise 3. ve 4. Kümelerin (Ankara Ostim hariç) normallik varsayımı altında elde edilen sonuçlarla aynı istasyonlara sahip olduğu görülürken diğer kümelerin farklılaştığı görülmektedir.

Tablo 4.5. t Dağılımı Altında Model Tabanlı Kümeleme Yöntemi ile Oluşturulan İstasyon Grupları

Küme	İstasyonlar	İstasyon Sayısı
1	Antalya Gazipaşa, Antalya Manavgat, Burdur, Aydın Efeler, Aydın Germencik, Aydın Nazilli, Denizli Civrill, İzmir Karabağlar, İzmir Ödemiş, İzmir Torbalı, Kütahya Kentpark, Manisa Akhisar, Manisa Kırkağaç, Manisa Salihli, Manisa Soma, Muğla Fethiye, Muğla Yatağan, Uşak, Bursa Kestel Hilal Parkı, Çanakkale Biga Icdas, Edirne, Edirne Keşan MTHM, İstanbul Arnavutköy, İstanbul Avcılar, İstanbul Başakşehir MTHM, İstanbul Maslak, İstanbul Sarıyer, İstanbul Şile MTHM, İstanbul Silivri MTHM, İstanbul Ümraniye, Kırklareli, Tekirdağ, Tekirdağ Çerkezköy MTHM, Yalova Armutlu MTHM, Ankara Yaşamkent, Çankırı, Kırşehir, Konya Karatay Sunaypark, Nevşehir, Artvin, Bayburt, Bolu Karaçayır Parkı, Giresun Gemilerçekeği, Kastamonu, Ordu Ünye, Rize, Rize Ardeşen, Samsun Atakum, Trabzon Akçaabat, Trabzon Uzungöl, Trabzon Valilik, Zonguldak Çatalağzı Kuzyaka, Zonguldak Çatalağzı Cuma yanı, Zonguldak Muslu Tepeköy, Ağrı, Ağrı Doğubeyazıt, Ağrı Patnos, Ardahan, Erzurum Palandöken, Iğdır, Iğdır Aralık,	62
2	Afyon Selçuklu Cami, İzmir Eğitim İstasyonu, İzmir Kemalpaşa, Kütahya Heymeana Cad, Muğla Milas, Bilecik Bozuyuk MTHM, Bursa Nilüfer, İstanbul Beşiktaş, İstanbul Esenyurt MTHM, İstanbul Kadıköy, İstanbul Kağıthane, İstanbul Kartal, İstanbul Tuzla, Kocaeli Körfez MTHM, Sakarya Ozanlar MTHM, Ankara Polatlı, Erzincan, Erzurum	18
3	Bursa Kestel MTHM, İstanbul Alibeyköy	2
4	Ankara Keçiören Sanatoryum, Ankara Sıhhiye, Ankara Ulus, Ankara Ostim	4

5.SONUÇ VE ÖNERİLER

Günümüzde nüfus artışı ve sanayileşmenin sonucu olarak halk sağlığına yönelik en büyük çevresel tehdit olmanın yanı sıra küresel ısınma ve asit yağmurlarına sebep olan hava kirliliği yaşanmaktadır. Bu nedenle ülkeler hava kirliliğine sebep olan kirleticileri ölçmekte ve ölçüm sonuçlarına bağlı olarak kirletici miktarlarını azaltacak tedbirler almaktadırlar. Ölçümlerin yapılmaya başlanması ile bu alanda pek çok çalışma da yürütölmeye başlanmıştır.

Konunun tüm dünyada gittikçe artan önemi nedeniyle bu konuya yönelerek bu alanda yaygın olarak kullanılan kümeleme analizine odaklanılmıştır. Kümeleme Analizi, hava kirliliği araştırılmalarında 1980'lerden beri sıklıkla hava kirleticilerinin mekansal ve zamansal kalıplarının tanımlanması, hava kirleticilerine maruz kalma, hava kalitesi yönetimi, hava kirletici davranışının yerel sinoptik meteoroloji ile ilişkilendirilmesi, taşıma yolları ve kaynak paylaşırma amacıyla uygulanan bir yöntemdir.

