

AKÜ FEMÜBİD 21 (2021) 047101 (958-970)

AKU J. Sci. Eng. 21 (2021) 047101 (958-970)

DOI: 10.35414/akufemubid.870649

Araştırma Makalesi / Research Article

## Havzaların Benzerliklerini Tanımlamada Alternatif Bir Yaklaşım: Hiyerarşik Kümeleme Yöntemi Uygulaması

Ersin AYTAÇ<sup>1</sup><sup>1</sup> Zonguldak Bülent Ecevit Üniversitesi, Mühendislik Fakültesi, Çevre Mühendisliği Bölümü, Zonguldak.e-posta: ersin.aytac@beun.edu.tr ORCID ID: <http://orcid.org/0000-0002-7124-4438>

Geliş Tarihi: 29.01.2021

Kabul Tarihi: 31.07.2021

### Öz

Makine öğrenmesi yöntemleri günümüzde birçok alanda kullanımını yaygınlaştırmış ve yerini sağlamlaştırmıştır. Denetimli, denetimsiz ve takviyeli öğrenme olmak üzere üç ana kola ayrılan makine öğrenmesi süreçleri, araştırmacıların gözle fark edemediği bağlantıları bulmada veya uzun süreli hesaplama gerektiren durumlarda ön plana çıkmaktadırlar. Denetimsiz öğrenme yöntemleri, etiketlerinin bulunmadığı verilerdeki kalıpları veya yapıyı keşfetmek için kullanılan makine öğrenmesi yöntemleridir. Hiyerarşik kümeleme süreci en önde gelen denetimsiz öğrenme yöntemlerinden birisidir. Bu çalışma havzaların benzerliklerini tanımlama da kullandığımız sürece alternatif bir yöntem sunmak amacıyla yürütülmüştür. Önerilen yöntemin avantajları arasında veri setinde yer alan tüm havzaların birbirleri ile olan ilişkilerini ortaya koyması, veri setindeki gürültüye daha az duyarlı olması, az havza içeren uygulamalarda daha kullanışlı olması ve küme içi tutarlılığı sağlamada araştırmacıya esneklik tanınmasıdır. Çalışmada Türkiye'nin kuzeyinde bulunan bazı havzaların hidrolojik müdahale birimleri (HRU) görüntüleri ve hiyerarşik kümeleme yaklaşımı kullanılarak kümelenmesi incelenmiştir ve birbirine en çok benzeyen iki havzanın Ereğli ve Çaykırı havzaları olduğu anlaşılmıştır. Havzaların birbirleri ile olan ilişkilerini ortaya çıkarmak için mesafe matrisi hazırlanmıştır. Ayrıca bağımsız kümeler oluşturmak için dendrogramın kesme mesafesi seçiminde dört farklı istatistiksel yaklaşım kullanılmıştır. İstatistiki yöntemlerin önerdiği küme sayıları içerisine kalmak şartı ve küme içi homojenliği korumak amacıyla 6 ayrı küme oluşturulmuş ve havzaların kümelere bağlı dağılımı gösterilmiştir. Bu çalışma havzaların HRU görüntülerine göre hidrolojik benzerliklerine dayanarak kümelenmelerinde alternatif bir bakış açısı sunmaktadır.

### Anahtar kelimeler

Hiyerarşik kümeleme;  
Hidrolojik müdahale birimi; Kesme yüksekliği; Denetimsiz öğrenme; Havza

## An Alternative Approach in Defining the Similarity of Catchments: Application of Hierarchical Clustering Method

### Abstract

Machine learning methods have widespread their use and strengthened their places in many areas. These procedures can be divided into three main branches as supervised, unsupervised, and reinforced learning and assist researchers to find connections in the data that cannot be seen or in situations that require long-term computation. Unsupervised methods are used to discover patterns or structures in data which does not have any labels. The hierarchical clustering process is one of the leading unsupervised methods. This study is an alternative approach to the process we used in defining the similarities of basins. The advantages of the proposed method are; reveals the relations between all basins, is less sensitive to noise in the data set, is more useful in applications with fewer basins, and is flexible in ensuring intra-cluster consistency. Some basins located in the north of Turkey based on their hydrological response units (HRU) images were hierarchically clustered and found out that and the two basins most similar to each other are the Ereğli and Çaykırı catchments. A distance matrix was prepared to reveal the relations of the basins. Besides, to create independent clusters, four different statistical approaches were used to select the cut-off height of the dendrogram. To stay within the cluster numbers suggested by statistical methods and to ensure cluster homogeneity, 6 separate clusters were created. The distribution of the catchments depending on the clusters was illustrated. This study provides an alternative perspective for the clustering of basins based on HRU images.

### Keywords

Hierarchical clustering;  
Hydrological response unit; Cut-off height;  
Unsupervised learning;  
Catchment

## 1. Giriş

Makine öğrenmesi (MÖ), tahmine dayalı analiz süreçleri için ortaya çıkan veri analiz yöntemleri alanıdır (Han *et al.* 2020). Bu yöntem, insanlara sürekli çaba göstermeden veya çok az çaba göstererek yardımcı olmak için eğitilebilen bilgisayarları kapsar (Ramkumar *et al.* 2019). Makine öğrenimi algoritmaları, sağlık, ulaşım, konuşma analizi, bilgisayarla görme, pazar analizi, yaşam bilimleri ve diğerleri dahil olmak üzere sayısız uygulama alanında geniş bir ölçekte yerini almıştır (Loreti *et al.* 2020). En basit haliyle MÖ, bir sonucu tahmin etmek veya öngörmek için gerçek dünya verilerinin çeşitli algoritmalar ile işlenmesidir (Haeberle *et al.* 2019). Makine öğrenmesi üç temel kategoriye ayrılır, denetimli öğrenme (supervised learning), denetimsiz öğrenme (unsupervised learning) ve takviyeli öğrenme (reinforcement learning,) (Castellanos-Garzón *et al.* 2019). Denetimli öğrenme yaklaşımları, çıktı değerlerini hesaplamak için çok değişkenli veri kümeleri kullanır. Denetimsiz prosedürler kullanıcı tarafından belirtilmeyen istatistiklerin birleştirilmesine izin veren modelleri geliştirme amacı güder (Sahu *et al.* 2020). Takviye öğrenmede ise, akıllı süreçler hedeflerine ulaşmak için dinamik bir ortamda birbirleriyle etkileşime girerler (Castellanos-Garzón, Costa *et al.* 2019). Karar ağaçları (decision trees), yapay sinir ağları (artificial neural networks), destek vektör makineleri (support vector machines), *k-ortalamalar* (*k-means*), hiyerarşik kümeleme (hierarchical clustering), gürültü ile uygulamaların yoğunluk temelli mekânsal kümelenmesi (Density-based spatial clustering of applications with noise) öne çıkan makine öğrenmesi yöntemleridir (Kotsiantis 2007, Wang *et al.* 2019).

Denetimsiz öğrenme, işlemlerin etiketi hakkında bilgi gerektirmeyen, veri dağılımını karakterize etmeyi amaçlayan ve son yıllarda giderek daha fazla araştırmacının odak noktası olmaya başlayan makine öğrenmesi tekniklerindedir (Carcillo *et al.* 2019, Tang *et al.* 2019). Hiyerarşik kümeleme algoritmaları, faktör analizi, gizli modeller ve sapkın gözlemlerin algılamasındaki son gelişmeler,

denetimsiz MÖ tekniklerinde önemli ölçüde ilerleme kaydedilmesine yardımcı olmuştur (Usama *et al.* 2019). Kümeleme yöntemleri denetimsiz öğrenmede öne çıkan yöntemlerdendir. Bu tekniklerin hedefi, benzer verileri kümelere ayırarak tanımlamaktır (Ferreira *et al.* 2019). Kümeleme algoritmaları sayesinde araştırmacılar bilinmeyen, ancak yararlı olan sınıflandırmaları keşfetme, sunulan bilgilerin altında yatan ve fark edilemeyen yapıyı ortaya çıkarma şansı bulmuştur (Kotsiantis 2007, Castro *et al.* 2018). Kümeler, benzerlik veya içsel özelliklere dayanarak sınırlandırılır, böylece aynı kümedeki örnekler benzerken, farklı kümelerdeki örnekler ayrı özellik gösterir (Chen *et al.* 2019). *k-ortalamalar*, kendi kendini düzenleyen harita (self-organizing map), karışım modeli (mixture model) hiyerarşik kümeleme (hierarchical clustering) (HK) ve bulanık C-ortalamaları (fuzzy C-means) denetimsiz öğrenmede kullanılan temel kümeleme aileleridir (Abdalla *et al.* 2019). Hiyerarşik kümeleme, kalıplar arasındaki yakınlığa göre değişken kümeleri (R-modu) veya gözlemler (Q-modu) hiyerarşisini ayarlar (Yang *et al.* 2019). Hiyerarşik kümeleme yöntemlerinde, veriler dendrogram adı verilen ağaçsı yapılar olarak kategorize edilir. Tüm veriler dendrogramdaki kök ve düğümler ile bağlanır, düğümlerdeki yapraklar verileri ifade eder ve orta düğümler verilerin birbirine benzerliğini gösterir (Jafarzadegan *et al.* 2019). Hiyerarşik kümelemenin avantajları arasında, önceden küme sayısının belirtilmesini gerektirmemesi, dendrogram adı verilen grafiksel ara yüze sahip olması ve değişken şekil ve büyüklükteki kümeleri tespit etme yeteneğine sahip olması gelmektedir (Govender and Sivakumar 2020).

