

Araştırma Makalesi - Research Article

Z-Skor ve Kutu Grafiğiyle Sapan Değer Belirleme

Outlier Detection by Z-Score and Box Plot

Ahu Dede^{1*}, Necati Ağırlioğlu²

Geliş / Received: 20/04/2024

Revize / Revised: 10/08/2024

Kabul / Accepted: 12/08/2024

ÖZ

Türkiye'deki 1926'dan 2012'ye kadar olan su kaynakları verileri hataya açıktır ve bu durum sıkı kalite kontrol prosedürlerini gerektirmektedir. 129 istasyondaki aylık ortalama sıcaklık, ortalama bağıl nem ve toplam yağış verilerindeki sapan değerler, z-skoru ve kutu grafikleri kullanılarak analiz edildi. Analiz sonuçları, mevsime özgü ön işleme prosedürlerinin uygulanmasının gerekliliğini ve sapan değer tespit sınırları arasındaki eşitsizliklerin belirgin olduğunu ortaya koymaktadır. Ayrıca, çeyrekler arası aralık için katsayı olarak dört değerinin kullanılmasının zorunluluğu vurgulanmıştır. Sapan değerlerinin, eksik verilere ve sapan olmayan değerlere, özellikle yaz aylarında sıfır olmayan yağış kayıtlarına yakın olma eğilimi göstermesi dikkat çekicidir. Sıcaklık ve yağış sapan değerlerinin mekansal ve zamansal açıdan benzerlikler gösterdiği belirlenmiştir. Bağıl nem ve yağış sapan değerlerini çıkaran komşu bölgeler arasında benzerlik gözlemlenmiştir. Çalışmada ayrıca, sapma miktarı ile sapan değer adedi arasında ters orantılı bir ilişki saptanmış olup, sapan değer içeren istasyonların bölgesel konumları ile bu istasyonlardaki verilerin dağılımları arasında anlamlı bir ilişki olduğu sonucuna varılmıştır.

Anahtar Kelimeler- *Hidroloji, Sapan Değer, Z-Skor, Kutu Grafiği*

ABSTRACT

The water resources data from Turkey, spanning the period from 1926 to 2012, are susceptible to errors, necessitating rigorous quality control procedures. Outliers in the monthly average temperature, relative humidity, and total precipitation data from 129 stations were analysed using z-scores and box plots. The analysis revealed the necessity of applying season-specific preprocessing procedures and highlighted significant inequalities in the thresholds used for outlier detection. Additionally, it was emphasized that a coefficient of four should be used for the interquartile range. Notably, outliers tended to be close to missing data and non-outlier values, particularly to non-zero precipitation records during the summer months. Spatial and temporal similarities were observed in the temperature and precipitation outliers, while relative humidity and precipitation outliers exhibited similarities among neighbouring regions that generated outliers. The study also identified an inverse relationship between the magnitude of deviation and the frequency of outliers, and it found a significant correlation between the regional locations of stations containing outliers and the distribution of data within these stations.

Keywords- *Hydrology, Outliers, Z-Score, Boxplot*

^{1*}Sorumlu yazar iletişimi: ahudedem@gmail.com (<https://orcid.org/0000-0002-0534-6823>)

Bağımsız Araştırmacı, İstanbul, Türkiye

²İletişim: necati.agiralioglu@antalya.edu.tr (<https://orcid.org/0000-0002-5336-9202>)

İnşaat Mühendisliği Bölümü, Mühendislik ve Doğa Bilimleri Fakültesi, Antalya Bilim Üniversitesi, Antalya, Türkiye

I. GİRİŞ

A. Genel

Su kaynaklarının yönetilmesi için ekosistemlerdeki değişiklikler ölçülür, toplanır, düzenlenir, karakteristikleri belirlenir, davranışları tahmin edilir, böylelikle yeni teknolojiler, araçlar ve hizmetler geliştirilir. Fakat ölçümler her ne kadar veri toplama merkezlerinde kontrol edilse de [1,2] çeşitli hatalarla yüklüdür. Bunlar teknikten, cihazdan, personelden ve çevreden kaynaklanabilir [3-6]. Hatalar analizlerin doğru sonuçlar vermesini engeller [3,6-11]. Bunun için kalite kontrol (quality control) ve türdeşlik (homogeneity) incelemesi yapılır. Kalite kontrol incelemelerinde sapan değer (outlier), iç tutarlılık (internal consistency) ve ardışık değerler arası farklar (persistence test /temporal coherence /time continuity) araştırılır. Türdeşlik incelemelerinde varyansın ve ortalamanın türdeşliği araştırılır [11-18]. 1620'den bu yana bahsi geçen sapan değere ilişkin literatür çok geniştir [19]. İlk olarak uyumsuz gözlemlerin göz ardı edilmesi eleştirilmiştir [20]. Ardından sapan değerlerle başa çıkmak için 1850'lerde yöntemler geliştirilmeye başlanmıştır [21]. Sapan değer dağılım aralığının belli bir mesafe uzağında kalan, örneklem ile aynı dağılımdan gelmeyen ve popülasyona uymadığı tespit edilen değerdir [22].

Su kaynaklarında sapan/aykırı değer (outlier), hatalı veya şüpheli veriyi gösterirken ekstrem değer (extreme) gerçekleşen aşırı olayları göstermek için kullanılır [2]. Kırılma/kesme noktası (breakpoint) veya değişim noktası (change point) veri serilerini nispeten keskin bir şekilde parçalara/segmentlere ayıran noktaları ifade eder. İklimsel anomali (anomaly) ise uzun dönemli ortalama değerden sapmayı ifade eder. Fakat bazı kaynaklarda anomali ve değişim noktası ifadelerinin sapan değer olarak kullanılmasına [23], sapan değer sadece hatalı olmadığı bilgi de içerdiği söylemine [24, 25] ve sapan değerlerin ekstrem değerler olduğu söylemine rastlanır [26, 27]. Hatalı sapan değerlerin tespit edilmesi durumunda bu değer veri setinden çıkarılabilir veya düzeltilir. Gerçek sapan değerler ise etkilerini anlamak için özel olarak değerlendirilir. Bu bağlamda sapan değerler karmaşık ve çok boyutlu etkiler yaratabilir. Su kaynaklarındaki sapan değerleri yaratan olaylar tarımda mahsul kayıplarına bu da gıda fiyatlarında artışa neden olabilir, enerjide taleplerin etkilenmesine dolayısıyla enerji piyasalarında dalgalanmalara yol açabilir. Ekstrem yağış olayları, kirleticilerin su sistemlerine karışmasına, ayrıca toprak erozyonuna ve bunun getirisi tarım arazilerinde verim düşüklüğüne, gıda arzında azalmaya ve fiyat artışlarına neden olabilir. Ekstrem sıcaklıklar, toprağın nemini azaltabilir dolayısıyla mikroorganizmaların faaliyetlerini olumsuz etkilenerek toprağın verimliliğini düşürür ve üretiminde sorunlara yol açabilir. Bunun yanında orman yangınlarına neden olabilir ve ormansızlaşma hızlanabilir, bu da iklim değişikliğiyle mücadele açısından kritik öneme sahip karbon döngüsünü etkileyerek atmosfere daha fazla karbondioksit salınımına neden olabilir. Su kaynaklarına bağımlı enerji üretim tesisleri kuraklık dönemlerinde enerji üretim kapasitesini kaybedebilir. Bu durum, enerji arzında dalgalanmalara ve enerji güvenliğinde risklere yol açabilir. Sapan değerler aşırılık içerdiğinden oluşan riskler sigorta taleplerinde artışa neden olabilir. Ekstrem hava koşulları, inşaat ve altyapı projelerinde maliyet artışlarına, tedarik zincirlerinde dolayısıyla üretim süreçlerinde aksamalara yol açabilir. Ayrıca sağlık üzerinde de etkili olabilir ve sosyal hizmetlerin yükünü artırabilir. Bunun yanında işgücü verimini düşürebilir ve iş güvenliği açısından risk oluşturabilir. Ek olarak göç hareketlerine neden olabilir. Bu durum, hem göç veren hem de göç alan bölgelerde ekonomik ve sosyal zorluklar yaratabilir. Son olarak kültürel ve sosyal dinamikleri de etkileyebilir. Örneğin yerel festivallerin veya turizm faaliyetlerinin aksamasına neden olabilir. Tüm bunlar dış ticaret dengesinden enflasyona makroekonomik istikrarı tehdit eder. Ve sadece ekonomik sonuçlar doğurmakla kalmaz, aynı zamanda ekosistemlerin ve insan topluluklarının sağlığı ve refahı üzerinde de kritik roller oynar. Bu açılarıdan bakıldığında, su kaynaklarındaki sapan değerlerin etkileri oldukça kapsamlı ve çok katmanlıdır. Bu durum, politika yapıcıların ve iş dünyasının bu etkileri anlamak ve yönetmek için daha bütünsel yaklaşımlar geliştirmelerini gerektirir. Bu etkiler için veri analistlerinin daha sağlam tahminler geliştirmesi ve bunların stratejilere dahil edilmesi, devletlerin inovasyona yatırım yapması ve su kaynaklarının daha verimli kullanılması için kaynak yönetimi stratejilerinin sürdürülebilirliğini artırması gerekir.

Sapan değeri tespit ederken veriye ve eşik değere bakılır. Verinin kendisinin ya da kendisi dışında kalanların bir referansa olan uzaklığı kullanılır. Bu referans, ortalama, medyan gibi bir merkez ya da en yakın veya belli bir yüzdelik dilim olabilir. Eşik; Z, F, ki-kare, t gibi çeşitli dağılımlardaki kritik değer ya da yaygın olarak kabul edilen bir sabit olabilir.

İstatistik tabanlı Z-skor, Weisberg t-testi, Walsh ve Grubbs t-testi gibi tekniklerde verinin kendisinin merkeze uzaklığı Z veya t dağılımına uyduğu için kritik değer üzerinde olan uzaklıklar sapan değer kabul edilir [22, 28-36]. Tek değişkenli bu yöntemleri çok değişkenli veri ortamına genişletmek için mesafe tabanlı yöntemler geliştirilmiştir. Çeşitli mesafe teknikleri vardır. Öklit, Mahalanobis, Manhattan, Minkowski gibi tekniklerle bulunan her verinin veri kümesinin merkezine mesafeleri, ki-kare dağılımına uyduğu için kritik değer üzerinde olan uzaklıklara ait değerler sapan değer olarak belirlenebilir [30, 35, 37, 38]. Gerçek veriye uydurulmuş bir regresyon doğrusunun gerçek veriyle farkları alındığında verinin kendisinin merkeze uzaklığı alınmış gibi olur. Student-türü artıklar denilen bu farklar Z dağılımına uyar ve kabul edilen kritik değeri aşanlar sapan değer kabul edilir [22, 28, 29, 30, 33]. Destek vektör veri açıklaması (SVDD) yönteminde verinin normal durumu temsil eden

örneklerle eğitilen, kabul edilen sapan değer toleransına bağlı olarak, yarıçapı, merkezi, hata marjı belirlenen hiperkürenin merkezine mesafesi yarıçap sınırlarının dışındaysa sapan değer olarak kabul edilir [38 – 42].

İstatistik tabanlı Dixon Testi, Tukey testi, kutu grafiği gibi tekniklerde verinin kendisinin en yakın veriye veya belli bir yüzdelik dilime uzaklığı kabul edilen bir sabitin dışındaysa sapan değer kabul edilir [22, 28-31, 33, 37, 43]. Derinlik tabanlı yöntemler bu tek değişkenli sıra istatistiklerini çok değişkenli veri ortamına genişletmek için geliştirilmiştir. Bu yöntemde derinlik fonksiyonları veri noktaları için bir merkezilik ölçüsü sağlar. Yüksek derinlik değerine sahip noktalar daha merkezi kabul edilirken düşük derinlik değerine sahip olanlar daha aykırı kabul edilir. Derinlikleri hesaplamak için birçok farklı algoritma geliştirilmiştir. Bunlar Tukey, Mahalanobis, projeksiyon, zonoid, uzaysal derinlik gibi fonksiyonlardır. Bunlar da ki-kare dağılımına uyduğu için kritik değer üzerinde olan derinliklere ait değerler sapan değer olarak belirlenebilir [38, 39, 44-50]. K-en yakın komşular algoritmasında, verinin kendisine k adet en yakın verilere olan uzaklığının ortalamaları kabul edilen bir sabitin dışındaysa sapan değer kabul edilir [28, 37-39, 51].

Çok değişkenli veriler farklı yoğunlukta dağılmış olabilir. Bu durumda kümeleme yardımıyla yoğunluğa göre gruplar bulunur ardından gruplara göre sapan değerler bulunabilir ya da sadece kümeleme yöntemiyle sapan değerler bulunabilir. Bunlar için hiyerarşik kümeleme, k-ortalama, Gauss dağılım tabanlı kümeleme, DBSCAN algoritmaları kullanılabilir [30, 34, 39, 48, 51-54]. İzolasyon ormanı yönteminde kullanılan dâhili düğümler değişen merkezler olarak düşünülebilir ve mesafe yerine dallanma sayısı hesaplanır. Oluşturulan ağaçların ortalamaları alınarak anomali skoru bulunur. Bire yakın skorların önceden kabul edilen sapan değer miktarına kadar sapan olduğu varsayılır [37, 38, 54-56].

Verilerin merkeze uzaklıklarının çarpımının, verilerin bir eksiğinin merkeze uzaklıklarının çarpımından farkının incelenmesi aslında verinin kendisi dışında kalanların incelenmesi ile bulunan sapan değerlerdir. Olabilirlik (likelihood) fonksiyonu kullanılarak bulunan bu farklar ki-kare dağılımına uyar ve kritik değeri aşanlar sapan değer kabul edilir. Sapan değer, verinin kendisi dışında kalanların incelenmesi ile bulunmasında diğer teknikler Cook uzaklığı, sapma artıkları (deviance residuals), DFFITS, DFBETAS gibi farklardır. Bu farklar F veya Z dağılımına uyar ve kritik değeri aşanlar sapan değer kabul edilir. Tüm bu tekniklerde gözlemin tahmin katsayılarına etkisini gösteren kaldıraç (leverage) noktası yardımcı olarak kullanılır [28, 30, 52, 57-59]. Lokal Sapan Değer Faktörü (LOF) algoritmasında verinin kendi komşularının ve komşularının komşularının diğer veriler tarafından erişilebilirliği kıyaslayarak her veriye bir erişilemezlik oranı atanır. Kritik değer olarak 1 kullanılır [37-39, 48, 51, 52, 54]. Katmanlı ağ yapısından meydana gelen autoencoders, GAN, DBN, RNN, LSTM gibi derin öğrenme teknikleri tahminle gerçek arasındaki hatanın büyüklüğüne göre sapan değeri belirler [34, 37-39, 55, 60].

Görüldüğü gibi sapan değer tespitinde verinin kendisinin bir referansa olan uzaklığının dikkate alındığı yöntemler temelde ikiye ayrılır. Bunlar Z-skor ve kutu grafiği yöntemleridir. Bu araştırmanın amacı, bu iki yöntemin uygulaması ve sonuçları hakkında daha önce incelenmemiş özellikleri hidrolojik veriler ile incelemektir.

B. Su Kaynaklarında Sapan Değer Tespiti için Kullanılan Yöntemler

Su kaynaklarında sapan değer analizlerinde birçok yöntem kullanılmıştır. Verinin merkeze uzaklığını normal dağılıma uyduğunu kabul edip, genellikle 3 olan bir sabitle karşılaştırılarak sapan değeri belirleyen z-skor yöntemi ile yağış [61 -63], yüzey sıcaklığı [63], deniz seviyesi [64], yeraltı su seviyesi [65], nehir su seviyesi [66], akım hızı verisi [67] gibi değişkenlerde sapan değerler belirlenmiştir. Bazı çalışmalarda 2 olan bir sabitle karşılaştırılarak yağış [63, 68], yüzey sıcaklığı [63], taşkın verileri [69], akım verileri [70] gibi değişkenlerde sapan değerler tayin edilmiştir. Bazı çalışmalarda 2,5 olan bir sabitle karşılaştırılarak sıcaklık [71], taşkın verileri [69] gibi değişkenlerde sapan değerler tespit edilmiştir. Nadiren 4 olan bir sabitle karşılaştırılarak da sıcaklık [72, 73] ve yağış [72] gibi değişkenlerde sapan değerler belirlenmiştir. Çok nadiren ise 5 olan bir sabitle karşılaştırılarak yüzey sıcaklığı ve yağış [63] gibi değişkenlerde sapan değerler tayin edilmiştir. Ve gene Çok nadiren 3,5 olan bir sabitle karşılaştırılarak sıcaklık [73] değişkeninde sapan değerler belirlenmiştir. Bazı çalışmalarda ise çeşitli iklimsel değerler [2], yağış [74, 77, 78], sıcaklık [75, 77], buharlaşma [76], basınç, kar derinliği, bağıl nem, güneşlenme süresi [77], hava kalitesinde incelenen PM_{2,5}, PM₁₀, SO₂, NO₂, CO ve O₃ [79] gibi değişkenlerdeki sapan değerleri belirlemek için kullanılan sabit belirtilmemiştir. Bazı çalışmalarda da veri adetinde göre değişen sabitler kullanılarak hidrolojik veri [80] ve akım hızı [67] gibi değişkenlerde sapan değerler belirlenmiştir. Z-skor yöntemine benzer şekilde verinin merkeze uzaklığını Log Pearson Tip 3 dağılımına uyduğunu kabul edip, %10'luk anlamlılık seviyesine göre sapan değeri belirleyen yöntem ile taşkın debilerindeki sapan değerler tayin edilmiştir [81]. Verinin merkeze uzaklığını Gumbel dağılımına uyduğunu kabul edip, kritik değere göre sapan değeri belirleyen yöntem ile yıllık maksimum yağışlarda sapan değerler belirlenmiştir [82].