Yapılan literatür taraması sonucunda, Türkiye'de genellikle bir veya iki kirletici üzerinden ve/veya

genellikle belli bölgelerdeki istasyonlar üzerinden kümeleme analizleri gerçekleştirildiği görölmüştür. Ayrıca bu alanda fonksiyonel veri kümeleme yöntemlerinin kullanıldığı bir çalışmaya rastlanamamıştır. Bu nedenle çalışmada, Ocak 2021-Aralık 2023 arasında 86 istasyondan elde edilen aylık ortalama SO₂, NO₂, NO, NO_x, O₃ and PM₁₀ değerleri (i) MFPCA sonucunda elde edilen skorlara k-ortalamlar kümeleme yöntemi uygulanan iki aşamada yapılan kümeleme yöntemi (ii) MFPCA sonucunda elde edilen skorların normal dağıldığı varsayımı altında geliştirilen model tabanlı kümeleme yöntemi ve (iii) çok değişkenli karışık t dağılım altında bir önceki yöntem olmak üzere üç çok değişkenli fonksiyonel kümeleme yöntemi ile gruplandırılmıştır.

İlk iki yöntemde 5, son yöntemde 4 küme oluşmakla birlikte istasyonların kümeler dağılımlarının farklılaştığı gözlemlenmiştir. 2. ve 3. Yöntem karşılaştırıldığında BIC kriterine göre normal dağılım altında yöntemin daha iyi olduğu sonucuna varılmıştır. Ancak normallik altında yakınsama problemi ile karşılaşıldığına dikkat etmek gerekir. Daha önceki çalışmalar istasyon sayısı, yıllar, kirleticiler ve yöntemler açısından farklılık göstermekle birlikte genel olarak hepsinde

bu çalışma ile tutarlı olarak küme sayısının 5 olduğu görülmektedir (Baltacıbaşı, 2014; Sel, 2023; Dogruparmak vd., 2014). Kümelerin oluşturulurken hem miktarsal büyüklüklerin hem de yapısal farklılıkların dikkate alındığı gözlemlenmiştir.

İncelendiği kadarıyla kümeleme yöntemlerini karşılaştıran pek çok farklı çalışma ve tez olmakla birlikte, bu tez önerisi, farklı amaç ve alanlarda kullanılmak üzere geliştirilmiş pek çok kümeleme yönteminin bu alandaki verilere uygulayarak karşılaştırma yapan en geniş kapsamlı çalışma olma özelliğini taşımaktadır. Ayrıca, gerek kümeleme analizi gerekse hava kirliliği çalışmalarında kümeleme analizinin yeri ve uygulanması konusunda güncel geniş bir literatür taraması sunması açısından da önem arz etmektedir.

Bu çalışmanın Türkiye özelinde 86 istasyon verisi ile 6 kirleticisi üzerinden aralarındaki ilişkiyi de göz önüne alarak fonksiyonel veri yaklaşımı ile gerçekleştirilen ilk çalışma olması itibarıyla literatüre önemli bir katkı sunacağı düşünülmektedir. Gerek kirleticiler gerekse zaman boyutundaki ilişkilerin göz önünde bulundurulması nedeniyle çok değişkenli fonksiyonel kümeleme yöntemleri ile yapılan kümelemelerin daha iyi sonuçlar vermesi beklenmektedir. Bu amaçla Türkiye’de hava kirliliği ve hava kalitesine yönelik yapılması planlanan sonraki çalışmalarda, hem iklim modelleme gibi karmaşık veri türlerinde fonksiyonel yapıyı koruyabilme noktasında daha faydalı olabileceği hem de analizi kolaylaştırdığı düşünüldüğünden çok değişkenli fonksiyonel kümeleme yöntemlerine daha fazla ağırlık verilmesi önerilmektedir. Ancak yöntemler de kendi içinde farklı gruplamalar yaratmaktadır. Bu nedenle oluşan son kümelerin alan uzmanları tarafından değerlendirilerek uygun olana karar verilmesini gerektirmektedir.