Son 30 yılda, Soil and Water Assessment Tool (SWAT) dünya çapında en yaygın kullanılan havza ölçekli modellerden biri haline gelmiştir (Wu *et al.* 2020). SWAT yazılımı, havza yönetimi uygulamalarında havzadaki suyun hem miktarını hem de kalitesini ölçebilir (Gungor and Goncu 2013). SWAT, günlük veya aylık bir zaman adımında ve uzun vadeli simülasyonlar için yıllık adımlarla çalışabilen,

mekansal olarak dağıtılmış, fiziksel tabanlı bir hidrolojik modeldir. 20 yıllık macerasından sonra yeni kodlama özellikleriyle SWAT+ olarak adlandırılan tamamen yeniden yapılandırılmış bir SWAT sürümü yayınlanmıştır. SWAT+, mekansal temsil açısından SWAT'tan daha esnektir ve ilgili modüler kodlar, gelecekte eklenecek uygulamalar ve genel kullanıcılar için geliştirmeyi kolaylaştırmak üzere tasarlanmıştır (Wu *et al.* 2020).

Sayısal yükseklik modelleri, arazi kullanımı ve toprak tipleri programın ihtiyaç duyduğu veri katmanlarıdır (Wangpimool *et al.* 2013). Bu üç haritanın kombinasyonları oluşturularak incelenecek olan havza, hidrolojik müdahale birimleri (hidrolojik response units) (HRUs) adı verilen birbirine benzemeyen parçalara ayrılır (Asres and Awulachew 2010, Rocha *et al.* 2015). Model, her bir HRU için yüzey akışı, taban akışı, evapotranspirasyon ve toprak nemi değişimi gibi ilgili hidrolojik bileşenleri tahmin ederek çalışır (Song and Zhang 2012). SWAT modelinde havzadaki suyun hareketi, HRU'lardan alt havza seviyesine ve daha sonra akarsu ağı yoluyla havza çıkışına simüle edilir (Zhou *et al.* 2013). SWAT modelinin bir havzayı HRU'lara bölmesinin bazı avantajları vardır. Bunlardan ilki hidrolojik bütçe, bitki büyümesi, erozyon, böcek ilacı dinamiği, tarımsal yönetim ve besin döngüsü simülasyonlarının daha kolay yapılabilmesidir. Yine bu küçük birimler sayesinde akarsulara su, besin ve tortu akışı günlük zaman ölçeğinde elde edilebilir ve kontrol edilebilir (Sheshukov *et al.* 2016). Ayrıca HRU'lar hidrolojik süreçlerin jeomorfolojik özelliklerini temsil eder. HRU oluşturmanın diğer bir avantajı ise HRU'ların enkaz akış riskini değerlendirmek için temel birimler olarak kullanılabilir olmalarıdır (Zou *et al.* 2019).

Bu çalışma, daha önce havzaların benzerliklerini tanımlama da kullandığımız yöntem (Aytaç 2020) alternatif olarak farklı bir denetimsiz öğrenme yaklaşımı olan hiyerarşik kümeleme yöntemi ile Türkiye'nin kuzeyinde bulunan bazı havzaların HRU görüntüleri kullanılarak kümelenmesinin yapılabirliği incelenmiştir. Literatürde yer alan çalışmamızdan farklı olarak hiyerarşik kümeleme yöntemin avantajları şu şekildedir; hiyerarşik

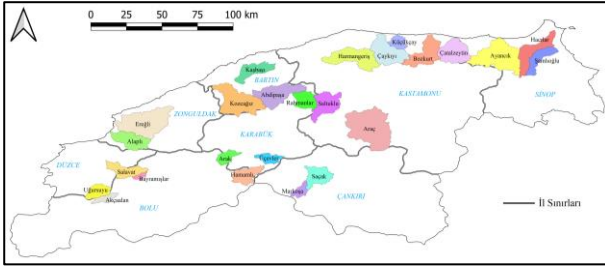
kümeleme yöntemi, yalnızca nesnelerin tek bir bölümünü değil, bir kümeleme hiyerarşisi oluşturur. Bu durum araştırmacılara veri setinde bulunan tüm verilerin birbirleri ile olan ilişkilerini gözlemleme şansı tanır ve verilerdeki benzerliği keşfetmeye yardımcı olabilecek daha küçük kümeler oluşturabilmelerini sağlar. Hiyerarşik kümeleme yöntemi veri setindeki gürültüye daha az duyarlıdır, böylece HRU'ların çok olduğu havzalarda daha iyi kümeleme yaklaşımı elde etmektedir. Yine bu yöntem küçük veri setlerinin işlenmesinde daha kullanışlıdır. Literatürde yer alan yöntemimizde kümeleme işlemi, önceden belirlenen bir küme sayısı ( $k$ ) gerektirirken, yeni önerilen yöntemde kümeleme süreci araştırmacı tarafından istenilen seviyede (veya kümelerde) yapılabilir. Bu durum araştırmacıya küme içi tutarlılığı artırmada esneklik sunmaktadır (Kaushik and Mathur 2014, Reddy *et al.* 2017).

Bu çalışmada sunulan yöntem ile ilgili dikkat edilmesi gereken en önemli nokta, bunun bir sınıflandırma (classification) işlemi olmadığı ve elde edilen kümelerin havzaların önceden belirlenmiş hidrolojik özelliklerine göre bir sıraya dizilmediğidir. Bu yöntem, gizli kalıplardan yeni bilgiler öğrenmeyi amaçlar ve havzalar arasında bir ilişki olup olmadığını bulmaya çalışır (Aytaç 2020).

## **2. Materyal ve Metot**

### **2.1 Çalışma Alanı**

Kümelemesi yapılacak olan havzaları kapsayan alan Türkiye'nin kuzeyinde, DDS 40°8'39" – 42°5'52" N ile 30°33'35" – 35°27'2" E koordinatları arasında yer almakta ve Düzce, Bolu, Zonguldak, Bartın, Karabük, Kastamonu, Çankırı ve Sinop illerini içermektedir. Çalışmada kullanılmak üzere 25 adet havza oluşturulmuştur. Havzalara, ya havza içerisinde bulunan akarsuyun adı ya da havza çıkışında bulunan yerleşim biriminin ismi verilmiştir. Şekil 1'de çalışma alanı ve çalışmada kullanılan havzalar görülmektedir.



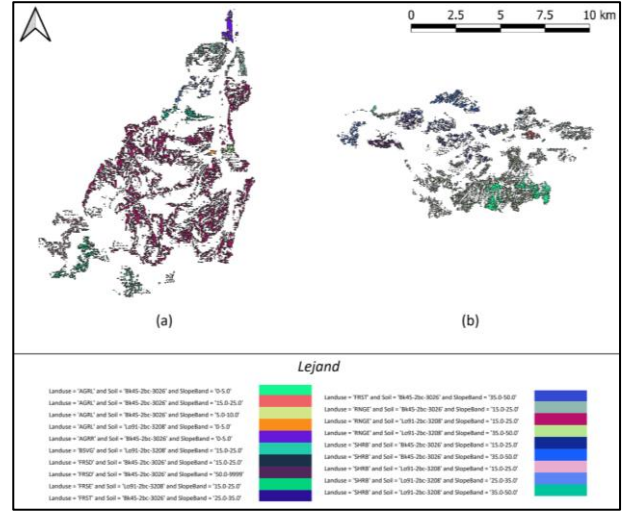
Şekil 1. Çalışma alanı ve çalışmada kullanılan havzaların konumları

## 2.2 Hidrolojik Müdahale Birimlerinin Görüntülerinin Oluşturulması

Havzaların HRU görüntülerinin oluşturulması süreci 3 basamakta gerçekleştirilmiştir. Birinci basamakta SWAT+ modeli kullanılarak her bir havzanın HRU'ları oluşturulmuştur. İkinci aşamada oluşturulan HRU tiplerine renk ataması yapılmıştır. Bu işlem için bir MS Excel makrosu kodlanmış ve kodlanan makro her bir özgün HRU'ya farklı bir renk atamıştır. Görüntü oluşturma sürecinin son aşaması HRU'larla ifade edilmiş ve renklendirilmiş havzaların şekil dosyalarının oluşturulmasıdır. Bu işlem için Quantum Geographic Information System (QGIS) programının baskı düzenleme seçeneği kullanılmıştır. Veri hazırlama sürecinin detaylarına Aytaç, E. (2020) [28] çalışmasından ulaşabilir. Standart hale getirilmiş görüntü dosyalarının özellikleri Çizelge 1'de verilmiştir. Şekil 2'de görüntü işlemede kullanılacak olan havzalardan ikisinin HRU şekilleri örnek olarak verilmiştir.

