

GNSS hızlarında kümelemeden topluluk kümelemesine: Meta-kümeleme odaklı bir yaklaşım

From clustering to ensemble clustering in GNSS velocities: A Meta CLustering-based approach

Seda ÖZARPACI *¹ , Batuhan KILIÇ ¹ , Mehmet KÖKÜM ² , Uğur DOĞAN ¹ 

¹Yıldız Teknik Üniversitesi, İnşaat Fakültesi, Harita Mühendisliği Bölümü, 34220, İstanbul

²Fırat Üniversitesi, Mühendislik Fakültesi, Jeoloji Mühendisliği Bölümü, 23119, Elazığ

• Geliş tarihi / Received: 23.02.2023

• Kabul tarihi / Accepted: 12.06.2023

Öz

Kıtasa deformasyonları anlayabilmek ve yorumlayabilmek için farklı yaklaşımlar ve modeller bulunmakta olup, bunlardan biri de blok modelleme yöntemidir. Blok modelleme yardımıyla plaka hareketleri, kayma hızları, faylardaki kilitleme derinlikleri, Euler kutbu gibi parametreler belirlenebilmektedir. Ancak, blok sınırları ne kadar iyi belirlenirse, modelleme sonuçları o kadar gerçeğe yaklaşmaktadır. Blok modellemenin en önemli adımlarından biri blok sınırlarının tespiti olup, kümeleme işlemi bunun için bir araç olarak kullanılabilir. Kümeleme analizi, kümelemeye konu verideki benzerlik ve farklılıklara dayanarak veriyi benzer gruplara atamaktadır. Bu çalışmada, çalışma alanı olarak Türkiye belirlenmiştir. Bu kapsamda Türkiye'nin en güncel Küresel Navigasyon Uydu Sistemi (Global Navigation Satellite Systems – GNSS) hız alanı topluluk kümeleme algoritması ile kümelendi ve bu hız alanına uygun blok sınırları belirlenmiştir. Türkiye için %22'si sürekli ve %78'i kampanya tipi verilerden oluşan GNSS gözlemlerinin birarada değerlendirilerek güncellenmiş hız alanı ilk defa bu çalışma ile kümelendi. Kümeleme öncesinde üç ayrı yöntemle, Davies-Bouldin, Gap (gap istatistiği) ve Silhouette ile, veriye en iyi uyum sağlayan optimum küme sayısı (GNSS hız alanına en uygun küme sayısı) tespit edilmiştir. Daha sonra, k-ortalamlar, HAC ve spektral kümeleme teknikleri kullanılarak güncel GNSS hızları kümelendi. Son olarak, Meta-Kümeleme Algoritması (Meta-CLustering Algorithm - MCLA) olan topluluk kümeleme tekniği ile güncel hız alanı yatay bileşenleri kümelendi ve sonuçlar paylaşılmıştır.

Anahtar kelimeler: Blok sınırlarının belirlenmesi, GNSS, Güncel hız alanı, Kümeleme, Meta-Kümeleme

Abstract

Although there are different approaches and models to understand and interpret the structures in crustal deformations, one of them is the block modeling method. Using block modeling, one can determine plate movements, parameters such as slip rates, locking depths or Euler poles on faults. However, the accuracy of the block modeling results is related to how well the block boundaries are determined. One of the most important steps of block modeling is the detection of block boundaries and clustering can be used as a tool for this. Cluster analysis assigns data to similar groups based on similarities and differences in the data subject to clustering. In this study, Türkiye was determined as the study area. In this context, we utilized the ensemble clustering algorithm to cluster recent Global Navigation Satellite Systems (GNSS) velocity field in Türkiye and determine block boundaries. Current GNSS velocity field, which consists of 78% survey and 22% continuous type GNSS data processed together, used for clustering analysis for the first time in this study. Before clustering, we employed three different methods - Davies-Bouldin, Gap statistics, and Silhouette - to determine the optimal number (cluster number that best fit to GNSS velocity field) of clusters. Then, k-means, HAC, and spectral clustering techniques were then applied to cluster current GNSS velocities. Finally, we utilized the Meta-Clustering Algorithm (MCLA) as an ensemble clustering technique to cluster the horizontal components of the current velocity domain and present our findings.