KAYNAKÇA

- Aghabozorgi, S., Shirkhorshidi, A. S., Soltanian, T. H., Herawan, U., Wah, T. Y. (2014). "Spatial and temporal clustering of air pollution in Malaysia: A review." International Conference on Agriculture, Environment and Biological Sciences, 213-219.
- Ahn, H., Lee, J., Hong, A. (2022). "Urban form and air pollution: Clustering patterns of urban form factors related to particulate matter in Seoul, Korea." Sustainable Cities and Society, 81, 103859.
- Althuwaynee, O. F., Balogun, A. L., Al Madhoun, W. (2020). "Air pollution hazard assessment using decision tree algorithms and bivariate probability cluster polar function: evaluating inter-correlation clusters of PM10 and other air pollutants." GIScience & Remote Sensing, 57(2), 207-226.
- Anıl, İ., Karaca, F., Alagha, O. (2009). "İstanbul'a uzun mesafeli atmosferik taşınım etkilerinin araştırılması:" solunabilen partikül madde epizotları." Ekoloji, 19(73), 86-97.
- Annas, S., Uca, U., Irwan, I., Safei, R. H., Rais, Z. (2022). "Using k-Means and Self Organizing Maps in Clustering Air Pollution Distribution in Makassar City, Indonesia." Jambura Journal of Mathematics, 4(1), 167-176.
- Anton, C., Smith, I. (2024). "Model-based clustering of functional data via mixtures of t distributions." Advances in Data Analysis and Classification, 18(3), 563-595.
- Arslan, H. (2020). Türkiye'nin Batisındaki Pm10 Konsantrasyonlarının Atmosferik Mekanizmaları İle Solunum Yolu Hastalıkları Üzerindeki Etkilerinin İncelenmesi. Doktora tezi. Marmara Üniversitesi Fen Bilimleri Enstitüsü, İstanbul, Türkiye.
- Austin, E., Coull, B. A., Zanobetti, A., Koutrakis, P. (2013). "A framework to spatially cluster air pollution monitoring sites in US based on the PM2. 5 composition." Environment international, 59, 244-254.
- Baltacıbaşı, S. (2014). Spatio-Temporal Analysis Of Particulate Matter Concentrations Of Turkey. Yüksek lisans tezi. İstanbul Teknik Üniversitesi Avrasya Yerbilimleri Enstitüsü, İstanbul, Türkiye.
- Bao, J., Yang, X., Zhao, Z., Wang, Z., Yu, C., Li, X. (2015). "The spatial-temporal characteristics of air pollution in China from 2001–2014." International journal of environmental research and public health, 12(12), 15875-15887.
- Bekar, E. (2017). Bulanık Kümeleme Yöntemlerinin Mevsimsel Bulanık Zaman Serilerinde Performanslarının Karşılaştırılması: Hava Kirliliği Verilerine Uygulama. Yüksek lisans tezi. Muğla Sıtkı Koçman Üniversitesi Fen Bilimleri Enstitüsü, Muğla, Türkiye.
- Bellinger, C., Mohamed Jabbar, M. S., Zaiane, O., Osorio-Vargas, A. (2017). "A systematic review of data mining and machine learning for air pollution epidemiology." BMC public health, 17, 1-19.
- Bougoudis, I., Iliadis, L., Spartalis, S. (2014). "Comparison of self organizing maps clustering with supervised classification for air pollution data sets." Artificial Intelligence Applications and Innovations: 10th IFIP WG 12.5 International Conference, AIAI 2014, Rhodes, Greece, September 19-21, 2014. Proceedings 10 (pp. 424-435). Springer Berlin Heidelberg.
- Bougoudis, I., Demertzis, K., Iliadis, L. (2016). "HISYCOL a hybrid computational intelligence system for combined machine learning: the case of air pollution modeling in Athens." Neural Computing and Applications, 27, 1191-1206.
- Carslaw, D. C., Beevers, S. D. (2013). "Characterising and understanding emission sources using bivariate polar plots and k-means clustering." Environmental modelling & software, 40, 325-329.
- Chen, Y., Wang, L., Li, F., Du, B., Choo, K. K. R., Hassan, H. vd. (2017). "Air quality data clustering using EPLS method." Information Fusion, 36, 225-232.
- Chen, H., Zhang, L., Zou, W., Gao, Q., Zhao, H. (2020). "Regional differences of air pollution in China: comparison of clustering analysis and systematic clustering methods of panel data based on gray relational analysis." Air Quality, Atmosphere & Health, 13, 1257-1269.
- Chuang, H. C., Shie, R. H., Chio, C. P., Yuan, T. H., Lee, J. H., Chan, C. C. (2018). "Cluster analysis of fine