Çizelge 1. Oluşturulan görüntülerin özellikleri

Boyut (mm)	Çözünürlük (dpi)	Koordinat Sistemi (ESPG)	Ölçek	Biçim
500 x 500	500	32636 - WGS 84 / UTM bölgesi 36N	1/90000	TIFF



Şekil 2. (a) Markoşa ve (b) Üçevler havzalarının standartlaştırılmış HRU görüntüleri

Şekil 2'de de görüldüğü gibi havzalarda bulunan bazı HRU tipleri silinmiştir. Bu durum SWAT+ programının filtreleme özelliği kaynaklıdır. SWAT+ programı modelleme sürecini basitleştirmek için önemsiz (çok küçük ve havza hidrolojisi üzerinde az bir etkisi olan) HRU tiplerini silmektedir. Bu filtreleme süreci nedeniyle havza sınırları içerisinde sadece baskın HRU desenleri kalmıştır. Filtreleme işlemi SWAT+ modeliyle yapılmış bütün çalışmalarda kullanılmaktadır ve sonuçlar üzerindeki etkileri çok düşük olduğu için standart bir prosedürdür (Aytaç 2020).

## 2.3 Hiyerarşik Kümeleme Süreci

Çalışmada Orange veri madenciliği aracı (Int Kyn. 3) kullanılarak görüntü yerleştirme (embedding), mesafe belirleme ve hiyerarşik kümeleme süreçleri yürütülmüştür. Orange programı ilk olarak Janez Demšar ve Blaž Zupan tarafından kodlanan, Ljubljana Üniversitesi, Bilgisayar ve Bilişim Bilimleri Fakültesi Biyoenformatik Laboratuvarı tarafından geliştirilen ve 20 yılı aşkın bir süredir yeni makine öğrenme algoritmaları için bir test platformu görevi gören bir modeldir (Janez Demšar 2013, Naik and Samant 2016). Kümeleme süreci Orange programının görüntü yerleştirme bileşenine aktarılmasıyla başlamıştır. Görüntü yerleştirme sürecinde resimler görüntü analiz eden bir derin ağ ile incelenir ve vektörlerine ayrılarak sayısal olarak ifade edilirler (Huang *et al.* 2018, Godec *et al.* 2019). Orange görüntü yerleştirme biriminde görüntüler

Google'ın InceptionV3 derin öğrenme ağı ile işlenmektedir. InceptionV3 ImageNet verileri üzerinde eğitilmiştir ve süreç sonucunda görüntü 2048 farklı boyut ile ifade edilmektedir (Int Kyn. 1, Godec *et al.* 2019).

Görüntü yerleştirme biriminden elde edilen veriler mesafe birimine gönderilmiştir. Bu birim sayesinde veri kümesindeki satırlar veya sütunlar arasındaki mesafeler hesaplanmaktadır (Int Kyn. 2). Mesafelerin hesaplanması kosinüs benzerlik ölçümü (KBÖ) ile yapılır. KBÖ yönteminde mesafe hesaplaması iki vektör arasındaki açının kosinüsünü ölçerek gerçekleştirilir (Chen *et al.* 2019). İki vektörün benzerliği N-boyutlu uzayda matematiksel olarak Eş. 1'deki gibi ifade edilebilir (Crosta *et al.* 1998);

$$Benzerlik = \cos(\theta) = \frac{A \cdot B}{\|A\|_2 \cdot \|B\|_2} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}} \quad (1)$$

Burada;  $A$  ve  $B$  görüntülerin piksellerinin vektörüdür. Son aşama ise görüntülerin kümelendirilmesidir. Kümeler arasındaki mesafe metriği, Tam Bağlantı (Complete linkage) olarak ayarlanmıştır. Birbirinden en uzak iki nesne arasındaki mesafe olarak tanımlanan tam bağlantı terimi, diğer bir deyişle, iki nesne arasındaki en büyük farklılığı hesaplar (Janert 2010, Govender and Sivakumar 2020). Tam bağlantı yöntemi, tek bağlantı (single linkage) yönteminin jeokimyasal yapıları maskeleyebilen zincirleme etkisi veya ortalama bağlantının (average linkage) ek varsayımlar gerektirmesi gibi sınırlamalarının üstesinden gelebildiği için tercih edilmiştir (Yang, Grunsky *et al.* 2019).  $G_1$  ve  $G_2$  gibi iki küme içerisinde bulunan verilerin tam bağlantı denklemi aşağıdaki gibi ifade edilir (Eş. 2).

$$d_{G_1 G_2} = \max_{i \in G_1, j \in G_2} (d_{i,j}) \quad (2)$$

Ayrıca verilerin birbirlerinden farklılıkları saptamak için Orange programı ile mesafe matrisi üretilmiştir. Mesafe matrisi iki boyutlu, simetrik, negatif

olmayan bir matristir ve bir kümedeki elemanlar arasında ikili mesafe tahminlerine karşılık gelir (University of Ljubljana , Weyenberg and Yoshida 2015).

Orange veri madenciliği programında hiyerarşik kümeleme yönteminde kümelerin oluşturulma süreci aşağıdaki adımlarla yürütülmektedir.

1. Öğeler arasındaki mesafelerin ölçümü için bir mesafe ölçüm yöntemi seçin (Bu çalışmada KBÖ yöntemi seçilmiştir)
2. Kümeler arası mesafeler için bir mesafe metriği seçin (Bu çalışmada tam bağlantı seçilmiştir)
3. Her bir verinin bir küme sayıldığı N adet küme oluşturun ve aşağıdaki döngüyü çalıştırın:
  - Mevcut mesafe matrisini kullanarak birbirine en yakın iki veriyi bulun ve bu verileri kümeleyin. Bu kümenin düğümü iki veri arası uzaklık olacaktır.
  - Bu işlemi diğer verilere de uygulayarak verilerin birincil kümeleri oluşturun.
  - Kümelerin birbirlerine olan mesafelerini içeren yeni mesafe matrisi oluşturmak için bağlantı kriterini (bu çalışma için tam bağlantı yöntemi kullanılmış yani kümeler arasındaki bağlantı kümelerdeki en uzak iki veri arasında yapılmıştır) ve kosinüs benzerlik ölçütünü kullanın.
  - Yeni hesaplanan mesafelere göre mesafe matrisini güncelleyin.
4. Tüm öğeler tek bir kümede birleştirilene kadar bu döngüyü tekrarlayın (de Aguiar Neto *et al.* 2020).

Hiyerarşik kümeleme yöntemi her ne kadar birbirinden mutlak olarak ayrılmış kümeler yaratmaya çalışmasa da (Dumont *et al.* 2018), HK'de elde edilen dendrogram optimal küme sayısının belirlenmesine yardımcı olabilir. Fakat dendrogramın kesme yüksekliğine karar vermek zorlu bir süreçtir (Li *et al.* 2018, Govender and Sivakumar 2020). Araştırmacılar temel kümeler elde etmek için bir kesme yüksekliği uygulayarak, dendrogramın dikey çizgilerinin kesilme sayısını

sayabilirler. Ancak burada en önemli sorun, kesme noktasını belirlemede kesin ve güvenilir bir yöntem bulunmamasıdır (Zambelli 2016). Kesme yüksekliğini belirlemek için önerilen bazı istatistiksel yöntemler vardır. Farklı araştırmalarda dirsek (elbow), en büyük fark (maximum difference) ve mod (mode) yöntemleri sunulmuştur (Zambelli 2016, Samarasinghe *et al.* 2019). Ayrıca Dumond ve ark. (2018), kesme noktasını belirlemek için x ekseninde küme sayıları ve y ekseninde kümeler arası mesafelerin olduğu bir çizgi grafik kullanmıştır. Bu çalışmada kesme yüksekliğinin seçimi, grafikte eğiminin azalmaya başladığı nokta (dirsek formasyonu) olarak belirlenmiştir (Dumont, Reninger *et al.* 2018).

