

Hava Kirlilik Düzeylerindeki Anomali Tespitinde Benford Kanunu ve Genelleştirilmiş Uç Değer Dağılımı: Ankara İli Çankaya İlçesi Örneği

Hatice Nur Karakavak^{*1}, Gamze Özel²

^{1,2}Hacettepe Üniversitesi, Fen Fakültesi, İstatistik Bölümü, 06800, Ankara, Türkiye

(Alınış / Received: 26.07.2024, Kabul / Accepted: 22.03.2025, Online Yayınlanma / Published Online: 25.04.2025)

Anahtar Kelimeler

Benford Kanunu,
Genelleştirilmiş Uç Değer
Dağılımı, Hava Kirliliği

Öz: Nüfus ve yüzölçümü bakımından Ankara'nın en büyük ilçelerinden biri olan Çankaya'da coğrafi yapısı, sanayi tesisleri, yoğun trafik ve ısınma amaçlı fosil yakıt kullanımı nedenleriyle sıklıkla hava kirliliği problemi ile karşılaşmaktadır. Bu çalışmada, hava kirliliğinin insan sağlığına olan olumsuz etkileri dikkate alınarak, Ankara ili Çankaya ilçesi için Çevre ve Şehircilik Bakanlığı'nın Ulusal Hava Kalitesi İzleme Ağı ölçümlerinden yararlanılarak 1 Ocak 2022 ile 1 Ocak 2024 tarihleri arasındaki günlük partiküler madde (PM₁₀) değerlerindeki anomali tespiti gerçekleştirilmiştir. Bu amaçla, bloktaki değerlerin en büyüğü (Block Maxima) ve Benford Kanunu'ndan faydalanılmıştır. Öncelikle verideki anomali tespiti Benford Kanunu ile değerlendirilmiştir. Daha sonra PM₁₀ değerlerinin uygun olduğu uç değer dağılımı uyum iyiliği testleri ile belirlenmiştir. Elde edilen bulgular, hava kirliliği yönetimi ve halk sağlığını koruma adına önemli ipuçları sunmaktadır.

Benford's Law and Generalised Extreme Value Distribution in Anomaly Detection in Air Pollution Levels: The Case of Ankara Province

Keywords

Benford's Law,
Generalised Extreme Value
Distribution, Air Pollution

Abstract: Çankaya, one of the largest districts of Ankara in terms of population and surface area, frequently encounters air pollution problems due to its geographical structure, industrial facilities, heavy traffic and fossil fuel use for heating purposes. In this study, considering the adverse effects of air pollution on human health, the anomaly in daily particulate matter (PM₁₀) values between 1 January 2022 and 1 January 2024 was determined by using the National Air Quality Monitoring Network measurements of the Ministry of Environment and Urbanisation for Çankaya district of Ankara province. For this purpose, the largest of the values in the block (Block Maxima) and Benford's Law were utilised. Firstly, the anomaly detection in the data was evaluated with Benford's Law. Then, extreme value distribution for which PM₁₀ values are appropriate was determined by goodness-of-fit tests. The obtained findings provide important clues for air pollution management and protection of public health.

1. Giriş

Hava kirliliği, insan sağlığını ve atmosfer dengesini bozabilecek düzeyde hava bileşenlerinin değişmesi olarak tanımlanmaktadır [1]. Hava kirliliği, solunum yolu hastalıklarından kalp rahatsızlıklarına kadar birçok sağlık sorununa yol açarak dünya genelinde ölümlere neden olmaktadır. Bu durum, hava kirliliğini önemli bir araştırma konusu haline getirmiştir. Dünya Sağlık Örgütü (WHO), hava kirliliğini kanserojen etkenlerin bulunduğu gruba dahil etmiştir. Hava kalitesini belirleyen bileşenler arasında partiküler

madde (PM₁₀, PM_{2.5}), kükürt dioksit (SO₂), karbon monoksit (CO), azot dioksit (NO₂), azot oksit (NO) ve ozon (O₃) bulunmaktadır. Özellikle PM₁₀, kanserojen etkenler arasında yer almakta olup, WHO tarafından günlük ortalama 0.05 µg/m³ ve yıllık ortalama 0.02 µg/m³ olarak önerilmektedir [2, 3]. PM₁₀, havada bulunan ve çapı 10 mikrometreden küçük olan zararlı toz parçacıklarıdır. Solunum yolu ile akciğerlere ulaşarak nefes darlığına yol açtığından havadaki diğer kirlenmeler arasında en zararlı ve tehlikeli bileşen olarak kabul edilmektedir [4]. Küçük çocuklarda astım ve bronşite neden olabilirken, bebeklerde solunum

*Sorumlu yazar: hatice.karakavak@hacettepe.edu.tr

kaynaklı ölümlere sebep olabilmektedir. Yapılan çalışmalarda solunum yolu ve kardiyovasküler hastalıkları olan kişilerde kaza olmaksızın ölüm oranı ile atmosferdeki partiküler madde (PM) konsantrasyonunun yakın ilişkili olduğu görülmüştür [5, 6]. Ayrıca, coroner kalp hastalığına bağlı ölüm oranının, havadaki karbon seviyelerindeki artışa bağlı olarak artabileceği sonucuna ulaşılmıştır [7].

Türkiye'de hava kirliliğinin temel nedenleri ısınma ve ulaşım amacıyla kullanılan fosil yakıtlardan kaynaklanmaktadır. Türkiye, Ortadoğu'dan gelen toz fırtınalarından yüksek ölçüde etkilenen ülkeler arasındadır. Genellikle çöllerden gelen toz fırtınaları yüksek PM konsantrasyonları ile sonuçlanır. Dolayısıyla, Türkiye'nin PM düzeyleri de bu fırtınalardan negatif yönde etkilenmektedir [8, 9, 10]. Partikül maddelere maruz kalmanın sağlık üzerinde önemli etkileri olduğunu gösteren birçok çalışma yapılmıştır [11, 12]. Bu çalışmalarda, PM'ye kısıtlı sürede maruz kalma çeşitli sağlık sorunları ile ilişkilendirilmiştir [13, 14, 15, 16, 17, 18, 19]. Sıkı önlemler ve gerekli düzenlemeler ile, PM'den kaynaklanan hava kirliliği düzeyleri son dönemlerde ülkenin genelinde düşürülmüştür ancak hala kirlilik düzeylerinin incelenmesi gerekmektedir. Ankara'da da hava kirliliğinin temel nedenleri arasında, trafik yoğunluğu, sanayi faaliyetleri, ısınma amaçlı fosil yakıt kullanımı, coğrafi ve meteorolojik faktörler, inşaat ve yol çalışmaları, tarım ve açık alan yakmaları ile endüstriyel atıklar ve kimyasallar yer almaktadır [20]. Şehirde artan araç sayısı ve yoğun trafik, motorlu taşıtlardan salınan egzoz gazları nedeniyle hava kalitesini olumsuz etkilemektedir [21]. Kış aylarında kömür ve diğer fosil yakıtların kullanımı, partikül madde ve kükürt dioksit gibi kirlleticilerin atmosfere salınmasına neden olmaktadır [22]. Türkiye'de PM₁₀ kaynaklı hava kirliliğinin temel nedeni konut ısıtmasında kullanılan yakıtlar, endüstriyel emisyonlar ve motorlu araçlardır [23]. Ayrıca, nüfusun artmasına bağlı olarak hava kirleticilerinin artması da hava kalitesindeki bozulmalara sebep olmaktadır [24]. Ankara'nın coğrafi yapısı ve iklim koşulları, kirleticilerin atmosferde birikmesine ve yoğunlaşmasına zemin hazırlamaktadır [25]. Bu nedenlerin kontrol altına alınması ve çevre dostu uygulamaların teşvik edilmesi, Ankara'da hava kalitesinin iyileştirilmesi açısından önem taşımaktadır.

Ankara'da hava kirliliği üzerine yapılan çalışmalar, genellikle PM₁₀ ve PM_{2.5} seviyelerinin WHO ve ulusal sınır değerlerini aştığını göstermektedir [26]. Bu çalışmalar arasında yer alan Çiçek vd. [27] tarafından yapılan araştırmada Ankara'nın çeşitli bölgelerindeki PM₁₀ konsantrasyonlarının özellikle kış aylarında önemli ölçüde arttığını göstermiştir. Bacak ve Toros [28], Ankara'da 2018-2020 yılları arasında meydana gelen bebek ölümleri ile PM₁₀ ve PM_{2.5} düzeyleri arasındaki ilişkiyi incelemişlerdir. Hesaplanan günlük ortalama PM₁₀ ve PM_{2.5} düzeyinin üç yıllık ortalaması

limit değerlerin üzerindedir ve belirtilen dönemde, 1295 bebek ölümü meydana gelmiştir. Yavuz [29] ise, Ankara'da Ulusal Hava Kalitesi İzleme Ağı'ndan elde edilen 1 Ocak 2013-31 Aralık 2017 tarihleri arasındaki yapılan PM₁₀ ölçümlerinde istasyonlara göre bazı farklılıklar olmakla birlikte genel kirlilik düzeyinin sınır değerlerin üstünde seyrettiğini tespit etmiştir. Ünal ve Özel [30] 2018-2021 yılları arasında Ankara'da çeşitli meteorolojik parametrelerin (hava basıncı, hava sıcaklığı, rüzgar yönü, rüzgar hızı ve bağıl nem), hava kirliliği göstergeleri (partiküler madde (PM₁₀), karbon monoksit (CO), kükürt dioksit (SO₂), azot (NO), azot dioksit (NO₂) ve azot oksit (NO_x)) üzerindeki etkilerini incelemiş ve meteorolojik parametreler ile hava kirliliği göstergeleri arasında negatif yönde ilişki olduğu sonucuna ulaşmışlardır.