- particulate matter (PM_{2.5}) emissions and its bioreactivity in the vicinity of a petrochemical complex." *Environmental Pollution*, 236, 591-597.
- de Boor, C. (2001) "A Practical Guide to Splines. Revised Edition." New York:Springer.
- Di Bernardino, A., Iannarelli, A. M., Casadio, S., Pisacane, G. ve Siani, A. M. (2023). "Spatial-temporal assessment of air quality in Rome (Italy) based on anemological clustering." *Atmospheric Pollution Research*, 14(2), 101670.
- Dincer, N. G., Akkuş, Ö. (2018). "A new fuzzy time series model based on robust clustering for forecasting of air pollution." *Ecological Informatics*, 43, 157-164
- Dogruparmak, S. C., Keskin, G. A., Yaman, S., Alkan, A. (2014). "Using principal component analysis and fuzzy c-means clustering for the assessment of air quality monitoring." *Atmospheric Pollution Research*, 5(4), 656-663
- D'Urso, P., Di Lallo, D., Maharaj, E. A. (2013a). "Autoregressive model-based fuzzy clustering and its application for detecting information redundancy in air pollution monitoring networks." *Soft Computing*, 17, 83-131.
- D'Urso, P., De Giovanni, L., Massari, R. ve Di Lallo, D. (2013b). "Noise fuzzy clustering of time series by autoregressive metric." *Metron*, 71, 217-243.
- D'Urso, P., De Giovanni, L., Massari, R. (2015). "Time series clustering by a robust autoregressive metric with application to air pollution." *Chemometrics and Intelligent Laboratory Systems*, 141, 107-124.
- Gao, H., Chen, J., Wang, B., Tan, S. C., Lee, C. M., Yao, X. vd. (2011). "A study of air pollution of city clusters." *Atmospheric Environment*, 45(18), 3069-3077.
- Govender, P., Sivakumar, V. (2020). "Application of k-means and hierarchical clustering techniques for analysis of air pollution: A review" (1980–2019). *Atmospheric pollution research*, 11(1), 40-56.
- Güler, M. (2011). *Surface Ozone Analysis İn Istanbul By Using Cluster Analysis And HYSPLIT Model*. Yüksek lisans tezi. Boğaziçi Üniversitesi Çevre Bilimleri Enstitüsü, Türkiye.
- Happ, C., Greven, S. (2018). "Multivariate functional principal component analysis for data observed on different (dimensional) domains." *Journal of the American Statistical Association*, 113(522), 649-659.
- Harishkumar, K. S., Gad, I., Yogesh, K. M. (2021, February). "Spatio-temporal clustering analysis for air pollution particulate matter (pm 2.5) using a deep learning model." In 2021 International conference on computing, communication, and intelligent systems (ICCCIS) (pp. 529-535). IEEE.
- Hemavani, B., Rao, G. S. (2020). "A study on air pollution trends in Sanathnagar, Hyderabad using principle component and cluster analysis." *International Journal of Science and Research*, 9(2), 844-847.
- Hulkkonen, M., Lipponen, A., Mielonen, T., Kokkola, H., Prisle, N. L. (2022). "Changes in urban air pollution after a shift in anthropogenic activity analysed with ensemble learning, competitive learning and unsupervised clustering." *Atmospheric Pollution Research*, 13(5), 101393.
- Im, U., Incecik, S., Guler, M., Tek, A., Topcu, S., Unal, Y. S. vd. (2013). "Analysis of surface ozone and nitrogen oxides at urban, semi-rural and rural sites in Istanbul, Turkey." *Science of the Total Environment*, 443, 920-931.