Dirsek yönteminde kümeler arası mesafe Eş. 3 ile ifade edilmektedir;

$$\{d_i \dots d_N\} = \{d_3 - 2d_2 - d_1, \dots, d_N - 2d_{N-1} - d_{N-2}\} \quad (3)$$

Burada;  $i$ ,  $i$ 'inci küme mesafesi ve  $N$  ise veri sayısıdır. Dirsek yöntemine göre küme sayısı ( $\hat{K}_E$ ) aşağıdaki eşitlikle hesaplanır (Eş. 4) (Samarasinghe, Gunawardena *et al.* 2019),

$$\hat{K}_E = N + 2 - \underset{i \in [3, N]}{\operatorname{argmax}} \{d_i - 2d_{i-1} - d_{i-2}\} \quad (4)$$

En büyük fark yönteminde ise dirsek yöntemine benzemektedir. Bu yöntemde kümeler arası mesafe Eş. 5 ile ifade edilir,

$$\{d_i \dots d_N\} = \{d_3 - d_2, \dots, d_N - d_{N-1}\} \quad (5)$$

Küme sayısı ( $\hat{K}_D$ ) ise aşağıdaki eşitlikteki gibi hesaplanır (Eş. 6),

$$\hat{K}_D = N + 2 - \underset{i \in [2, N]}{\operatorname{argmax}} \{d_i - d_{i-1}\} \quad (6)$$

Mod yöntemi, kümeler arasındaki mesafenin ampirik dağılımına bağlıdır. İlk olarak dağılımın modu ( $\hat{D}$ ) hesaplanır. Küme sayısı aşağıdaki eşitlikle belirlenebilir (Eş. 7).

$$\hat{K}_M = \hat{D} + \alpha \sigma_D \quad (7)$$

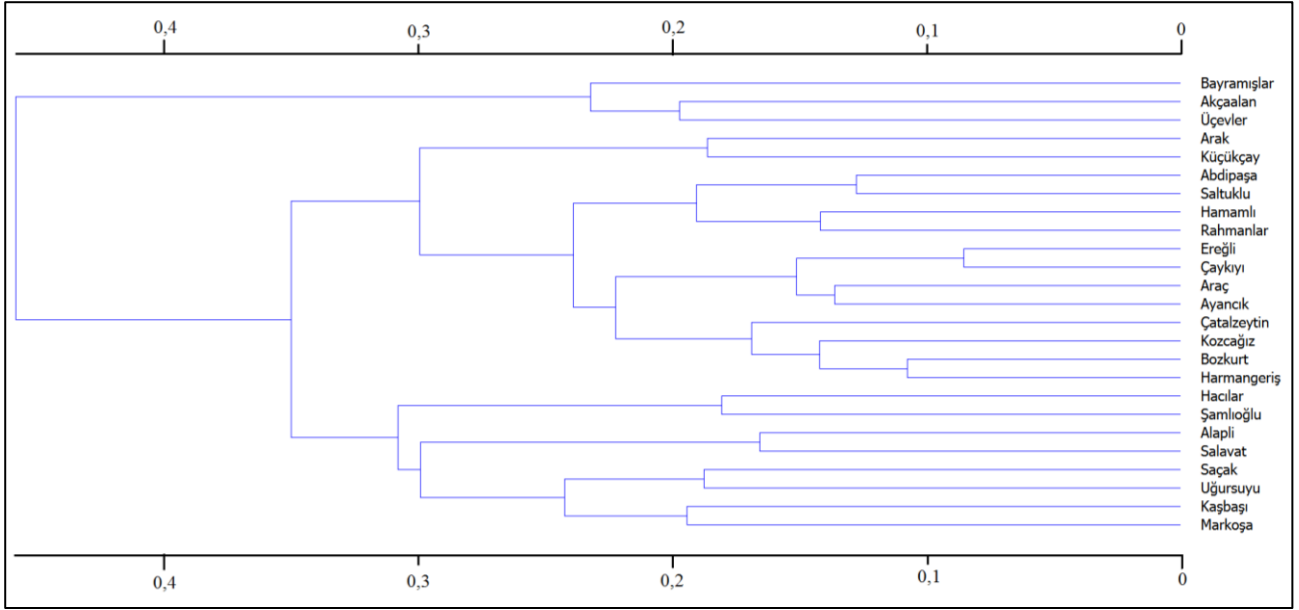
Burada;  $\sigma_D$  standart sapmadır ve  $\alpha$ , araştırmacı tarafından keyfi olarak seçilebilen bir parametredir (Zambelli 2016).

### 3. Bulgular ve Tartışma

Orange veri madenciliği aracı, hiyerarşik kümelemenin işleminin sonuçlarını bir dendrogramla ifade eder. Dendrogram, düğümlerin hiyerarşinin bölgelerini temsil ettiği ve küme yapısını görselleştirmeye yardımcı olan ağaç benzeri bir yapıdır (Li, Ma *et al.* 2018, Perret *et al.* 2019). Bir dendrogramda, düğümler arasındaki benzerlik, onları birleştiren bağlantının mesafeleriyle ifade edilir ve her düğümün koordinatı, nesnelere arasındaki farklılığı gösterir (Manning and Schütze 1999, Weyenberg and Yoshida 2015). Her düğüm kendi kümesinde başlar ve en küçük mesafeye sahip düğümler kademeli olarak birleştirilir (Yang *et al.* 2019). Ortaya çıkan dendrogram Şekil 3'de görülmektedir.

Şekil 3'de görüldüğü gibi, x eksenindeki birimler veri örnekleri arasındaki farkları sayısal olarak göstermektedir. Daha yüksek bir değer daha az benzerlik anlamına gelir. İki düğümü bağlayan kolun uzunluğu ne kadar küçük olursa, iki veri de birbirine o kadar benzerdir. Şekil 4'de en küçük düğümle birbirlerine bağlanmış ve birbirine en çok benzeyen iki havza Ereğli ve Çaykırı havzaları olarak karşımıza çıkmaktadır. Bu iki havza arasındaki mesafe 0,085 olarak ölçülmüştür. Bu havzaları benzerlik bakımından sırasıyla 0,107 ve 0,127 değerleriyle Bozkurt – Harmangeriş ve Abdipaşa – Saltuklu havzaları izlemektedir. Dendrogram ilk olarak iki kola daha sonra üç, dört ve en son her bir veri örneğinin bir grup oluşturacağı şekilde 22 kola ayrılarak kümelenebilir. Kolların ayrılma noktalarından orijine olan uzaklıkları, o kümelerin birbirlerine olan mesafelerini belirtmektedir.

Dendrogram ilk olarak 0,458 mesafesinde iki kola ayrılmıştır. Birinci kolda Bayramışlar, Akçaalan ve Üçevler havzaları küme oluşturmuş, diğer havzalar ise ikinci kolda ikinci kümeyi oluşturmuşlardır.



Şekil 3. Hiyerarşik kümeleme işlemi sonucu ortaya çıkan dendrogram

Orange veri madenciliği aracının veri setindeki her bir verinin birbirlerinden farklılığını ifade ettiği başka bir birimi ise mesafe matrisi birimidir. Şekil 4'de

Orange tarafından oluşturulan mesafe matrisi görülmektedir.