Benford Kanunu, verilerin ilk basamaklarının dağılımını inceleyen bir matematik kuralıdır ve genellikle doğal olayların ve ekonomik verilerin analizinde kullanılır. Hava kirliliği değerleri üzerinde yapılan çalışmalarda da Benford Kanunu'nun uygulanması, verilerin doğruluğunu ve olası anormalliklerin tespitini sağlamak için etkili bir yöntem olarak kullanılmaktadır. Bilimsel araştırmalarda, özellikle çevre bilimleri ve astrofizik gibi alanlarda, veri manipülasyonunu tespit etmek için Benford Kanunu yaygın olarak kullanılmaktadır [31]. Çevre bilimleri alanında, iklim değişikliği verileri veya çevre kirlilik seviyeleri gibi çeşitli veri setlerinin analizi sırasında Benford Kanunu'ndan yararlanılmıştır [32]. Özkan [33] çalışmasında, Benford Kanunu'nu ekosistemlerin doğallığını değerlendirmek amacıyla kullanmıştır. Bu araştırmada, Benford Kanunu ekosistem verilerindeki sayısal desenleri inceleyerek insan müdahalesi veya doğal olmayan değişikliklerin tespit edilmesine katkı sağlamıştır. Bu geniş kullanım alanı, Benford Kanunu'nun farklı disiplinlerde önemli bir analitik araç olduğunu ortaya koymaktadır. Uç değer teorisi, istatistiksel analizlerde nadiren rastlanan ancak potansiyel olarak büyük etkileri olan olayları incelemek için kullanılır. Bu teori, özellikle doğal afetler, finansal krizler gibi aşırı olayların tahmin edilmesinde kritik bir rol oynar. Uç değer dağılım modelleri, veri setlerindeki en küçük veya en büyük değerler gibi sıradışı veriler üzerinde yoğunlaşır. Bu modellerin iki temel yaklaşımı vardır: "Bloktağı Değerlerin En Büyüğü" ve "Eşik Üstü Yöntemi." Bloktağı Değerlerin En Büyüğü (Block Maxima) yöntemi, uç değer teorisi kapsamında hava kirliliği verilerinin analizi için sıklıkla kullanılan bir yaklaşımdır. Bu yöntem, belirli zaman aralıklarında (genellikle aylık veya yıllık) maksimum değerleri seçerek, bu ekstrem değerlerin dağılımlarını analiz etmeyi amaçlar. Hava kirliliği üzerine yapılan çeşitli çalışmalarda, bu yöntem kullanılarak önemli bulgular elde edilmiştir. Masseran ve Safari [34] tarafından yapılan bir çalışmada, Malezya'nın Klang şehrinde 1997-2020 yılları arasında saatlik Hava Kirliliği İndeksi (API) verileri kullanılarak, blok maksima

yöntemi ile analizler gerçekleştirilmiştir. Çalışma, muson mevsimleri bilgilerini kullanarak ekstrem hava kirliliği olaylarının şiddetini analiz etmeyi amaçlamıştır. Elde edilen sonuçlar, mevsimsel muson dönemlerine göre belirlenen dönüş periyotlarının hava kirliliği şiddetinin artışıyla ilişkilendirildiğini göstermiştir. Bu çalışma, hava kirliliği olaylarının şiddetini ve tekrarlama riskini değerlendirmede blok maksima yönteminin etkinliğini ortaya koymuştur. Masseran ve Safari [35] tarafından yapılan bir diğer çalışmada, blok maksima yöntemi ile eşik aşan değerler (POT) yönteminin kombinasyonu kullanılarak ekstrem hava kirliliği olayları modellenmiştir. Bu çalışma, ekstrem olayların bağımsız özelliklerini tatmin eden veri noktalarını belirlemiş ve model uyumunu değerlendirmiştir. Çalışmanın sonuçları, bu karma yaklaşımın ekstrem değer modellemesinde önyargı ve varyans arasında bir denge sağladığını göstermiştir. Örneğin, Klang şehrindeki hava kirliliği olayları için belirlenen eşik değeri 100 olarak kabul edilmiştir ve bu değer üzerinde gerçekleşen olaylar analiz edilmiştir. Martins vd. [36], São Paulo ve Rio de Janeiro metropol bölgelerindeki hava kirliliği davranışlarını araştırmak için Genelleştirilmiş Uç Değer (GUD) ve Genelleştirilmiş Pareto Dağılımı (GPD) modellerini kullanmışlardır. Bu çalışmada, blok maksima yaklaşımı ile elde edilen maksimum değerler analiz edilerek, bu bölgelerdeki hava kirliliği olaylarının dağılım özellikleri incelenmiştir. Bu çalışmalar, blok maksima yönteminin hava kirliliği analizlerinde nasıl kullanıldığını ve bu yöntemin ekstrem olayların modellenmesinde sağladığı avantajları göstermektedir. Ancak uç değer yaklaşımı ile Ankara ili için hava kirlilik düzeyi modellemesini amaçlayan bir çalışmanın literatürde yer almadığı görülmüştür.

Bu çalışmada Ankara ili Çankaya ilçesi için Çevre ve Şehircilik Bakanlığı Hava Kalitesi İzleme İstasyonlarından elde edilen 01.01.2022-01.01.2024 yılları arasındaki günlük PM₁₀ verileri üzerinde istatistiksel dağılım analizinin yapılması amaçlanmıştır. Veri üzerinde Bloktaki Değerlerin En Büyüğü (Block Maxima) yöntemi ve Benford Kanunu kullanılarak R program üzerinde uyum iyiliği testleri

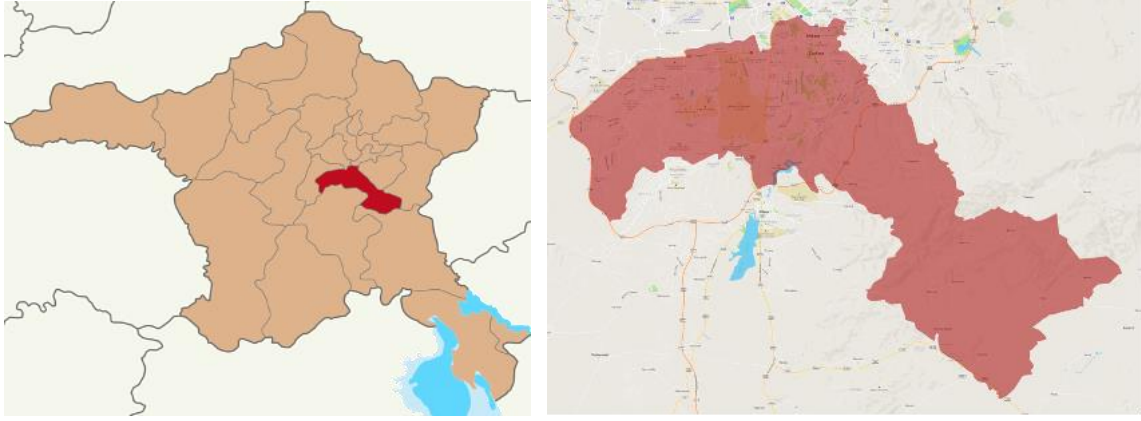
yapılmıştır. Elde edilen sonuçlar p-değeri ve grafikler analiz edilerek yorumlanmıştır. Çalışma Çankaya ilçesine yönelik olması ve son 2 yıla ait gerçek/güncel veri seti kullanılması ile özgün bir araştırma olma ayrıcalığına sahiptir. Çalışma ikinci bölümde kullanılan yöntemlerin anlatımı, üçüncü bölümde uygulama ve sonuçlar ve son olarak dördüncü bölümde değerlendirme yapıp sonlandırılacaktır.

2. Materyal ve Yöntem

2.1. Materyal

Ankara, 39° ve 57' kuzey enlemleri ile 32° ve 53' doğu boylamları arasında, İç Anadolu bölgesinin kuzeybatısında yer almaktadır. Kırşehir ve Kırıkkale doğusunda; Eskişehir batısında; Çankırı kuzeyinde; Aksaray ve Konya güneyinde bulunmaktadır. Yüzölçümü 25,706 km² ve deniz seviyesinden yüksekliği 938 metredir. Hatip Ovası'nda, tepelerle çevrili bir havzada yer alan şehir, genişleyerek çevresindeki sırtlara yayılmıştır. Bunun sonucunda, hem engebeli hem de düz arazilere sahip bir yerleşim alanı oluşmuştur. Şehir merkezi, çoğunlukla Ankara Nehri ve kollarının oluşturduğu ova ve çevresindeki dağlık, engebeli alanda bulunmaktadır. Bu nedenle, kâse benzeri topografik bir yapıya sahiptir. Ankara'daki yerleşim yerlerinin dağılışı bu topografik özellikler sayesinde belirlerken, bölgede karasal iklim hakimdir [27]. Şehirde yeşil alanlar homojen ve sistematik bir şekilde dağılmamıştır. Her bir kişi için yeşil alan miktarı ilçelere göre farklılık göstermektedir: Altındağ'da 0,6m², Gölbaşı'nda 1.1m², Mamak'ta 1.3m², Yenimahalle ve Keçiören'de 1.6 m², Çankaya'da 2,8m², Etimesgut'ta 2,9m² ve Sincan'da 6.0m² şeklindedir [37]. Veriler, şehirdeki her bir kişi için yeşil alan miktarının yeterli olmadığı göstermektedir.