Z-skor yöntemine çok benzer verinin merkeze uzaklığını Grubbs tablosuna uyduğunu kabul edip sapan değeri belirleyen Grubbs yöntemi ile hidrolojik veri [23], yağış [83, 85-87], kar derinliği [84], sıcaklık, bağıl nem, rüzgar hızı, güneşli saatler ve buharlaşma [85], nehir su seviyesi [66], taşkın debisi [88, 89], yıllık maksimum akım [90] gibi değişkenlerde sapan değerler tespit edilmiştir. Gene Z-skor yöntemine çok benzer verinin ortancaya uzaklığını 2,5 ile 7,5 arasında değişen bir sabitle karşılaştırılarak sapan değeri belirleyen ortanca mutlak sapma

yöntemi ile yeraltı su seviyesi [65], sıcaklık [91, 92], yağış [91, 92], rüzgar hızı, güneş radyasyonu, bağıl nem [91], basınç, bulutluluk oranı [92], radyasyon verisi [95], akım hızı [67], hava kalitesinde incelenen PM_{2,5}, PM₁₀, SO₂, NO₂, CO ve O₃ [79], atmosferdeki eser gazların miktarı [94] gibi değişkenlerde sapan değerler belirlenmiştir.

Regresyon yöntemiyle iklim değişikliği analizlerinde kullanılan bitki fenoloji verileri [95], sıcaklık [96, 97], hidrolojik veriler [23], taşkın verileri [69], yağış [97], tuzluluk ve bor izleme verileri [98], deniz seviyesi [64] gibi değişkenlerde sapan değerler tayin edilmiştir.

Destek vektör veri açıklaması (SVDD) yönteminin bir alt modeli olan tek sınıf SVM ile toprak tipini belirten uydu verisi [99] ve geofiziksel veriler [54] gibi değişkenlerde sapan değerler belirlenmiştir.

Verinin ilk %25. ve %75. değere uzaklığını, genellikle 1,5 olan bir sabitle karşılaştırılarak sapan değeri belirleyen kutu grafiği yöntemi ile yağış [100 -104], sıcaklık [103-105], deniz seviyesi [106], yeraltı su seviyesi [65], nehir debisi [107,108], bitki örtüsü indeksi [109] gibi değişkenlerde sapan değerler tespit edilmiştir. Bazı çalışmalarda 3 olan bir sabitle karşılaştırılarak da yağış [110-114], sıcaklık, su buharı basıncı, rüzgar hızı [114] gibi değişkenlerde sapan değerler belirlenmiştir. Nadiren 5 olan bir sabitle karşılaştırılarak da sıcaklık, çiğlenme noktası sıcaklığı, deniz seviyesi basıncı, rüzgar hızı [115] gibi değişkenlerde sapan değerler tayin edilmiştir. Bazı çalışmalarda ise sıcaklık [116-118], basınç [117], bağıl nem [118], soğutma – ısıtma derece günleri [119] gibi değişkenlerdeki sapan değerleri belirlemek için kullanılan sabit belirtilmemiştir. Verinin ilk ve son değere uzaklığını Dixon sabitiyle karşılaştırılarak sapan değeri belirleyen Dixon yöntemi ile yağış [85, 120], sıcaklık, bağıl nem, rüzgar hızı ve güneşli saatler, buharlaşma [85], hidrolojik veriler [23] gibi değişkenlerde sapan değerler ortaya konmuştur. Bazı çalışmalarda ise istatistiksel yöntemlerle sıcaklık, yağış, bağıl nem [121], nehir debisi [122, 123] gibi değişkenlerde sapan değerler belirlenmiştir. Ama hangi istatistiksel yöntemler olduğu belirtilmemiştir. Çeyrekler aralığındaki gibi yüzdelerin kullanıldığı başka çalışmalar da mevcuttur [74, 124]. Yağışın göl seviyesinde sebep olduğu değişimi gösteren grafikteki sapan değerler görsel inceleme ile bulunmuştur [125]. Sapan debi değerleri için su seviyesi-debi eğrisinin modelleneni ve ölçüleni arasındaki farkın deneyime dayalı karar verilen bir değeri geçmesine göre sapan değer belirlenmiştir [126, 127]. Bir başka çalışmada uç (ekstrem) değerler için çizilen yağış -dönüş aralığı grafikleriyle sapan değer belirlenmiştir [26]. Maksimum nehir akım verilerindeki sapan değerler taşkın frekans eğrilerine dayalı kritik değerlerle ortaya konmuştur [128].

Derinlik tabanlı yöntemlerle yıllık maksimum akım [90], taşkın verisi [129, 134], deniz seviyesi basıncı, yüzey rüzgarının bölgesel ve meridyen bileşenleri, deniz yüzeyi sıcaklığı, yüzey hava sıcaklığı, gökyüzü bulutluluğu [130], hava kalitesinde incelenen ozon ve nitrojen dioksit [131], sıcaklık [132], yağış [113], nehir debisi [47, 133] gibi değişkenlerde sapan değerler belirlenmiştir.

Kümeleme yöntemi ile iklim değişikliği analizlerinde kullanılan bitki fenoloji verisi [95], elektrik sayaçlarındaki ölçüm verisi [135], hava sıcaklığı, buhar basıncı [136, 137], atmosfer basıncı, bağıl nem ve rüzgar hızı [137], yağış, güneşlenme süresi, güneş ışıması, kar suyu eşdeğeri [136]; kümeleme algoritmalarından DB-SCAN ile sıcaklık, yağış, basınç ve bulutluluk oranı [92], geofiziksel veriler [54]; başka bir kümeleme algoritması K-Ortalamalar yöntemiyle su ürünleri yetiştiriciliğindeki pH, çözünmüş oksijen ve sıcaklık [138] gibi değişkenlerde sapan değerler tespit edilmiştir.

İzolasyon ormanı yöntemi ile nehir su seviyesi [139], geofiziksel veriler [54], su ürünleri yetiştiriciliğindeki pH, çözünmüş oksijen ve sıcaklık [138] gibi değişkenlerde sapan değerler belirlenmiştir.

Temel bileşenler analiziyle taşkın verilerindeki [140] sapan değerler, dalgacık tabanlı yöntemle, yüzey modelleri üretmek için kullanılan havadan lazer tarama verisinde [141] sapan değerler tayin edilmiştir.

Verinin kendisini değil de kendisi dışındaki verileri inceleyerek sapan değeri belirleyen yöntemlerden biri olan olabilirlik fonksiyonu ile bazı iklimsel gözlemler [2], nehir debisi [30], aylık ortalama ozon miktarı [142], güneşlenme süresi [143]; olabilirlik fonksiyonu ve Bayes ağları ile yıllık maksimum yağışlar [144], rüzgar hızı [52]; Cook uzaklığı ile yağış [57]; lokal sapan değer faktörüyle (LOF) toprak tipini belirten uydu verisi [99], geofiziksel veriler [54], sıcaklık, basınç ve bağıl nem [145], su ürünleri yetiştiriciliğindeki pH, çözünmüş oksijen ve sıcaklık [138], debi değerleri [146] gibi değişkenlerde sapan değerler belirlenmiştir. Bazı çalışmalarda da yeni yöntemler önerilmiştir [70, 102, 122, 147].

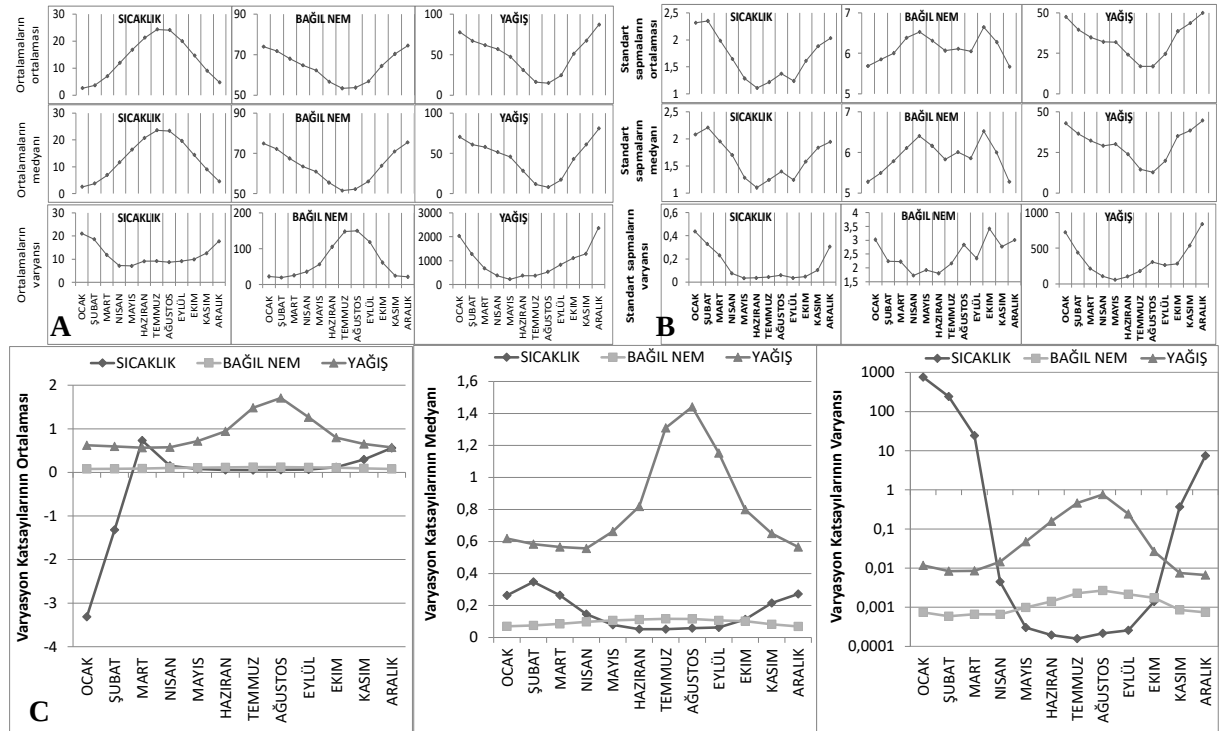
II. VERİLERİN ANALİZİ

Bu çalışmada Devlet Meteoroloji İşlerine (DMİ) bağlı Türkiye Meteorolojik Veri Arşiv ve Yönetim Sisteminden ve University of East Anglia (İngiltere) kurumunun genel ağ açık veri tabanından temin edilen ve Dede & Ağırlioğlu [148] çalışmasında da kullanılan 1926 – 2012 arası Türkiye genelinde 129 istasyonda aylık ortalama sıcaklık (°C), 128 istasyonda aylık ortalama bağıl nem (%) ve 128 istasyonda aylık toplam yağış (mm) olmak üzere 4620 adet dizi kullanılmıştır. Verilerin korelasyon özellikleri önceki çalışmada verilmiştir [149].

Şekil 1.A'daki ortalamaların ortalamaları normal olarak yazın yükselen sıcaklıklara, düşen bağıl nem ve yağışlara, kışın düşen sıcaklıklara, yükselen bağıl nem ve yağışlara işaret eder. 128 istasyonun her birinin standart

sapması her birinin ayrı ayrı 1926 – 2012 arası yayılımını gösterir. Şekil 1.B'deki ilk iki satır bu zamansal değişimin genel değerinin aylık olarak farklılaşmasını göstermektedir. En alt satır ise bu zamansal değişimin coğrafi değişiminin yani istasyondan istasyona değişiminin aylık olarak farklılaşmasını gösterir. Şekil 1.B'de görüldüğü gibi sıcaklık ve yağışta standart sapmalarının ortalaması birbirine benzer; yazın azalır, kışın artar. Yani 87 yıl boyunca yaz aylarında genelde yüksek sıcaklıklar ve düşük yağışlar görülür, kış aylarında ise yüksek veya düşük sıcaklıklar ve az veya çok yağışlar görülür. Standart sapmalarının varyansı ise sıcaklıkta ocak, şubat ve aralıkta yükselir. Nisan, mayıs, haziran, temmuz, ağustos, eylül ve ekimde en düşük seviyeye gelir. Bağıl nemde en az nisan, mayıs, haziran, ağustos, eylül, ekim, kasım, aralık ve ocakta en yüksek değerde olduğu görülür. Yağışta mart, nisan, mayıs, haziran ve temmuzda en düşükken ağustos, eylül ve ekimde ortalamasının altında, şubat ve kasımda ortalamasının üstünde, ocak ve aralıkta en yüksektir.

Standart sapmayı ve ortalamayı birlikte değerlendirmek için birim ortalama başına düşen standart sapma yani varyasyon katsayısı (Cv) alınmıştır. Bunun için Şekil 1.B'deki standart sapmaların ortalamaları Şekil 1.A'daki ortalamaların ortalamalarına bölünmüştür. Bu hem her ikisini basitçe değerlendirme açısından iyidir, hem de farklı toplumlar olan sıcaklık, bağıl nem ve yağış birbirleriyle karşılaştırma olanağı sunar. Bunları bütün olarak yorumlayabilmek için yani 12 ayın hepsini görebilecek şekle getirmek için her ay için ayrı ayrı olmak üzere 128 istasyonun ortalaması alınmıştır. Şekil 1.C'deki ilk grafik bunu göstermektedir. Ortalama aşırı değerlerden etkilenebilir diye medyanı da alınmıştır. Şekil 1.C'deki ikinci grafik de bunu göstermektedir. Yani ilk grafikteki tek bir değer 1926-2012 arası birim ortalama standart sapmaların tüm istasyonlardaki genel değeridir. Grafiğin tümü ise 1926-2012 arası birim ortalama standart sapmaların tüm istasyonlardaki genel değerlerini her bir ay için gösterir. Böylece bunların mevsimlik nasıl değiştiği görülebilir. Buna göre, yıllık değişimlerin genel coğrafi değeri sıcaklık için yazın daha az kışın daha fazladır. Bağıl nem ve yağış içinse bunun tam tersidir. Dolayısıyla bağıl nem ve yağışın histogramları yaz aylarında kış aylarına göre yayvandır yani 1926-2012 arası değişimler genel olarak yazın kışa göre daha fazladır. Şekil 1.C'deki üçüncü grafikte görüldüğü gibi, varyasyon katsayısının (Cv) varyansı için de aynı durum geçerlidir. Bu, aslında bir nevi ortalamanın sağlaması gibidir. Çünkü yıllık değişimlerin genel coğrafi değeri bağıl nem ve yağış için yazın daha fazla oluyorsa varyansta bir nevi farklılığı, çeşitliliği ifade ettiği için değişimin fazla olduğu aylarda çeşitliliğinde fazla olması uyumlu bir durumdur. Yani bağıl nem ve yağış için 1926-2012 arası değişimlerin istasyonlar arası çeşitliliği yazın kışa göre daha fazladır. Zaman içindeki artış ve azalışın dışında bir de bunlar miktar olarak birbirleriyle kıyaslanabilirler. Ortalama ve varyansta, bağıl nem ve sıcaklığın büyüklükleri birbirine yakındır. Yağış onlardan uzaktır.



Şekil 1. Tüm istasyonlardaki ortalama (A), tüm istasyonlardaki standart sapma (B), varyasyon katsayısı(C).

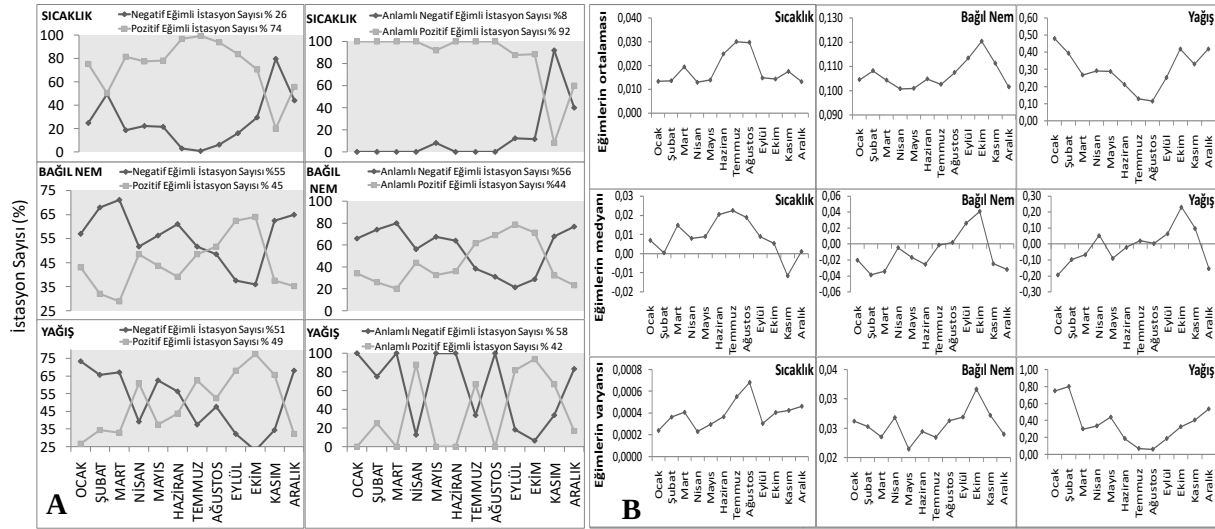
Şekil 2.A'da ilk sütunda tüm eğimler, ikinci sütunda sadece anlamlı eğimler dikkate alınmıştır. Bunlar eğimlerindeki artışın ve azalışın aylık değişimlerini gösterir. Sıcaklıkta pozitif eğimli istasyon sayısı negatif eğimli istasyon sayısına göre fazla, bağıl nem ve yağışta sonbahar hariç olmak üzere tam tersidir. Sonbaharda 1926'dan 2012'ye doğru istasyonların yarısından fazlasında artış eğilimi vardır. Sıcaklıktaki artış eğilimindeki istasyon sayıları

azalış eğilimindekilere göre oldukça fazladır. Bağıl nem ve yağıştaki azalış eğilimindeki istasyon sayıları ile artış eğilimindeki istasyon sayıları birbirine yakındır. Kısaca sıcaklık için eğimler yazın artar, kışın daha az artar. Bağıl nem ve yağış için eğimler sonbaharda artar, ilkbaharda ve kışın azalır. İstasyonların 1926-2012 arasındaki değişimini gösteren eğimler doğrusal regresyon denklemi Denklem (1) ile hesaplanmış; t istatistiğiyle, Denklem (2), kontrol edilmiştir. Bu bağıntılarda x bağımsız değişken y bağımlı değişkendir. Ayrıca σ , r ve n sırasıyla standart sapma, korelasyon katsayısı ve veri adetidir.

$$eğim = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2} \quad \begin{array}{l} x: \text{bağımsız değişken} \\ y: \text{bağımlı değişken} \end{array} \quad (1)$$

$$t_değeri = \frac{eğim \cdot \sigma_x \cdot \sqrt{n-1}}{\sigma_y \sqrt{(1-r^2) \left(\frac{n-1}{n-2}\right)}} \quad (2)$$

Tüm eğimleri pozitif kabul edip sadece büyüklüklerine bakıldığında Şekil 2.B’de sıcaklık için yazın eğimlerin değişimi ki bu değişim artış şeklindedir Şekil 2.A’da görüldüğü gibi, kışın eğimlerin değişimine göre ki bu da çoğunlukla artış şeklindedir, daha fazladır. Yağış için yazın eğimlerin değişimi ki bu değişim, biraz artış biraz azalış şeklindedir, kışın eğimlerin değişiminden yani azalışından daha azdır. Bağıl nem için sonbaharda eğimlerin değişimi yani artışı, ilkbahardaki eğimlerin değişiminden yani azalışından daha fazladır. Kısaca Şekil 2.B’de ilk iki satırda görüldüğü gibi sıcaklık için eğimlerin yazın artışı, kışın artışından fazladır. Yağış için tam tersi yani eğimlerin yazın artışı, kışın azalışından azdır. Bağıl nem ise ikisi arasında bir geçiş özelliğindedir. Sonbahardaki artış ilkbahardaki azalıştan fazladır. Basitçe eğimler genel olarak sıcaklık için yaz aylarında kış aylarına göre diktir, bağıl nem için sonbaharda ilkbahara göre diktir, yağış için kışta daha göre diktir. Yani 1926-2012 arası meydana gelen değişim sıcaklık için daha çok yazdan gelir, bağıl nem için daha çok sonbahardan gelir, yağış için bu değişim daha çok kıştan gelir. Şekil 2.B’de son satırda görülen eğimlerin varyansı bir nevi ortalamanın sağlaması gibidir. Çünkü yıllık değişimlerin genel coğrafi değeri sıcaklık ve bağıl nem için yazın daha fazla oluyorsa varyans da bir nevi farklılığı, çeşitliliği ifade ettiği için değişimin fazla olduğu aylarda çeşitliliğin de fazla olması uyumlu bir durumdur. Zaman içindeki artış ve azalışın dışında bir de miktar olarak bunlar birbirleriyle kıyaslanabilirler. Ortalama ve varyansta bağıl nem ve sıcaklığın büyüklükleri birbirine yakındır. Yağış onlardan uzaktır.