	Abdipaşa	Akçaalan	Alaplı	Arak	Araç	Ayancık	Bayramışlar	Bozkurt	Ereğli	Hacılar	Hamamlı	Harmangeriş	Kaşbaşı	Kozcağız	Küçükçay	Markoşa	Rahmanlar	Salavat	Saltuklu	Saçak	Uğursuyu	Çatalzeytin	Çaykırı	Üçevler	Şamlıoğlu
Abdipaşa		0.310	0.223	0.229	0.129	0.136	0.429	0.173	0.138	0.208	0.139	0.149	0.219	0.186	0.282	0.217	0.190	0.265	0.127	0.280	0.287	0.207	0.181	0.295	0.228
Akçaalan	0.310		0.391	0.341	0.331	0.340	0.214	0.252	0.341	0.309	0.299	0.308	0.312	0.270	0.351	0.282	0.277	0.311	0.323	0.283	0.239	0.316	0.298	0.197	0.282
Alaplı	0.223	0.391		0.282	0.264	0.178	0.441	0.175	0.190	0.277	0.274	0.173	0.240	0.152	0.293	0.280	0.267	0.166	0.219	0.236	0.247	0.201	0.220	0.354	0.292
Arak	0.229	0.341	0.282		0.257	0.257	0.387	0.240	0.245	0.233	0.227	0.271	0.241	0.267	0.186	0.197	0.263	0.343	0.195	0.223	0.309	0.212	0.151	0.294	0.255
Araç	0.129	0.331	0.264	0.257		0.136	0.458	0.154	0.111	0.215	0.193	0.185	0.265	0.222	0.299	0.243	0.176	0.307	0.154	0.290	0.309	0.212	0.151	0.294	0.255
Ayancık	0.136	0.340	0.178	0.257	0.136		0.442	0.144	0.108	0.227	0.194	0.140	0.274	0.150	0.282	0.242	0.187	0.203	0.195	0.225	0.300	0.175	0.146	0.293	0.282
Bayramışlar	0.429	0.214	0.441	0.387	0.458	0.442		0.379	0.449	0.343	0.382	0.448	0.332	0.359	0.414	0.262	0.424	0.387	0.407	0.294	0.267	0.398	0.432	0.232	0.348
Bozkurt	0.173	0.252	0.175	0.240	0.154	0.144	0.379		0.114	0.238	0.218	0.107	0.262	0.129	0.263	0.216	0.177	0.176	0.183	0.220	0.239	0.167	0.101	0.235	0.238
Ereğli	0.138	0.341	0.190	0.245	0.111	0.108	0.449	0.114		0.213	0.196	0.143	0.274	0.155	0.248	0.239	0.183	0.231	0.165	0.291	0.324	0.178	0.085	0.308	0.286
Hacılar	0.208	0.309	0.277	0.233	0.215	0.227	0.343	0.238	0.213		0.194	0.263	0.235	0.201	0.242	0.213	0.254	0.285	0.229	0.252	0.308	0.264	0.252	0.301	0.180
Hamamlı	0.139	0.299	0.274	0.227	0.193	0.194	0.382	0.218	0.196	0.194		0.223	0.223	0.199	0.200	0.179	0.142	0.294	0.161	0.251	0.289	0.226	0.239	0.236	0.227
Harmangeriş	0.149	0.308	0.173	0.271	0.185	0.140	0.448	0.107	0.143	0.263	0.223		0.303	0.142	0.279	0.272	0.189	0.185	0.221	0.261	0.279	0.160	0.137	0.284	0.258
Kaşbaşı	0.219	0.312	0.240	0.241	0.265	0.274	0.332	0.262	0.274	0.235	0.223	0.303		0.239	0.340	0.194	0.300	0.299	0.184	0.242	0.226	0.263	0.324	0.318	0.246
Kozcağız	0.186	0.270	0.152	0.267	0.222	0.150	0.359	0.129	0.155	0.201	0.199	0.142	0.239		0.242	0.197	0.215	0.166	0.197	0.240	0.246	0.169	0.165	0.276	0.232
Küçükçay	0.282	0.351	0.293	0.186	0.299	0.282	0.414	0.263	0.248	0.242	0.200	0.279	0.340	0.242		0.238	0.224	0.310	0.270	0.268	0.307	0.249	0.241	0.316	0.350
Markoşa	0.217	0.282	0.280	0.197	0.243	0.242	0.262	0.216	0.239	0.213	0.191	0.272	0.194	0.197	0.238		0.266	0.281	0.189	0.206	0.216	0.229	0.258	0.224	0.245
Rahmanlar	0.190	0.277	0.267	0.263	0.176	0.187	0.424	0.177	0.183	0.254	0.142	0.189	0.300	0.215	0.224	0.266		0.300	0.183	0.253	0.287	0.215	0.192	0.208	0.266
Salavat	0.265	0.311	0.166	0.343	0.307	0.203	0.387	0.176	0.231	0.285	0.294	0.185	0.299	0.166	0.310	0.281	0.300		0.286	0.219	0.256	0.240	0.242	0.323	0.307
Saltuklu	0.127	0.323	0.219	0.195	0.154	0.195	0.407	0.183	0.165	0.229	0.161	0.221	0.184	0.197	0.270	0.189	0.183	0.286		0.231	0.248	0.218	0.231	0.278	0.244
Saçak	0.280	0.283	0.236	0.223	0.290	0.225	0.294	0.220	0.291	0.252	0.251	0.261	0.242	0.240	0.268	0.206	0.253	0.219	0.231		0.187	0.268	0.287	0.214	0.272
Uğursuyu	0.287	0.239	0.247	0.247	0.309	0.300	0.267	0.239	0.324	0.308	0.289	0.279	0.226	0.246	0.307	0.216	0.287	0.256	0.248	0.187		0.258	0.300	0.223	0.252
Çatalzeytin	0.207	0.316	0.201	0.225	0.212	0.175	0.398	0.167	0.178	0.264	0.226	0.160	0.263	0.169	0.249	0.229	0.215	0.240	0.218	0.268	0.258		0.153	0.301	0.300
Çaykırı	0.181	0.298	0.220	0.253	0.151	0.146	0.432	0.101	0.085	0.252	0.239	0.137	0.324	0.165	0.241	0.258	0.192	0.242	0.231	0.287	0.300	0.153		0.290	0.293
Üçevler	0.295	0.197	0.354	0.299	0.294	0.293	0.232	0.235	0.308	0.301	0.236	0.284	0.318	0.276	0.316	0.224	0.208	0.323	0.278	0.214	0.223	0.301	0.290		0.274
Şamlıoğlu	0.228	0.282	0.292	0.307	0.255	0.282	0.348	0.238	0.286	0.180	0.227	0.258	0.246	0.232	0.350	0.245	0.266	0.307	0.244	0.272	0.252	0.300	0.293	0.274	

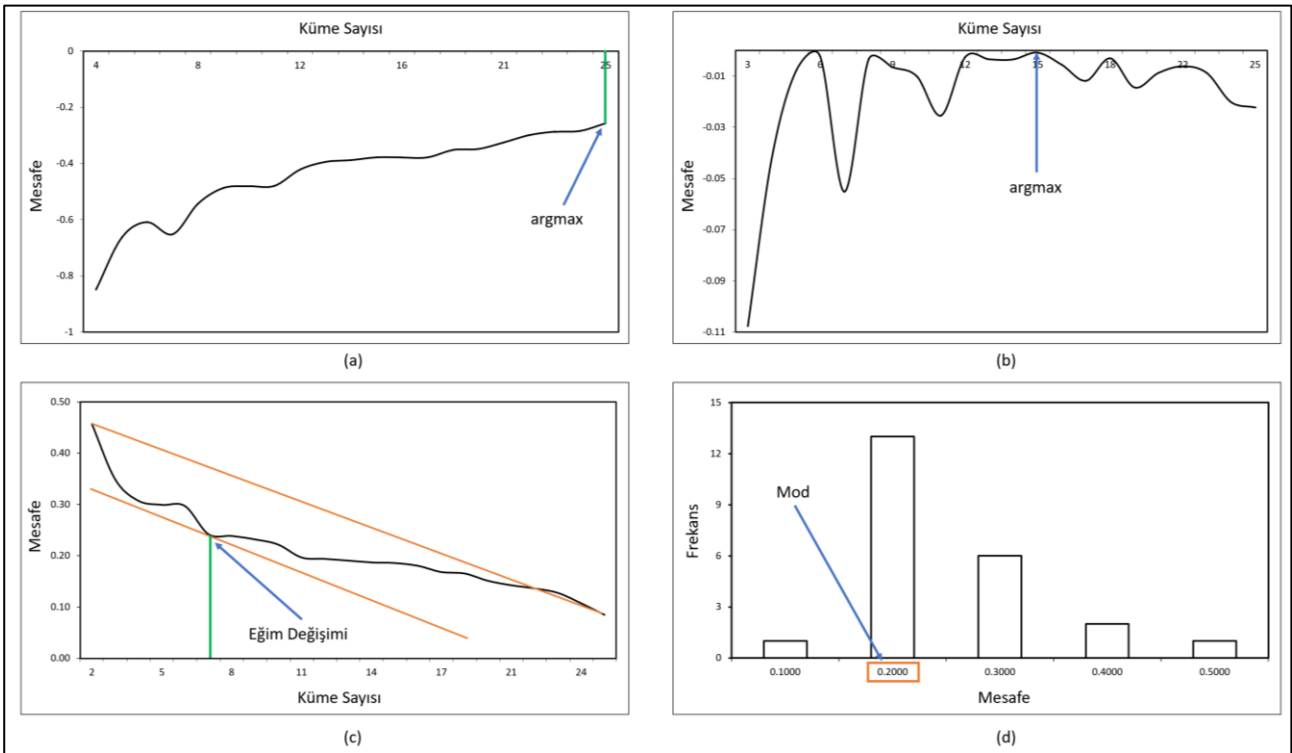
Şekil 4. Mesafe matrisi

Şekil 4 incelendiğinde 0,085 en düşük değeri ile Çaykırı – Ereğli eşleşmesi birbirine en yakın havzalar olarak görülmektedir. Bu durum hali hazırda oluşturulan dendrogramda da ortaya konmuştur. Fakat mesafe matrisinde Çaykırı – Bozkurt havzalarının 0,101 mesafe değeri ile birbirine en

benzer ikinci sıradaki havzalar olduğu göze çarpmaktadır. Bu durum Şekil 4’de oluşturulan dendrogramda ifade edilmemiş ve dendrogramda Çaykırı ve Bozkurt havzaları farklı şubelerde yer almıştır. Fakat daha sonra bu iki havza aynı kolda birleşmişlerdir. Bu durum veri setindeki havzaların

ikişerli karşılaştırmasının ele alınmak istenmesi durumunda mesafe matrisi kullanımının önemini ortaya koymaktadır. Burada unutulmaması gereken en önemli nokta, mesafe matrisinin sadece iki boyutlu karşılaştırma yaptığı ve dendrogramda olduğu gibi bir hiyerarşik kümeleme oluşturmadığı gerçeğidir. Şekil 4 aynı zamanda birbirine en az benzeyen havzalarında hangileri olduğunu görsel olarak bizlere aktarır. Çalışılan havzalar arasında birbirine en az benzeyenler sırasıyla 0,458, 0,449, ve 0,448 mesafe değerleriyle Bayramışlar – Araç, Bayramışlar – Ereğli ve Bayramışlar – Harmangeriş ikilileri olarak saptanmıştır. Bu durum dendrogram incelenirken Bayramışlar havzasını içeren kolun veri setini iki gruba ayırmasıyla da bağlantılıdır.