- Jacques, J., Preda, C. (2014). "Model-based clustering for multivariate functional data." *Computational Statistics & Data Analysis*, 71, 92-106.
- Jorquera, H., Villalobos, A. M. (2020). "Combining cluster analysis of air pollution and meteorological data with receptor model results for ambient PM_{2.5} and PM₁₀." *International journal of environmental research and public health*, 17(22), 8455.
- Josse, J., Husson, F. (2012). "Handling missing values in exploratory multivariate data analysis methods." *Journal de la société française de statistique*, 153(2), 79-99.
- Karaca, F., Camci, F. (2010). "Distant source contributions to PM₁₀ profile evaluated by SOM based cluster analysis of air mass trajectory sets." *Atmospheric Environment*, 44(7), 892-899.
- Kasparoglu, S., Incecik, S., Topcu, S. (2018). "Spatial and temporal variation of O₃, NO and NO₂ concentrations at rural and urban sites in Marmara Region of Turkey." *Atmospheric pollution research*, 9(6), 1009-1020.
- Khedairia, S. ve Khadir, M. T. (2012). "Impact of clustered meteorological parameters on air pollutants concentrations in the region of Annaba, Algeria." *Atmospheric research*, 113, 89-101.
- Kışlak, İ. (2009). *Veri Madenciliği Yöntemi ile Türkiye'deki Hava Kirliliği Durum Analizi. Yüksek Lisans Tezi. Abant İzzet Baysal Üniversitesi Sosyal Bilimler Enstitüsü, Bolu.*
- Kingsy, G. R., Manimegalai, R., Geetha, D. M., Rajathi, S., Usha, K., Raabiathul, B. N. (2016). "Air pollution analysis using enhanced K-Means clustering algorithm for real time sensor data." In 2016 IEEE Region 10 Conference (TENCON) (pp. 1945-1949). IEEE.
- Koçak, E. (2018). "Aksaray kentinin PM₁₀ ve SO₂ konsantrasyonlarının zamansal değişimi: Koşullu iki değişkenli olasılık fonksiyonu ve k-means kümeleme." *Mühendislik Bilimleri ve Tasarım Dergisi*, 6(3), 471-478.
- Li, S. T., Shue, L. Y. (2004). "Data mining to aid policy making in air pollution management." *Expert Systems with Applications*, 27(3), 331-340.
- Ljungman, P. L., Wilker, E. H., Rice, M. B., Austin, E., Schwartz, J., Gold, D. R. vd. (2016). "The impact of multi-pollutant clusters on the association between fine particulate air pollution and microvascular function." *Epidemiology (Cambridge, Mass.)*, 27(2), 194.
- Maciąg, P. S., Kasabov, N., Kryszkiewicz, M., Bembenik, R. (2019). "Air pollution prediction with clustering-based ensemble of evolving spiking neural networks and a case study for London area." *Environmental Modelling & Software*, 118, 262-280.
- Ojeda-Magaña, B., Cortina-Januchs, M. G., Barrón-Adame, J. M., Quintanilla-Domínguez, J., Hernandez, W., Vega-Corona, A. vd. (2010). "Air pollution analysis with a PFCM clustering algorithm applied in a real database of Salamanca (Mexico)." In 2010 IEEE International Conference on Industrial Technology (pp. 1297-1302). IEEE.
- Omar, Z. A., Hashim, S. R. M., Sentian, J., Chin, S. N. (2022). "Application of K-Means Clustering and Calendar View Visualisation for Air Pollution Index Analysis." In IOP Conference Series: Earth and