Köppen iklim sınıflandırmasına göre Ankara ili ve çevresinde soğuk step iklimi (BSk) hakimdir. Bu sebeple, yarı kurak bir iklim söz konusu olup, yıllık ortalama sıcaklık 18.0 °C'den düşüktür. Yıllık yağış miktarı ise 250-400 mm arasında değişmekte olup oldukça düşüktür [38].



Şekil 1. Çalışma alanı konumu

Ülkemizde de hava kirliliği takibi genel olarak Çevre ve Şehircilik Bakanlığı'nın kurmuş olduğu bir sistem üzerinden yapılmaktadır. Ulusal Hava Kalitesi İzleme Ağı olarak adlandırılan bu sistem Başkanlık verilerine göre 2016 yılında ülke geneline kurulan 249 istasyon ile hava kalitesini takip etmektedir [39]. Ankara'da hava kalitesinin ölçülmesi amacıyla kurulan 18 adet (Bahçelievler, Batıkent, Çankaya, Çubuk, Demetevler, Etimesgut, Etlik, Kayaş, Mamak, OSTİM, Polatlı, Sanatoryum, Sıhhiye, Sincan, Siteler, Törekent, Ulus, Ümitköy) hava kalitesi ölçüm istasyonunun 9'undan (Bahçelievler, Çankaya, Çubuk, Demetevler, Kayaş, Sanatoryum, Sıhhiye, Sincan, Siteler) alınan ham veriler, İlimiz Temiz Hava Eylem Planı çerçevesinde online olarak izlenebilmektedir. Bu çalışmada Ankara'nın en kalabalık ilçesi olan Çankaya ele alınmıştır. Yaklaşık 937.000 kişinin yaşadığı ilçe, yoğun kentleşmesi ve kalabalık trafiğiyle ön plana çıkmaktadır. Çankaya, başkentteki birçok kamu kurumuna ve iş merkezine ev sahipliği yapması nedeniyle hem yüksek enerji tüketiminin hem de ticari hareketliliğin merkezi durumundadır. Kış aylarında artan ısınma ihtiyacı ve yaygın katı yakıt kullanımı, ilçedeki hava kirliliğinin başlıca nedenleri arasında yer almaktadır. Çankaya'nın sosyal ve ekonomik çeşitliliği, buradan elde edilecek sonuçların diğer şehir bölgelerine de örnek olmasını sağlar. Bu nedenle, ilçedeki hava kirliliğinin kaynaklarını incelemek ve etkilerini değerlendirmek oldukça önemlidir. Çalışma kapsamında Ankara ili Çankaya ilçesi için Çevre ve Şehircilik Bakanlığı Hava Kalitesi İzleme İstasyonlarından elde edilen 01.01.2022-01.01.2024 yılları arasındaki günlük PM₁₀ verilerinden yararlanılmıştır. Veri setinde toplam 731 gözlem değeri yer almaktadır.

2.2. Yöntem

2.2.1. Benford kanunu

Tablo 1. Benford'a Göre Rakamların Basamaklarda Yer Alma Olasılıkları

Rakamlar	İlk Basamak	İkinci Basamak	Üçüncü Basamak	Dördüncü Basamak
0		0.11968	0.10178	0.10018
1	0.30103	0.11389	0.10138	0.10014
2	0.17609	0.10882	0.10097	0.10010

Bir veri setinde yer alan değerlerin istatistiksel dağılımıyla ilgilenen bir matematik kuralı olarak ifade edilebilir. Benford Kanunu'nun ilk keşfi Simon Newcomb tarafından yapılmıştır. Newcomb [40], logaritma tablolarında sıkça kullanılan sayfaların daha çabuk aşındığını fark etmiş ve bunun sayıların doğal dağılımıyla ilgili bir olgu olduğunu öne sürmüştür. Kanun 1938 yılında Amerikalı fizikçi Frank Benford [41], tarafından sayıların logaritmik bir dağılıma sahip olduğu yönünde kuramlaştırılmıştır. Ardından Carslaw [42], Benford kanunu finansal olaylardaki hilelerin tespit edilmesi için kullanmıştır.

Benford Kanunu, bir veri kümesindeki sayıların ilk basamaklarının belirli bir dağılımda olacağını öngörür. Kanuna göre, bir veri kümesindeki sayıların ilk basamağının 1 olması olasılığı 2 olması olasılığından, 2 olması olasılığı ise 3 olması olasılığından daha yüksektir. Sayıların ilk basamağının 9 olması olasılığı ise en düşüktür. Newcomb, sıfırdan farklı herhangi bir rakamın bir sayının ilk basamağı olma olasılığını hesaplamış ve bu olasılıklar için aşağıdaki formülü elde etmiştir:

$$P(d) = \log_{10}\left(1 + \frac{1}{d}\right), d = 1, 2, \dots, 9.$$

Newcomb, bu formüle dayanarak rakamların birinci ve ikinci basamakta olması olasılıklarını hesaplamıştır. 1938 yılında ise 20229 veri setini analiz ederek Newcomb'un çalışmalarına ek olarak Benford, rakamların ilk iki basamaktaki olasılıklarının yanı sıra üçüncü ve dördüncü basamaklardaki olasılıklarını da hesaplamıştır. Benford'a göre rakamların basamaklarda yer alma olasılıkları Tablo 1'de gösterilmektedir:

3	0.12494	0.10433	0.10057	0.10006
4	0.09691	0.10031	0.10018	0.10002
5	0.07918	0.09668	0.09979	0.09998
6	0.06695	0.09337	0.09940	0.09994
7	0.05799	0.09035	0.09902	0.09990
8	0.05115	0.08757	0.09864	0.09986
9	0.04576	0.85000	0.09827	0.09982

Tablo 1'de görüldüğü gibi, ikinci basamaktaki rakamların dağılım olasılıkları, birinci basamaktaki rakamların dağılım olasılıklarına göre birbirine daha yakındır. Bu durum, sayıların ilk basamaklarının daha belirgin olasılık farklarına sahip olduğunu ve soldan sağa doğru gidildikçe bu farkların azaldığını gösterir. Bu nedenle, sayıların oluşumunda ilk basamaklar daha anlamlıdır.

Benford Kanunu kapsamında, veri kümesindeki sayısal anormallikleri tespit etmeyi amaçlayan çeşitli sayısal analiz testleri mevcuttur. Bu testler aşağıdaki gibi özetlenebilir [43, 44, 45]:

- **Birinci Basamak Testi:** Veri kümesindeki sayıların ilk rakamlarının dağılımlarını inceleyerek genel bir uygunluk testi yapar. İlk rakamı sıfır olan sayılar bu testte dikkate alınmaz.
- **İkinci Basamak Testi:** Veri kümesindeki sayıların ikinci basamaktaki rakamlarının dağılımlarını inceleyerek genel bir uygunluk testi yapar.
- **İlk İki Basamak Testi:** Veri kümesindeki sayıların ilk iki basamağındaki rakamların kombinasyonlarını inceleyerek diğer iki testten daha ayrıntılı bir uygunluk testi yapar.
- **İlk Üç Basamak Testi:** Veri kümesindeki sayıların ilk üç rakamının dağılımlarını inceleyerek anormallikleri tespit etmeye odaklanır. Bu test, genellikle 10.000'den fazla veri kaydı olduğunda uygulanır.
- **Mükerrer Sayılar Testi:** Veri kümesindeki en fazla tekrar eden sayıları belirler ve bu sayıları ilk iki ve üç basamak testleri ile karşılaştırarak sapmaları tespit eder.
- **Son İki Basamak Testi:** Veri kümesindeki kusuratu olan fakat yuvarlanan sayıları tespit eder. Bu test, genellikle 10.000'den az veri kaydı olduğunda kullanılır.

Benford Kanunu'na göre yapılan analizlerde, elde edilen sonuçlar çeşitli istatistiksel testlerle

değerlendirilir. Bu değerlendirme sürecinde, gözlemlenen ve beklenen oranlar arasındaki farklar incelenir. Bu sapmalar incelenirken kullanılan testler aşağıdaki gibi özetlenebilir [46, 47]:

- **Z-İstatistiği Testi:** Z istatistiği testi, bir rakam veya rakam kombinasyonlarının Benford Kanunu'na göre beklenen oranından sapmasını ölçer. Z istatistiği formülü aşağıdaki gibidir:

$$Z = \left[|p_0 - p_e| - \frac{1}{2n} \right] / \sqrt{\left(p_e * \frac{1-p_e}{n} \right)}.$$

Burada p_0 , Benford Kanunu'na göre beklenen oranı; p_e , veri kümesindeki gözlemlenen oranı ve n , gözlem sayısını ifade etmektedir. Z istatistik testinin hesaplanan değeri, belirlenen güvenilirlik derecesine göre olasılık tablolarındaki kritik değerlerle karşılaştırılır. Testin hassasiyeti, incelenen veri miktarı arttıkça yükselir, bu yüzden tüm verilerin analize dâhil edilmesi önerilir. Bu şekilde, örnekleme yapılmadan direkt tüm veri seti üzerinden analiz yapmak daha doğru sonuçlar verebilir.