Çalışmada kullanılan hiyerarşik kümeleme yöntemi her ne kadar mutlak olarak birbirinden ayrılmış kümeler oluşturma amacı gütmese de bu çalışmada dirsek, en büyük fark, mod istatistiki yöntemleri ve eğimin azalışının gözlenmesini sağlayan grafik kullanılarak dendrogram belirli bir noktasından kesilerek havzalar HRU desenlerinin benzerliklerine göre kümelere ayrılmıştır. Şekil 5’de kesme yüksekliğinin belirlenmesinde kullanılan istatistiki yöntemler için çizilen grafikler ve Çizelge 2’de yöntemlerin önerdiği küme sayıları sırasıyla verilmiştir.



**Şekil 5.** Kesme yüksekliğinin belirlenmesinde kullanılan grafikler a) Dirsek yöntemi, b) En büyük fark yöntemi, c) Eğim değişimi yöntemi, d) Mod yöntemi

**Çizelge 2.** Farklı yöntemlerin önerdiği küme sayıları

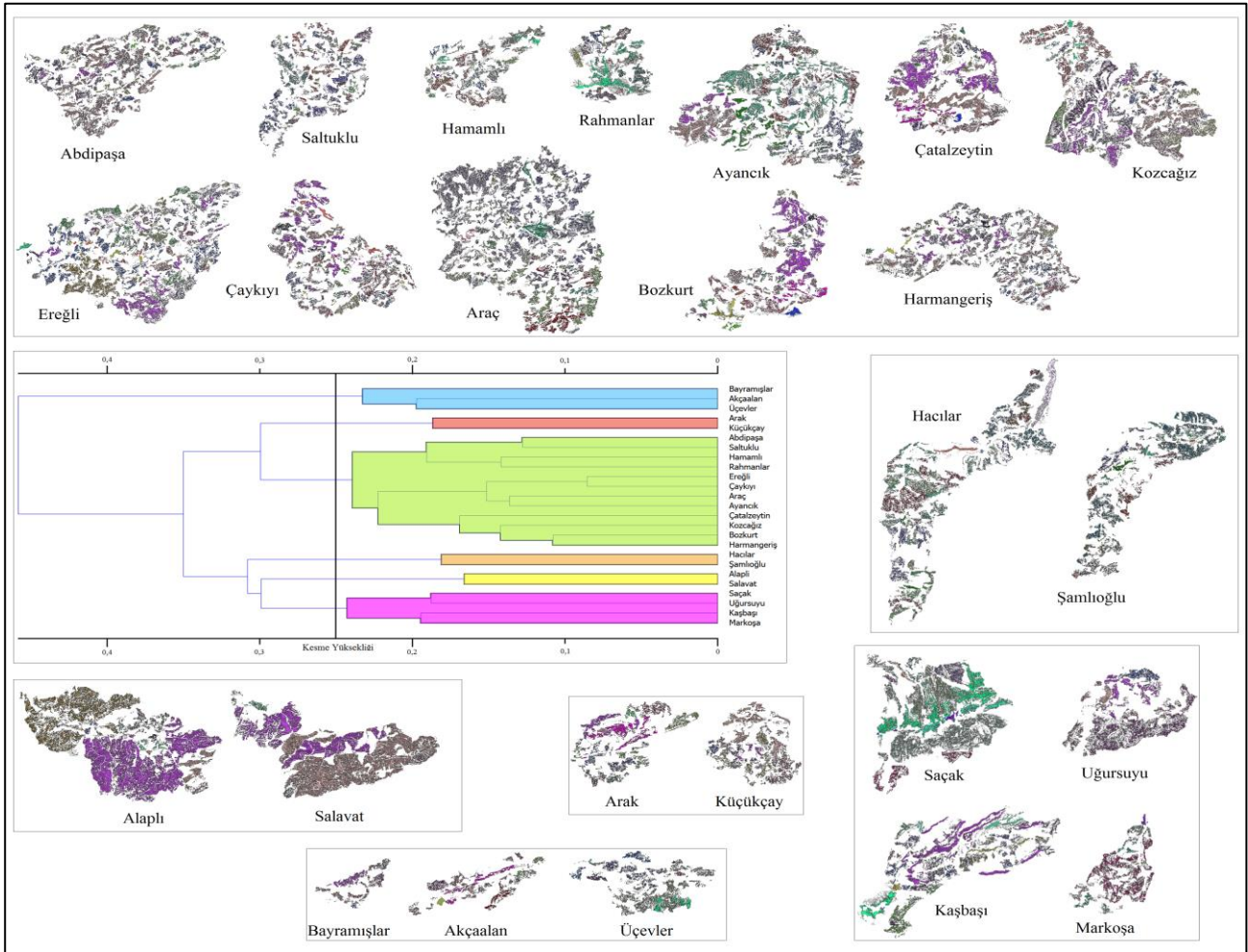
	Yöntem			
	Dirsek	En Büyük Fark	Eğim Değişimi	Mod
Küme Sayısı	2	12	7	1

Çizelge 2’de görülebileceği gibi, farklı yöntemler farklı sayıda küme üretmektedir. Bu çalışmada kullanılan veri setini mutlak kümelere ayırmak için kullanılan yöntemlerce önerilen küme sayıları 1 – 12 arasında değişmektedir. Bu yöntemler dışında kesme seviyesi araştırmacılar tarafından keyfi olarak da belirlenebilir ve dendrogram bir dizi ayrık küme oluşturulacak şekilde istenilen seviyede kesilebilir (Samarasinghe, Gunawardena *et al.* 2019, Govender and Sivakumar 2020). Birçok araştırmacı



dendrogramı inceleyerek kendi keyfi kararlarına göre kesme seviyesini seçmektedir (Zambelli 2016). Kesme yüksekliği seçimi dikkatle bir şekilde yürütülmelidir çünkü sonuçların yorum üzerinde büyük etkisi vardır. Bu seçim, gerekli ayrıntı düzeyiyle tutarlı olmalı ve ayrıca kümeler arasındaki mesafenin en aza indirilmiş olması göz önünde bulundurulmalıdır. Kesme mesafesi ne kadar yüksek olursa, kümeler içindeki heterojenlik o kadar fazla olur (Dumont *et al.* 2018). Bir analistin veri setinde kaç küme bulması gerektiğine dair kesin bir cevap yoktur, önemli olan konu bu kümeler ile ne yapılacağıdır (Haynes 2012). Bu nedenle, araştırmacılar hem istatistiksel önerileri dikkate almalı hem de anlamlı bir kesme yüksekliği seçmek için dendrogramı incelemelidir. Bu çalışmada, daha homojen kümeler oluşturmak ve ayrıca istatistiksel olarak belirlenen küme değerleri içinde kalmak koşuluyla, istatistiksel yöntemlerin aritmetik ortalaması olan 6 küme oluşturulmuş ve kesme

yüksekliği 0,250 olarak seçilmiştir. Belirtilen yükseklik değeri ile kesilen dendrogram ve 6 farklı kümeye ayrılan havzalar Şekil 6'da görülmektedir. Şekil 6'da en fazla havza içeren küme yeşil renk ile ifade edilmiştir ve içerisinde 12 adet veri örneği bulunmaktadır. Geri kalan kümeler ise iki ila dört arasında havza bulundurmaktadırlar. Ayrıca Şekil 6 kümelerine ayrılmış havzaların küme içi örneklerinin diğer kümelerle oranla birbirlerine daha fazla benzedikleri olgusunu da ortaya koymaktadır. Havzaların HRU bazlı benzerlikleri ortaya konurken sadece oluşturulan dendrogramın incelenmesi işlemi ile sınırlı kalınmayıp bağımsız kümeler oluşturulması, veri setinde bulunan havzaların daha büyük veri grupları şeklinde ele alınıp incelenebilmesi olanağını da sunmaktadır. Bu noktada göz önünde bulundurulması gereken önemli bir husus ise, kesme yüksekliği orijine ne kadar yakın seçilirse küme içinde kalan verilerin benzerliğinin o kadar daha fazla olacaktır.