Keywords: Determining block boundaries, GNSS, Current velocity field, Clustering, Meta-Clustering

*Seda ÖZARPACI; ozarpaci@yildiz.edu.tr

1. Giriş

1. Introduction

Türkiye'nin tektoniği, Avrasya ve Arap levhalarının çarpışmaları, bu çarpışma sonucu oluşan kaçış, Kıbrıs ve Helen yayları boyunca Anadolu ve Afrika levhalarının oluşturduğu yitim zonları gibi etkiler nedeniyle karmaşık bir yapıya sahiptir (McClusky vd., 2000). Bu süreçlerin sonucu olarak meydana gelen Kuzey Anadolu Fayı (KAF) ve Doğu Anadolu Fayı (DAF) bölgenin en önemli doğrultu atımlı faylarını oluşturmaktadır. Bu yapıları anlamak ve yorumlayabilmek için birçok model kullanılmakla birlikte, blok modelleme yüzey deformasyonlarına en yakın sonuçları üretmektedir (Thatcher, 2009). Modelleme sonuçlarının gerçeğe yaklaşması blok sınırlarının ne kadar iyi belirlendiği ile alakalıdır. Birçok araştırmacı blok sınırlarını belirlerken haritalanmış fay yüzey izlerini, sismoloji verilerini ve/veya GNSS hız alanını kullanmış olmakla birlikte blok sınırlarının belirlenmesi yoruma açık ve öznel (Reilinger vd., 2006; Thatcher, 2009).

Blok sınırları belirlenirken kullanılacak bir yöntem de kümeleme yöntemidir. Kümeleme analizi herhangi kümelemeye konu verideki benzerlik ve farklılıklara dayanarak veriyi benzer gruplara atamaktadır. 2000'li yıllardan itibaren bu yöntem GNSS hızlarının kümeleneğinde de kullanılmakta olup, blok modelleme öncesinde, blok sınırlarını belirlemeye yardımcı olarak da kullanılabilir.