- **Ki-Kare Testi:** Ki-kare testi, verilerin beklenen dağılımlara uygunluğunu test etmek için kullanılır ve testin adımları gözlemlenen ve beklenen oranların farkının karesinin alınması ve beklenen oranlara bölünmesi şeklinde ilerler. Ki-kare testi aşağıdaki gibi hesaplanır:

$$\chi^2 = \sum \frac{(p_0 - p_e)^2}{p_e}.$$

Bu formülde p_0 , gözlemlenen oranı ve p_e , beklenen oranı temsil eder. Bu testte elde edilen ki-kare değeri, belirlenen bir kritik değerle karşılaştırılır. Bu değer, kritik değeri aşarsa, bu durum gözlemlenen oranların beklenen oranlardan istatistiksel olarak önemli bir farklılık gösterdiğini ve hipotezin reddedilmesi gerektiğini gösterir. Bu sonuç, verilerde olası yolsuzluk veya hata olabileceğine işaret eder ve daha detaylı bir

inceleme gerektirir. Eğer ki-kare değeri kritik değerin altında kalırsa, verilerin beklenen dağılıma uyduğu kabul edilir.

- **Kolmogorov-Smirnov Testi:** Kolmogorov-Smirnov testi de Benford Kanunu'na verinin uyumunun incelenmesinde kullanılmaktadır. Bir veri kümesinin normal dağılıma uygun olup olmadığını belirlemek için kullanılır. Bu testte, gözlemlenen dağılım ile bir referans dağılım arasındaki farkı ölçen bir test istatistiği hesaplanır. Test istatistiği hesaplandıktan sonra, belirli bir anlamlılık düzeyi için kritik değerlere bakılır. Test istatistiği, belirlenen anlamlılık düzeyi için kritik değerden büyükse, yokluk hipotezi reddedilir ve veri setinin normal dağılıma uymadığı sonucuna varılır.

- **Ortalama Mutlak Sapma:** Mutlak sapma, bir rakamın ya da rakam kombinasyonlarının gözlemlenen frekanslarının Benford Kanunu tarafından öngörülen beklenen frekanslardan olan sapmasının mutlak değeri olarak tanımlanmaktadır. Ortalama mutlak sapma ise, bu mutlak sapmaların her biri için ayrı ayrı hesaplandıktan sonra bunların ortalamasının alınmasıyla elde edilen bir değerdir. Ortalama mutlak sapma aşağıdaki biçimde hesaplanır:

$$\sum_{i=1}^K \frac{|AP-EP|}{K}.$$

Burada K, basamak sayısını; AP, veri kümesindeki gözlemlenen oranı; EP, Benford Yasası'nın beklenen değerini göstermektedir. Hesaplanan mutlak sapma değeri, Tablo 2'de verilen sınırlara göre değerlendirilir:

Tablo 2. Benford Kanunu Uyum Sınırları

İlk Basamak	İkinci Basamak	İlk İki Basamak	Sonuç
0.000-0.006	0.000-0.008	0.0000-0.0012	Yakın Uyumlu
0.006-0.012	0.008-0.010	0.0012-0.0018	Kabul Edilebilir Uyumlu
0.012-0.015	0.010-0.012	0.0018-0.0022	Marjinal Kabul Ed. Uyumlu
>0.015	>0.012	>0.0022	Uyumsuz

2.2.2. Bloktaki değerlerin en büyüğü (block maxima) yöntemi

Uç değer dağılım modelleri bir veri setinde yer alan çok küçük ve çok büyük değerlere sahip sıradışı olanlara ilişkin istatistiksel analizler ile ilgilenirler. Uç değer dağılım modellemesinde temel olarak kullanılan iki yaklaşım bloktaki değerlerin en büyüğü ve eşik seviyesini aşan değerler olmak üzere iki yaklaşım kullanılmaktadır. Bloktaki değerlerin en büyüğü yaklaşımı ile örneklem içerisinde yer alan en büyük değerlerin davranışlarının belirlenmesi amaçlanmaktadır. Yöntemin adımları öncelikle örneklem en büyük değerlerin ayrılması, her bir blokta yer alan en büyük değerlerin belirlenmesi ve parametre tahminlerinin yapılması şeklinde özetlenebilir.

Bloktaki Değerlerin En Büyüğü Yöntemi, veri setini eşit zaman aralıklarına ayırarak her bir dönemin en yüksek değerini analiz etmeyi hedeflemektedir. Bloktaki en büyük değer yöntemi, aşırı olayların sıklığı ve şiddeti hakkında önemli bilgiler sağlamaktadır. Bu yöntem, genellikle hidroloji, meteoroloji ve finans sektörlerinde gözlemlenen aşırı olayların modellenmesinde kullanılır. Örneğin, bir şehrin sel riskinin değerlendirilmesinde, tarihsel yağış verilerinin maksimumları kullanılarak gelecekteki potansiyel risklerin tahmininde veya maksimum

kirlilik düzeylerinin tespitinde faydalanılabilir. Yöntemin adımları aşağıdaki gibi özetlenebilir:

- **Veri Bölümlendirme:** İlk adım, gözlemlenen veri serisini eşit büyüklükteki bloklara ayırmaktır. Bu bloklar genellikle bir yıl, bir ay gibi belirli zaman aralıklarını temsil eder.
- **Maksimum Değerlerin Belirlenmesi:** Her bir bloktan, o blok içindeki en yüksek değer seçilir. Bu, blok maksimumları olarak bilinir ve bu değerlerin dağılımı daha sonra analiz edilir.
- **Parametre Tahmini:** Seçilen maksimum değerler üzerinden, genellikle Genelleştirilmiş Uç Değer (GUD) dağılımı kullanılarak parametre tahminleri yapılır. GUD dağılımı, konum (location), ölçek (scale) ve şekil (shape) biçiminde üç parametrelidir. Bu parametreler, maksimum değerlerin dağılımının genel karakteristiğini tanımlamaktadır. GUD'un olasılık yoğunluk fonksiyonu aşağıdaki gibidir:

$$f(x) = \frac{1}{\sigma} t(x)^{\xi+1} e^{-t(x)},$$

$$\text{burada, } t(x) = \begin{cases} \left[1 + \xi \left(\frac{x-\mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} & \text{if } \xi \neq 0 \\ \exp \left(-\frac{x-\mu}{\sigma} \right) & \text{if } \xi = 0. \end{cases}$$

Burada μ , konum parametresi dağılımın başlangıç noktasını belirler. σ , ölçek parametresi dağılımın yayılımını, yani değerlerin ne kadar geniş bir aralıkta dağıldığını gösterir. ξ (biçim) ise kuyruk indeks parametresidir. Dağılımın kuyruğunun ne kadar "ağır" veya "hafif" olduğunu ifade eder.

- **Model Uygunluğu ve Değerlendirme:** Modelin veriye uygunluğu çeşitli istatistiksel testler ve grafiksel yöntemler kullanılarak değerlendirilir. Bu süreç, modelin gerçek verileri ne kadar iyi temsil ettiğini belirlemeye yardımcı olur.

3. Bulgular

Çalışmada, öncelikle Ankara, Çankaya ilçesi için toplanan günlük PM₁₀ verileri üzerinde aşırı değer değerlendirmesi için Benford Kanunu'ndan faydalanılmıştır. Veri son iki yıla ait günlük PM₁₀ değerlerini içeren 731 gözlemden oluşmaktadır. Benford Kanunu'na uyum analizi R Studio programı

üzerinde "gtrendsR, reshape2, dplyr, benford.analysis" paketleri kullanılarak gerçekleştirilmiştir [48, 49, 50, 51].

Tablo 3'te Mantissa istatistiklerine yer verilmiştir. Ortalama ve varyans, verilerin merkezi eğilimi ve dağılımını gösterirken, eksiklik ve çarpıklık, verilerin simetrik olup olmadığını ve kuyruğun hangi yönde olduğunu göstermektedir. Mantissa değerlerinin ortalaması 0.431 olarak bulunmuştur. Bu değer, verilerin genel olarak pozitif bir dağılıma sahip olduğunu göstermektedir. 0.058'lik bir varyans değeri, verilerin ortalama etrafında nispeten sıkı bir şekilde toplandığını göstermektedir. Düşük varyans, verilerin homojen olduğunu ve ortalamaya yakın değerler içerdiğini ifade etmektedir. Negatif basıklık değeri (-0.353), verilerin normal dağılıma göre daha yassı bir dağılım gösterdiğini ifade etmektedir. Bu, verilerin uç değerlerden ziyade merkezi değerlere daha yakın olduğunu göstermektedir. Pozitif çarpıklık değeri (0.547), verilerin sağa doğru uzun bir kuyruk gösterdiğini belirtmektedir. Bu durum, verilerin çoğunlukla düşük değerlere sahip olduğunu ve bazı yüksek değerlerin bulunduğunu göstermektedir. Genel olarak, PM₁₀ verilerinin, merkezde yoğunlaşan ve sağa doğru uzun bir kuyruk gösteren bir dağılıma sahiptir.

