

K-Ortalamalar Yöntemi ile Yıllık Yağışların Sınıflandırılması ve Homojen Bölgelerin Belirlenmesi

Mahmut FIRAT*

Fatih DİKBAŞ**

Abdullah Cem KOÇ***

Mahmud GÜNGÖR****

ÖZ

Hidrolojik veri ve bilginin sınırlı olduğu bölgelerde hidrolojik ve meteorolojik süreçlerin güvenilir ve doğru tahmininin gerçekleştirilmesi önemli uğraşlardan biridir. Hidrolojik değişkenlerin sınıflandırılması ve homojen bölgelerin belirlenmesi bölgesel tahmin çalışmalarının ilk ve en önemli adımlarıdır. Bu çalışmada, K-Ortalamalar yöntemi ile yıllık toplam yağışların sınıflandırılması ve homojen bölgelerin belirlenmesi amaçlanmıştır. En basit ve en çok uygulanan kümeleme yöntemi olan K-Ortalamalar yöntemi her bir özellik vektörü ile ona en yakın merkez arasındaki Öklit mesafesinin toplamını en küçükleyerek veri setini kümelere ayırmaktadır. Kümeleme analizi için Türkiye genelinde Devlet Meteoroloji İşleri (DMİ) tarafından işletilen 188 yağış gözlem istasyonuna ait yıllık yağışlar ve enlem, boylam ve yükselti verileri dikkate alınmıştır. Kümeleme analizi sonunda küme sayısı 7 olarak belirlenmiştir. Ayrıca, kümeleme ile belirlenen bölgelerin homojenliğini test etmek için bölgesel homojenlik testi uygulanmıştır.

Anahtar Kelimeler: kümeleme analizi, k-ortalamalar, yıllık yağış, homojenlik testi

ABSTRACT

Classification of Annual Precipitations and Identification of Homogeneous Regions using K-Means Method

Reliable and correct estimation of hydrological and meteorological processes is one of the major problems in regions with insufficient hydrologic information and data. The classification of the hydrological variables and determination of homogeneous regions are the most important steps of regional studies. The purpose of this study is to classify the

Not: Bu yazı

- Yayın Kurulu'na 19.04.2010 günü ulaşmıştır.
- 30 Eylül 2012 gününe kadar tartışmaya açıktır.

* İnönü Üniversitesi, İnşaat Mühendisliği Bölümü, Malatya - mahmut.firat@inonu.edu.tr

** Pamukkale Üniversitesi, İnşaat Mühendisliği Bölümü, Denizli - f_dikbas@pau.edu.tr

*** Pamukkale Üniversitesi, İnşaat Mühendisliği Bölümü, Denizli - a_c_koc@pau.edu.tr

**** Pamukkale Üniversitesi, İnşaat Mühendisliği Bölümü, Denizli - mgungor@pau.edu.tr

annual total precipitation series and to identify the homogeneous regions by K-Means method. The K-means method, which is the simplest and most commonly used clustering method, divides a data set into clusters by minimizing the sum of the Euclidean distance between each feature vector and its closest cluster centre. The annual precipitation records and longitude, latitude and altitude values obtained of 188 stations operated by the National Meteorology Works (DMI) in Turkey were considered for clustering analysis. The number of clusters was determined as 7 by means of clustering analysis. Moreover, the regional homogeneity test was applied for testing the homogeneity of regions.

Keywords: clustering, k-means method, annual precipitation, homogeneity test

1. GİRİŞ

Bir bölgede su kaynaklarının planlanması ve yönetimi, baraj, dolu savak ve diğer su yapılarının güvenilir bir şekilde tasarımı ve işletilmesi için hidrolojik ve meteorolojik süreçlerin doğru bir şekilde modellenmesi oldukça önemlidir. Bunun için bölgede gözlenmiş yeterli uzunluğa sahip verilere ihtiyaç duyulmaktadır. Ancak, bazen çeşitli sebeplerden dolayı veriler eksik ya da gözlem süreleri kısa olmaktadır. Bölgesel tahmin çalışmalarının en önemli ve en zor adımı homojen bölgelerin belirlenmesidir. Çoğu zaman bölgelerin sınıflandırılması coğrafi olarak birbirine yakın istasyonların aynı bölgede yer alması şeklinde yapılmaktadır. Ancak bu tür sınıflandırma ile oluşturulan bölgelerin hidrolojik olarak homojen olduğu söylenemez. Bu şekilde sınıflandırılan bölgelerde bilginin bir bölgeden diğer bölgeye aktarılması mümkün değildir. Bu nedenle daha güvenilir bir bölgesel çalışma için, hidrolojik olarak birbirine benzerlik gösteren bölgelerin belirlenmesi gerekmektedir. Homojen bölgelerin belirlenmesi ve sınıflandırma amacıyla hiyerarşik ve hiyerarşik olmayan kümeleme yöntemleri yaygın bir şekilde uygulanmaktadır. Kümeleme yöntemleri özellikle, bölgesel taşkın frekans analizi için homojen bölgelerin belirlenmesi [1, 2, 3, 4, 5, 6, 7] ve yağış değişkenin sınıflandırılması ve homojen yağış bölgelerinin belirlenmesi [8, 9] amacıyla kullanılmıştır. Demirel [10], hiyerarşik kümeleme yöntemi ile Türkiye akarsu havzalarının sınıflandırılmasını ve benzer özelliklere sahip bölgelerin oluşturulmasını amaçlamıştır. Turan [11], tarafından yapılan çalışmada, Ward yöntemi ile Türkiye akarsu verimlerinin sınıflandırılması amaçlanmıştır. Kahya vd. [12], K-Ortalamalar yöntemi ile Türkiye’de akarsu akımlarının konumsal olarak sınıflandırılmasını amaçlamıştır. Demirel vd. [13], Türkiye akarsu havzalarında temel kuraklık bileşenlerinin sınıflandırılmasında K-Ortalamalar yöntemini uygulamıştır. Isik ve Singh [14], aylık ortalama akım verilerini dikkate alarak Ward ve K-ortalamalar yöntemleri ile Türkiye havzalarını sınıflandırmıştır. Bu çalışmada, hiyerarşik olmayan (K-Ortalamalar) kümeleme yöntemi ile Türkiye yıllık yağışların sınıflandırılması ve homojen bölgelerin belirlenmesi amaçlanmaktadır. Bunun için Türkiye genelinde ölçüm değerine sahip 188 yağış gözlem istasyonunda alınan yıllık yağışlar ve istasyonlara ait veriler kullanılmış ve kümeleme analizi gerçekleştirilmiştir. Ayrıca, belirlenen bölgelerin hidrolojik olarak homojenliğinin kontrolü için L-momentler yöntemini temel alan homojenlik testi uygulanmıştır.