Şekil 6. Bağımsız kümelerine ayrılan havzaların HRU görüntüleri

HRU'lar havzaların hidrolojik dinamikleri kontrol eder ve benzer HRU modellerine sahip havzalar, hava olayları sonucu oluşan girdiler ve havzadaki su akışıyla ilgili olarak aynı hidrolojik davranışa sahiptir. Bu çalışmada sunulan yöntem bir eleme işlemi olarak düşünülebilir. Denetimsiz öğrenme yaklaşımları (örneğin kümeleme), denetimli öğrenme yaklaşımlarında olduğu gibi (örneğin sınıflandırma) önceden tanımlanmış etiketler kullanmamaktadır. Eğer bu çalışmada kullanılan havzaların hidrolojik verileri (su girdisi, toprak yapısı, infiltrasyon oranı, iklim şartları vb.,) saptanmış olsaydı bu verilere denetimli öğrenme yaklaşımları uygulanabilir ve aynı yağış şartlarında benzer havzaların aynı tepkileri verebilecekleri düşünülebilirdi. Fakat sunulan çalışma da denetimsiz öğrenme algoritmasına bir grup resim verilmiş (HRU desenleri) ve yönteminin havzalar arasında bir bağlantı saptaması istenmiştir. Denetimsiz öğrenme yöntemlerinin verdiği sonuçlar denetimli öğrenme yöntemlerinin sonuçları gibi kesin değildir. Ayrıca denetimsiz öğrenme yaklaşımlarının anlaşılması zor yapısı gereği sonuçların kalibrasyonu ve doğrulaması da yapılamaz. Denetimsiz öğrenme yöntemleri araştırmacılara sadece benzerlik ön bilgisi sağlar. Bu çalışmanın amacı da havzaların HRU desenlerinin benzerlik gösterip göstermediğini denetimsiz öğrenme yaklaşımı ile ölçmektir. Her ne kadar denetimsiz öğrenme yöntemi doğası gereği içerisinde bir muğlaklık bulundursa da, çalışmamızda sunduğumuz bazı ölçütler sayesinde (mesela mesafe matrisi yada dendrogram) bu muğlaklık belirli noktaya kadar ortadan kalkmaktadır. Örneğin dendrogramda aynı kolda birleşen havzaların benzer hidrolojik tepkiler vermesi beklenirken diğer kollardaki havzaların ise hidrolojik süreçlerinin farklı olacağı aşikardır. Fakat havzaların HRU desenlerindeki benzerlikler yüksek oranda çıksa da gerçek hidrolojik süreçlerin ölçümünün yine de araştırmacılar tarafından yapılması gerekmektedir (Aytaç, 2020).

#### **4. Sonuç**

Bu çalışma havzaların benzerliklerinin ortaya konması hususunda literatürde yer alan çalışmamıza

alternatif bir yöntem olarak hiyerarşik kümeleme sürecini içermektedir. Hiyerarşik kümeleme yöntemi sayede veri setinde bulunan tüm havzaların birbirleri ile olan ilişkilerini gözlemleme şansı ortaya çıkmakta (mesafe matrisi) ve havzaların benzerliğini keşfetmeye yardımcı olabilecek daha küçük kümeler oluşturabilmelerini sağlamaktadır. Bu yöntem az sayıda havzanın karşılaştırmasında daha kullanışlı bir yöntem olarak sunulmuştur. Ayrıca hiyerarşik kümeleme yöntemi çok fazla özgün HRU içeren havzalarda daha iyi sonuçlar elde etmektedir. Hiyerarşik kümeleme yönteminin en önemli artlarından biri, kullanıcılara mutlak olarak birbirinden ayrılmış kümeler sunmaması ve kümeleme için bir kesme yüksekliğinin belirlenmesi gerekliliğidir. Araştırmacı isterse dendrogramı yorumlayıp uygun bulduğu bir kesme yüksekliğini kullanarak verileri kümeleyebilir fakat dendrogramda kesme seviyesini seçmek zordur. Mutlak kümeleme işlemi için farklı istatistiksel yöntemler kullanılabilir. Bu çalışma birbirinden ayrı kümeler oluşturmak için dört farklı yöntem incelemiştir. Yöntemler bir ila 12 arasında değişen küme sayıları önermektedir. Bu nedenle küme sayısını belirlemek için hem istatistiksel hesaplamalar dikkate alınmış hem de küme içi homojenliği korunarak kesme noktası altı küme oluşturacak şekilde seçilmiştir. Aynı zamanda bu çalışma havzaların kümelenebilirliği işleminde makine öğrenmesi yöntemleri kullanımı için yeni bir perspektif sunmuş ve farklı havzaların hidrolojik benzerlikleri hakkında ön bilgi almak için hiyerarşik yönteminin kullanımına örnek teşkil etmiştir.

Havzaların HRU farklılıklarının veya benzerliklerinin hiyerarşik kümeleme yöntemi ile ortaya konulması sayesinde, aynı şubeden gelen ve bağlantı mesafeleri kısa olan havzalarda, daha önce hidrolojik olarak modellenen bir havzanın sonuçları göz önüne alınarak, modellemesi yapılmayan havzalar hakkında öngörülerde bulunulabilir. Bu öngörüler içerisinde havza koruma planları hazırlamak veya uygulamak, havzalardaki akarsulara su, besin maddesi ve tortu akışını tahmin etmek, enkaz akışı riski hakkında değerlendirme yapmak ve önlem almak, havzaların jeomorfolojik özelliklerinin hesaplanması ve havzaların hidrolojik bütçe,

tarımsal yönetim, bitki büyümesi ve erozyon gibi hidrolojik özellikleri öngörmek bulunabilir. Ayrıca kümeleme işlemi gelecekte yapılacak çalışmalarda istenen özelliklere uyan havza seçiminde zamandan kazanç da sağlayabilir.

Bu yöntemde dikkat edilmesi gereken en önemli nokta, bunun bir sınıflandırma işlemi olmaması ve elde edilen kümelerin havzaların önceden belirlenmiş hidrolojik özelliklerine göre bir sıraya dizilmemesidir. Bu yöntem, ilişkileri daraltmanın yanı sıra gizli kalıplardan yeni bilgiler öğrenmeyi amaçlar ve havzalar arasında bir ilişki olup olmadığını bulmaya çalışır.

## 5. Kaynaklar

- Abdalla, A., Cen, H. Y., El-manawy, A. and He, Y. 2019. Infield oilseed rape images segmentation via improved unsupervised learning models combined with supreme color features. *Computers and Electronics in Agriculture*, **162**, 1057-1068.
- Asres, M. T. and Awulachew, S. B. 2010. SWAT based runoff and sediment yield modelling: a case study of the Gumera watershed in the Blue Nile basin. *Ecohydrology & Hydrobiology*, **10**(2), 191-199.
- Aytaç, E. 2020. Unsupervised learning approach in defining the similarity of catchments: Hydrological response unit based k-means clustering, a demonstration on Western Black Sea Region of Turkey. *International Soil and Water Conservation Research*, **8**(3), 321-331.
- Carcillo, F., Le Borgne, Y.-A., Caelen, O., Kessaci, Y., Oblé, F. and Bontempi, G. 2019. Combining unsupervised and supervised learning in credit card fraud detection. *Information Sciences*, **557**, 317-331.
- Castellanos-Garzón, J. A., Costa, E., Jaimes S, J. L. and Corchado, J. M. 2019. An evolutionary framework for machine learning applied to medical data. *Knowledge-Based Systems*, **185**, 1-8.
- Castro, L., Wasserman, E. A. and Lauffer, M. 2018. Unsupervised learning of complex associations in an animal model. *Cognition*, **173**, 28-33.
- Chen, J., Yousefi, M., Zhao, Y., Zhang, C. P., Zhang, S., Mao, Z. X., Peng, M. S. and Han, R. P. 2019. Modelling ore-forming processes through a cosine similarity measure: Improved targeting of porphyry copper deposits in the Manzhouli belt, China. *Ore Geology Reviews*, **107**, 108-118.
- Chen, J., Zeng, Y., Li, Y. and Huang, G.-B. 2019. Unsupervised feature selection based extreme learning machine for clustering. *Neurocomputing*, **386**, 198-207.
- Crosta, A. P., Sabine, C. and Taranik, J. V. 1998. Hydrothermal alteration mapping at Bodie, California, using AVIRIS hyperspectral data. *Remote Sensing of Environment*, **65**(3), 309-319.
- de Aguiar Neto, Fernando S., da Costa, Arthur F., Manzato, Marcelo G., Campello, Ricardo J. G. B. 2020. Pre-processing approaches for collaborative filtering based on hierarchical clustering. *Information Sciences*, **532**, 172-191.
- Dumont, M., Reninger, P. A., Pryet, A., Martelet, G., Aunay, B. and Join, J. L. 2018. Agglomerative hierarchical clustering of airborne electromagnetic data for multi-scale geological studies. *Journal of Applied Geophysics*, **157**, 1-9.
- Ferreira, A. D., Freitas, D. M., da Silva, G. G., Pistori, H. and Folhes, M. T. 2019. Unsupervised deep learning and semi-automatic data labeling in weed discrimination. *Computers and Electronics in Agriculture*, **165**, 1-11.
- Godec, P., Pancur, M., Ilenic, N., Copar, A., Strazar, M., Erjavec, A., Pretnar, A., Demsar, J., Staric, A., Toplak, M., Zagar, L., Hartman, J., Wang, H., Bellazzi, R., Petrovic, U., Garagna, S., Zuccotti, M., Park, D., Shaulsky, G. and Zupan, B. 2019. Democratized image analytics by visual programming through integration of deep models and small-scale machine learning. *Nature Communications*, **10**, 1-7.
- Govender, P. and Sivakumar, V. 2020. Application of k-means and hierarchical clustering techniques for analysis of air pollution: A review (1980-2019). *Atmospheric Pollution Research*, **11**(1), 40-56.