Tablo 3. Mantissa İstatistikleri

İstatistik	Değer
Ortalama	0.431
Varyans	0.058
Basıklık	-0.353
Çarpıklık	0.547

Tablo 4'te en büyük beş sapma değeri verilmiştir. Bu basamaklar (22, 23, 24, 28 ve 21) Benford dağılımından en büyük sapmaları gösteren basamaklardır. Sapmaların büyüklüğü, bu

basamakların verilerde Benford yasasının öngördüğünden çok farklı bir sıklıkta yer aldığını göstermektedir. Bu, verilerin doğal bir düzene uymadığını göstermektedir.

Tablo 4. En Büyük Beş Sapma

Basamaklar	Mutlak Fark
22	27.64
23	23.17
24	15.65
28	13.25
21	12.07

Tablo 5. Pearson Ki-Kare Test Sonuçları

Ki-Kare Değeri	sd	p-değeri
290.92	89	< 2.2e-16

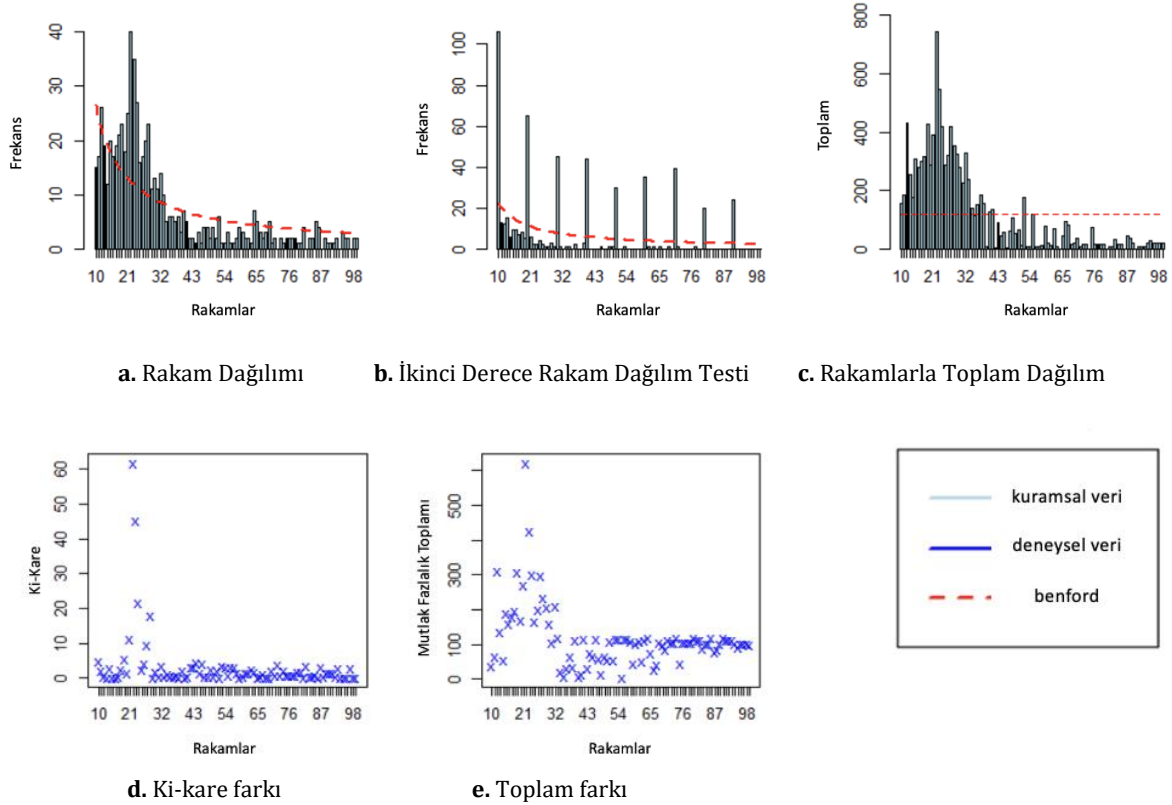
Pearson'ın Ki-kare testi sonuçlarına göre, verilerin Benford dağılımına ne kadar uyduğunu test etmektedir. Ki-kare değeri 290.92 ve p-değeri (<2.2e-16), verilerin Benford dağılımına uymadığını göstermektedir. Bu, verilerde anormallik olabileceğine işaret etmektedir. Tablo 6'da Mantissa Arc testi sonuçları verilmiştir. Ki-kare testi ile benzer bulgular tespit edilmiştir. Tablo 6'da verilen ortalama

sapma değeri de verilerin Benford dağılımından ortalama olarak ne kadar saptığını ölçmektedir. Düşük ortalama sapma değeri, verilerin Benford dağılımına daha yakın olduğunu gösterir. Ancak burada belirtilen düşük değer, genel olarak verilerin Benford Yasası'na uymadığını belirtmektedir. Benzer olarak bozulma faktörü de, verilerin Benford dağılımına göre ne kadar sapma gösterdiğini belirtir. Negatif değer, verilerin

önemli ölçüde Benford Yasası'ndan saptığını işaret etmektedir.

Tablo 6. Mantissa Arc Testi Sonuçları

Metrik	Değer
L2	0.11853
Sd	2
p-değeri	< 2.2e-16
Ortalama Mutlak Sapma	0.00572
Bozulma Faktörü	-40.84547



Şekil 2. Benford kanununa ait grafikler

Bu grafikler, veri setindeki rakamların Benford Yasası'na uyumunu değerlendirmek için kullanılan çeşitli görselleştirmeleri içermektedir. Aşağıda her bir grafiğin yorumu bulunmaktadır. Şekil 2.a'da, veri setindeki rakamların frekans dağılımını göstermektedir. Grafikteki çubuklar, her bir rakamın verilerde kaç kez görüldüğünü temsil ederken, kırmızı kesikli çizgi Benford dağılımına göre beklenen frekansları gösterir. Veri setindeki rakamların dağılımı, Benford Yasası'ndan belirgin sapmalar göstermektedir. Özellikle düşük rakamlar (10-30 arası) beklenenden daha sık görülmektedir. Şekil 2.b'deki grafik, ikinci dereceden rakamların dağılımını göstermektedir. Benford Kanunu yalnızca birinci basamağın değil, ikinci ve diğer basamakların da belirli bir dağılıma sahip olacağını öngörmektedir. Bu bağlamda, İkinci Derece Rakam Dağılım Testi, veri setindeki sayıların ikinci basamaklarının gözlemlenen

frekanslarını, Benford Kanunu'na göre beklenen frekanslarla karşılaştırır. Çalışmada bu test, PM10 verilerindeki olası tutarsızlıkların ve anormalliklerin tespit edilmesinde kullanılmıştır. İlk grafikte olduğu gibi, gözlemlenen frekansların çubuklarla, beklenen Benford frekanslarının ise kırmızı kesikli çizgiyle gösterildiği grafik, bu testin sonuçlarını sunmaktadır. Grafik, ikinci basamaktaki belirli rakamların beklenen sıklıktan sapma gösterdiğini açıkça ortaya koymaktadır. İkinci dereceden rakamlar arasında bazı rakamların beklenenden çok daha fazla veya az sıklıkta görüldüğü anlaşılmaktadır. Bu tür analizler, veri manipülasyonu, ölçüm hataları veya doğal olmayan düzenliliklerin belirlenmesinde önemli bir rol oynamaktadır. Şekil 2.c'deki grafik, her bir rakamın toplamda kaç kez görüldüğünü gösterir. Çubuklar, her bir rakamın toplam görünme sayısını, kırmızı kesikli çizgi ise Benford Yasası'na göre beklenen toplamı

gösterir. Bazı rakamlar beklenenden daha yüksek toplam değerlere sahip olup, bu da Benford Yasası'na uyumsuzluğu işaret etmektedir. Özellikle 10-30 arası rakamlar bu durumu net bir şekilde göstermektedir. Şekil 2.d'deki grafik, her bir rakam için gözlemlenen ve beklenen değerler arasındaki ki-kare farkını göstermektedir. Mavi "x" işaretleri, her bir rakam için hesaplanan ki-kare farkını temsil eder. Yüksek ki-kare farkları, belirli rakamların Benford Yasası'ndan önemli ölçüde saptığını gösterir. Bu sapmalar, veri setinde anormalliklerin veya olası hataların göstergesi olabilir. Şekil 2.e'deki grafik, her bir rakam için gözlemlenen ve beklenen toplam değerler arasındaki farkı göstermektedir. Yüksek toplam farkları, belirli rakamların beklenen Benford dağılımından önemli ölçüde saptığını gösterir. Bu, veri setinde tutarsızlıkların veya anormalliklerin olduğuna işaret eder. Bu grafikler, veri setinin Benford Yasası'na uyumunu değerlendirmede kullanılmıştır. Genel olarak, grafikler verilerin Benford Yasası'ndan önemli ölçüde saptığını göstermektedir. Bu sapmalar, veri setinde anormalliklerin, tutarsızlıkların veya olası veri manipülasyonunun göstergesi olabilir. Bu durumun kış ayları ve bazı günlerdeki aşırı yüksek değerlerden kaynaklandığı, Çankaya ilçesinin hava kirlilik düzeylerinin belirli bir düzen içermediği biçiminde

yorumlanabilir. Bu nedenle, aşırı değerler içeren PM₁₀ düzeylerinin incelenmesi için uç değer dağılımlar yardımıyla verinin incelenmesine karar verilmiştir.