Şekil 2. Pozitif ve negatif eğimli istasyon sayıları (A), tüm istasyonlardaki eğim (B).

III. KULLANILAN YÖNTEM

Bu çalışmada kullanılan yöntem önceki çalışmada [149] yöntem bölümünde “histogram ve kutu grafiği yönteminin kullanılma biçimi” başlığı altında açıklanmıştır. Sapan değerleri tespit etmek için sıcaklık ve bağıl nemde $\mu \pm 3\sigma$ (μ : ortalama, σ : standart sapma) sınırlarıyla Z-Skor, yağışta $Q_{3,1} \pm 2IQR$, $Q_{3,1} \pm 3IQR$, $Q_{3,1} \pm 4IQR$ ve $Q_{3,1} \pm 5IQR$ ($Q_{3,1}$: üçüncü ve birinci çeyrek, $IQR: Q_3 - Q_1$ çeyrekler açıklığı) sınırlarıyla kutu grafiği yöntemi kullanılmıştır. Z-Skoru yöntemi, veri noktasının ortalamadan ne kadar uzakta olduğunu gösterir. Z-Skorlarının mutlak değerlerinin -3/+3 aralığı dışında kalanları sapan değer olarak kabul edilir. Tukey kutu grafiği yöntemi

olarak bilinen kutu grafiği yöntemi ise 75. yüzdalık dilim ve 25. yüzdalık dilim arasındaki aralığın iki katının 75. yüzdalık dilime eklenmesi ve 25. yüzdalık dilimden çıkarılmasıyla bulunan sınırların dışında kalanları sapan değer olarak kabul edilir. Kullanılan yöntemle yapılan ekler aşağıda kısaca açıklanmıştır.

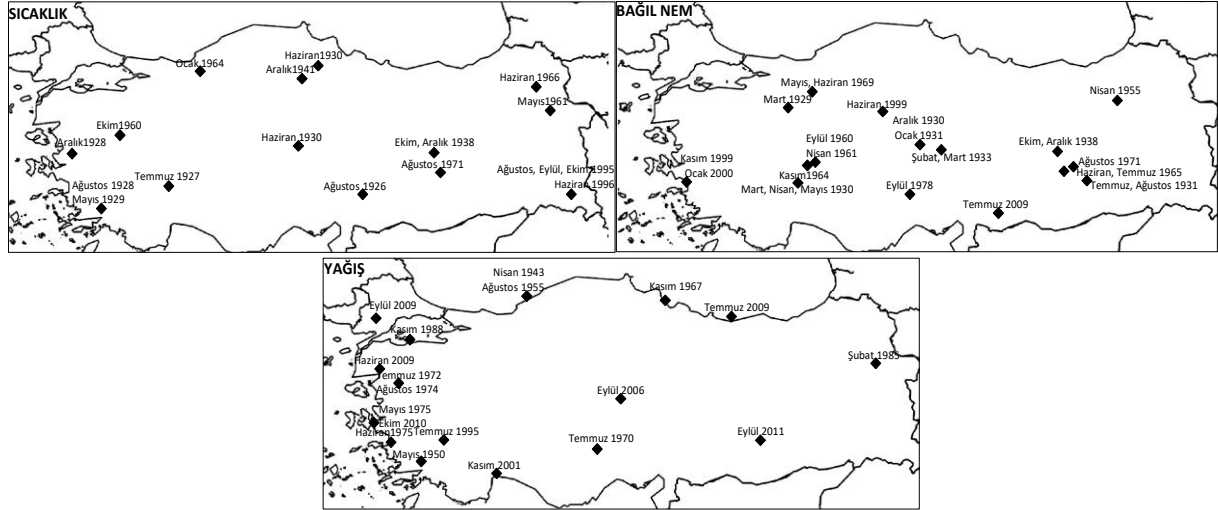
Sapan değerleri daha detaylı değerlendirebilmek için hem komşularıyla hem kendi istasyonunda sapsız olan değerler, yüksek güvenilirliğe sahip sapan değerler (YGSD) olarak kabul edilmiştir. YGSD'lere sadece komşularıyla sapan, ama kendi içinde sapsız değerler eklenince tüm sapan değerler olarak kabul edilmiştir.

Diziler genel olarak ortalama, varyans, çarpıklık, basıklık/sivrilik parametrelerine ve doğrusal düzenli (eğim), dalgalı düzenli (period), doğrusal düzensiz, dalgalı düzensiz bileşenlerine sahiptir. İstasyonların sahip olduğu bu parametreler ve bileşenler istasyondan istasyona farklılık gösterir. Genellikle algoritmalar uygulanmadan önce tüm özelliklerin hesaplamalara eşit şekilde katkıda bulunabilmesi için veriyi ölçeklendirmek tavsiye edilir. Ölçeklendirme, istasyonların sahip olduğu bu parametreleri ve bileşenleri ortak bir değere getirmek için ve dolayısıyla veri değerlerini belirli bir aralığa yeniden boyutlandırmak için yapılan ön işlemlerdir. Bu şekilde veri setindeki her değer aynı ölçekte olur, birbirleriyle karşılaştırılabilir hale gelir ve daha tutarlı sonuçlar üretildiği düşünülür. Çarpıklık ve sivrilik parametresi sadece yağışta vardır [149] ve ortak bir değere getirmek sapan değerlerin kaçırılmasına sebep olur. Doğrusal veya dalgalı düzensiz bileşenler ortak bir değere getirilemezler. Veriler aylık olarak ayrıldığı için dalgalı düzenli bileşen (period bileşeni) bulunmaz. Ortak bir değere getirilenler ortalama, standart sapma, ve eğimdir. Eğimler sıfıra getirilince ortalama da getirilmiş olur. O yüzden sadece standart sapma ve eğim ortak değerlere getirilmiştir. Bazı çalışmalarda sadece ortalama ortak bir değere getirilirken [150], bazı çalışmalarda sadece eğim [139], bazılarında ise standart sapma ve ortalama birlikte (standartizasyon) [30, 57, 107, 114, 117, 118, 151-153] ortak bir değere getirilmiştir. Bazı çalışmalarda ortalama yerine ortanca [75, 115] veya ilk çeyrek [54], standart sapma yerine çeyrekler açıklığı aralığı kullanılmıştır. Bu çalışmada 4 ön işlem denenmiştir: eğimleri sıfıra eşitlemek, standart sapmaları birleştirmek, her ikisini de eşitlemek, hiçbirini eşitlememek. Bu şekilde tek tek denenmesinin sebebi, istasyonların eğimleri veya standart sapmaları arasında fark çok ise bunları eşitlemek gerekliliğidir. Aksi takdirde sapan değer kaçırılabilir. Ama istasyonların eğimleri veya standart sapmaları arasında fark az ise bunları ortak bir değere getirmek avantajlı da olabilir dezavantajlı da olabilir. Veriyi inceleyip ona göre karar vermek gerekir. Bu durumların hepsi belli bir veri adedinin üzerinde böyledir. Belirli bir veri adedinin altında farklar azsa da çoksa da durum belirsizdir.

Belirlenen sapan değerler eşit önem seviyesine sahip olmadıklarından belli bir sıralamada olmalıdırlar. Bu sıralamanın nasıl yapılacağını araştırmacı tasarlamalıdır. Bunun için referansları belirlemek gerekir. Bunlar sırasıyla; sapma durumu, komşu güvenilirliği, sapma miktarı, korelasyon katsayısı, komşusuyla mesafesi ve sapan değerlerin yılıdır. Sapma durumuna göre sapan değerler iki gruba ayrılır. Bunlar kendi istasyonunda ve komşularıyla uyumlu olarak sapsız olanlar üst sırada, sadece komşularıyla uyumlu olarak sapsız olanlar bu grubun altında listelenir. Hem kendi istasyonunda sapsız olanlar hem komşularıyla uyumlu olarak sapsız olanların kendi istasyonundaki sapma miktarları yüksek olanları üst sırada listelenir. Eğer kendi istasyonunda eşit sapma miktarına sahipse yüksek güvenilirlikli gruplara mensup olan komşulara sahip sapan değerler daha üst sıradadır. Bunlar önceki çalışmada ayrıntılı açıklanmıştır [149]. Bu gruplarda da eşit güvenilirlik varsa komşularıyla gösterdiği sapma miktarı fazla olan üst sıradadır. Komşularıyla olan sapma miktarlarında da eşitlik varsa komşularıyla olan korelasyon katsayıları yüksek olanlar üst sıradadır. Bunda da eşitlik varsa komşularıyla olan mesafeleri düşük olanlar üst sıradadır. Bunda da eşitlik varsa daha eski yıllarda ölçüm ve kayıt tekniklerinin daha düşük hata oranının daha yüksek olabileceği düşünüldüğünden senesi küçük olan daha üst sıradadır. Bunun altında sadece komşularıyla uyumlu olarak sapsız olanların yüksek güvenilirlikli gruplara mensup olan komşulara sahip olanları vardır. Bu gruplarda eşit güvenilirlik varsa komşularıyla gösterdiği sapma miktarı fazla olan üst sıradadır. Bu gruplandırmalar bu şekilde devam eder.

IV. BULGULAR

Sıcaklık ve bağıl nem için Z-Skor ve yağış için kutu grafiğiyle tespit edilen sapan değerlerin yüksek güvenilirliğe sahip olanları Şekil 3'te gösterilmiştir. Bu değerler ve tüm sapan değerler üzerinden yapılan incelemeler sonucu ortaya çıkan bulgular dört başlık altında toplanmıştır. Bunlar sapan değeri tespit etmek için kullanılan sınırların, ön işlemlerin, komşuların ortaya çıkardığı ve sapan değerlerin kendisinin ortaya çıkardığı bulgulardır.



+

Şekil 3. Yüksek güvenilirliğe sahip sapan değerlerin (YGSD) coğrafi dağılımı.

A. Sapan Değer Belirlerken Kullanılan Sınırların Ortaya Çıkardığı Bulgular

İlk bulgu normal dağılımlı olmayan yağış için kullanılan $Q_{3,1} \pm 2IQR$ sınırının uyumluluğuyla ilgilidir. Tablo 1.A'da YGSD'e baktığımızda sıcaklık ve bağıl nemde bütün ön işlemlerde on binde 2 – 3 değer sınırlarının dışındadır. Yağışta da bütün ön işlemlerde on binde 1 – 2 değer $Q_{3,1} \pm 4IQR$ sınırlarının dışındadır. Yani sıcaklık ve bağıl nem değişkenleriyle benzerdir. $2IQR$, $3IQR$ ve $5IQR$ sınırlarında sırasıyla yaklaşık on binde 20, 7 ve 1 değer sınırlarının dışındadır. Tüm sapan değerlere baktığımızda sıcaklık değil ama bağıl nemde bütün ön işlemlerde on binde 3 – 4 değer sınırlarının dışındadır. Yağışta da bütün ön işlemlerde on binde 1 – 2 değer $Q_{3,1} \pm 4IQR$ sınırlarının dışındadır. Yani bağıl nem değişkenleriyle benzerdir. $2IQR$, $3IQR$ ve $5IQR$ sınırlarında sırasıyla yaklaşık on binde 30, 8 ve 0,1 değer sınırlarının dışındadır. Bu çalışmada diğer çalışmalarda sıklıkla kullanıldığı için $Q_{3,1} \pm 2IQR$ ile ve tüm değişkenlerin uyumlu olması açısından $Q_{3,1} \pm 4IQR$ ile değerlendirme yapılmıştır.

Diğer bir bulgu veriler normal dağılımlı olmadığında $\mu \pm x\sigma$ sınırı yerine kullanılan $Q_{3,1} \pm xIQR$ sınırının $\mu \pm x\sigma$ sınırını temsil edebilmesi ile ilgilidir. Sıcaklık ve bağıl nem değişkenlerinde $\mu \pm x\sigma$ ve $Q_{3,1} \pm xIQR$ sınırları kullanılarak bulunan sapan değerler Tablo 1.B'de görülmektedir. Dört ön işlem yöntemi için de denenmiştir. Bu sınırların yakalayabildikleri sapan değerler birbirleriyle birebir örtüşmezler. Ortak olarak gördükleri değerler vardır. Birinin sapan değer olarak gördüğü bir değeri, diğerinin sapan değer olarak görmediği durumlar da vardır. Sınırların birbirleriyle % 40 - 50 arası bir uyumsuzluğu vardır. Dolayısıyla birbirleri yerine kullandıklarında yarı yarıya birbirlerini temsil ederler.

Tablo1. Yağışta dört (A), sıcaklık ve bağıl nemde iki (B) farklı sınır için sapan değer adetleri.

A		Normal Sınırlar		Çeyrekler Sınırları				B
		(μ±3σ)		(IQ3,1 ± 5IQR)	(IQ3,1 ± 4IQR)	(IQ3,1 ± 3IQR)	(IQ3,1 ± 2IQR)	
		Bağıl						
YGSD		Sıcaklık	Nem	Yağış	Yağış	Yağış	Yağış	
		0,02%	0,03%	0,01%	0,02%	0,08%	0,25%	
	(İstasyon– Komşu)	14 adet	25 adet	8 adet	19 adet	70 adet	231 adet	
		0,02%	0,03%	0,01%	0,01%	0,07%	0,23%	
	E* (İstasyon– Komşu)	19 adet	26 adet	5 adet	13 adet	64 adet	208 adet	
		0,01%	0,02%	0,003%	0,01%	0,06%	0,22%	
	S** (İstasyon– Komşu)	11 adet	20 adet	3 adet	10 adet	54 adet	201 adet	
		0,02%	0,02%	0,001%	0,01%	0,05%	0,19%	
S(E(İstasyon– Komşu))	16 adet	19 adet	1 adet	8 adet	44 adet	175 adet		
Tüm Sapan Değerler		0,07%	0,04%	0,01%	0,02%	0,09%	0,36%	
	(İstasyon– Komşu)	66 adet	36 adet	8 adet	19 adet	84 adet	324 adet	
		0,07%	0,04%	0,01%	0,01%	0,08%	0,31%	
	E* (İstasyon– Komşu)	61 adet	33 adet	5 adet	13 adet	70 adet	283 adet	
		0,08%	0,04%	0,003%	0,02%	0,08%	0,34%	
	S** (İstasyon– Komşu)	70 adet	32 adet	3 adet	15 adet	68 adet	307 adet	
		0,07%	0,03%	0,002%	0,01%	0,06%	0,28%	
	S(E(İstasyon– Komşu))	64 adet	28 adet	2 adet	9 adet	55 adet	254 adet	
YGSD		(İstasyon – Komşu)	9	2	5	22	14	3
	E* (İstasyon – Komşu)	12	3	7	21	7	5	
	S** (İstasyon – Komşu)	8	2	3	15	6	5	
	S(E(İstasyon – Komşu))	11	3	5	15	4	4	
	Tüm Sapan Değerler	(İstasyon – Komşu)	56	44	10	26	21	8
		E* (İstasyon – Komşu)	55	33	6	23	13	10
		S** (İstasyon – Komşu)	56	34	14	22	10	10
		S(E(İstasyon – Komşu))	51	37	13	18	10	10

*E: eğimlerin sıfıra eşitlendiği durum.