2. HİYERARŞİK OLMAYAN (K-ORTALAMALAR) KÜMELEME YÖNTEMİ

Kümeleme analizi, çok değişkenli istatistiksel bir teknik olup, verilerin benzerliklerine göre sınıflandırılması ve homojen alt gruplara ayrılması için kullanılmaktadır. K-Ortalamalar yöntemi, en çok bilinen kümeleme yöntemlerinden biri olup hiyerarşik olmayan bir yapıya sahiptir [7, 15]. K-Ortalamalar yöntemi, bir X veri setine ait d adet değişkeni ve N adet özellik vektörünü C adet kümeye ayırma ve sınıflandırma özelliğine sahiptir [6, 16]. Bu yöntemde, önceden belirlenen C adet kümeye ait merkezlerin belirlenmesi ile başlanır ve her bir değişken benzerlik ölçütü yardımıyla en yakın küme merkezine atanır. Giriş veri setindeki her bir değişkenin bir kümeye atanmasından sonra, her bir küme için küme merkezi yeniden hesaplanarak değişkenler bu yeni küme merkezlerinin yerleşimine bağlı olarak yeni farklı kümelere atanabilir. Bu işlem küme üyeliklerinde herhangi bir değişim olmayana kadar tekrarlanır [16]. İncelenen bir problemde, N adet özellik vektörüne ve d adet değişkene sahip bir X veri seti $X = \{x_k | k = 1, 2, \dots, N\}$ şeklinde tanımlanabilir.

Bu veri setinde k . özellik vektörü $x_k = [x_{k1}, x_{k2}, \dots, x_{kd}]$, $x_k \in R^d$ şeklinde yazılabilir [17]. K-Ortalamalar yönteminde, veri seti alt kümelere ayrılırken denklem (1)'de verilen amaç fonksiyonunun en küçüklenmesi amaçlanmaktadır. Bu denklemde verilen amaç fonksiyonun en küçüklenmesi, her bir özellik vektörü ile en yakın küme merkezi arasındaki mesafe ölçütünün en küçüklenmesini ve dolayısıyla benzer yapıya sahip verilerin aynı kümede toplanmasını sağlamaktadır. Literatürde mesafe ölçütü için genellikle denklem (2)'de verilen Öklit mesafe ölçütü kullanılmaktadır [16].

$$J(S : X) = \sum_{i=1}^C \sum_{k=1}^N d_{ik}^2(x_k, s_i) \quad (1)$$

Denklem (1)'de verilen d_{ik}^2 aşağıdaki gibi tanımlanmıştır.

$$d_{ik}^2 = \|x_k^{(i)} - s_i\|^2 \quad (2)$$

$$s_i = \frac{1}{N} \sum_{k=1}^N x_k^{(i)} \quad (3)$$

Öklit mesafe ölçütü matrisinin açılımı aşağıda verilmiştir.

$$\begin{bmatrix} d(1,1) & d(1,2) & d(1,3) & \dots & d(1,C) \\ d(2,1) & d(2,2) & & & \\ d(3,1) & & & & \\ & & & & \\ & & & & \\ d(N,1) & d(N,2) & d(N,3) & \dots & d(N,C) \end{bmatrix} \quad (4)$$

Burada, C küme sayısı, S küme merkezlerini içeren matris ($S = \{s_1, s_2, \dots, s_C\}$), s_i i . kümenin merkezi, $x_k^{(i)}$ i . kümeye ait k . özellik vektörü, d_{ik}^2 i . kümeye ait k . özellik vektörü ile i . küme merkezi arasındaki mesafe şeklinde tanımlanabilir. K-Ortalamalar yönteminin en önemli avantajı uygulamasının basit olması olarak gösterilebilir. Diğer taraftan, bu yöntemin performansının başlangıçta seçilen küme sayısına ve küme merkezlerine bağlı olarak değişmesi bu yöntemin en önemli dezavantajı olarak verilebilir. K-Ortalamalar yönteminin işlem adımları; (i) Küme sayısının ve merkezlerinin başlangıç değerlerinin rastgele belirlenmesi, (ii) Her bir değişken için belirlenen küme merkezlerinden olan mesafelerin hesaplanması, (iii) Her bir değişkenin en yakın küme merkezine göre bir kümeye atanması, (iv) Amaç fonksiyonunun en küçüklenmesi, (v) Küme merkezlerinin yeniden hesaplanması ve yeni kümelerin belirlenmesi, (vi) Küme üyeliklerinde herhangi bir değişim olmayana kadar analizin devam ettirilmesi şeklindedir.