Uç değer dağılımları ile veri kümesinin incelenebilmesi için öncelikle veri setindeki eksik değerler, Kalman filtresi kullanılarak tahmin edilmiştir. Bu yöntem zaman serisi verilerindeki eksiklikleri doldurmak için etkili bir yaklaşım olarak kabul edilmektedir. Aylık maksimum PM₁₀ değerlerini belirlemek için R programında "extremes", "evd", "EnvStats" paketlerinden yararlanılmıştır [52, 53, 54]. Bu paketlerde yer alan "blockMaxima" fonksiyonu ile aylık maksimum PM₁₀ değerleri belirlenmiştir. Bu değerler, çevresel verilerdeki aşırı olayların belirlenmesi için kritik öneme sahiptir. Elde edilen aylık maksimumlar, GUD dağılımı ile modellenmiştir. GUD modeli, atmosfer bilimleri ve çevre mühendisliği gibi alanlarda yaygın olarak kullanılan güçlü bir istatistiksel araçtır [55, 56]. Modelin uygunluğu, Kolmogorov-Smirnov (KS) uyum iyiliği testi ile değerlendirilmiştir. KS değeri 0.174 ve p-değeri 0.435 olarak elde edilmiştir. p-değeri 0.05'ten büyük olduğu için, GUD dağılımının veriye uygun olduğu sonucuna ulaşılmıştır.

Tablo 7. GUD Dağılımı için Parametre Tahmini

Parametre	Tahmin	Standart Hata
Konum	29.906	3.889
Ölçek	18.005	2.558
Şekil	-0.060	0.082

Konum parametresi yaklaşık 29.91 olarak tahmin edilmiştir. Bu değer, veri setindeki olayların ortalama büyüklüğünü temsil etmektedir. Ölçek parametresi 18.01 olup, veri setindeki değerlerin yayılımını

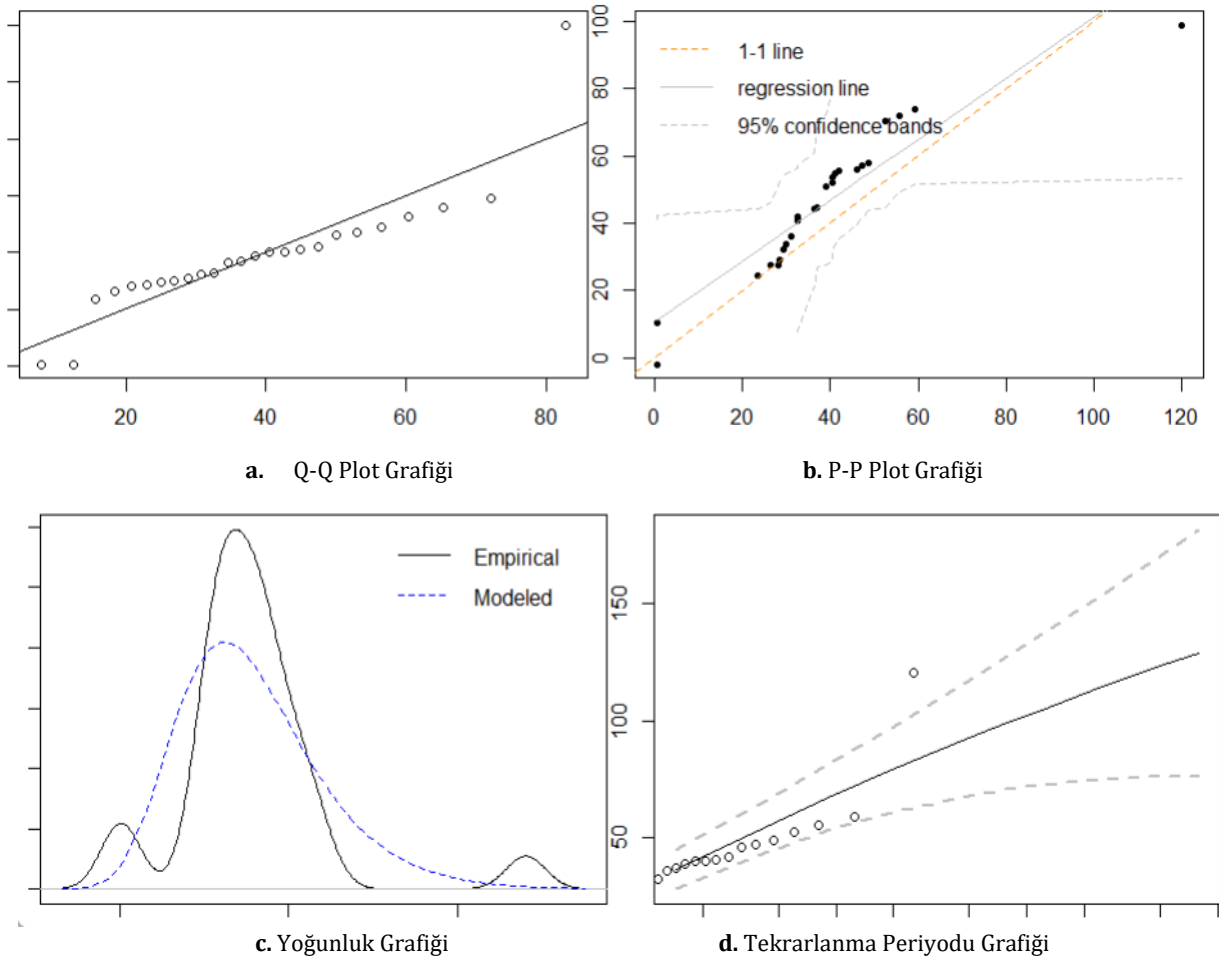
göstermektedir. Şekil parametresi ise, -0.06 olup, bu değer dağılımın kuyruğunun şekli hakkında bilgi vermektedir. Negatif değer, dağılımın sol kuyruklu olduğunu göstermektedir.

Tablo 8. PM₁₀ Aşım Düzeylerine ait Tahmin ve Güven Aralıkları

Tekrarlanma Periyodu	%95 Alt Güven Aralığı	Tahmin	%95 Üst Güven Aralığı
2 ay	28.23851	36.43317	44.62783
20 ay	60.72578	78.84279	96.95980
100 ay	71.76999	102.17620	132.58241

GUD kullanılarak yapılan tahminler, PM₁₀ konsantrasyonlarının belirli tekrarlanma (geri dönüş) seviyelerinde ne kadar sık ulaşabileceğini belirlemek için kullanılmıştır. Analiz sonuçlarına göre, 2, 20 ve 100 aylık dönemler için aşırı PM₁₀ değer tahminleri yapılmıştır. 2 aylık geri dönüş seviyesi tahmini 36.43, 20 aylık geri dönüş seviyesi tahmini 78.84 ve 100 aylık geri dönüş seviyesi tahmini 102.18 olarak

hesaplanmıştır. Diğer bir deyişle Çankaya ilçesinde her 2 ayda bir kez ortalama olarak 36.43 değerini aşması beklenmektedir. Buna ek olarak 0.95 olasılıkla, her 100 ayda ortalama olarak bir kez ölçülen PM₁₀ değerleri 102.18 değerini aşacağı ifade edilebilir. Bu seviyeler, belirli bir süre boyunca gözlemlenebilecek en ekstrem olayların büyüklüğünü temsil etmektedir.



Şekil 3. Bloktaki değerlerin en büyüğü yöntemine ait grafikler

Şekil 3'te verilen grafikler, verilen veri setinin GUD ile modellemesi üzerine oluşturulmuştur. Her bir grafik, modelin uygunluğunu ve verinin dağılımını değerlendirmek için farklı bir bakış açısı sunmaktadır. Şekil 3.a'da verilen Q-Q plot grafiği, gözlemlenen değerlerin teorik dağılıma ne kadar uyduğunu göstermektedir. Noktalar, teorik kantillere karşılık gelen gözlemleri temsil etmektedir. Noktaların çoğu, 45 derece eğik çizgiye (1-1 çizgisi) yakın olduğunda, modelin veriye iyi uyduğunu göstermektedir. Bu grafikte, noktaların büyük bir kısmı çizgiye yakın seyrettiğinden modelin veriye oldukça iyi uyduğunu söylenebilir. Şekil 3.b'de verilen P-P plot grafiğinde gözlemlenen olasılıkların teorik olasılıklarla karşılaştırılması sunulmuştur. 1-1 çizgisi, gözlemlenen ve teorik olasılıkların eşit olduğu noktaları temsil etmektedir. Noktalar bu çizgiye ne kadar yakınsa, model o kadar iyidir. Grafik, modelin gözlemlenen verilerle iyi bir uyum içinde olduğunu göstermektedir. Şekil 3.c'de verilen grafik, gözlemlenen verilerin ve modelin yoğunluk fonksiyonlarını karşılaştırmaktadır. Siyah çizgi gözlemlenen verilerin yoğunluğunu, mavi kesik çizgi ise modelin yoğunluğunu temsil etmektedir. İki çizginin birbirine yakın olması, modelin veriyi iyi bir şekilde temsil ettiği anlamına gelmektedir. Bu grafikte, iki çizgi birbirine yakın ve benzer bir dağılım göstermektedir. Şekil 3.d'de verilen tekrarlanma

periyodu grafiği, belirli geri dönüş dönemleri için tahmini aşırı değer seviyelerini göstermektedir. Çizgiler, modelin tahminlerini ve güven aralıklarını temsil etmektedir. Noktalar, gözlemlenen aşırı değerleri göstermektedir. Noktaların çoğunun tahmin çizgisine ve güven aralıklarına yakın olması, modelin doğruluğunu işaret etmektedir. Bu grafikte, gözlemlenen değerlerin büyük çoğunluğu tahmin edilen değerler ve güven aralıklarıyla uyumludur.