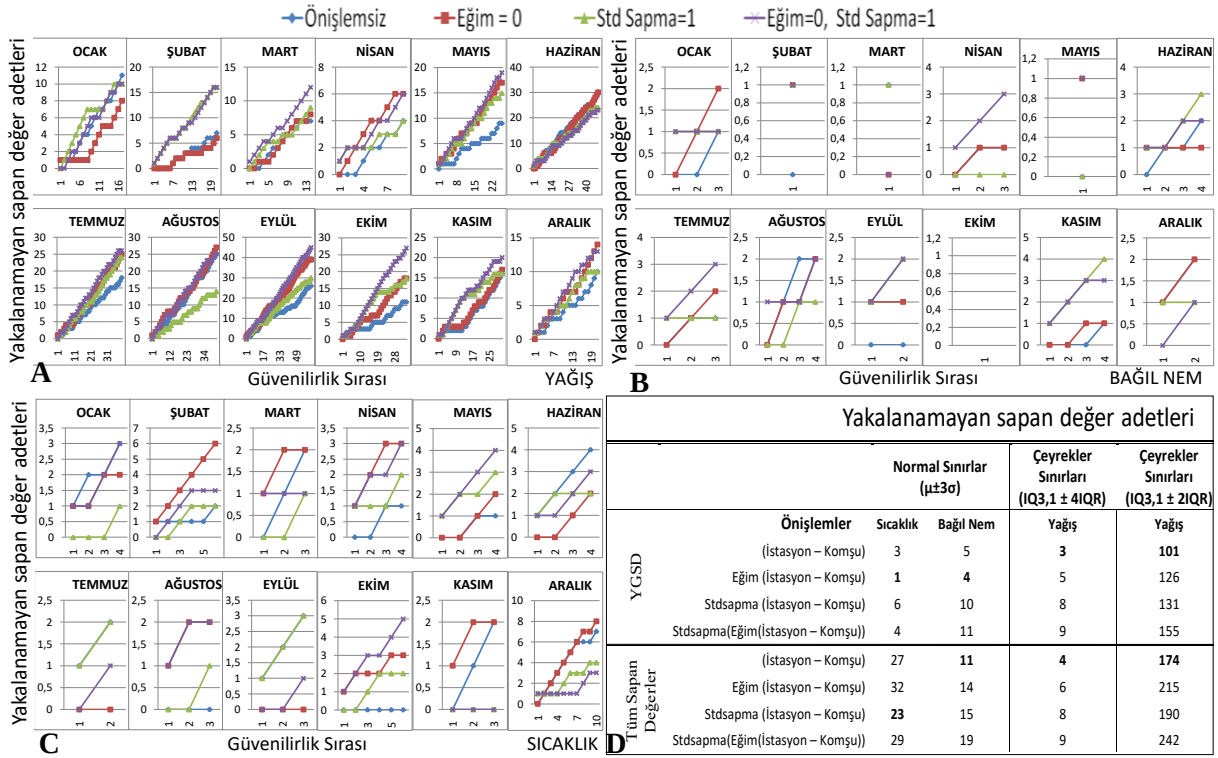
**S: standart sapmaların bire eşitlendiği durum.

*E: eğimlerin sıfıra eşitlendiği durum.

**S: standart sapmaların bire eşitlendiği durum.

B. Sapan Değer Belirlerken Kullanılan Ön İşlemlerin Ortaya Çıkardığı Bulgular

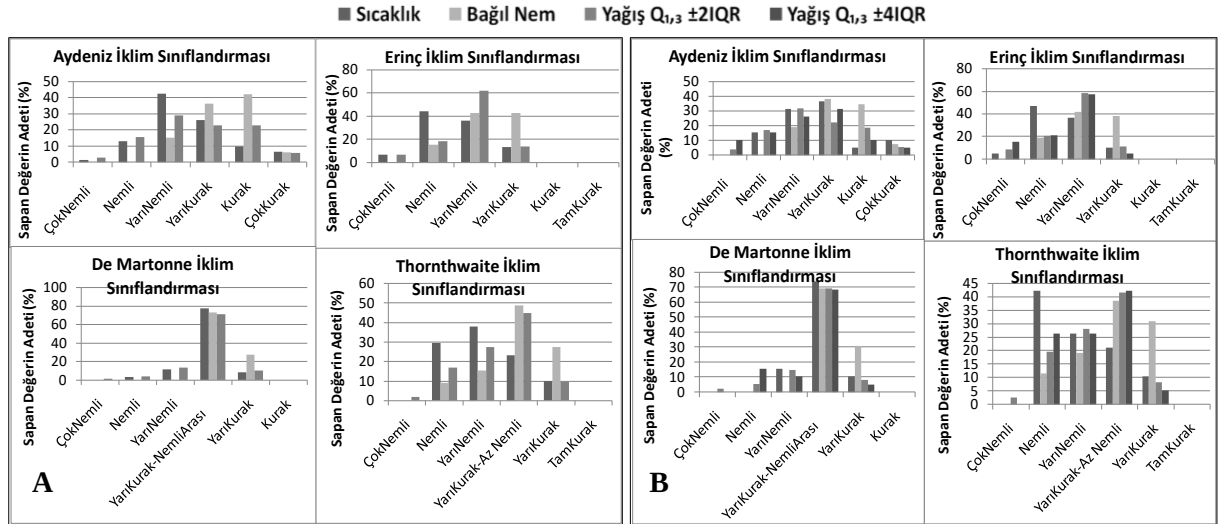
Yöntem bölümünde belirtilen dört ön işlemin etkinliğine karar vermek için her ön işlemde geçen verilerin ayrı ayrı sapan değerleri bulunmuştur. Ortak olmayan sapan değerler hangi ön işlemde daha fazla yakalanmışsa bu ön işlem daha verimli kabul edilmiştir. Bu değerlerin yakalanamadığı ön işlem gruplarının verimi daha düşük kabul edilmiştir. Bu yüzden farklı ön işlemlerde yakalanamamış sapan değer adetlerinin grafiği çizilmiştir. Şekil 4'te farklı ön işlemlerde yakalanamamış sapan değer adetleri görülmektedir. En alttaki çizgiler yakalanamamış sapan değerlerin en az olduğu ön işlemleri gösterir. Şekil 4.A'da yağış için mavi çizgi 12 ayın 8'inde aşağıdadır. Haziran ve ağustosta yeşil çizgi aşağıdadır. Yani yağış için yazın varyansları eşitlemek, diğer aylarda ön işlem yapmamak yani mavi çizgi ağır basmaktadır. Şekil 4.B'de bağıl nem için kırmızı çizginin ağırlıklı olarak daha aşağıda kaldığını söyleyebiliriz. Kış ayları için ise mavi çizgi aşağıdadır. Yani bağıl nem için kışın ön işlem yapmamak, diğer aylarda eğimleri eşitlemek ağır basmaktadır. Şekil 4.C'de sıcaklık için özellikle güvenilirlik sırasına göre başta olan değerler içinde kırmızı çizginin ağırlıklı olarak daha aşağıda kaldığını söyleyebiliriz. Ocak şubat mart ayları için yeşil çizgi daha aşağıdadır. Yani sıcaklık için kışın varyansları eşitlemek, diğer aylarda eğimleri eşitlemek ağır basmaktadır. Aylara göre değişen ön işlem seçmek yerine her değişken için bir tane ön işlem seçmek istenirse ağırlıklı olan aylardaki ön işlem alınır. Yağış için bu ön işlem, ön işlem yapmamaktır. Bağıl nem için 6 ay eğimleri eşitlemek 6 ay ön işlem yapmamak, sıcaklık için 5 ay eğimleri eşitlemek 5 ay ön işlem yapmamak baskın görüldüğünden ilgili ön işlemlerde yakalanamayan sapan değer adetlerinin en az nerede olduğuna bakılır. Şekil 4.D'de görüldüğü gibi YGSD'de sıcaklık ve bağıl nem için standart sapmaları eşitlememek eğimleri eşitlemek en az yakalanamayan sapan değer adetlerine sebep olur. Sapan değerlerin hepsinde sıcaklık için standart sapmaları eşitlemek ama eğimleri eşitlememek, bağıl nem için her ikisini de eşitlememek en az yakalanamayan sapan değer adetlerine sebep olur. Yüksek güvenilirlikli olması sebebiyle ön işlem olarak sıcaklık ve bağıl nem değişkeni için sadece eğimler eşitlenmesi, yağış değişkeni için ön işlem yapılmaması uygundur.



Şekil 4. Yağışta (A), bağıl nemde (B), sıcaklıkta (C) yakalanamamış aylık sapan değerler ve toplam adetleri (D).

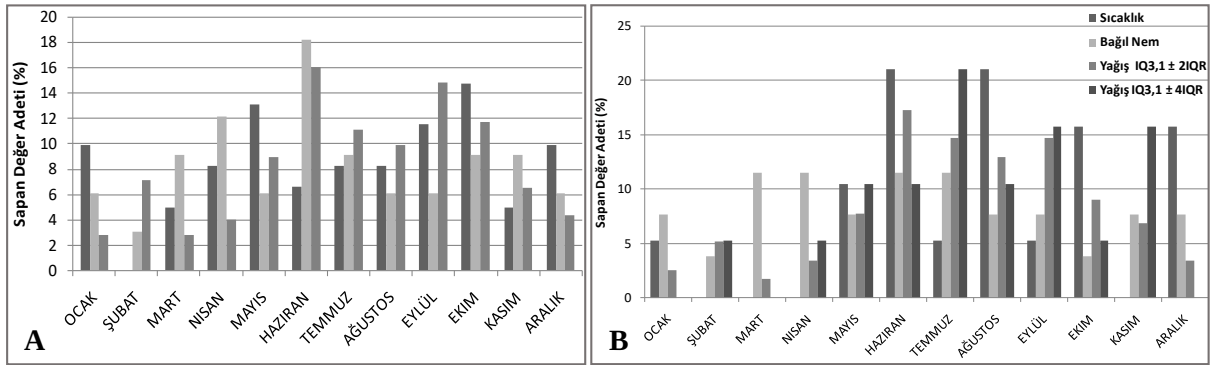
C. Belirlenen Sapan Değerlerin Ortaya Çıkardığı Bulgular

Şekil 5.A'da tüm sapan değerler sıcaklık için nemliye meyilli yarı nemli, yağış için yarı nemli, bağıl nem için yarı nemliye meyilli yarı kurak bölgelerde bulunur. Şekil 5.B'de YGSD'ler sıcaklık için nemli ve yarı nemli bölgelerde öncelikli, yarı kurak bölgelerde ikincil, bağıl nem için tam tersi yarı kurak bölgelerde öncelikli nemli ve yarı nemli bölgelerde ikincil, yağış için yarı nemli bölgelerde öncelikli yarı kurak ve nemli bölgelerde ikincil olarak görülmüştür. Özetle belirlenen sapan sıcaklık ve yağış değerleri bağıl neme göre biraz daha nemli bölgelerde çıkmıştır.



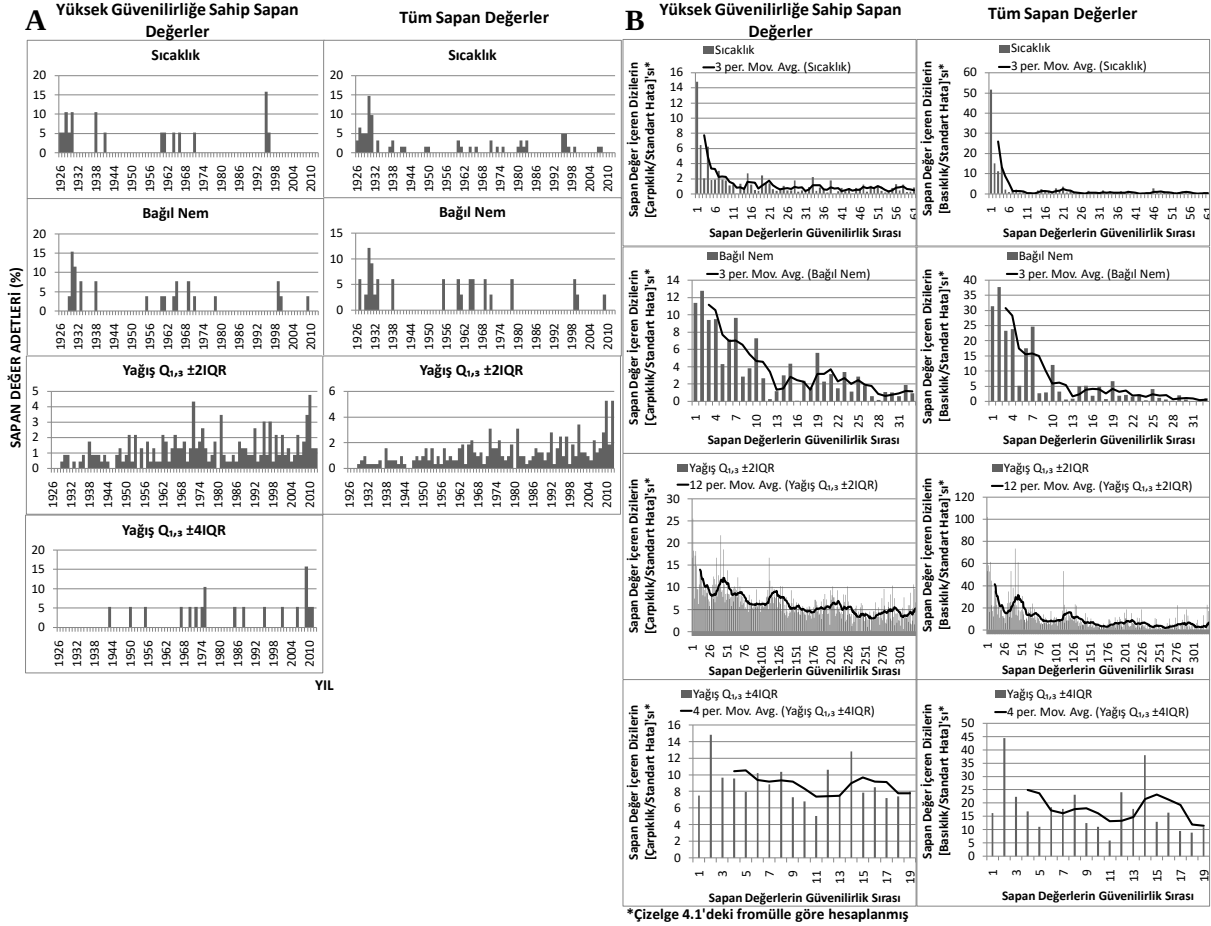
Şekil 5. İklim bölgelerine göre tüm sapan değerlerin (A) ve YGSD (B) dağılımı.

Şekil 6.A’da tüm sapan değerler genel olarak sıcaklıkta şubat hariç tüm yıla, bağlı nemde haziran ayında diğer aylara göre biraz daha fazla olmak üzere tüm yıla, yağışta kış ve ilkbaharda daha az olmak üzere tüm yıla yayılmıştır. Şekil 6.B’de YGSD’ler sıcaklıkta ağırlıklı olarak yılın son sekiz ayına, bağlı nemde genel olarak tüm yıla, yağışta kış ve ilkbaharda daha az olmak üzere genel olarak tüm yıla yayılmıştır. Ortak durum tüm mevsimlere yayılmakla birlikte, sapan sıcaklık ve yağış değerlerinin kış mevsiminde biraz daha az çıkmasıdır.



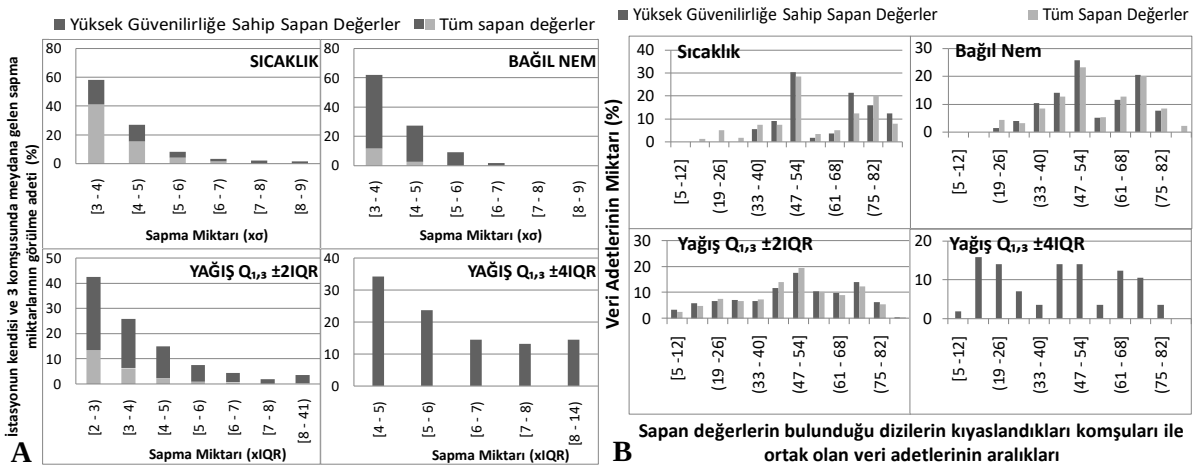
Şekil 6. Farklı aylara göre tüm sapan değerlerin (A) ve YGSD (B) dağılımı.

Şekil 7.A’da sapan sıcaklık ve bağlı nem değerleri yağışınkilere göre daha benzer yıllarda çıkmıştır. Sıcaklık ve bağlı nem değişkenlerinde ortak olarak gözlemlenen 3 adet sapan değeri (İst No:17201 Ekim ve Aralık 1938 ve İst No:17874 Ağustos 1971) mevcuttur. Şekil 7.B’de sapan değeri içeren dizilerden normal dağılımlı olmayanları çarpıklık ve basıklık katsayısına göre sıcaklık için %5-21, bağlı nem için %36-54, yağış için %87-96’dır.



Şekil 7. Yıllara göre sapan değerlerin dağılımı (A). Sapan değer içeren dizilerin normalite sıralaması (B).

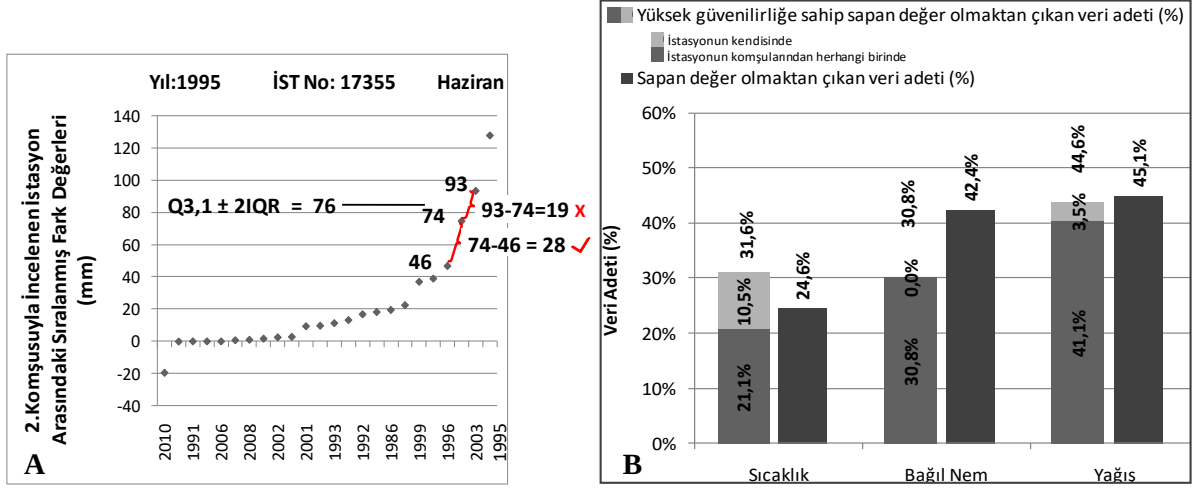
Şekil 8.A'da görüldüğü gibi $3\sigma - 7\sigma$ sapma miktarı ile bağıl nem, $3\sigma - 9\sigma$ sapma miktarı sıcaklık ve $2IQR - 4IQR$ sapma miktarı yağış değişkenine göre daha düşük sapma miktarları göstermiştir. Şekil 8.B'de görüldüğü gibi sapan değerleri içeren dizilerin veri adetleri çoğunlukla 47-54 adet arasındadır.