3. UYUMSUZLUK VE BÖLGESEL HOMOJENLİK TESTİ

Kümeleme yöntemi ile belirlenen bölgelerde yer alan uyumsuz istasyonların belirlenmesi için denklem (5)'te verilen uyumsuzluk ölçütü kullanılmaktadır [18, 19].

$$D_i = \frac{1}{3} N_i (u_i - \bar{u})^T A^{-1} (u_i - \bar{u}) \quad (5)$$

$$u_i = [t^{(i)}, t_3^{(i)}, t_4^{(i)}] \quad (6)$$

$$\bar{u} = \frac{1}{N} \sum_{i=1}^N u_i \quad (7)$$

$$A = \sum_{i=1}^N (u_i - \bar{u})(u_i - \bar{u})^T \quad (8)$$

Yukarıda, D_i uyumsuzluk ölçütü, N özellik vektör sayısı, u_i i istasyonuna ait L-moment oranlarını içeren vektör, \bar{u} bölgesel ortalama, T vektör ya da matrisinin transpoze işlemi, A kovaryans matrisi, $t^{(i)}$ i istasyonuna ait $L-Cv$ moment oranı, $t_3^{(i)}$ i istasyonuna ait $L-Cs$ moment oranı, $t_4^{(i)}$ i istasyonuna ait $L-Ck$ moment oranı olarak kullanılmaktadır. Denklem (5) ile hesaplanan D_i ölçütünün değeri büyük olursa bu durumda i istasyonu ihmal edilebilir. İstasyon sayısı 15'ten fazla olan bölgelerde, istasyon için hesaplanan D_i değerinin 3'ten büyük olması durumunda bu istasyonun tanımlanan bölgeden çıkarılması önerilmektedir [18, 19]. Kümeleme analizi ile tanımlanan bölgelerin homojenliğinin test edilmesinde L-moment oranlarını ($L-Cv$ değişkenlik, $L-Cs$ çarpıklık ve $L-Ck$ kurtosis) temel alan homojenlik testi (H testi) kullanılmaktadır. Bu test yardımıyla, kümeleme analizi ile belirlenen grupların homojen olup olmadığı istatistiksel olarak değerlendirilmektedir.

Bunun için, belirlenen bir bölgedeki istasyonların oluşturduğu grup için L-momentlerdeki değişimler karşılaştırılır. Bu değişim, bölgesel L-moment oranları değerlerinin ağırlıklı standart sapmasının (V) her bir istasyondaki veri uzunluğuna oranı şeklinde ifade edilebilir. $L-Cv$ oranındaki değişimin büyüklük tahminlerinin varyansı üzerindeki etkisinin $L-Cs$ ve $L-Ck$ oranlarındaki değişimlere göre daha büyük olmasından dolayı $L-Cv$ oranının kullanılması önerilmektedir [18, 19]. Bir bölgede, N adet gözlem istasyonu için hesaplanan L-moment oranları $(t^{(i)}, t_3^{(i)}, t_4^{(i)})$ ve bölgesel ortalama L-moment oranları $(t^R = \sum_{i=1}^N n_i t^{(i)} / \sum_{i=1}^N n_i, t_3^R = \sum_{i=1}^N n_i t_3^{(i)} / \sum_{i=1}^N n_i, t_4^R = \sum_{i=1}^N n_i t_4^{(i)} / \sum_{i=1}^N n_i)$ şeklinde yazılabilir.

Bir bölgede $L-Cv$ oranları temel alınarak ağırlıklı standart sapma (V_1) denklem (9) yardımıyla belirlenebilir [18, 20, 21].

$$V_1 = \left[\frac{\sum_{i=1}^N n_i (L-Cv^{(i)} - \overline{L-Cv})^2}{\sum_{i=1}^N n_i} \right] \tag{9}$$

Burada, $L-Cv^{(i)}$ i istasyonu için L-moment oranı, $\overline{L-Cv}$ bölgesel ortalama L-moment oranı, n_i i istasyonundaki veri uzunluğu, V_1 $L-Cv$ oranlarını temel alan ağırlıklı standart sapma olarak ifade edilebilir. Bu çalışmada homojenlik ölçütünü değerlendirmek için, iki ve üç parametrelili dağılımlar yerine hidrolojik olayların frekans analizlerinde birçok dağılımı temsil etmesinden dolayı, güçlü bir dağılım olan dört parametrelili (dağılıma ait bu parametreler; konum parametresi, ölçek parametresi ve 2 adet şekil parametresi) kappada dağılımı kullanılmış ve bölgesel ortalama L-moment oranlarına uydurulmuştur. Kappada dağılımının avantajı, diğer iki veya üç parametrelili dağılımlardan daha az kısıtlayıcı olmasıdır. Kappada dağılımına ait toplam dağılım fonksiyonu $F(x)$ ve olasılık yoğunluk fonksiyonu $f(x)$ sırasıyla denklem (10) ve (11)'de verilmiştir [18].