4. Tartışma ve Sonuç

Bu çalışmada Ankara ili Çankaya ilçesi için Çevre ve Şehircilik Bakanlığı Hava Kalitesi İzleme İstasyonlarından elde edilen 01.01.2021-01.01.2023 yılları arasındaki günlük partikül madde (pm₁₀) verileri üzerinde istatistiksel dağılım analizi yapılmıştır. Veri seti 731 gözlem (gün) sayısından oluşmaktadır. Veri seti üzerinde Bloktaki Değerlerin En Büyüğü (Block Maxima) yöntemi ve Benford Kanunu kullanılarak uyum iyiliği testleri yapılmıştır. Elde edilen sonuçlar ile pm₁₀ değerlerinin modellenmesinde kullanılan bloktaki değerlerin en büyüğü yöntemi için uyum iyiliği test sonucu p-değeri 0.4354457 olarak elde edilmiş ve p-değeri 0.05'den büyük olduğu için ilgili yöntem ile kurulan modelin veri ile uyum sağladığı ifade edilmiştir. Öte yandan pm₁₀ değerlerinin modellenmesinde kullanılan

Benford kanunu yaklaşımı için uyum iyiliği test sonucu p-değeri 0.05'den küçük olduğu için ilgili yöntem ile kurulan modelin veri ile uyum sağlamadığı ifade edilmiştir.

Aşırı Değer Dağılımı (GEV) modelinin, veri seti için iyi bir uyum sağladığı görülmektedir. QQ ve PP grafikleri, modelin gözlemlenen verilerle iyi bir uyum sağladığını gösterirken, yoğunluk grafiği ve geri dönüş seviyesi grafiği de modelin tahminlerinin güvenilir olduğunu desteklemektedir. Modelin, verinin dağılımını ve aşırı değerlerini başarılı bir şekilde temsil ettiği söylenebilir. Bu sonuçlara dayanarak, hava kirliliği seviyelerindeki anormallikleri tespit etmek için Benford Kanunu ve Genelleştirilmiş Uç Değer Dağılımı (GUD) yöntemlerinin bu alanda etkili bir şekilde kullanılabileceği ortaya konulmuştur.

Sonuç olarak, Benford Kanunu ve GUD yöntemleri, hava kirliliği verilerindeki anormallikleri belirlemede ve aşırı olayları öngörmeye birbirini tamamlayan güçlü araçlardır. Bu iki yöntemin bir arada kullanılması, hem kirlilik verilerindeki düzenliliklerin hem de ekstrem olayların daha iyi anlaşılmasını sağlayarak çevresel risk yönetimi ve halk sağlığı açısından kritik çıkarımlar sunmaktadır. Çalışmanın sonuçları, bu yöntemlerin hava kirliliği analizlerinde etkili bir şekilde uygulanabileceğini göstermekte ve bu alandaki literatüre değerli bir katkı sağlamaktadır.

Bu çalışmada elde edilen bulgular, hava kirliliği ve çevre yönetimi açısından son derece değerli sonuçlar ortaya koymaktadır. Benford Kanunu'nun yardımıyla, hava kirliliği verilerindeki düzensizlikler daha erken tespit edilerek, kirliliği kontrol altına almak için zamanında ve etkili müdahaleler yapılabilir. Aynı şekilde, GUD yöntemi ile yapılan tahminler, gelecekte karşılaşılabilecek aşırı kirlilik olaylarına dair somut öngörüler sunarak, uzun vadeli çevre politikalarının şekillendirilmesine destek olmaktadır. Bu iki yöntemin bir arada kullanılması, yalnızca mevcut sorunların analiziyle sınırlı kalmayıp, geleceğe dair proaktif adımlar atılmasına olanak tanımaktadır. Çalışmanın ortaya koyduğu sonuçlar, hava kirliliğinin izlenmesi ve kontrolü kadar, halk sağlığını koruma ve çevresel sürdürülebilirlik hedeflerine ulaşma yolunda da bilimsel bir temel sunarak, somut katkılar sağlamaktadır. Ayrıca, çalışmanın Çankaya ilçesine yönelik olması ve son 2 yıla ait gerçek/güncel veri seti kullanılması ile özgün bir araştırma olma ayrıcalığına sahiptir. Çalışma verinin farklı yaklaşımlara uyumu açısından araştırılıp geliştirilebilir. Aynı veri setinde farklı yaklaşımların kullanıldığı analizler ile bu çalışmadan elde edilen sonuçlar kıyaslanabilir. Çalışmada kullanılan veri seti tarih aralıkları ve çalışılan ilçe bazında genişletilerek Ankara geneli için benzer istatistiksel analizler yapılabilir.

Teşekkür

Bu makalenin hazırlanmasında, kullanılan yöntem ve analizlerin seçiminde 1129B372400336 numaralı TÜBİTAK 2237-A "Doğa Bilimleri Temelli Keşifsel Veri Analizi ve Veri Görselleştirme" ve 123B756 numaralı TÜBİTAK 4004 "İstatistiği Doğada Öğren-IV" adlı etkinliklerden yararlanılmıştır. İlgili projelere teşekkür ederiz.

Etik Beyanı

Bu çalışmada, "Yükseköğretim Kurumları Bilimsel Araştırma ve Yayın Etiği Yönergesi" kapsamında uyulması gerekli tüm kurallara uyulduğunu, bahsi geçen yönergenin "Bilimsel Araştırma ve Yayın Etiğine Aykırı Eylemler" başlığı altında belirtilen eylemlerden hiçbirinin gerçekleştirilmediğini taahhüt ederiz.

Kaynakça

- [1] Zencirci, S. A., Işıklı, B. 2017. Hava kirliliği. ESTÜDAM Halk Sağlığı Dergisi, 2(2), 24-36.
- [2] International Agency for Research on Cancer, & World Health Organization. 2013. IARC: Outdoor air pollution a leading environmental cause of cancer deaths. No. 221. World Health.
- [3] WHO (World Health Organization), 2018. Ambient (outdoor) air pollution. [https://www.who.int/en/news-room/factsheets/detail/ambient-\(outdoor\)-air-quality-and-health](https://www.who.int/en/news-room/factsheets/detail/ambient-(outdoor)-air-quality-and-health), (Erişim Tarihi: 14.01.2023).
- [4] Zeydan, Ö. 2021. 2019 Yılında Türkiye'deki Partikül Madde (PM10) Kirliliğinin Değerlendirilmesi. Journal of the Institute of Science and Technology, 11(1), 106-118.
- [5] Bayram, H., Dörtbudak, Z., Fişekçi Evyapan, B. F., Kargin, M., & Bülbül, B. 2006. "Hava kirliliğinin insan sağlığına etkileri, dünyada, ülkemizde ve bölgemizde hava kirliliği sorunu" paneli ardından. Dicle Tıp Dergisi.
- [6] Yu, R., Yang, Y., Yang, L., Han, G. ve Oguti, M. 2016, RAQ—a random forest approach for predicting air quality in urban sensing systems, Sensors, 16, 86–104.
- [7] Tai, A.P.K., Mickley, L.J. ve Jacob, D.J. 2010, Correlations between fine particulate matter (PM2.5) and meteorological variables in the United States: Implications for the sensitivity of PM2.5 to climate change, Atmospheric Environment, 44, 3976–3984.
- [8] Wang, S., Wang, J., Zhou, Z., & Shang, K. 2005. Regional characteristics of three kinds of dust storm events in China. Atmospheric Environment, 39(3), 509-520.
- [9] Miller, S. D., Kuciauskas, A. P., Liu, M., Ji, Q., Reid, J. S., Breed, D. W., ... & Mandoos, A. A. 2008. Haboob dust storms of the southern Arabian