Şekil 8. Sapma miktarlarının dağılımı (A). Sapan değer içeren veri adetlerinin dağılımı (B).

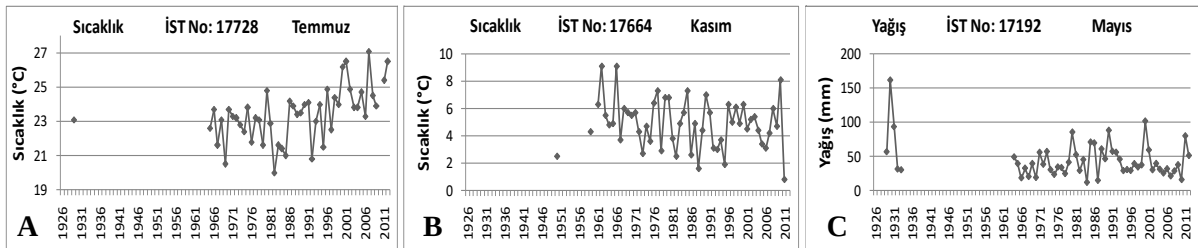
Belirlenen sapan değerler sapan olmayan değerlere bazen çok yakın çıkar. Çünkü hem sapan değerler ile sapan değer kendine en yakın sapan olmayan değerler arasındaki fark bu sınırlarla denetlenen bir durum değildir, hem de birden fazla sapan değerler sahte sapan veya maskelenmiş sapan değer problemlerine sebep olabilir. Şekil 9.A'da Osmaniye merkez istasyonunu gösteren bir grafik y-ekseni istasyonun değerlerinin 2. komşusunun değerleriyle olan küçükten büyüğe sıralanmış bir şekilde farkını gösterir. x-ekseni de bu sıralanışa denk gelen zamanları gösterir. Bu dizide son 3 değer diziden uzaklaştığı görülmektedir. Sondan 3. değerle sondan 4. değer

arasında 28 veri yoktur. Sonndan 2. değerle sonndan 3. değer arasında 19 veri yoktur. Sapan çıkma sınırı ise 76 mm'dir. Yani 93 mm değeri normal değildir ama 74 mm değeri normaldir. Ama baktığımızda 93mm'ye göre 74 mm en yakınındaki veriye daha uzaktır. Bu çalışmada kullanılan yöntemde en yakın 3 komşusundan biriyle bu durumun meydana gelmesi zaten değerın sapan çıkmasına engeldir. Kendi içinde böyle bir durumun meydana gelmesi yüksek güvenilirliğe sahip sapan değer çıkmasına engeldir. Sonuç olarak tüm sapan çıkan değerler incelendiğinde bu problemin görülme durumu %30 – 45 arasındadır. Yüksek güvenilirliğe sahip sapan değerler için Şekil 9.B'de soldaki sütunlar bunu göstermektedir. Sağdaki sütunlarda görüldüğü gibi %25 ila %45 arası veri de sapan değer olmaktan çıkar.



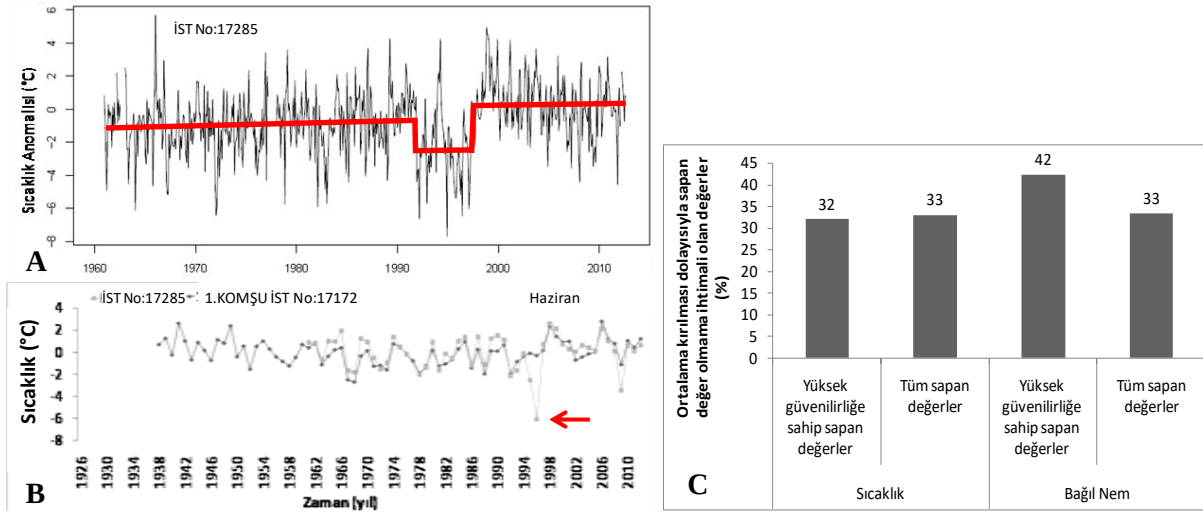
Şekil 9. 17355 no.lu istasyonun komşusuyla farkı (A), sapan belirlenen ama sapan olmayabilecek veri adeti (B).

Eksik veri tamamlama hesaplarında yararlanılabilecek büyük bir zaman diliminde tek bir veri olarak duran verileri sapan değer olarak almamak bir avantaj olabilir. Fakat YGSD'in bilgisinin eksik veri tahmininde kullanılması ihtimali düşüktür. Fakat yüksek güvenilirliğe sahip olmayan sapan değerler eksik verilere yakın çıkarsa bunun bir dezavantaj olacağını düşünülebilir. Çünkü bu sapan değer zaten çok fazla sapmadığı için veri bilgisi önemli olabilir ve eksik veri tahmininde yararlanılabilir. Şekil 10.A'da Ankara Polatlı istasyonu (No:17728) 1929 yılı temmuz ayı için 23°C'yi kendi içinde değil ama komşuları ile uyumsuzluğundan sapan değer olarak göstermiştir. Fakat 1930'daki değerden sonra 1965'e kadar veri kaydı yoktur. 1925 – 1965 arası veriyle ilgili fikir verebilecek tek değer budur. Aynı şekilde Şekil 10.B'de Ankara Kızılcahamam istasyonunda da (No:16664) 1952 yılı kasım ayı için 2,6°C'yi sapan değer göstermiştir. Ama 1925 – 1959 arası veriyle ilgili fikir verebilecek tek değer budur. Gene benzer şekilde Şekil 10.C'de Aksaray merkez istasyonunda da (No:17192) 1930 yılı mayıs ayı için 162 mm yağışı sapan değer göstermiştir. Burada da 1925 – 1964 arası veriyle ilgili fikir verebilecek 5 değerden biridir.



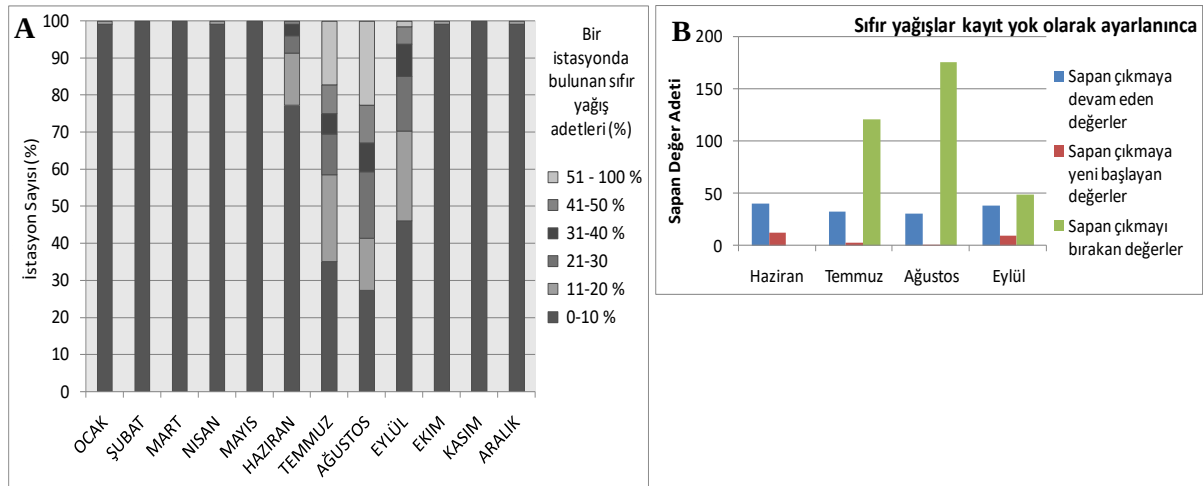
Şekil 10. 17728 no.lu istasyonun verisi (A), 17664 no.lu istasyonun verisi (B), 17192 no.lu istasyonun verisi (C).

Yıllık verilerde RHtestV3 yazılımı ile yapılan incelemelere göre sıcaklık için 86, bağıl nem için 115 ve yağış için 27 istasyonda ortalama kırımlar mevcuttur [154]. Bunlar aylık verilerde sapan değer olarak yakalanabilir. Şekil 11.A'daki Hakkâri merkez istasyonunda (No:17285) 1995-96 arası ortalama kırımla görülmektedir. Bu istasyonda 1995 yılının ağustos ayından 1996 yılının haziran ayına kadar 6 değer sapan olarak dışarıda kalarak sapan değer gözükmemektedir. Şekil 11.B'deki haziran ayına ait grafikte 1996 değeri bu sapanlardan biridir. Sıcaklık için olduğu gibi bağıl nem içinde böyle durumlar vardır. Ama yağış için ortalama kırımları olsa da sapan çıkan değerlerle paralellik göstermez. Şekil 11.C'de görüldüğü gibi belirlenen sapan sıcaklık değerlerinin % 33'ü RHtestV3 ile uyumlu olarak ortalama kırımla olmasından şüphelenilen değerlerdir. YGSD için bu %32'dir. Belirlenen sapan bağıl nem değerlerinin % 33'ü RHtestV3 ile uyumlu olarak ortalama kırımla olmasından şüphelenilen değerlerdir. YGSD için bu %42'dir



Şekil 11.17285 no.lu istasyonun yıllık verisi (A), 17285 no.lu istasyonun ve komşusunun aylık verisi (B), sapan değerlerin ortalama kılınmasına denk gelenlerinin miktarı (C).

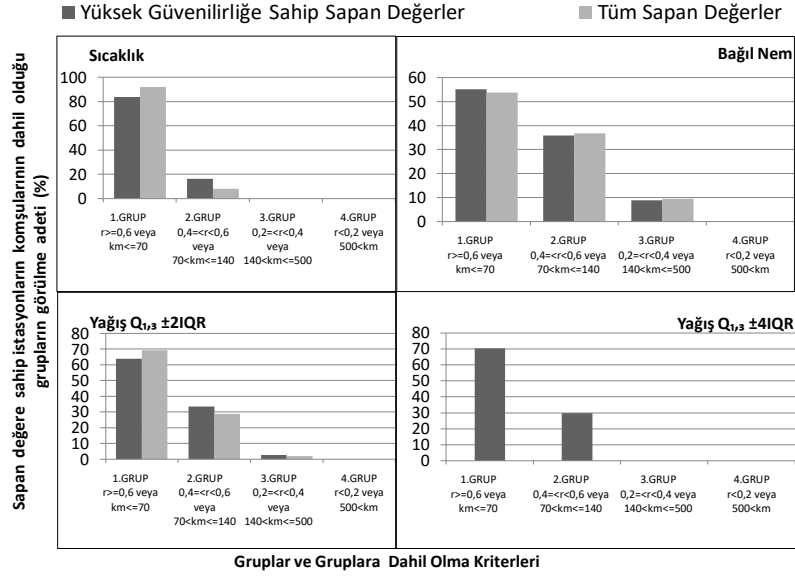
Sıfır yağış kayıtlarının fazlaca olduğu aylarda sapan değerlerin tespit edilmesi daha güçtür. Çünkü nadir yağışlı durumlar çoğunlukla sapan değer olarak gözükmemekte Şekil 12.A'da görüldüğü gibi haziran, temmuz, ağustos ve eylül hariç diğer aylarda istasyonların sıfır yağış adeti %10'un altındadır. Temmuz ve ağustosta yarısından fazlası sıfır yağış içeren istasyonlar mevcut istasyonların %20'sidir. Eylülde istasyonların yarısında, haziranda ise sadece % 20'sinde, sıfır yağış miktarı %10'un üstündedir. Dolayısıyla Eylül ayı yapısı itibarıyla haziran ayından daha çok temmuz ve ağustosa benzerdir. Sıfır yağışların kayıt yok yapılmasının bir değişiklik ortaya çıkarıp çıkarmayacağını anlaşılmaması amacıyla, bu 4 ay için sıfır yağışlar kayıt yok yapılmıştır. Sonuç olarak Şekil 12.B'de görüldüğü gibi önceden sapan çıkan değerlerin hepsi haziranda halen sapan çıkmaya devam etmiş, eylülde bu değer %44'e gerilemiş, fakat esas gerileme ağustos (%15) ve temmuzda (%21) olmuştur.



Şekil 12. Sıfır yağış adetleri (A), sıfır yağışlar kayıt yok yapılınca değişen sapan değer miktarı (B).

D. Belirlenen Sapan Değerlerin Komşularının Ortaya Çıkardığı Bulgular

Belirlenen sapan değerlerin sapan çıkmasına işaret eden komşular Dede & Ağırlioğlu çalışmasında [149] açıklanan güvenilirlik değerlendirmesine tabi tutulduğunda, Şekil 13'de görüldüğü üzere, düşük güvenilirlikli gruplar (3.ve 4.Grup) sıcaklık değişkeni için yoktur, bağıl nem değişkeni için %10 ve yağış değişkeni için $Q_{3,1} \pm 2IQR$ sınırlarına göre %2 görülmüştür, $Q_{3,1} \pm 4IQR$ sınırlarına göre yoktur.



Şekil 13. Sapan değere sahip istasyonların komşularının güvenirliliğini belirten grupların görülme sıklığı.

V. TARTIŞMA

Birçok çalışmada, yağış için $Q_{3,1} \pm 1,5IQR$ veya $Q_{3,1} \pm 3IQR$ sınırları kullanılmasına karşın [62, 100-104, 110-114], sıcaklık ve bağıl nemle uyumlu olabilmesi için sınırları $Q_{3,1} \pm 4IQR$ almak gerekir. Yağış normal dağılımlı olmadığı ve normal hale getirme dönüşümü yapmanın sapan değerleri normal sınırlara geriletmeye ve/veya veride bulunmayan başka sapan değerler oluşmasına sebep olacağından $Q_{3,1} \pm xIQR$ sınırlarıyla incelenmesi, $\mu \pm \sigma$ sınırlarıyla arasında uyumsuzluk yaratmıştır.

En uygununu seçebilmek için ön işlemler arasında karşılaştırma yapılması yada aylara göre değişen ön işlemler kullanılması fazladan işlem yüküne sebep olduğu için zayıf bir yön olarak düşünülebilir.

Sapan değerlerin eksik verilere ve sapan olmayan değerlere yakın çıkması yada ortalama kırılmalarıyla aynı anda ortaya çıkması yöntemin güvenirliliğini düşürür. Şekil 12'deki aylardan hangilerinde değişiklik yapılacağını belirlemek için Şekil 12. A ve B'nin çizilmesi gerekir. Ve sonucunda sıfır yağışları kayıt yok durumuna çevirmek gerekir. Tüm bunlar ikinci bir işlem yükü çıkarmaktadır.

Bahar ayları gibi geçiş mevsimleri ve yarı kurak, yarı nemli gibi geçiş bölgeleri diğer kararlı durumlara göre daha düzensiz olurlar. Sapan değerler kararlı bir durumla düzensiz bir durumun bir araya gelmesiyle ortaya çıkarken, iki kararsız durumun bir araya geldiği zamanlarda pek görülmemiştir. Şekil 5'te görüldüğü gibi sıcaklık ve bağıl nem normal dağılımlı değişkenler olarak geçiş bölgeleri olan yarı kurak ve yarı nemi bölgelerde daha çok sapan değer çıkarmışken, çarpık dağılımlı olan yağış değişkeni nemli bölgelerde daha çok sapan değer çıkarmıştır. Şekil 6'te görüldüğü gibi çarpık olan yağış verileri ise geçiş mevsimi olan ilkbaharda pek sapan değer göstermemiştir.

Şekil 7.B'deki verilere göre sapan değer diziyi tek başına çarpık dağılımlı yapamaz ama dizinin normallliğini azaltabilir. Bu durumu desteklemeyen bir çalışma olmasına rağmen [74] destekleyen [2, 24, 155-157] çalışmalar çoğunluktadır.

Tablo 1.A'da ve Şekil 8'deki verilere dayanarak sapan değer adeti fazlaysa sapma miktarının az, sapan değer adeti az ise sapma miktarının fazla olduğu yorumu yapılır. Bağıl nem, sıcaklık ve yağış değişkenine göre daha düşük sapma miktarları göstermiştir. Fakat YGSD'ler dikkate alındığında bağıl nem diğer iki değişkene göre daha çok sapan değer barındırır.