$$F(x) = \left\{ 1 - \alpha \left[1 - \frac{\omega}{\delta} (x - \xi) \right]^{1/\omega} \right\}^{1/\alpha} \quad \omega \neq 0 \quad \alpha \neq 0 \tag{10}$$

$$f(x) = \frac{1}{\delta} \left[1 - \frac{\omega}{\delta} (x - \xi)^{1/k} \right]^{1/\omega - 1} \left\{ 1 - \alpha \left[1 - \frac{\omega}{\delta} (x - \xi) \right]^{1/\omega} \right\}^{1/\alpha - 1} \tag{11}$$

$$\begin{aligned}
 \xi + \delta / \omega \leq x < \infty & \quad \omega \leq 0 \text{ ve } \alpha < 0 \\
 \xi + \delta(1 - \alpha^{-\omega}) / \omega \leq x \leq \xi + \delta / \omega & \quad \omega > 0 \text{ ve } \alpha > 0 \\
 -\infty < x \leq \xi + \delta / \omega & \quad \omega \leq 0 \text{ ve } \alpha > 0 \\
 \xi + \delta(1 - \alpha^{-\omega}) / \omega \leq x < \infty & \quad \omega > 0 \text{ ve } \alpha < 0
 \end{aligned} \tag{12}$$

Burada, olasılık yoğunluk fonksiyonu, ξ kapa dağılımına ait konum parametresi, δ kapa dağılımına ait ölçek parametresi, ω ve α kapa dağılımına ait şekil parametrelerini göstermektedir. Bu çalışmada kapa dağılımı, bölgesel ortalama L-moment oranları yardımıyla 500 homojen veri bölgesi elde etmek için kullanılır ve simülasyon gerçekleştirilir. Üretilen her bir bölge için değişim hesaplanır ve bu değişimlerin standart sapması ve ortalaması belirlenir. Bu hesaplamalar sonucunda homojenliğin test edilmesi için H ölçütü denklem (13) yardımıyla elde edilir.

$$H_k = \frac{V_k - \mu_{V_k}}{\sigma_{V_k}} \quad k = 1, 2, 3 \tag{13}$$

Burada, H_k bölgesel homojenlik ölçütü (H_1, H_2, H_3), H_1 L-değişkenlik için bölgesel homojenlik ölçütü, H_2 L-çarpıklık için bölgesel homojenlik ölçütü, H_3 L-kurtosis için bölgesel homojenlik ölçütü, σ_{V_k} simülasyon ile elde edilen değerlerin standart sapması, μ_{V_k} simülasyon sonucu elde edilen değerlerin ortalaması, V_k bölgesel istatistikler ve verilerden hesaplanan ağırlıklı standart sapma (V_1, V_2 ve V_3) şeklinde yazılabilir. Hesaplanan H değerleri sonuçları; (i) Eğer $H < 1$ ise küme “*homojen kabul edilir*”, (ii) Eğer $1 \leq H < 2$ ise, küme “*homojen olabilir*”, (iii) Eğer $H \geq 2$ ise küme “*kesinlikle heterojen kabul edilir*” olarak değerlendirilmektedir [18,19].

4. ÇALIŞMA ALANI VE VERİ

Bu çalışmada, kümeleme analizi için Türkiye genelinde Devlet Meteoroloji İşleri (DMİ) Genel Müdürlüğü tarafından işletilen yağış gözlem istasyonlarından alınan yıllık toplam yağış verileri kullanılmıştır. Gözlem istasyonlarının ve kullanılacak verilerin belirlenmesi kümeleme analizi sonuçları üzerinde oldukça etkili olduğu için ölçüm yapılan yağış gözlem istasyonlarında, eksik veri durumu, gözlem süresi ve verilerin güvenilirliği incelenmiştir. Literatürde yapılan çalışmalar incelendiğinde, bölgesel tahmin çalışmalarında ve analizlerinde kullanılacak istasyonların istatistik olarak anlamlı veri sayısına sahip olması önerilmektedir [18]. Bu durum dikkate alınarak analizler için, 01.10.1967–30.09.1998 yılları arasında gözlem değerine sahip 188 yağış gözlem istasyonu seçilmiş ve bu istasyonların dağılımı Şekil 1’de gösterilmiştir. Literatürde, farklı ölçeklere sahip değişkenlerin kümeleme sonuçlarını etkilediği belirtilmiş ve verilerin uygun dönüşüm fonksiyonları ile normalize edilmesi gerektiği önerilmiştir [20, 21]. Bu çalışmada da, veriler

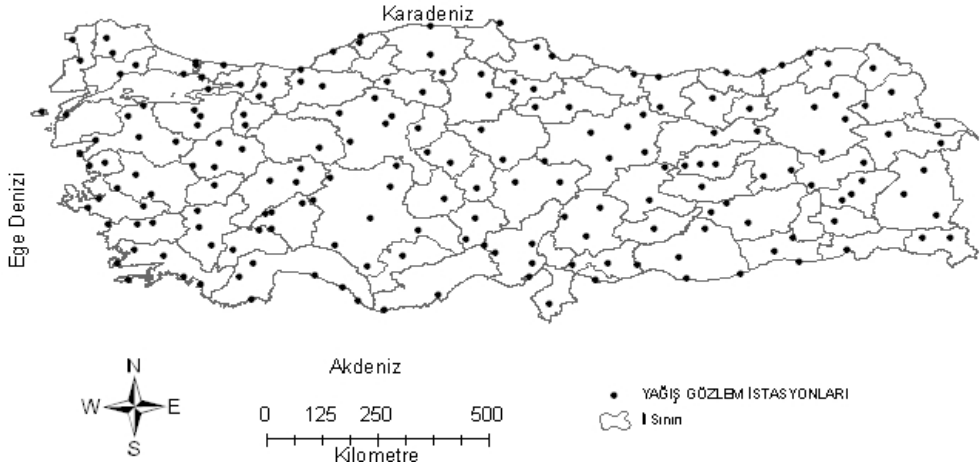
kümeleme analizinde kullanılmadan önce denklem (14), (15) ve (16) 'da verilen dönüşüm fonksiyonları kullanılarak normalize edilmiştir.