- Environmental Science (Vol. 1103, No. 1, p. 012004). IOP Publishing.
- Özel, H. (2022). Hava Kalitesinin İncelenmesi Ve Eğitilebilir Algoritmalarla Modellenmesinin Araştırılması; Çerkezköy Örneği. Doktora Tezi. İstanbul Üniversitesi, Cerrahpaşa Lisansüstü Eğitim Enstitüsü, İstanbul.
- Paulose, B., Sabitha, S., Punhani, R., Sahani, I. (2018, February). "Identification of Regions and Probable Health Risks Due to Air Pollution Using K-Mean Clustering Techniques." In 2018 4th International Conference on Computational Intelligence & Communication Technology (CICIT) (pp. 1-6). IEEE.
- Qian, Z., Zhang, J., Wei, F., Wilson, W. E., Chapman, R. S. (2001). "Long-term ambient air pollution levels in four Chinese cities: inter-city and intra-city concentration gradients for epidemiological studies." *Journal of Exposure Science & Environmental Epidemiology*, 11(5), 341-351.
- Ramsay, J. O., Silverman, B.W. (2005) "Functional Data Analysis, 2nd ed.," New York: Springer, ISBN 0-387-40080-X
- Ren, Z., Ji, X. (2023). "On prediction of air pollutants with Takagi-Sugeno models based on a hierarchical clustering identification method." *Atmospheric Pollution Research*, 14(4), 101731.
- Sahafizadeh, E., Ahmadi, E. (2009, December). "Prediction of air pollution of Boushehr city using data mining." In 2009 Second International Conference on Environmental and Computer Science (pp. 33-36). IEEE.
- Sarı, D. (2023). Çanakkale'nin Biga Ve Lapseki Bölgesi Özelinde Kırsal Alanlardaki Troposferik Ozon Seviyelerinin Değerlendirilmesi Ve Meteorolojik Fotokimyasal Modellerin Uygulanmasıyla Dağılımının Belirlenmesi. Doktora Tezi. İstanbul Teknik Üniversitesi Lisansüstü Eğitim Enstitüsü, Atmosfer Bilimleri Doktora Programı, İstanbul.
- Sarı, T. (2023). Türkiye'deki illerin Havadaki Partiküler Madde Ve Kükürt Dioksit Değerlerine Göre Kümeleme Analizi İle Sınıflandırılması. Yüksek Lisans Tezi. Gazi Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
- Schmutz, A., Jacques, J., Bouveyron, C., Cheze, L., Martin, P. (2020). "Clustering multivariate functional data in group-specific functional subspaces." *Computational Statistics*, 35(3), 1101-1131.
- Sel, A. (2023). "Kirlenici parametrelere göre Türkiye'deki illerin hava kalitesinin sınıflandırılması." *Journal of Business and Communication Studies*, 1(1 (1)), 13-26.
- Shafi, J., Waheed, A. (2020). "K-means clustering analysing abrupt changes in air quality." In 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA) (pp. 26-30). IEEE.
- Sunori, S. K., Negi, P. B., Juneja, P., Niranjnamurthy, M., Prakash, P. O., Mittal, A. vd. (2021). "Unsupervised and Supervised Learning based Classification Models for Air Pollution Data." In 2021 2nd Global Conference for Advancement in Technology (GCAT) (pp. 1-5). IEEE.
- Tian, D., Fan, J., Jin, H., Mao, H., Geng, D., Hou, S. vd. (2020). "Characteristic and spatiotemporal variation of air pollution in Northern China based on correlation analysis and clustering analysis of five air pollutants." *Journal of Geophysical Research: Atmospheres*, 125(8), e2019JD031931.
- Wang, T., Qin, L., Dai, C., Wang, Z., Gong, C. (2023). "Heterogeneous Learning of Functional Clustering Regression and Application to Chinese Air Pollution Data." *International Journal of Environmental Research and Public Health*, 20(5), 4155.
- Wu, E. M. Y., Kuo, S. L. (2013). "A study on the use of a statistical analysis model to monitor air pollution status in an air quality total quantity control district." *Atmosphere*, 4(4), 349-364.
- Wu, X., Wang, Q., Liu, Y. ve Li, Y. (2017). "Data-driven regional analysis of urban atmosphere pollution based on density clustering." In 2017 IEEE 2nd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC) (pp. 412-417). IEEE.
- Yan, R., Liao, J., Yang, J., Sun, W., Nong, M., Li, F. (2021). "Multi-hour and multi-site air quality index forecasting in Beijing using CNN, LSTM, CNN-LSTM, and spatiotemporal clustering." *Expert Systems with Applications*, 169, 114513.
- Yaşa, S. (2019). Kocaeli Atmosferinde Ağır Metal Kirliliğinin Likenler Kullanılarak İzlenmesi. Yüksek Lisans Tezi. Kocaeli Üniversitesi, Fen Bilimleri Enstitüsü, Kocaeli.
- Ye, W. F., Ma, Z. Y., Ha, X. Z. (2018a). "Spatial-temporal patterns of PM2. 5 concentrations for 338 Chinese cities." *Science of The Total Environment*, 631, 524-533.
- Ye, W. F., Ma, Z. Y., Ha, X. Z., Yang, H. C., Weng, Z. X. (2018b). "Spatiotemporal patterns and spatial clustering characteristics of air quality in China: A city level analysis." *Ecological Indicators*, 91, 523-530.
- Yusuf, K. A., Oluwale, S., Abdusalam, I. O., Adewusi, G. R. (2013). "Spatial patterns of urban air pollution in an industrial estate, Lagos, Nigeria." *International Journal of Engineering Inventions*, 2(4), 1-9.
- Zareba, M., Dlugosz, H., Danek, T., Weglinska, E. (2023). "Big-Data-Driven Machine Learning for Enhancing Spatiotemporal Air Pollution Pattern Analysis." *Atmosphere*, 14(4), 760.
- Zhang, J., Zhang, L. Y., Du, M., Zhang, W., Huang, X., Zhang, Y. Q. vd. (2016). "Identifying the major air pollutants base on factor and cluster analysis, a case study in 74 Chinese cities." *Atmospheric Environment*, 144, 37-46.
- Zhang, L. ve Yang, G. (2022). "Cluster analysis of PM2. 5 pollution in China using the frequent itemset clustering approach." *Environmental Research*, 204, 112009.