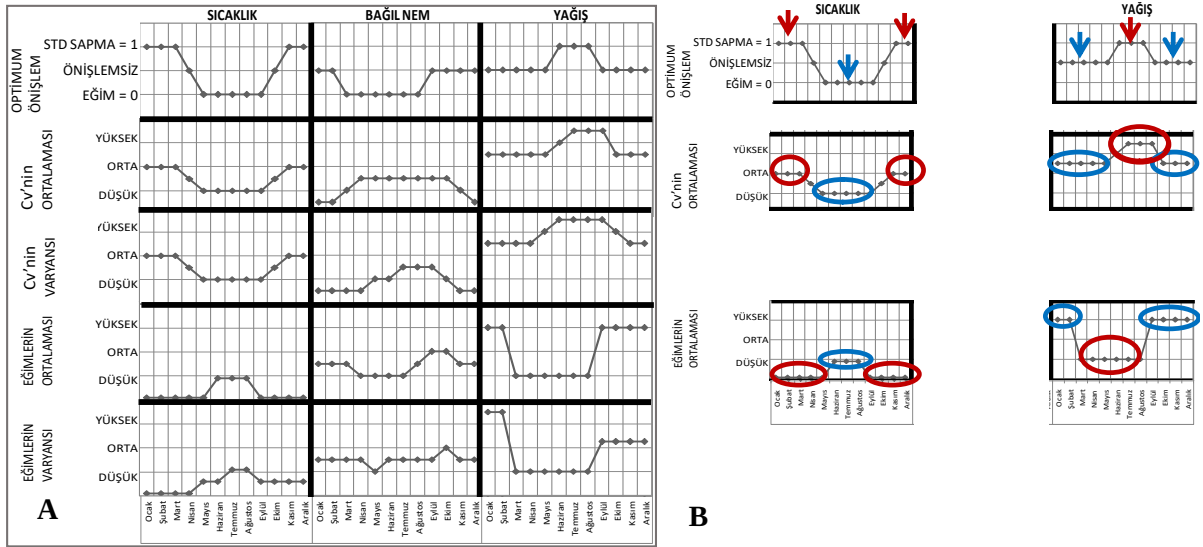
Şekil 14.A'da optimum ön işlemlerin basitleştirilmiş grafiklerinde görüldüğü gibi bağıl nem 6 ay yağışa, 5 ay sıcaklığa benzeyerek bir geçiş özelliği göstermektedir. Bu geçiş özelliği verinin özellikleriyle de uyumludur. Şekil 1.B'de yaz ve kış değişen sıcaklık ve yağışa kıyasla bağıl nem bahar aylarında değişim göstermiştir. Şekil 2.B'de görüldüğü gibi sıcaklıkta pozitif bir eğim hakim olduğundan bunun sıcaklığın daha fazla olduğu yaz aylarında olması normaldir. Yağışta da değişimin daha çok kışın olması yağışlar kışın olduğu için normaldir. Bağıl nemde en kuvvetli değişimlerin sonbaharda olması bağıl nemin sıcaklık ve yağış arasında bir geçiş özelliğinde olmasından kaynakladığı düşünülebilir. Miktar olarak da bağıl nemin eğimleri sıcaklık ve yağış arasındadır. Bunların haricinde önceki çalışmada [149] görüldüğü üzere sıcaklık ve yağışta bir istasyonun diğer istasyonlarla

korelasyonu yazın kışa göre daha düşüktür. Bağlı nemde ise bahar aylarında daha fazladır. Ek olarak mesafe arttıkça azalan korelasyonun azalışı genelde sıcaklık için kışın yaza göre, yağış için yazın kışa göre daha düşüktür. Bağlı nem için yarı yarıya her iki durum da geçerlidir [149]. Son olarak çarpıklık/basıklık sınamasında %99,9 güven seviyesine göre istasyonların sıcaklık için % 0 – 5'i, bağlı nem için % 1 – 11'i ve yağış için % 9 – 91'i çarpık veya basıktır [148]. Burada da bağlı nem sıcaklık ve yağış arasında kalmaktadır.

Şekil 7.A'da yıllara göre sapan değer miktarlarında artış görülmektedir. Bu sapan değerler, gerçekte var olmayan bir durumu yansıtan, veri setine yanlış bir şekilde dahil edilmiş değerler olabileceği gibi gerçekten olmuş olaylar da olabilir. Şekil 7.A'da kısmen sıcaklık için ama net olarak yağış için günümüze yaklaştıkça sapan değerlerin miktarındaki artış iklim değişikliği ile ilişkilendirilebilir. İklim değişikliği nedeniyle daha sık ve şiddetli hale gelen aşırı hava olayları bu artışların sebebi olarak yorumlanabilir.

Şekil 6'te görüldüğü gibi sapan sıcaklık ve yağış değerlerinin kış mevsiminde daha az çıkması Şekil 1.B'nin ilk iki satırındaki grafiklerle uyumludur. Çünkü istasyonların standart sapmalarının ortalamasını almak hem tüm coğrafya hakkında genel bir fikir verir hem de bölgesel farklılıkları kapatır ve değişkenlerin birbiriyle olan farklarını öne çıkarır. Böylelikle standart sapmanın sıcaklık ve yağış değişkenlerinde aynı aylık seyri takip ettiği görülmüş olur. Mevsim olarak ortak hareket eden değişkenlerin sapan değerlerinin de mevsim olarak ortak hareket ettiği düşünülebilir.

Şekil 5'te görüldüğü gibi sapan sıcaklık ve yağış değerlerinin bağlı neme göre daha nemli bölgelerde çıkması Şekil 1.A'nın ve Şekil 1.B'nin son satırındaki grafiklerle uyumludur. Çünkü istasyonların ortalamalarının ve standart sapmalarının varyansını almak istasyonların bölgesel olarak ne kadar benzer olduklarının anlaşılmasını sağlar. Böylelikle sıcaklık ve yağış değişkenlerinde istasyonların bölgesel olarak benzediği görülmüş olur. Bölgesel olarak benzer değişkenlerin sapan değerlerinin de bölgesel olarak ortak hareket ettiği düşünülebilir.



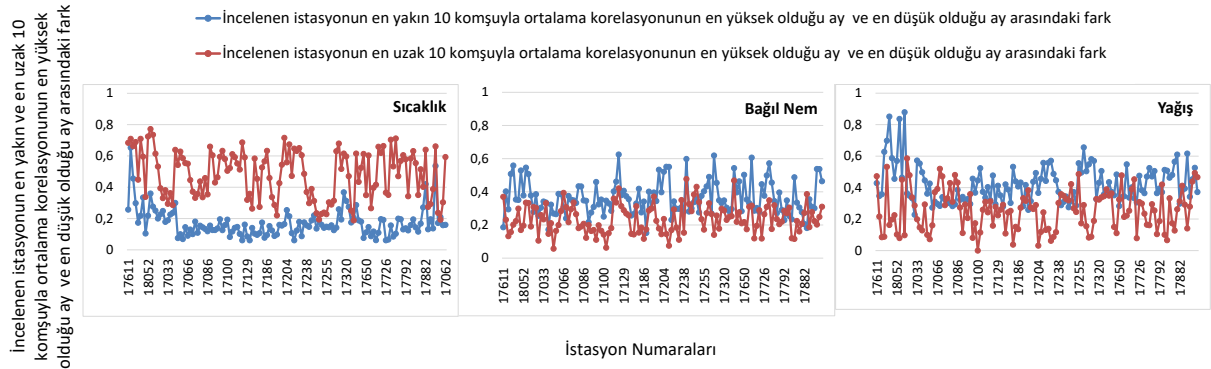
Şekil 14. Ön işlem, Cv ve eğimin basitleştirilmiş grafikleri (A), Cv'nin ve eğimin özellikleri (B).

Sapan sıcaklık ve yağış değerleri kışın benzerlik göstermesine rağmen sıcaklık ve bağlı nemde aynı senede çıkan 3 sapan değer görülmesi bu iki değişkenin istasyonlardaki ölçüm aletlerinin benzerliğine bağlı olabilir. Sıcaklık ve bağlı nem aynı kuru termometreye veya aynı sensöre bağlı olarak ölçülür. Yağış değerleri ise plüviyometreye ölçüldüğü için diğer ikisinden bağımsızdır [158]. Bu durum Şekil 1.C ile de uyumlu olabilir. Çünkü sıcaklık ve bağlı nemin birim ortalama başına düşen standart sapması miktar olarak birbirine yakındır. Bu da ortalamadan sapma miktarlarının yakın olduğunu gösterir. Sapan değer de ortalamadan bir sapma olduğundan uyumlu olması beklenebilir.

Şekil 14.A optimum ön işlemin, birim ortalama başına düşen standart sapmanın ve eğimin aylık değişimlerini belirten basitleştirilmiş grafikleri göstermektedir. Şekil 14.B'de varyasyon katsayısı (Cv) ve eğimin ortalamasına baktığımızda, varyansın yüksek, eğimin düşük olduğu durumlarda optimum ön işlem sıcaklıkta ve yağışta varyans eşitleme ön işlemi olarak bulunmuştur. Varyansın düşük, eğimin yüksek olduğu durumlarda ise sıcaklıkta eğim eşitleme ön işlemi olarak bulunmuştur. Dolayısıyla yükselen özellik, istasyonlar arasında eşitlendiğinde yakalanan sapan değerde artış olabilir. Varyans parametresi, eğim bileşeninden daha fazla bu duruma uyar.

Şekil 14.A'da ikinci satırda görülen birim ortalama başına düşen standart sapma yükseliş seyrindeyken ve Dede & Ağırlioğlu çalışmasındaki [149] Şekil 5'te mesafe arttıkça korelasyonun azalışı düşüş seyrindeyken Şekil 14.A'da ilk satırdaki optimum ön işlem sıcaklık ve yağış için standart sapmanın eşitlenmesidir. Yani ortalamadan sapışlar artınca istasyonlar birbirlerine daha çok benzer ve istasyonları aynı standart sapmaya getirmek daha çok sapan değer yakalanmasını sağlar.

Dede & Ağırlioğlu çalışmasındaki [149] Şekil 5'te mesafe arttıkça korelasyonun azalışı yükseliş seyrinden düşüş seyrine geçmesinin sebebi sıcaklıkta uzak korelasyonların daha da düşmesi, bağıl nemde ve yağışta yakın korelasyonların daha da yükselmesidir. Yani mevsimden mevsime sıcaklıkta yakın mesafedeki korelasyonlar az değişim gösterirken bağıl nem ve yağışta yakın mesafedeki korelasyonlar çok değişim gösterir. Başka bir deyişle bağıl nem ve yağış kısa mesafede çok değişirken sıcaklık kısa mesafelerde çok değişmez. Şekil 15'de görülen bağıl nem ve yağışta yakın korelasyonlardaki mevsimsel farklılığın sıcaklıktakinden daha yüksek olması bu durumu destekler. Tüm bunlar Şekil 13'deki sapan değer içeren istasyonların komşularının bağıl nem ve yağışta daha düşük güvenilirlikte olduğunu açıklayabilir.



Şekil 15. Tüm istasyonlarda incelenen istasyonun en yakın ve en uzak 10 komşuyla korelasyonunun en yüksek olduğu ay ve en düşük olduğu ay arasındaki fark.

VI. SONUÇ

$Q_{3,1} \pm xIQR$ sınırları ile $\mu \pm x\sigma$ sınırları arasında uyumsuzluk olmasına rağmen, $Q_{3,1} \pm 4IQR$ sınırlarının kullanılması daha tutarlı ve karşılaştırılabilir sonuçlar elde edilmesine yardımcı olur. Uyumlu sınırların kullanılması, sonuçların güvenilirliğini artırır.

En uygun ön işlemi seçmek için karşılaştırma yapılması veri işleme sürecine ek bir karmaşıklık ve zaman alıcı adım ekler. Başka bir ek işlem sıfır yağışları “kayıt yok” durumuna çevirmektir. Bunlar yöntemin etkinliğini azaltabilir ve pratik uygulanabilirliğini zorlaştıran bir faktör olarak görülebilir. Su kaynakları yönetiminde, verilerin hızlı ve verimli bir şekilde işlenmesi önemlidir. Ek işlem yükü, bu süreci yavaşlatabilir ve su yönetimi kararlarının alınmasını geciktirebilir.

Sapan değerlerin eksik verilere, sapan olmayan değerlere ya da ortalama kırılmalarına yakın çıkmasının, veri analizi yönteminin güvenilirliğini düşürür ve su yönetimi stratejilerinin doğruluğunu etkileyebilir.

Normal dağılımlı sıcaklık ve bağıl nem, yarı kurak ve yarı nemli bölgelerde daha fazla sapan değer üretirken, çarpık dağılımlı yağış değişkeninin nemli bölgelerde daha fazla sapan değer üretmesi, sapan değer üretme eğilimlerinde görünürde bölgeler ve dağılımların daha derinde kararlı ve kararsız durumların kombinasyonunun rol oynadığını gösterebilir. Su kaynakları yönetiminde bu kombinasyonların izlenmesi göz önünde bulundurulmalıdır.

Sapan değerlerin, veri dağılımının simetrisini bozmasa bile, birçok istatistiksel analizin geçerliliği için önemli olan normal dağılıma uygunluğunu olumsuz etkilemesi, analizlerin sonuçlarını etkileyebilir. Bu durum, su kaynakları yönetiminde daha doğru ve güvenilir kararların alınmasını geciktirebilir.

Değişkenlerde sapan değer adeti fazla ise, sapma miktarının az olduğu; sapan değer adeti az ise sapma miktarının fazla olduğu gözlemi, su kaynakları analizlerine farklı bir veri yorumlama stratejisi sunabilir. Hangi parametrelerin daha dikkatle izlenmesi gerektiğini belirlemek için kullanılabilir.

Bağıl nemin, hem sıcaklığa hem de yağışa benzer özellikler sergilemesi, dolayısıyla sıcaklık ve yağış arasında bir geçiş özelliği göstermesi, su kaynakları yönetimi açısından bu iki farklı değişken arasındaki bağlantıyı anlamada ve etkileşimi analiz etmede kullanışlı bir gösterge olabilir.

Sapan değerlerin zamanla artışı, aşırı hava olaylarının sıklığındaki artış olarak da düşünülebilir. Bu, iklim değişikliğinin tehlikeli etkilerinden biridir ve hem doğal ekosistemler hem de insan toplulukları üzerinde büyük bir baskı oluşturur. Öngörülemeyen hava koşulları, ekonomik kayıplara, altyapı hasarlarına ve insanların yaşamlarını tehlikeye atabilecek krizlere yol açabilir.

Kışın atmosferik koşulların daha değişken ve dengesiz olması, sıcaklık ve yağışta daha geniş bir değişime yol açar. Bu değişimler genellikle birbirini dengeler ve uç noktalar daha az belirgin olur. Yazın atmosfer daha stabil ve homojen olduğundan, sıcaklık ve yağışta daha az değişkenlik beklenir. Ancak bu stabilite, aynı zamanda ekstrem olayların daha belirgin olmasına yol açabilir. Bu durum iklim değişikliği ile bağlantılı olabilir. İklim değişikliği, farklı hava olaylarının birbirini dengelemesine neden olabilir. Örneğin, bir bölgeye gelen aşırı soğuk hava kütlesi kısa süre sonra bir sıcak hava dalgası ile dengelenebilir, bu da uzun vadeli verideki uç noktaların daha az belirgin olmasına yol açabilir. Küresel ısınma, özellikle yaz aylarında ortalama sıcaklıkların artmasına ve bu nedenle daha sık ekstrem sıcaklık olaylarının görülmesine neden olur. Yazın atmosfer stabil olduğunda bile, ani ve şiddetli sağanak yağışlar meydana gelebilir. İklim değişikliği, atmosferdeki su buharı miktarını artırarak bu tür olayların sıklığını ve şiddetini artırabilir. Stabil hava koşulları altında bu tür ekstrem yağış olayları, uç noktaların belirginleşmesine yol açabilir.

Kışın kutupsal soğuk hava kütleleri, Akdeniz üzerinden gelen siklonlar ve Balkanlar üzerinden gelen soğuk hava dalgaları gibi farklı hava olayları, farklı bölgelerde farklı sıcaklık ve yağış koşullarına neden olabilir. Yazın Azor yüksek basıncı gibi yüksek basınç sistemleri, hava koşullarını stabil ve homojen hale getirebilir. Yüksek basınç sistemleri geniş alanlarda benzer sıcaklık ve yağış koşullarına yol açabilir, bu da yazın bölgeler arasındaki benzerlikleri artırabilir. Sapan sıcaklık ve yağış değerlerinin nemli bölgelerde yaygın görülmesi, bu bölgelerde havanın genellikle daha doygun olması, bunun da büyük yağış olaylarına ve bu olaylara bağlı sıcaklıkta ani değişimlere yol açması olabilir. Bağlı nemin, yarı kurak bölgelerde daha fazla sapan değer göstermesi, bu bölgelerde güneşlenme, rüzgar ve düşük nem etkisiyle daha belirgin olan sıcaklık dalgalanmaları bağlı nemin aniden çok yüksek veya çok düşük değerlere ulaşmasına neden olabilir. Sapan yağış ve sıcaklık değerleri genellikle büyük ölçekli hava olayları veya belirli coğrafi faktörlerle oluşabilir, ancak bağlı nem daha yerel ve anlık sıcaklık değişimlerine tepki veriyor olabilir. Yarı kurak bölgelerde, gündüz-gece sıcaklık farkları ve aniden gelen nemli hava kütleleri bağlı nemde sapan değerlere yol açabilir.

Sıcaklık ve bağlı nem ölçümleri genellikle birbirine benzer veya aynı cihazlar kullanılarak yapılır. Bu cihazlar aynı çevresel koşullardan etkilenir ve benzer kalibrasyon hatalarına maruz kalabilir. Bu, sıcaklık ve bağlı nemin standart sapmalarının miktar olarak yakın olmasının bir nedeni olabilir. Aynı anda sıcaklık ve bağlı nemde sapan değerlerin görülmesi, bu iki değişkenin aynı atmosferik olaylardan etkilenmesinden kaynaklanabileceği gibi, aynı cihaz tarafından ölçülüyor olmalarından da kaynaklanabilir. Eğer bir sensörde bir hata veya sapma varsa, bu hata hem sıcaklık hem de bağlı nem ölçümlerinde benzer bir sapmaya neden olabilir. Bu durum, iki değişkende de aynı dönemde sapan değerlerin görülmesine yol açabilir. Eğer bu sapan değerler gerçekten atmosferik olaylardan kaynaklanıyorsa, su yönetimi planlarında bu olaylara hazırlıklı olmak gerekir. Eğer atmosferik olaylar yerine cihaz hatalarından kaynaklanıyorsa, bu durum su kaynakları yönetimini doğrudan etkileyebilir. Ölçüm hatalarından kaynaklanan yanlış veriler, su kaynaklarının aşırı veya yetersiz kullanılmasına neden olabilir, bu da uzun vadede kaynakların tükenmesine veya bozulmasına yol açabilir. Cihazlardan kaynaklanan hataların tespit edilmesi ve düzeltilmesi, su kaynaklarının sürdürülebilir yönetimi açısından kritik önemdedir. Bu, su yönetimi için kullanılan verilerin güvenilirliğini artırır ve yanlış kararların önüne geçer.