$$P_{yi} = (P_i - P_{\min}) / (P_{\max} - P_{\min}) \quad (14)$$

$$Y_{yi} = (Y_i - Y_{\min}) / (Y_{\max} - Y_{\min}) \quad (15)$$

$$Z_{yi} = Z_i / Z_{\max} \quad (16)$$

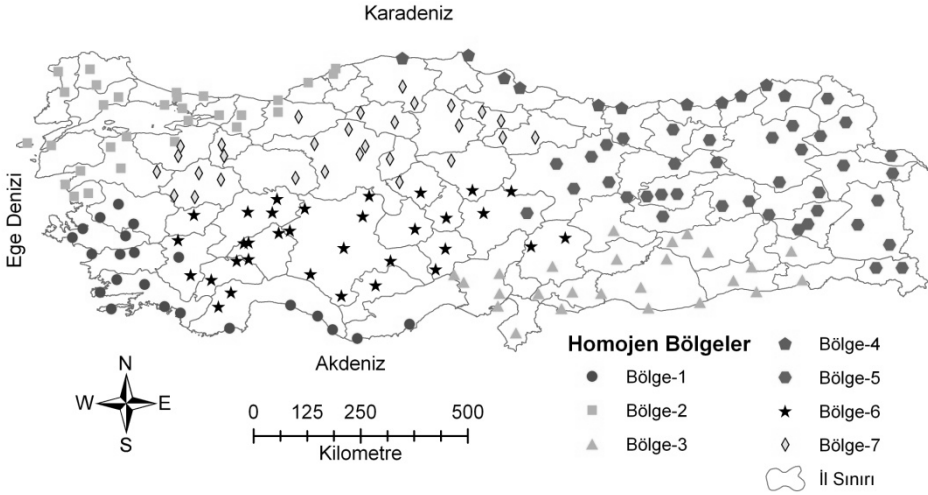
Burada P_i i istasyonundaki yağış, P_{yi} i istasyonunda normalize edilmiş yağış, P_{\max} maksimum yağış, P_{\min} minimum yağış, Y_i i istasyonuna ait enlem ya da boylam Y_{yi} i istasyonuna ait normalize edilmiş enlem ya da boylam, Y_{\max} maksimum enlem ya da boylam, Y_{\min} minimum enlem ya da boylam, Z_i i istasyonuna ait yükselti, Z_{yi} i istasyonuna ait normalize edilmiş yükselti, Z_{\max} maksimum yükseltidir.



Şekil 1. Kümeleme analizinde kullanılan yağış gözlem istasyonları

5. BULGULAR VE TARTIŞMA

K-Ortalamalar yöntemi ile yıllık toplam yağışların sınıflandırılmasında kullanılan veri seti, 188 adet özellik vektörü ve 4 adet değişkenden (188x4 matris) oluşmaktadır. Bu veri seti kullanılarak yapılan sınıflandırma sonucunda küme sayısı 7 olarak belirlenmiştir. Yağış serilerini sınıflandırılması ve homojen bölgelerin belirlenmesi için Matlab R.14 programı kullanılmıştır. Şekil 2'de K-Ortalamalar yöntemi ile belirlenen kümelerdeki istasyonların dağılımı gösterilmektedir.



Şekil 2. K-Ortalamalar yöntemi ile belirlenen bölgelerdeki istasyonların dağılımı

Şekil 2’de verilen sonuçlar incelendiğinde takriben, Bölge 1 Akdeniz bölgesinin orta ve batı kesimi ile Ege bölgesini, Bölge 2 Marmara bölgesi ile Karadeniz bölgesinin batı kesimini, Bölge 3 Güneydoğu Anadolu bölgesini, Bölge 4 Orta ve Doğu Karadeniz bölgesini, Bölge 5 Doğu Anadolu bölgesi ile Doğu Karadeniz bölgesinin iç kesimlerini, Bölge 6 İç Anadolu bölgesinin Güney kesimini ve Bölge 7 ise İç Anadolu bölgesinin kuzey kesimi ile Batı Karadeniz bölgesinin iç kesimlerini içermektedir. Belirlenen bölgelerde yer alan istasyonlara ait yıllık yağışların, bölgesel ortalaması $P_{ort}^R(mm)$ ve maksimum değeri $P_m^R(mm)$, standart sapmasının bölgesel ortalaması $\sigma_{ort}^R(mm)$ ve maksimum değeri $\sigma_m^R(mm)$ ve değişkenlik katsayısının bölgesel ortalaması C_{vort}^R ve maksimum C_{vm}^R hesaplanmış ve Çizelge 1’de verilmiştir. Kümeleme yöntemi ile belirlenen bölgelerin ve bu bölgelerde yer alan istasyonların bundan sonraki çalışmalarda kullanılabilmesi için bölgesel homojenlik ve için uyumsuzluk testinin uygulanması gerekmektedir. Belirlenen 7 bölgenin homojenliğinin test edilmesinde ve uyumsuzluk testinin uygulanmasında Hosking ve Wallis [18, 19] tarafından önerilen $H (H_1, H_2 \text{ ve } H_3)$ ve D değerleri hesaplanmış ve Çizelge 2’de verilmiştir.