- Peninsula. Journal of Geophysical Research: Atmospheres, 113(D1).
- [10] Jaafari, J., Naddafi, K., Yunesian, M., Nabizadeh, R., Hassanvand, M. S., Ghoskhal, M. G., ... & Yaghmaeian, K. 2018. Study of PM₁₀, PM_{2.5}, and PM₁ levels in during dust storms and local air pollution events in urban and rural sites in Tehran. Human and ecological risk assessment: An International Journal, 24(2), 482-493.
- [11] Ozel, G., & Cakmakyapan, S. 2015. A new approach to the prediction of PM₁₀ concentrations in Central Anatolia Region, Turkey. Atmospheric Pollution Research, 6(5), 735-741.
- [12] Cekim, H. O. 2020. Forecasting PM₁₀ concentrations using time series models: a case of the most polluted cities in Turkey. Environmental Science and Pollution Research, 27(20), 25612-25624.
- [13] Brunekreef, B., & Forsberg, B. 2005. Epidemiological evidence of effects of coarse airborne particles on health. European respiratory journal, 26(2), 309-318.
- [14] Host, S., Larrieu, S., Pascal, L., Blanchard, M., Declercq, C., Fabre, P., ... & Lefranc, A. 2008. Short-term associations between fine and coarse particles and hospital admissions for cardiorespiratory diseases in six French cities. Occupational and Environmental Medicine, 65(8), 544-551.
- [15] Graff, D. W., Cascio, W. E., Rappold, A., Zhou, H., Huang, Y. C. T., & Devlin, R. B. 2009. Exposure to concentrated coarse air pollution particles causes mild cardiopulmonary effects in healthy young adults. Environmental health perspectives, 117(7), 1089-1094.
- [16] Malig, B. J., & Ostro, B. D. 2009. Coarse particles and mortality: evidence from a multi-city study in California. Occupational and environmental medicine, 66(12), 832-839.
- [17] Linares, C., Tobías, A., & Díaz, J. 2010. Is there new scientific evidence to justify reconsideration of the current WHO guidelines for particulate matter during dust intrusions?. The Science of the total environment, 408(10), 2283-2284.
- [18] Qiu, H., Yu, I. T. S., Tian, L., Wang, X., Tse, L. A., Tam, W., & Wong, T. W. 2012. Effects of coarse particulate matter on emergency hospital admissions for respiratory diseases: a time-series analysis in Hong Kong. Environmental health perspectives, 120(4), 572-576.
- [19] Goudarzi, G., Mohammadi, M., Angali, K.A., Mohammadi, B., Soleimani, Z., Babaei, A., Neisi, A. ve Geravandi, S. 2013, Estimation of number of cardiovascular death, myocardial infarction and chronic obstructive pulmonary disease (COPD) from NO₂ exposure using air Q model in Ahvaz City during 2009, Iranian Journal of Health and Environment, 6, 91-102.
- [20] Toros, H., Bağış, S., & Gemici, Z. 2018. Ankara'da hava kirliliği mekânsal dağılımının modellenmesi. Ulusal Çevre Bilimleri Araştırma Dergisi, 1(1), 20-53.
- [21] Coşkun, A. 2008. Şehir atmosferinde taşıt emisyonlarından kaynaklanan hava kirliliğinin belirlenmesi (Master's thesis, Marmara Üniversitesi (Turkey)).
- [22] İmal, M., Karapinar, Ç., & Doğan, O. 2014. Hava kalitesine doğalgazın etkisi: Kahramanmaraş örnek çalışması. Kahramanmaraş Sütçü İmam Üniversitesi Mühendislik Bilimleri Dergisi, 16(2), 22-28.
- [23] Garipağaoğlu, N. 2022. Türkiye'de Hava Kirliliğinin Partiküler Madde (PM₁₀) Açısından İncelenmesi.
- [24] Toros, H., & Bağış, S. 2017. Hava kirlilik modellerinde kullanılacak emisyon envanteri oluşturulması için yaklaşımlar ve İstanbul hava kirliliği dağılımı örneği. Çukurova Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi, 32(2), 1-12.
- [25] Barış, M. E., & Koç, N. 1996. Ankara Kentinde Hava Kirliliği Sorununun Çözümünde Peyzaj Mimarlığı Açısından Alınması Gerekli Önlemler. Journal of Agricultural Sciences, 3(02), 1-8.
- [26] Zeydan, Ö., & Karakaya, B. N. 2017. Assessment of PM₁₀ limit exceedances in Turkish cities. Journal of Young Scientist, 5, 115-120.
- [27] Çiçek, İ., Türkoğlu, N. ve Gürgen, G. 2004, Ankara'da Hava Kirliliğinin İstatistiksel Analizi, Fırat Üniversitesi Sosyal Bilimler Dergisi, 14(2), 1-18.
- [28] Bacak, T. N., & Toros, H. 2021. Impact of the COVID-19 event on PM₁₀ air pollution in Istanbul and Ankara. Journal of Research in Atmospheric Science, 3(1), 1-7.
- [29] Yavuz, C. I. 2020, Kentsel Alanda Hava Kirliliği: Hava Kirliliği İzleme Ağı Ankara İstasyonlarının Beş Yıllık PM.
- [30] Ünal, C., & Özel, G. 2023. Ankara İli Meteoroloji Parametrelerinin Hava Kirliliği Üzerindeki Etkilerinin Regresyon Analizi ile İncelenmesi. Nicel Bilimler Dergisi, 5(2), 135-150.
- [31] Diekmann, A. 2007. Not the first digit! Using benford's law to detect fraudulent scientific data. Journal of Applied Statistics, 34(3), 321-329.
- [32] Nigrini, M. J., & Mittermaier, L. J. 1997. The use of Benford's Law as an Aid in Analytical Procedures. Auditing: A journal of practice & theory, 16(2).

- [33] Özkan, K. 2021. Estimating ecosystem naturalness using Benford's law and generalized Benford's law. *Turkish Journal of Forestry*, 22(2), 73-82.
- [34] Masseran, N., & Safari, M. A. M. 2022. Statistical modeling on the severity of unhealthy air pollution events in Malaysia. *Mathematics*, 10(16), 3004.
- [35] Masseran, N., & Safari, M. A. M. 2021. Mixed POT-BM approach for modeling unhealthy air pollution events. *International Journal of Environmental Research and Public Health*, 18(13), 6754.
- [36] Martins, M. C. H., Fatigati, F. L., Vespoli, T. C., Martins, L. C., Pereira, L. A. A., Martins, M. D. A., ... & Braga, A. L. F. 2004. Influence of socioeconomic conditions on air pollution adverse health effects in elderly people: an analysis of six regions in Sao Paulo, Brazil. *Journal of Epidemiology & Community Health*, 58(1), 41-46.
- [37] Yeşil, A. 2006, Ankara Metropolitan Alanının Yeşil Alan Sisteminin Analizi, Fen Bilimleri Enstitüsü, Yıldız Teknik Üniversitesi, İstanbul.
- [38] Akman, Y. 1990. İklim ve biyoiklim. *Palme Yayınları*, Ankara, 1, 319.
- [39] ÇŞB (Çevre ve Şehircilik Bakanlığı), 2023. Ulusal Hava Kalite İzleme Ağı. http://sim.csb.gov.tr/STN/STN_Report/StationDataDownloadNew, (Erişim Tarihi: 10.01.2023).
- [40] Newcomb, S. 1881. Note on the frequency of use of the different digits in natural numbers. *American Journal of mathematics*, 4(1), 39-40.
- [41] Benford, F. 1938. The law of anomalous numbers. *Proceedings of the American philosophical society*, 551-572.
- [42] Carslaw, C. A. 1988. Anomalies in income numbers: Evidence of goal oriented behavior. *Accounting Review*, 321-327.
- [43] Kaşıkçı, B., & Dalgıç, H. 2021. Muhasebe hilelerinin ortaya çıkarılmasında benford yasası'nın kullanılmasına yönelik bir uygulama. *Süleyman Demirel Üniversitesi Vizyoner Dergisi*, 12(32), 1196-1211.
- [44] Güner, M., & Kurnaz, E. 2021. Muhasebe Denetiminde Benford Kanununun Kullanımı Bir Devlet Üniversitesi Uygulaması. *Gümüşhane Üniversitesi Sosyal Bilimler Dergisi*, 12(1), 81-96.
- [45] Kıvraklar, M. K., & Can, R. E. 2023. Denetimde Güçlü Bir Araç "Benford Yasası": Kamu İdaresi Hesaplarında Bir Uygulama. *Muhasebe ve Denetim Bakış*, 23(70), 183-208.
- [46] Öncü, M. A., Yücel, R., & Özevin, O. 2018. Benford analizi ile muhasebe denetimi: kamu hastaneleri üzerine bir uygulama. *Muhasebe ve Finansman Dergisi*, (80), 1-22.
- [47] İlkdoğan, S. 2020. İç denetimin hileye yaklaşımında Benford Kanunu'nun uygulanması (Master's thesis, Balıkesir Üniversitesi Sosyal Bilimler Enstitüsü).
- [48] Massicotte, P., & Eddelbuettel, D. (2023). *gtrendsR: Perform and display Google Trends queries*. R package version 1.5.0. Retrieved from <https://CRAN.R-project.org/package=gtrendsR>
- [49] Wickham, H. (2007). *reshape2: Flexibly reshape data: A reboot of the reshape package*. R package version 1.4.4. Retrieved from <https://CRAN.R-project.org/package=reshape2>
- [50] Wickham, H., François, R., Henry, L., & Müller, K. (2023). *dplyr: A grammar of data manipulation*. R package version 1.1.3. Retrieved from <https://CRAN.R-project.org/package=dplyr>
- [51] Cinelli, C. (2021). *benford.analysis: Benford analysis for data validation and forensic analytics*. R package version 0.1.6. Retrieved from <https://CRAN.R-project.org/package=benford.analysis>
- [52] Gilleland, E., & Katz, R. W. (2016). *extRemes 2.0: An extreme value analysis package in R*. *Journal of Statistical Software*, 72(8), 1-39. <https://doi.org/10.18637/jss.v072.i08>
- [53] Stephenson, A. G. (2002). *evd: Extreme value distributions*. *R News*, 2(2), 31-32. Retrieved from <https://CRAN.R-project.org/doc/Rnews/>
- [54] Millard, S. P. (2013). *EnvStats: An R package for environmental statistics*. Springer. <https://doi.org/10.1007/978-1-4614-8456-1>
- [55] Fontolan, M., Xavier, A. C. F., Pereira, H. R., & Blain, G. C. 2019. Using climate change models to assess the probability of weather extremes events: a local scale study based on the generalized extreme value distribution. *Bragantia*, 78, 146-157.
- [56] Rypkema, D., & Tuljapurkar, S. 2021. Modeling extreme climatic events using the generalized extreme value (GEV) distribution. In *Handbook of Statistics* (Vol. 44, pp. 39-71). Elsevier