Varyasyon katsayısının (Cv) yüksek olduğu mevsimlerde, istasyonların standart sapmalarını eşitlemek, veriler arasındaki yayılımın daha homojen hale gelmesini ve sapmaların daha belirginleşmesini sağlıyor ve tespit edilmesini kolaylaştırıyor. Eğimlerin yüksek olduğu mevsimlerde, istasyonların eğimlerini eşitlemek, eğimden sapmaların daha net bir şekilde ortaya çıkmasını sağlıyor ve sapan değerlerin tespit edilmesini kolaylaştırıyor. Verilerin mevsimsel varyasyonları ve eğimleri dikkate alınarak yapılan analizler, sapan değer tespitinin doğruluğunu artırır ve bu da veri analizi süreçlerinde daha sağlam ve güvenilir sonuçlara ulaşılmasını sağlar. Bu tür bulgular, veri bilimi ve istatistiksel analiz alanlarında uygulamalı stratejiler geliştirmek için temel oluşturabilir.

Kışın, atmosferde kutupsal soğuk hava kütleleri, Akdeniz üzerinden gelen siklonlar gibi büyük ölçekli hava sistemleri hakimdir. Bu sistemler geniş alanları etkileyebilir, bu nedenle iki uzak istasyonun sıcaklıkları yazıya göre daha çok birlikte artıp azalabilir. Yazın atmosfer daha stabildir, ancak yerel etkiler daha belirgin hale gelebilir. Örneğin, bir bölgedeki yüksek basınç sistemi geniş bir alanı etkilerken, deniz esintisi, topografya, yerel rüzgarlar gibi yerel faktörler iki uzak istasyon arasında farklı sıcaklık değişikliklerine yol açabilir. Yakın istasyonlar ise her iki mevsimde de daha benzer sıcaklık değişimleri gösterir. Yağış miktarının yaz aylarında kış aylarına kıyasla belirgin ölçüde düşük olması, yakın konumda bulunan iki istasyonun yağış verilerinde daha az benzerlik göstermesine yol açmaktadır. Uzak konumdaki istasyonlar ise, genel olarak düşük bir benzerlik sergilediklerinden, her iki mevsimde de birbirlerine yakın düzeyde benzerlikler göstermektedir.

Sonuç olarak, sıcaklık, bağıl nem ve yağış verilerinde sapan değer belirlemek için kullanılan $\mu \pm x\sigma$ ve $Q_3, 1 \pm xIQR$ sınırları, x sırasıyla 3 ve 4 kullanıldığında birbirine daha çok benzese de bu benzerlik %60'ı geçmez. Bu sınırlar bazen ortalama kırılmalarını ve sıfır yağışın çok olduğu mevsimlerdeki sıfır olmayan yağışları sapan değer çıkarabilir. Ayrıca bu sınırlar eksik verilere ya da sapan olmayan değerlere çok yakın sapan değerler belirleyebilir. Bu sınırları kullanan araştırmacıların bunlara dikkat etmesi önerilir. Bu sınırlarla belirlenen sıcaklık ve yağış sapanları bölgesel ve zamansal olarak benzerlik gösterirken, bağıl nem ve yağış sapanlarını, sapan çıkaran komşular benzerlik gösterir. Belirlenen sapan değerlerin kendi içinde bir dengesi vardır. Normal dağılımı, nemli ve kurak bölgeleri kararlı, normal olmayan dağılımı, yarı nemli ve yarı kurak bölgeleri kararsız kabul edersek; Kararsız dağılımdan gelen sapanlar kararlı bölgelerde yada tam tersi olacak şekilde çıkarken, sapma miktarları yüksekse sapan değer adetleri düşük yada tam tersi şeklinde çıkar. Sapan değer belirlerken ön işlem kullanmak gerekebilir. Bağıl nemin geçiş özelliğiyle ve standart sapmanın benzerliği etkilemesiyle de desteklendiği gibi ön işlemler mevsimle birlikte değişebilir. Ön işlem olarak dizilerin yükselen istatistiksel özelliğini istasyonlar arasında eşitlemek daha çok sapan değer yakalanmasını sağlayabilir. Bu açıdan varyans parametresi eğim bileşeninden daha çok baskındır. Mevsimsel olarak farklı ön işlemler kullanılarak yapılacak bir çalışmayla sonuçlar karşılaştırılabilir.

TEŞEKKÜR

Bu çalışma, birinci yazarın ikinci yazar danışmanlığında hazırladığı doktora tezi esas alınarak üretilmiştir. Tez çalışmasına katkıda bulunan jüri üyeleri Prof. Dr. Ercan Kahya'ya ve Prof. Dr. C. Melek Kazezyılmaz Alhan'a teşekkür ederiz.

Bu doktora tezi, yazar İTÜ'de öğrenciyken yapılmıştır. Makale sürecinde yazar, bağımsız araştırmacıdır.

KAYNAKLAR

- [1] WMO. (1993). *Guide on the global data-processing system*. Geneva –Switzerland WMO, 305.
- [2] WMO. (2011). *Guide to Climatological Practices. 3rd Edition*, World Meteorological Organization, Geneva.
- [3] Bayazıt, M. (1999). *Hidroloji*. İTÜ İnşaat Fakültesi Matbaası, İstanbul.
- [4] SimavM., TürkezerA., SezenE., AkyolS., İnamM., CingözA., KılıçoğluA. (2011). Data quality control & management system of Turkish sea level monitoring network. *Harita Dergisi*, 145, 15 -28.
- [5] RissanenP., JacobssonC., MadsenH., MoeM., Pálsdóttir Þ. & VejenF. (2004). *Nordic methods for quality control of climate data*. Norwegian Meteorological Institute.
- [6] Gürnlü AlmaÖ. (2009). *Genetic algorithm based outlier detection using information criterion*. Phd thesis, DEU The Graduate School of Natural and Applied Sciences, İzmir.
- [7] BannayanM., HoogenboomG. (2008). Weatheranalogue: A toolforreal-time prediction of dailyweather data realizationsbased on a modified k- nearestneighborapproach. *Environmental Modelling & Software*. 23 (6), 703-713.
- [8] Government of thePeople'sRepublic of BangladeshMinistry of WaterResources. (2015). *Time Series Data Quality Control Guideline (For Hydro-meteorological data)*. WaterResources Planning Organization.
- [9] BarberC., LamontagneJ.R., VogelR. M. (2020). Improvedestimators of correlationand R2 forskeewedhydrologic data. *Hydrological Sciences Journal*. 65 (1), 87-101.
- [10] DikbaşF. (2018). A New Two-DimensionalRankCorrelationCoefficient. *Water Resource Manage*. 32, 1539–1553.
- [11] WMO. (2016). *Guidelines on Best Practices for Climate Data Rescue*. World Meteorological Organization. Geneva. 1182.
- [12] MillerP. A., StanleyG. B. (1992). A system for the hourly assimilation of surface observations in mountainous and flat terrain. *Monthly Weather Review*. 120, 2342–2359.
- [13] National Climatic Data Center. (1993). *Report of the international workshop on quality control of monthly climate data*. The international workshop on quality control of monthly climate data. Asheville, NC (United States).
- [14] WestcottN., AndsagerK., StoeckerL., SpinarM., SmithR., ObrechtR. & O'ConnellD. (2011). *Quality control of 19th century weather data*. Midwestern Regional Climate Center Publication.
- [15] ToretiA., KuglitschF. G., XoplakiE., LuterbacherJ. (2012). A Novel Approach for the Detection of Inhomogeneities Affecting Climate Time Series. *Journal of Applied Meteorology and Climatology*. 51 (2), 317–326.
- [16] National Oceanic and Atmospheric Administration. (2009). *Handbook of automated data quality control checks and procedures*. Stennis Space Center, Mississippi: National Data Buoy Center.
- [17] AguilarE., AuerI., BrunetM., PetersonT. C., WieringaJ. (2003). *Guidelines on climate metadata and homogenization*. Geneva: WMO.
- [18] FiratM., DikbaşF., KocA. C., GungorM. (2012). Analysis of temperature series: estimation of missing data and homogeneity test. *Meteorological Applications*. 19 (4), 1469-8080.

- [19] Bacon F. (1902). *Novum Organum*, 1620, ed. by J. Devey, New York: P. F. Collier.
- [20] Bernoulli D. (1961). The Most Probable Choice Between Several Discrepant Observations and the Formation. There from of the Most Likely Induction. 1777, in C. G. Allen, *Biometrika*. 48, 3 – 13.
- [21] Beckman R. J., Cook R. D. (1983). “Outlier.....s”, *Technometrics*, 25(2), 119–149.
- [22] Hawkins D. M. (1980). Identification of Outliers. Springer Science Business Media. B.V, Chapman & Hall.
- [23] Singh V. P. (2017). *Handbook of Applied Hydrology. Second Edition*, McGraw-Hill Education: USA.
- [24] Maity R. (2018). *Statistical Methods In Hydrology And Hydroclimatology*. Springer Nature, Singapore.
- [25] Laio F., Allamano P., Claps P. (2010). Exploiting the information content of hydrological ‘outliers’ for goodness-of-fit testing. *Hydrol. Earth Syst. Sci.* 14 (10), 1909–1917.
- [26] Hershfield D. M. (1973). “On the probability of extreme rainfall events”, *Bulletin of the American Meteorological Society*. 54 (10), 1013-1018.
- [27] Van den Brink H. W., Können G. P. (2008). The statistical distribution of meteorological outliers. *Geophysical Research Letters*. 35, 23702.
- [28] Barnett V., Lewis T. (1978). *Outliers in Statistical Data*. Great Britain: John Wiley & Sons Ltd.
- [29] Iglewicz B., Hoaglin D. C. (1993). *Volume 16: How to Detect and Handle Outliers*. USA: American Society for Quality Control Statistics Division.
- [30] Rousseeuw P. J., Leroy A. M. (1987). *Robust Regression And Outlier Detection*. USA: John Wiley & Sons.
- [31] Ovla H. D., Taşdelen B. (2012). Aykırı Deger Yönetimi. *Mersin Univ Saghk Bilim Derg.* 5 (3), 1-8.
- [32] Grubbs F. E. (1969). Procedures for Detecting Outlying Observations in Samples. *Technometrics*. 11(1), 1-21.
- [33] Prescott P. (1978). Examination of the Behaviour of Tests for Outliers when More than One Outlier is Present. *Journal of the Royal Statistical Society. Applied Statistics*. 27(1), 10-25.
- [34] Chandola V., Banerjee A., Kumar V. (2009). *Anomaly Detection : A Survey*. ACM Computing Surveys.
- [35] Wilks D. S. (2006). *Statistical Methods in the Atmospheric Sciences. Second Edition*, Elsevier Inc, USA.
- [36] Abbott P. F. & World Meteorological Organization Commission for Climatology. (1986). *Guidelines on the quality control of surface climatological data*. World Meteorological Organization WMO/TD. 85 (111).
- [37] Aggarwal C. C., Sathe S. (2017). *Outlier Ensembles*, Switzerland: Springer.
- [38] Aggarwal C. C. (2017). *Outlier Analysis*, Switzerland: Springer.
- [39] Wang X., Wang X., Wilkes M. V. (2021). *New Developments in Unsupervised Outlier Detection*. Singapore: Springer, Xi'an Jiaotong University Press.
- [40] Swersky L. (2018). *A Study of Unsupervised Outlier Detection for One-Class Classification*. Master thesis, Department of Computing Science, University of Alberta.
- [41] Tax D. M. J., Duin R. P. W. (2004). Support Vector Data Description. *Machine Learning*. 54, 45–66.
- [42] Ruff L., Vandermeulen R. A., Gornitz N., Deecke L., Siddiqui S. A., Binder A., Müller E., Kloft M. (2018). Deep One-Class Classification. *International Conference on Machine Learning*, Stockholm, Sweden.
- [43] Dawson R. (2011). How Significant is a Boxplot Outlier?. *Journal of Statistics Education*. 19(2), 1-14.
- [44] Ruts I., Rousseeuw P. J. (1996). Computing depth contours of Bivariate point clouds. *Computational Statistics & Data Analysis*. 23, 153-168.
- [45] Cárdenas-Montes M. (2014). Depth-Based Outlier Detection Algorithm. M. Polycarpou et al. (Eds.): HAIS 2014, LNAI 8480, 122–132.
- [46] Chen Y., Dang X., Peng H., Barth L. (2009). Outlier Detection with the Kernelized Spatial Depth Function. *Jr. IEEE transactions on pattern analysis and machine intelligence*. 31(2).
- [47] Hussain I. (2020). Outlier Detection using Nonparametric Depth-Based Techniques in Hydrology. *International Journal of Advanced Computer Science and Applications*. 11(9), 456-462.
- [48] Ilango V., Subramanian R., Vasudevan V. (2012). A Five Step Procedure for Outlier Analysis in Data Mining. *European Journal of Scientific Research*. 75(3), 327-339.
- [49] Mozharovskiy P. (2022). Anomaly detection using data depth: Multivariate case. arXiv:2210.02851v1 [stat.ML].
- [50] Rutst I. & Rousseeuw P. J. (1996). ISODEPTH: A Program for Depth Contours. *Proceedings in computational statistics: 12th symposium*, Barcelona, Spain.
- [51] Ranga Suri N. N. R., Narasimha M. M., Athithan G. (2019). *Outlier Detection: Techniques and Applications*. Switzerland: Intelligent Systems Reference Library, Springer. 155.
- [52] Gupta M., Gao J., Aggarwal C., Han J. (2014). *Outlier Detection for Temporal Data*. Morgan & Claypool Publishers.
- [53] Kırıl G., Billor N., Türkmen A. (2013). Robust kümeleme yöntemi ile grup sapan değerlerin belirlenmesi. *Ç.Ü. Sosyal Bilimler Enstitüsü Dergisi*. 22(1), 189-212.
- [54] Misra S., Li H., He J. (2020). *Machine Learning for Subsurface Characterization*. Gulf Professional Publishing.
- [55] Tran K. P. (2022). *Control Charts and Machine Learning for Anomaly Detection in Manufacturing*. Springer Series in Reliability Engineering, Switzerland.
- [56] Liu F. T., Ting K. M. & Zhou Z. H. (2008). Isolation Forest. *Eighth IEEE International Conference on Data Mining*, Pisa, Italy. 413-422.

- [57] Pegram G. (1997). Patching rainfall data using regression methods. 3. Patching and outlier detection. *Journal of Hydrology*, 198, 319–334.
- [58] Türkan S., Çetin M. C., Toktamış O. (2012). Outlier detection by regression diagnostics based on robust parameter estimates. *Hacettepe Journal of Mathematics and Statistics*. 41(1), 147 – 155.
- [59] Dilmaç H., Şişman Y. (2023). New approaches for outlier detection: The least trimmed squares adjustment. *International Journal of Engineering and Geosciences*. 8(1), 26-31.
- [60] Alla S., Adari S. K. (2019). *Beginning Anomaly Detection Using Python-Based Deep Learning: With Keras and PyTorch*. Apress; 1st ed.
- [61] Acar R. (2003). Çoruh hidrolojik havzasında yıllık yağış verilerinin istatistiksel modellemesi. *Pamukkale University Engineering College Journal of Engineering Science*. 9(3), 313-317.
- [62] Yeşilirmak E., Akçay S., Dağdelen N., Gürbüz T., Sezgin F. (2009). Quality control and homogeneity of annual precipitation data in Büyük Menderes Basin, Western Turkey. *Fresenius Environmental Bulletin*. 18, 1748-1757.
- [63] Hunt B. (2007). Climatic outliers. *International Journal of Climatology*. 27, 139-156.
- [64] Radzi A.B.A. (2009). *Trend analysis of sea level rise for west coast of peninsular Malaysia*. Master Thesis, Civil – Hydrology and Water Resources, Universiti Teknologi Malaysia.
- [65] Jeong J., Park E., Han W. S., Kim K.Y., Choung S., Chung Il-M. (2017). Identifying Outliers of Non-Gaussian Groundwater State Data Based on Ensemble Estimation for Long-Term Trends. *Journal of Hydrology*. 548, 135-144.
- [66] Ekeu-wei I.T., Blackburn G.A., Pedruco P. (2018). Infilling Missing Data in Hydrology: Solutions Using Satellite Radar Altimetry and Multiple Imputation for Data-Sparse Regions. *Water*. 10, 1483.
- [67] Köse Ö. (2010). Characteristics of Various Filters Used to Eliminate Outliers in Velocity Time Series Obtained Using ADVs. VI. *Ulusal Hidroloji Kongresi*, Pamukkale Üniversitesi, Denizli. 360-368.
- [68] Noone S., Murphy C., Coll J., Matthews T., Mullan D., Wilby R., Walsh S. (2015). Homogenization and analysis of an expanded long-term monthly rainfall network for the Island of Ireland (1850–2010). *International Journal of Climatology*. 36, 2837-2853.
- [69] Gottschalk L., Kundzewicz Z. (1995). Analysis of outliers in Norwegian flood data. New Uncertainty Concepts in Hydrology and Water Resources, *Cambridge University Press*. 245-251.
- [70] Kirk A., McCuen R. (2008). Outlier Detection in Multivariate Hydrologic Data. *Journal of Hydrologic Engineering*. 13(7), 641-646.
- [71] Hansen J., Ruedy R., Sato M., Lo K. (2010). Global surface temperature change. The American Geophysical Union. *Reviews of Geophysics*. 48.
- [72] Domonkos P. (2006). Testing of homogenization methods purposes, tools and problems of implementation. *Proceedings of the Fifth Seminar for Homogenization and Quality Control in Climatological Databases*, Budapest, Hungary, Climate Data and Monitoring WCDMP-No. 71.
- [73] Domonkos P. (2011). Adapted Caussinus-Mestre Algorithm for Networks of Temperature series (ACMANT). *International Journal of Geosciences*. 2(3), 293-309.
- [74] Vicente-Serrano S. M., Beguería S., López-Moreno J. I., García-Vera and Petr Štepanek M. A. (2010). A complete daily precipitation database for northeast Spain: reconstruction, quality control, and homogeneity. *International Journal Climatology*. 30, 1146–1163.
- [75] Yakunin A., Mogahed H. (2018). Improvement of statistical methods for detecting anomalies. *Climate and Environmental Monitoring Systems*. 289.
- [76] Shabani S., Samadianfard S., Sattari M. T., Mosavi A., Shamshirband S., Kmet T., Várkonyi-Kóczy A. R. (2020). Modeling Pan Evaporation Using Gaussian Process Regression K-Nearest Neighbors Random Forest and Support Vector Machines; Comparative Analysis. *Atmosphere*. 11(66).
- [77] Klok E.J., Klein T. A.M.G. (2009). Updated and extended European dataset of daily climate observations. *International Journal of Climatology*. 29(8), 1182–1191.
- [78] Hofstra N., New M., McSweeney C. (2010). The influence of interpolation and station network density on the distributions and trends of climate variables in gridded daily data. *Clim Dyn*. 35, 841–858.
- [79] Wu H., Tang X., Wang Z., Wu L., Lianfang W., Zhu J. (2018). Probabilistic Automatic Outlier Detection for Surface Air Quality Measurements from the China National Environmental Monitoring Network. *Advances in Atmospheric Sciences*. 35, 1522-1532.
- [80] Chow V., Maidment D., Mays L. (2013). *Applied Hydrology*. 2nd Edition, McGraw-Hill Companies.
- [81] Oğuz E.B., Öñöz B. (2003). Assessment of outliers in statistical data analysis. *Nato Science Series: IV: Earth and Environmental Sciences*. 23, 173-180.
- [82] Amin M.T., Rizwan M., Alazba A.A. (2015). Comparison of mixed distribution with EV1 and GEV components for analyzing hydrologic data containing outlier. *Environmental Earth Sciences*. 73(3), 1369-1375.
- [83] Ciupak M., Ozga-Zielinski B., Tokarczyk T., Adamowski J. (2021). A Probabilistic Model for Maximum Rainfall Frequency Analysis. *Water*. 13, 2688.