İstasyonlar için hesaplanan uyumsuzluk testi sonuçlarının değerlendirilirken, kümede istasyon sayısının 15’ten fazla olması durumunda, 3’ten büyük D değerine sahip olan istasyonların uyumsuzluk testini geçemediği için bundan sonraki bölgesel çalışmalar için değerlendirme dışı bırakılabileceği önerilmektedir [18, 19]. Çizelgede verilen sonuçlara göre, 17088, 17110, 17232, 17668, 17748 ve 17926 nolu istasyonlara ait D değerleri sınır değerden daha büyük çıkmıştır. Kümeler için bölgesel homojenlik testi sonuçları karşılaştırıldığında, sadece küme 7 için hesaplanan H_1 değeri 2’den büyük çıkarken, diğer kümelere ait H_1 değerlerinin 2’den küçük çıktığı görülmektedir. Küme 2, 3 ve 6 için

hesaplanan H_1 değeri 1'den küçük çıkmış ve bu kümeler “homojen kabul edilir” şeklinde değerlendirilmiştir. Küme 1, 2 ve 5 için H_1 değeri 1 ile 2 arasında değer aldığı için bu kümeler “*homojen olabilir*” şeklinde değerlendirilmiştir. Ancak küme 7 için hesaplanan H_1 değeri 2'den büyük olduğu için bu kümenin yeterli homojenliğe sahip olmadığı görülmektedir. Bu sonuçlara göre yıllık toplam yağışların sınıflandırılmasında ve homojen hidrolojik bölgelerin belirlenmesinde sadece bir küme için homojenlik sağlanamamış olup K-Ortalamlar yönteminin yıllık toplam yağışların sınıflandırılmasında genel anlamda başarılı olduğu söylenebilir. Bu çalışmada yağış serilerinin sınıflandırılmasında ve homojen bölgelerin belirlenmesinde K-Ortalamlar yönteminin sonuçlarını değerlendirmek ve karşılaştırma yapmak amacıyla hiyerarşik kümeleme yöntemi olan Ward yöntemi ile de çözüm yapılmıştır. Yapılan çözümlerde en uygun küme sayısı 7 olarak belirlenmiş ve her bir küme için bölgesel homojenlik testi uygulanmıştır. Ward yöntemi ile belirlenen bölgeler için bölgesel homojenlik testi ve uyumsuzluk testi sonuçları Çizelge 3'te verilmiştir.

Çizelge 1. K-Ortalamlar yöntemi ile belirlenen kümeler için hesaplanan istatistikler

Bölge	İstasyon Sayısı	$P_{ort}^R (mm)$	$P_m^R (mm)$	$\sigma_{ort}^R (mm)$	$\sigma_m^R (mm)$	C_{vort}^R	C_{vm}^R
1	22	745.25	1167.99	254.80	169.53	0.23	0.29
2	30	729.37	1211.48	216.95	145.78	0.20	0.30
3	25	645.04	1099.73	267.50	166.76	0.26	0.34
4	11	1204.04	2203.28	286.16	158.83	0.14	0.21
5	38	562.19	1287.51	311.19	117.71	0.20	0.29
6	33	452.94	745.00	151.37	90.29	0.20	0.31
7	29	517.92	1490.18	274.93	93.05	0.18	0.23

Çizelge 2. K-Ortalamlar Yöntemi ile Belirlenen Kümeler için Bölgesel Homojenlik Testi

Bölge	İstasyon Sayısı	İstasyon-Yıl Sayısı	H Testi			Uyumsuzluk İstasyon No(D)
			H_1	H_2	H_3	
1	22	682	-1.532	-2.330	-1.872	17232 (3.40)
2	30	930	-0.462	-1.056	-0.884	17110 (3.08)
3	25	775	-0.600	-1.672	-1.588	-
4	11	341	1.854	-0.781	-0.726	-
5	38	1178	1.173	0.667	0.155	17088 (3.65) 17668 (3.43)
6	33	1023	-0.092	-0.799	-0.912	17926 (4.69)
7	29	899	-3.300	-3.556	-4.024	17748 (4.05)

Çizelge 3. Ward Yöntemi ile Belirlenen Kümeler için Bölgesel Homojenlik Testi

Bölge	İstasyon Sayısı	İstasyon-Yıl Sayısı	H Testi			Uyumsuzluk İstasyon No(D)
			H_1	H_2	H_3	
1	14	434	2.370	-0.825	-0.962	-
2	42	1302	1.042	0.393	-0.029	17088 (3.58) 17172 (3.09) 17668 (3.53)
3	26	806	-0.303	-0.925	-0.703	-
4	19	589	-3.044	-3.015	-3.312	-
5	41	1271	-0.636	-1.704	-1.775	17824 (3.13) 17926 (5.48)
6	27	837	-1.517	-2.168	-1.795	-
7	19	589	-1.503	-2.173	-2.115	-

Ward yöntemi sonuçları incelendiğinde, bölge 1 ve 4 için hesaplanan H_1 değerinin sınır değerden büyük çıktığı için bu kümeler “kesinlikle heterojen kabul edilir” şeklinde değerlendirilmiştir. Diğer bölgelere ait H_1 değerlerine bakıldığında, bölge 3 ve 5 “homojen kabul edilir”, bölge 2 ve 6 ise “homojen olabilir” şeklinde değerlendirilmiştir. Çizelge 2’de verilen sonuçlara göre, K-Ortalamalar yönteminde, sadece bir bölge “heterojen kabul edilir”, 3 bölge “homojen kabul edilir” ve 3 bölge ise “homojen olabilir” şeklinde değerlendirilmiştir. Diğer taraftan Ward yönteminde, 2 bölge “homojen kabul edilir”, 2 bölge “heterojen kabul edilir” 3 bölge ise “homojen olabilir” olarak belirlenmiştir. Bu sonuçlara göre, K-Ortalamalar yönteminde sadece 1 bölge “heterojen” olarak değerlendirilirken, Ward yönteminde ise 2 bölge “heterojen” olarak tespit edilmiştir.