- Gungor, O. and Goncu, S. 2013. Application of the soil and water assessment tool model on the Lower Porsuk Stream Watershed. *Hydrological Processes*, **27**(3), 453-466.
- Haeberle, H. S., Helm, J. M., Navarro, S. M., Karnuta, J. M., Schaffer, J. L., Callaghan, J. J., Mont, M. A., Kamath, A. F., Krebs, V. E. and Ramkumar, P. N. 2019. Artificial Intelligence and Machine Learning in Lower Extremity Arthroplasty: A Review. *Journal of Arthroplasty*, **34**(10), 2201-2203.
- Han, D. H., Lee, S. and Seo, D. C. 2020. Using machine learning to predict opioid misuse among US adolescents. *Preventive Medicine*, **130**, 1-7.
- Haynes, P. 2012. Case based methods and national comparisons. Public Policy Beyond the Financial Crisis: An International Comparative Study. London, Routledge, 120-125.
- Huang, F. R., Zhang, X. M., Li, Z. J., Zhao, Z. H. and He, Y. Y. 2018. From content to links: Social image embedding with deep multimodal model. *Knowledge-Based Systems*, **160**, 251-264.
- Jafarzadegan, M., Safi-Esfahani, F. and Beheshti, Z. 2019. Combining hierarchical clustering approaches using the PCA method. *Expert Systems with Applications*, **137**, 1-10.
- Janert, P. K. 2010. *Data Analysis with Open Source Tools*, O'Reilly Media, Inc, Cambridge, 319-324.
- Janez Demšar, B. Z. 2013. Orange: Data Mining Fruitful and Fun - A Historical Perspective. *Informatica*, **37**, 55-60.
- Kaushik, M. and Mathur, B. 2014. Comparative Study of K-Means and Hierarchical Clustering Techniques. *International Journal of Software & Hardware Research in Engineering*, **2**(6), 93-98.
- Kotsiantis, S. B. 2007. Supervised Machine Learning: A Review of Classification Techniques. *Emerging Artificial Intelligence Applications in Computer Engineering*, **160**, 3-24.
- Li, K. H., Ma, Z. J., Robinson, D. and Ma, J. 2018. Identification of typical building daily electricity usage profiles using Gaussian mixture model-based clustering and hierarchical clustering. *Applied Energy*, **231**, 331-342.
- Loreti, D., Lippi, M. and Torroni, P. 2020. Parallelizing Machine Learning as a service for the end-user. *Future Generation Computer Systems*, **105**, 275-286.
- Manning, C. D. and Schütze, H. 1999. *Foundations of statistical natural language processing*, MIT Press, Cambridge, 417-452.
- Naik, A. and Samant, L. 2016. Correlation review of classification algorithm using data mining tool: WEKA, Rapidminer, Tanagra, Orange and Knime. *International Conference on Computational Modelling and Security (Cms 2016)*, **85**, 662-668.
- Perret, B., Cousty, J., Guimaraes, S. J. F., Kenmochi, Y. and Najman, L. 2019. Removing non-significant regions in hierarchical clustering and segmentation. *Pattern Recognition Letters*, **128**, 433-439.
- Ramkumar, P. N., Haeberle, H. S., Bloomfield, M. R., Schaffer, J. L., Kamath, A. F., Patterson, B. M. and Krebs, V. E. 2019. Artificial Intelligence and Arthroplasty at a Single Institution: Real-World Applications of Machine Learning to Big Data, Value-Based Care, Mobile Health, and Remote Patient Monitoring. *The Journal of Arthroplasty*, **34**(10), 2204-2209.
- Reddy, M. V., Vivekananda, M. and Satish, R. U. V. N. 2017. Divisive Hierarchical Clustering with K-means and Agglomerative Hierarchical Clustering. *International Journal of Computer Science Trends and Technology*, **5**(5), 6-11.
- Rocha, J., Roebeling, P. and Rial-Rivas, M. E. 2015. Assessing the impacts of sustainable agricultural practices for water quality improvements in the Vouga catchment (Portugal) using the SWAT model. *Science of the Total Environment*, **536**, 48-58.
- Sahu, S., Ghosh, S. K., Kalita, J. M., Ginjupalli, M. C. and K, K. R. 2020. Discovery of potential 1,3,5-Triazine compounds against strains of Plasmodium falciparum using supervised machine learning models. *European Journal of Pharmaceutical Sciences*, **144**, 1-8.

- Samarasinghe, T., Gunawardena, T., Mendis, P., Sofi, M. and Aye, L. 2019. Dependency Structure Matrix and Hierarchical Clustering based algorithm for optimum module identification in MEP systems. *Automation in Construction*, **104**, 153-178.
- Sheshukov, A. Y., Douglas-Mankin, K. R., Sinnathamby, S. and Daggupati, P. 2016. Pasture BMP effectiveness using an HRU-based subarea approach in SWAT. *Journal of Environmental Management*, **166**, 276-284.
- Song, L. R. and Zhang, J. Y. 2012. Hydrological Response to Climate Change in Beijiing River Basin Based on the SWAT Model. *2012 International Conference on Modern Hydraulic Engineering*, **28**, 241-245.
- Tang, C., Bian, M., Liu, X., Li, M., Zhou, H., Wang, P. and Yin, H. 2019. Unsupervised feature selection via latent representation learning and manifold regularization. *Neural Networks*, **117**, 163-178.
- Usama, M., Qadir, J., Raza, A., Arif, H., Yau, K. L. A., Elkhatib, Y., Hussain, A. and Al-Fuqaha, A. 2019. Unsupervised Machine Learning for Networking: Techniques, Applications and Research Challenges. *Ieee Access*, **7**, 65579-65615.
- Wang, B., Ning, L. J. and Kong, Y. 2019. Integration of unsupervised and supervised machine learning algorithms for credit risk assessment. *Expert Systems with Applications*, **128**, 301-315.
- Wangpimool, W., Pongput, K., Sukvibool, C., Sombatpanit, S. and Gassman, P. W. 2013. The effect of reforestation on stream flow in Upper Nan river basin using Soil and Water Assessment Tool (SWAT) model. *International Soil and Water Conservation Research*, **1(2)**, 53-63.
- Weyenberg, G. and Yoshida, R. (2015). Chapter 12 - Reconstructing the Phylogeny: Computational Methods. *Algebraic and Discrete Mathematical Methods for Modern Biology*. R. S. Robeva. Boston, Academic Press: 293-319.
- Wu, J., Yen Primary, H., Arnold, J. G., Yang, Y.-C. E., Cai, X., White, M. J., Santhi, C., Miao, C. and Srinivasan, R. 2020. Development of Reservoir Operation Functions in SWAT+ for National Water Quantity and Quality Assessments. *Journal of Hydrology*, **583**, 1-21.
- Yang, J., Grunsky, E. and Cheng, Q. M. 2019. A novel hierarchical clustering analysis method based on Kullback-Leibler divergence and application on dalaimiao geochemical exploration data. *Computers & Geosciences*, **123**, 10-19.
- Zambelli, A. E. 2016. A data-driven approach to estimating the number of clusters in hierarchical clustering. *F1000Res*, **5**, 1-6.
- Zhou, F., Xu, Y. P., Chen, Y., Xu, C. Y., Gao, Y. Q. and Du, J. K. 2013. Hydrological response to urbanization at different spatio-temporal scales simulated by coupling of CLUE-S and the SWAT model in the Yangtze River Delta region. *Journal of Hydrology*, **485**, 113-125.
- Zou, Q., Cui, P., He, J., Lei, Y. and Li, S. S. 2019. Regional risk assessment of debris flows in China-An HRU-based approach. *Geomorphology*, **340**, 84-102.

#### İnternet Kaynakları

- 1 - <https://orange3-imageanalytics.readthedocs.io/en/latest/widgets/imageembedding.html> (02.01.2020)
- 2 - <http://docs.biolab.si/3/visual-programming/widgets/unsupervised/distances.html> (03.01.2020)
- 3 - <https://orange.biolab.si> (20.09.2020)