- [84] Sallabanda A. (1996). Homogeneity evaluation of snow cover records in Albania. *Proceedings of the First Seminar for Homogenization of Surface Climatological Data*, Budapest. 6–12.
- [85] Shabani S., Samadianfard S., Sattari M., Mosavi A., Band S., Kmet T., Varkonyi-Koczy A. (2020). Modeling Pan Evaporation Using Gaussian Process Regression K-Nearest Neighbors Random Forest and Support Vector Machines; Comparative Analysis. *Atmosphere*. 11, 1-17.
- [86] Fernández-Palomino C.A., Lavado-Casimiro W.S. (2017). Regional maximum rainfall analysis using L-moments at the Titicaca Lake drainage, Peru. *Theor Appl Climatol*. 129, 1295–1307.
- [87] Aslan S. (2008). Ege Bölgesi günlük maksimum yağışlarının bölgesel analizi. Dokuz Eylül Üniversitesi FBE, yüksek lisans tezi.
- [88] Lamontagne J., Stedinger J. (2015). Examination of the Spencer-McCuen Outlier-Detection Test for Log-Pearson Type 3 Distributed Data. *Journal of Hydrologic Engineering*. 21, 04015069.
- [89] Cohn T., England J., Berenbrock C., Mason R., Stedinger J., Lamontagne J. (2013). A generalized Grubbs-Beck test statistic for detecting multiple potentially influential low outliers in flood series. *Water Resources Research*. 49, 5047-5058.
- [90] Kaynar L. (2005). *Doğu Karadeniz bölgesi yıllık maksimum akımlarının istatistik methodlarla analizi*. İTÜ, FBE, yüksek lisans tezi.
- [91] Faybishenko B., Versteeg R., Pastorello G. (2022). Challenging problems of quality assurance and quality control (QA/QC) of meteorological time series data. *Stoch Environ Res Risk Assess*. 36, 1049–1062.
- [92] Altay L. (2011). *Konum-zaman bağlamında aykırı gözlem tespit yöntemlerinin incelenmesi ve bulanık Kümelendirme tabanlı bir metod*. Yüksek lisans tezi. Hava harp okulu, havacılık ve uzay teknolojileri enstitüsü.
- [93] Lanzante J.R. (1996). Resistant, Robust and Non-Parametric Techniques for the Analysis of Climate Data: Theory and Examples, Including Applications to Historical Radiosonde Station Data. *International Journal of Climatology*. 16, 1197-1226.
- [94] Sheese P., Boone C., Walker K. (2015). Detecting physically unrealistic outliers in ACE-FTS atmospheric measurements. *Atmospheric Measurement Techniques*. 8, 741-750.
- [95] Li J.S., Hamann A., Beaubien E. (2020). Outlier detection methods to improve the quality of citizen science data. *Int J Biometeorol*. 64(11), 1825-1833.
- [96] Estévez J., Gavilán P., García-Marín A. (2018). Spatial regression test for ensuring temperature data quality in southern Spain. *Theoretical and Applied Climatology*. 131, 309-318.
- [97] Tellez B., Cernocky T., & Terradellas E. (2008). Calculation of climatic reference values and its use for automatic outlier detection in meteorological datasets. *Advances in Science and Research*. 2, 1-4.
- [98] Lenwood W. H. JR., Killen W. D., Anderson R. D., Perry E. (2004). Analysis of historical salinity and boron surface water monitoring data from the San Joaquin River watershed: 1985-2002. *Environmental monitoring and assessment*. 95(1-3), 125–151.
- [99] Erdoğan G. (2012). *Spectral methods for outlier detection in machine learning*. The Institute for Graduate Studies in Science and Engineering, Master of Science, Boğaziçi University.
- [100] Longman R. J., Newman A. J., Giambelluca T. W., Lucas M. (2020). Characterizing the Uncertainty and Assessing the Value of Gap-Filled Daily Rainfall Data in Hawaii. *American Meteorological Society*. 1261 – 1276.
- [101] Zhao C., Yang J. (2019). A Robust Skewed Boxplot for Detecting Outliers in Rainfall Observations in Real-Time Flood Forecasting. *Advances in Meteorology*. 1-7.
- [102] Bhattacharya S., Beirlant J. (2019). Outlier detection and a tail-adjusted boxplot based on extreme value theory.
- [103] Sharif M., Burn D. H. (2006). Simulating climate change scenarios using an improved K-nearest neighbor model. *Journal of Hydrology*. 325, 179–196.
- [104] Boulanger J.P., Aizpuru J., Leggieri L., Et al (2010). A procedure for automated quality control and homogenization of historical daily temperature and precipitation data. (APACH): part 1: Quality control and application to the Argentine weather service stations, *Climatic Change*. 98, 471–491.
- [105] Freitas L., Pereira M., Caramelo L., Mendes M., Nunes L. (2013). Homogeneity of monthly air temperature in Portugal with HOMER and MASH”, *Idojaras*. 117, 69-90.
- [106] Murav'ev A.V., Kulikova I.A., Kruglova E.N. (2009). Distribution of extreme characteristics of atmospheric circulation from reanalysis data and hydrodynamic modeling. *Russ. Meteorol. Hydrol*. 34, 431–441.
- [107] Naghettini M. (2017). *Fundamentals of Statistical Hydrology*. Springer, Switzerland.
- [108] Farmer W. H., Vogel R. M. (2013). Performance-weighted methods for estimating monthly stream flow at ungauged sites. *Journal of Hydrology*. 477, 240–250.
- [109] Liu D., Liu H., Meng X. (2023). *Advanced Hydrologic Modeling in Watershed Scales*. MDPI – Multidisciplinary Digital Publishing Institute.
- [110] González R., Fidel J. & Jiménez J., Quesada V., Valero F. (2001). Quality Control and Homogeneity of Precipitation Data in the Southwest of Europe. *Journal of Climate*. 14, 964-978.

- [111] Meena H. M., Machiwal D., Santra P., Moharana P., Singh Dr. (2019). Trends and homogeneity of monthly, seasonal, and annual rainfall over arid region of Rajasthan, India. *Theoretical and Applied Climatology*. 136, 795-811.
- [112] Göktürk O. M., Bozkurt D., Sen O., Karaca M. (2008). Quality control and homogeneity of Turkish precipitation data. *Hydrological Processes*. 22, 3210 – 3218.
- [113] Zaifoglu H., Akintug B., Yanmaz A. (2018). Quality and Homogeneity Analysis of Precipitation Data in North Cyprus. *Conference: 13th International Congress on Advances in Civil Engineering*, İzmir, Turkey.
- [114] Štěpánek P., Zahradníček P. (2008). Experiences with quality control and homogenization of daily series of various meteorological elements in the Czech Republic 1961-2007. *Proceedings of the sixth Seminar for Homogenization and Quality Control in Climatological Databases*, Budapest, Hungary, Climate Data and Monitoring WCDMP-No. 76.
- [115] Dunn R. J. H., Willett K. M., Thorne P. W., Woolley E. V., Durrel I., Dai A., Parker D. E., Vose R. S. (2012). HadISD: a quality-controlled global synoptic report database for selected variables at long-term stations from 1973–2011. *Clim. Past*, 8, 1649–1679.
- [116] Toret A., Desiato F. (2006). Homogenization and validity controls for temperature trend estimates over Italy. *Proceedings of the Fifth Seminar for Homogenization and Quality Control in Climatological Databases*, Budapest, Hungary, Climate Data and Monitoring WCDMP-No. 71.
- [117] Štěpánek P., Řezníčková L., Brázdi R. (2006). Homogenization of daily air pressure and temperature series for Brno (Czech Republic) in the period 1848-2005. *Proceedings of the Fifth Seminar for Homogenization and Quality Control in Climatological Databases*, Budapest, Hungary, Climate Data and Monitoring WCDMP-No. 71.
- [118] Štěpánek P., Mikulová K. (2006). Homogenization of air temperature and relative humidity monthly means of individual observation hours in the area of The Czech and Slovak Republic. *Proceedings of the Fifth Seminar for Homogenization and Quality Control in Climatological Databases*, Budapest, Hungary, Climate Data and Monitoring WCDMP-No. 71.
- [119] Shi J., Yang L. (2020). A Climate Classification of China through k-Nearest-Neighbor and Sparse Subspace Representation. *Journal of Climate*. 33(1), 243-262.
- [120] Borges P., Franke J., Anunciação Y., Weiss H., Bernhofer C. (2015). Comparison of spatial interpolation methods for the estimation of precipitation distribution in Distrito Federal, Brazil. *Theoretical and Applied Climatology*.
- [121] Estévez J., Gavilán P., García-Marín A. P. (2011). Data validation procedures in agricultural meteorology – a prerequisite for their use. *Adv. Sci. Res.* 6, 141–146.
- [122] Aydın M., Tuna A. (2018). Aykırı Değerlerin Tespiti ve Taşkın Debilerinin Hesabı için Bir Yöntem Geliştirilmesi. *Teknik Dergi*. 8339-8363.
- [123] Aydın M. (2015). Hidrolojik Verilerdeki Aykırı Değerlerin Taşkın Frekans Analizi Üzerine Etkisinin İncelenmesi: Van Gölü Havzası Örneği. *Journal of The Institute of Natural & Applied Sciences*. 20(1-2), 47-55.
- [124] El Kenawy A., López-Moreno J. I., Stepanek C. P., Vicente-Serrano S. M. (2013). An assessment of the role of homogenization protocol in the performance of daily temperature series and trends: application to north eastern Spain. *International Journal Climatology*. 33, 87–108.
- [125] Kuhn N. J., Baumhauer R., Schütt B. (2011). Managing the impact of climate change on the hydrology of the Gallocanta Basin, NE-Spain. *Journal of Environmental Management*. 92(2), 275-283.
- [126] McMahon T. A., Peel M. C. (2019). Uncertainty in stage–discharge rating curves: application to Australian Hydrologic Reference Stations data. *Hydrological Sciences Journal*. 64(3), 255-275.
- [127] DeGagne M. P. J., Douglas G. G., Hudson H. R., Simonovic S. P. (1996). A decision support system for the analysis and use of stage-discharge rating curves. *Journal of Hydrology*. 184, 225-241.
- [128] Aydın M. (2012). Hidrolojik verilerdeki aykırı değerlerin tahmini için bir method geliştirilmesi ve Dicle ile Fırat havzalarına uygulanması. Yüksek Lisans Tezi, Fırat Üniversitesi Fen Bilimleri Enstitüsü.
- [129] Chebana F., Dabo-Niang S., Ouada T. (2012). Exploratory functional flood frequency analysis and outlier detection. *Water Resources Research*. 48, 4514.
- [130] Suhaila J. (2021). Functional Data Visualization and Outlier Detection on the Anomaly of El Niño Southern Oscillation. *Climate*. 9(118).
- [131] Chebana F., Ouada T. (2011). Multivariate extreme value identification using Depth functions. *Environmetrics*. 2, 441–455.
- [132] Romano E., Mateu J. (2013). Outlier Detection for Geostatistical Functional Data: An Application to Sensor Data. A. Giusti et al. (eds.), Classification and Data Mining, Studies in Classification, Data Analysis and Knowledge Organization, Springer-Verlag, Berlin Heidelberg. 131-138.
- [133] Chebana F., Ouada T. (2011). Depth-based multivariate descriptive statistics with hydrological Applications. *Journal of geophysical research*. 116, 1-19.

- [134] Wazneh H. (2016). Identification of hydrological neighborhoods for regional flood frequency analysis using statistical depth function. *Advances in Water Resources*.
- [135] Eichinger F., et al. (2013). Data Analysis Challenges in the Future Energy Domain. *International Symposium on Intelligent Data Analysis*.
- [136] Jennifer L., Andreas H., Elisabeth B. (2020). Outlier detection methods to improve the quality of citizen science data. *International Journal of Biometeorology*. 64, 1825-1833.
- [137] Lu Y., Kumar J., Collier N., Krishna B. & Langston M. A. (2018). Detecting Outliers in Streaming Time Series Data from ARM Distributed Sensors. *IEEE International Conference on Data Mining Workshops (ICDMW)*, Singapore. 779-786.
- [138] Petkovski A., Shehu V. (2023). Anomaly Detection on Univariate Sensing Time Series Data for Smart Aquaculture Using K-Means, Isolation Forest, and Local Outlier Factor. *12th Mediterranean Conference on Embedded Computing*.
- [139] Yu Q., Yuan Sheng L. (2019). Hydrological Time Series Anomaly Pattern Detection based on Isolation Forest. *IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference*.
- [140] Amirpouya S. (2015). Flood Outlier Detection Using PCA and Effect of How to Deal with Them in Regional Flood Frequency Analysis via L-Moment Method. *Water Resources*. 42, 448-459.
- [141] Akyay T., Karşoğlu M. O. (2009). Wavelet-Based Outlier Detection of Airborne Laser Scanning Data. *4th International Conference on Recent Advances in Space Technologies*, İstanbul, Turkey. 426-427.
- [142] Ih C., George T., Chung C. (1988). Estimation of Time Series Parameters in the Presence of Outliers. *Technometrics*. 30, 193-204.
- [143] Marinova T., Alexandrov V. (2006). On homogenization of climate long-term series in Bulgaria. *Proceedings of the Fifth Seminar for Homogenization and Quality Control in Climatological Databases*, Budapest, Climate Data and Monitoring WCDMP-No. 71.
- [144] Panagiota G., Panayotis P. (2007). Outliers and trend detection tests in rainfall extremes. *Proceedings of the 32nd IAHR World Congress*, Venice.
- [145] Akar A. U. (2021). *GNSS gözlemlerindeki troposferik gecikmenin destek vektör makineleri algoritması ile keşirimi*. Yüksek Lisans Tezi, Harita Mühendisliği Anabilim Dalı, Konya Teknik Üniversitesi.
- [146] Giustarini L., Parisot O., Ghoniem M., Hostache R., Trebs I., Otjacques B. (2016). A user-driven case-based reasoning tool for infilling missing values in daily mean river flow records. *Environmental Modelling & Software*. 82, 308-320.
- [147] Wakdikar M. P. & Deo M. C. (1999). Outlier detection and effects in dynamic modeling. *ISH Journal of Hydraulic Engineering*. 5(1), 33-41.
- [148] Dede A. (2023). *Hidrolojik sapan değer tespitine yeni bir yaklaşım*. Doktora Tezi, İTÜ.
- [149] Dede A., Ağırlioğlu N. (2023). Hidrolojik Sapan Değer Tespitinde Komşu İstasyon Seçimi. *Fırat Üniversitesi Mühendislik Bilimleri Dergisi*. 35(1), 81-99.
- [150] González-Hidalgo J. C., Arrillaga M. de L., Stepanek P., Lanjeri S. (2006). Quality control of monthly precipitation data from Mediterranean areas of Spain. *Proceedings of the Fifth Seminar for Homogenization and Quality Control in Climatological Databases*, Budapest, Hungary, Climate Data and Monitoring WCDMP-No. 71.
- [151] Lund R., Wang X. L., Reeves J., Feng Y., Lu Q., Gallagher C. (2006). Change point Detection in Periodic and Autocorrelated Time Series. *Proceedings of the Fifth Seminar for Homogenization and Quality Control in Climatological Databases*, Budapest, Hungary, Climate Data and Monitoring WCDMP-No. 71.
- [152] Guijarro J. A. (2006). Homogenization of a dense thermo-pluviometric monthly database in the Balearic Islands using the free contributed R package 'Climatol'. *Proceedings of the Fifth Seminar for Homogenization and Quality Control in Climatological Databases*, Budapest, Hungary, Climate Data and Monitoring WCDMP-No. 71.
- [153] Yang J., Chen Y., Rahardja S. (2022). Neighborhood Representative for Improving Outlier Detectors. *Information Sciences*. 625.
- [154] Wang X. L., Feng Y. (2010). *RHtests V3 User Manual*. Science and Technology Branch, Environment Canada, Climate Research Division, Atmospheric Science and Technology Directorate. Toronto: Canadian Centre for Climate Modelling and Analysis.
- [155] Water Resources Planning Organization. (2015). *Time Series Data Quality Control Guideline (For Hydro-meteorological data)*. Government of the People's Republic of Bangladesh Ministry of Water Resources.
- [156] Sevruck B., Geiger H. (1981). *Selection of distribution types for extremes of precipitation*. World Meteorological Organization (WMO), Operational hydrology report (OHR) No. 15.
- [157] Harvey J. E. R., Max A. L. (2015). *Understanding Mathematical And Statistical Techniques in Hydrology An Examples-Based Approach*. John Wiley & Sons, UK.
- Yalçın G. D. (2005). *Klimatoloji – I*. Ankara: DMİ Genel Müdürlüğü Matbaası.