6. SONUÇ

Bu çalışmada, Türkiye genelinde yıllık toplam yağışların sınıflandırılması ve homojen bölgelerin belirlenmesi amacıyla K-Ortalamalar yöntemi uygulanmıştır. Bunun için Türkiye genelinde yeterli ölçüm uzunluğuna sahip 188 adet yağış gözlem istasyonuna ait veriler kullanılarak yapılan sınıflandırma sonucunda küme sayısı 7 olarak belirlenmiştir. Belirlenen kümeler için uyumsuzluk testini ve L-momentler yöntemini temel alan bölgesel homojenlik testi uygulanmıştır. Uyumsuzluk testi sonucunda 6 istasyon için hesaplanan D değerlerinin sınır değerden daha büyük çıktığı görülmüştür. Bölgesel homojenlik testi sonuçlarına göre, sadece küme 7 için hesaplanan H_1 değeri 2’den büyük çıktığı için bu küme “heterojen kabul edilir” şeklinde değerlendirilmiş ve bu kümenin yeterli homojenliğe sahip olmadığı görülmüştür. Diğer kümelere ait H_1 değerleri 2’den küçük çıkmış ve küme 2, 3 ve 6 “homojen kabul edilir”, küme 1, 2 ve 5 ise “homojen olabilir” şeklinde değerlendirilmiştir. Bu çalışmada ayrıca, Ward yöntemi ile yağış serileri sınıflandırılmış ve belirlenen bölgeler için bölgesel homojenlik testi uygulanmıştır. Her iki kümeleme yöntemi

ile belirlenen bölgeler için uygulanan bölgesel homojenlik testi sonuçlarına göre, yağış serilerinin sınıflandırılmasında K-Ortalamalar yönteminde sadece 1 bölge “heterojen” olarak değerlendirilirken, Ward yönteminde ise 2 bölge “heterojen” olarak tespit edilmiştir.

Bu değerlendirmelere göre, yıllık toplam yağışların sınıflandırılmasında K-Ortalamalar yöntemi ile elde edilen sonuçların kabul edilebilir seviyede olduğu söylenebilir.

Teşekkür

Bu çalışma 107Y318 nolu proje kapsamında TÜBİTAK tarafından desteklenmiştir. Yazarlar, desteğinden dolayı TÜBİTAK’a, veri ve bilgi noktasında katkılarından dolayı Devlet Meteoroloji İşleri Genel Müdürlüğüne (DMİ) ve değerli görüş ve önerilerinden dolayı Editör ve danışmanlara teşekkür etmektedir.

Semboller

- A : Kovaryans matrisi
 C : Küme sayısı
 d : Değişken sayısı
 d_{ik}^2 : i . kümeye ait k . özellik vektörü ile i . küme merkezi arasındaki mesafe
 D_i : Uyumsuzluk ölçütü
 $F(x)$: Toplam dağılım fonksiyonu
 $f(x)$: Olasılık yoğunluk fonksiyonu
 H_k : Bölgesel homojenlik ölçütü (H_1, H_2, H_3)
 H_1 : L-Cv için bölgesel homojenlik ölçütü
 H_2 : L-Cs için bölgesel homojenlik ölçütü
 H_3 : L-Ck için bölgesel homojenlik ölçütü
 $L - Cv$: L-Moment oranı (L-değişkenlik)
 $L - Cs$: L-Moment oranı (L-çarpıklık)
 $L - Ck$: L-Moment oranı (L-Kurtosis)
 $L - Cv^{(i)}$: i istasyonu için L-moment oranı (L-değişkenlik),
 $\overline{L - Cv}$: Bölgesel ortalama L-moment oranı
 N : Özellik vektör sayısı
 n : İstasyondaki ait veri uzunluğu
 P_i : i istasyonundaki toplam yağış

- P_{yi} : i istasyonunda normalize edilmiş yağış
- P_{mak} : Maksimum yağış
- P_{min} : Minimum yağış
- P_{ort}^R : Kümeleme ile belirlenen bölgelerde yer alan istasyonlara ait yıllık yağışların bölgesel ortalaması
- P_m^R : Kümeleme ile belirlenen bölgelerde yer alan istasyonlara ait yıllık yağışların en büyük değeri
- S : küme merkezlerini içeren matris
- s_i : i . Kümenin merkezi
- T : Vektör ya da matrisinin transpozesi
- $t^{(i)}$: i istasyonuna ait $L-Cv$ L-moment oranı
- $t_3^{(i)}$: i istasyonuna ait $L-Cs$ L-moment oranı
- $t_4^{(i)}$: i istasyonuna ait $L-Ck$ L-moment oranı
- t^R : $L-Cv$ L-moment oranının bölgesel ortalaması
- t_3^R : $L-Cs$ L-moment oranının bölgesel ortalaması
- t_4^R : $L-Ck$ L-moment oranının bölgesel ortalaması
- u_i : i istasyonuna ait L-moment oranlarını içeren vektör
- \bar{u} : L-moment oranlarının bölgesel ortalaması
- V_k : Bölgesel istatistikler ve verilerden hesaplanan ağırlıklı standart sapma
- V_1 : $L-Cv$ oranlarını temel alan ağırlıklı standart sapma
- V_2 : $L-Cs$ oranlarını temel alan ağırlıklı standart sapma
- V_3 : $L-Ck$ oranlarını temel alan ağırlıklı standart sapma
- Y_i : i istasyonuna ait enlem ya da boylam
- Y_{yi} : i istasyonuna ait normalize edilmiş enlem ya da boylam
- Y_{mak} : Maksimum enlem ya da boylam
- Y_{min} : Minimum enlem ya da boylam
- Z_i : i istasyonuna ait yükselti
- Z_{yi} : i istasyonuna ait normalize edilmiş yükselti

- Z_{mak} : Maksimum yükselti
- X : Veri seti
- x_k : Veri setinde k . özellik vektörü
- $x_k^{(i)}$: i . kümeye ait k . özellik vektörü
- σ_{V_k} : Simülasyon ile elde edilen değerlerin standart sapması
- σ_{ort}^R : Kümeleme ile belirlenen bölgelerde yer alan istasyonlara ait yıllık yağışların standart sapmasının bölgesel ortalaması
- σ_m^R : Kümeleme ile belirlenen bölgelerde yer alan istasyonlara ait yıllık yağışların standart sapmasının en büyük değeri
- C_{vort}^R : Kümeleme ile belirlenen bölgelerde yer alan istasyonlara ait yıllık yağışların değişkenlik katsayısının bölgesel ortalaması
- C_{vm}^R : Kümeleme ile belirlenen bölgelerde yer alan istasyonlara ait yıllık yağışların değişkenlik katsayısının en büyük değeri
- μ_{V_k} : Simülasyon sonucu elde edilen değerlerin ortalaması
- ω ve α : Kappa dağılımına ait şekil parametreleri
- ξ : Kappa dağılımına ait konum parametresi
- δ : Kappa dağılımına ait ölçek parametresi

Kaynaklar

- [1] Mosley M. P., Delimitation of New Zealand Hydrologic Regions. Journal of Hydrology, 49, 173–192, 1981.
- [2] Acreman, M.C., Sinclair, C. D., Classification of Drainage Basins According to Their Physical Characteristics, An Application for Flood Frequency Analysis in Scotland. Journal of Hydrology, 84(3-4), 365-380, 1986.
- [3] Burn, D.H., Cluster Analysis as Applied to Regional Flood Frequency. Journal of Water Resources Planning and Management, 115, 567–582. 1989.
- [4] Burn, D. H., Catchment Similarity for Regional Flood Frequency Analysis using Seasonality Measures. Journal of Hydrology, 202, 212–230, 1997.
- [5] Lecce, S.A., Spatial Variations in the Timing of Annual Floods in the Southeastern United States. Journal of Hydrology, 235, 151–169, 2000.
- [6] Burn, D.H., Goel, N.K., The Formation of Groups for Regional Flood Frequency Analysis. Hydrological Sciences Journal, 45(1), 97–112, 2000.

- [7] Burn, D. H., Zrinji, Z., and Kowalchuk, M., Regionalization of Catchments for Regional Flood Frequency Analysis. *Journal of Hydrologic Engineering*, 2(2), 76–82, 1997.
- [8] Guttman, N.B., The use of L-Moments in the Determination of Regional Precipitation Climates. *Journal of Climate* 6, 2309–2325, 1993.
- [9] Soltani, S., Modarres, R., Classification of Spatio-Temporal Pattern of Rainfall in Iran Using A Hierarchical and Divisive Cluster Analysis. *Journal of Spatial Hydrology*, 6(2), 1-12, 2006.
- [10] Demirel, M.C. 2004. Cluster Analysis of Streamflow Data over Turkey. Master of Science Thesis. İstanbul Technical University, 119p.
- [11] Turan, A., Türkiye Akarsu Verimlerinin Küme Analizi ile Sınıflandırılması. Sakarya Üniversitesi, Fen Bilimleri Enstitüsü Yüksek Lisans Tezi, 155s, 2005.
- [12] Kahya, E., Demirel, M.C., Piechota, T.C., Spatial Grouping of Annual Streamflow Patterns in Turkey. *Hydrology Days*, 169-176, 2007.
- [13] Demirel, M.C., Mariano, A.J., Kahya, E., Performing K-Means Analysis to Drought Principal Components of Turkish Rivers. *Hydrology days*, 145-151, 2007.
- [14] Isik, S., Singh, V.P., Hydrologic Regionalization of Watersheds in Turkey. *Journal of Hydrologic Engineering*. 13(9), 824-834, 2009.
- [15] Lin, G-F., Chen, L-H., Identification of Homogeneous Regions for Regional Frequency Analysis using the Self-Organizing Map. *Journal of Hydrology*, 324, 1-9, 2006.
- [16] Rao, A., Srivinas, V.V., Regionalization of Watersheds by Fuzzy Cluster Analysis. *Journal of Hydrology*, 318, 57-79, 2006.
- [17] Hall, M.J., Minns, A.W., The Classification of Hydrologically Homogeneous Regions. *Hydrological Sciences Journal*, 44(5), 693-704, 1999.
- [18] Hosking, J.R.M., Wallis, J.R., *Regional Frequency Analysis: An Approach Based on L-Moments*. Cambridge University Press, Cambridge.1997.
- [19] Hosking, J.R.M., Wallis, J.R., Some Statistics Useful in Regional Frequency Analysis. *Water Resources Research*, 29 (2), 271–281, 1993
- [20] Cannarozzo, M., Noto, L.V., Viola, F., La Loggia, G., Annual Runoff Regional Frequency Analysis in Sicily. *Physics and Chemistry of the Earth*, 34, 679–687, 2009.
- [21] Lim, Y.H., Voeller, D.L., Regional Flood Estimations in Red River using L-Moment-Based Index-Flood and Bulletin 17B Procedures. *Journal of Hydrologic Engineering*, 14(9), 1002-1016, 2009.