Kümeleme analizi, aynı bloklar üzerinde GNSS ile elde edilen yatay hızların birbirleri ile olan benzerliklerini dikkate alarak herhangi bir öncül bilgi olmaksızın küme/blok sınırlarını belirlemek amacıyla, bir çözüm olarak kullanılabilir. Küme analizi, ilk olarak 1932 yılında Driver ve Kroeber tarafından Antropoloji alanında gerçekleştirdikleri ufuk açıcı çalışma ile başlamıştır (Driver & Kroeber, 1932). Son 20 yıllık süreçte ise, veri madenciliği, örüntü tanıma, istatistik ve makine öğrenimi gibi çok sayıda gerçek dünya probleminin çözümünde ele alınmaktadır (Vega-Pons & Ruiz-Shulcloper, 2011; Golalipour vd., 2021). Buna ek olarak, birçok çalışma disiplini olduğu gibi jeodezik çalışmalarda blok modellemeler öncesinde blok sınırlarının belirlenmesinde de kullanılmıştır (Savage & Simpson, 2013a; Özarpacı vd., 2023). Simpson vd. (2012), San Francisco Körfez bölgesinde yer alan 168 sürekli Küresel Konumlandırma Sistemi (Global Positioning System - GPS) hız verisinin kümeleneşi için Birleştirici Hiyerarşik Kümeleme (Hierarchical Agglomerative Clustering – HAC) ve k -medoidler yaklaşımlarını karşılaştırmalı olarak incelemiştir. Her iki yöntem için de kümeleme öncesinde kümelerin beklenen değerlerinin karelerinin toplamının, beklenen değerlerine göre log olarak farkını karşılaştırmaya dayanan Gap istatistik tekniği kullanılmıştır ve en uygun küme sonuçlarının 4'lü kümeler ile sağlandığı görülmüştür. Kümeleme ile küme/blok sınırları boyunca bölge bölge hem kilitli ve hem de deprem üretmeden yavaş kayma hareketi (krip) olan üç büyük fay ile uyumlu sonuçlar elde edilmiştir. Savage ve Simpson (2013a;b), k -medoidler ve Euler-vektör kümeleme yöntemlerini kullanarak sırasıyla Mojave bloğu, Sierra Nevada bloğu, Kuzey Walker Lane Kuşağı, Orta Nevada Sismik Kuşağı ve Kaliforniya-Nevada bölgelerini kapsayan bir GPS ağına hızları kümelemiştir. Savage ve Simpson (2013a) ayrıca kümeleme işlemine GPS istasyonlarının konum bilgisini de eklemiş ancak sonuçlar üzerinde bir değişiklik görmemiştir. Savage ve Wells (2015), Kuzeybatı Pasifik bölgesinde bulunan 319 GPS istasyona ait hızlar için ilk olarak k -medoidler yöntemi ile en uygun kümeleme ($k = 2 - 10$) sonuçlarını tespit etmiş ve elde edilen sonuçları Euler-vektör kümelemesi için başlangıç olarak kullanmıştır. Elde edilen bulgulara göre, McCaffrey vd. (2013) tarafından aynı bölgede önerilen ana blok sınırları ile tutarlı kümeleme sonuçları ortaya koymuştur. Savage (2018), Japonya'nın güneybatısındaki 469 GEONET istasyonuna ait (ortalama olarak 1996-2006 yılları arasında) hızları Euler-vektör yöntemi kullanarak kümelemiş ve ana blok sınırlarını tanımlamada bu kümeleme yönteminin yararlı olduğunu belirlemiştir. Özdemir ve Karşoğlu (2019), Türkiye'deki GPS hızlarını k -ortalamlar, HAC ve Gauss Karışım Modeli (Gaussian Mixture Model - GMM) yaklaşımlarını kullanarak kümelemiştir. Özellikle, bölgede yer alan KAF ve DAF gibi en önemli aktif faylar boyunca kümeleme yöntemlerinin uyumlu sonuçlar verdiğini tespit etmiş ve bu yöntemlerde optimum küme sayısının belirlenmesi için Silhouette kullanılmıştır. Takahashi vd. (2019), Tayvan'da bulunan (2007 – 2013 yılları arasında) 281 sürekli GNSS istasyonuna ait hızları HAC yöntemi ile kümelemiştir ($k = 2 - 11$). Sonuç olarak, küme sonuçlarının tektonik özellikleri yansıttığını belirtmiştir. Granat vd. (2021), Kaliforniya ve çevresinde bulunan 919 istasyona ait yatay hızların yanı sıra düşey hızlar ve coğrafi koordinat bilgilerini bir arada k -ortalamlar yöntemi ile kümelemiştir ($k = 2 - 10$). Bu işlemleri gerçekleştirirken bölgedeki deprem anı, deprem sonrası deformasyonları ve yer değiştirmeleri dikkate almıştır. Özarpacı vd. (2023), Türkiye'deki üç farklı veri setine ait sürekli ve kampanya GNSS istasyonlarının hızlarını k -ortalamlar yöntemi ile kümelemiştir. Ayrıca, diğer çalışmalardan farklı olarak, kümeleme işlemi öncesinde Davies-Bouldin indeksi, Elbow yöntemi, Gap istatistiği ve Silhouette yöntemi gibi çeşitli iç göstergeler kullanarak optimum küme sayılarının belirlenmesi ve

bu veri setleri üzerindeki etkilerini incelemiştir. Son olarak, [Kılıç ve Özarpacı \(2022\)](#), Türkiye’de bulunan 188 sürekli GNSS istasyonuna ait hızlar üzerinde ilk kez Hibrit Çift Parçalı Graf Formülasyonu, Meta-Kümeleme Algoritması ve Negatif Olmayan Matris Faktörizasyonu (Hybrid Bipartite Graph Formulation – HBGF, Meta-CLustering Algorithm – MCLA ve Non-Negative Matrix Factorization-based clustering- NMF) odaklı topluluk kümeleme yöntemleri kullanarak küme/blok sınırlarının uyumunu incelemiştir. Tekli kümeleme algoritmalarına göre küme/blok sınırları hakkında daha iyileştirilmiş çözümler sunabileceklerini ortaya koymuş ve MCLA yönteminin GNSS verileriyle elde edilen hız alanının kümelenmesi analizinde kullanılabilirliğini göstermiştir.

GNSS verilerinden elde edilen hızlarının kümelenmesinde tekli kümeleme uygulamaları çok çeşitli olması ve yaygın bir şekilde kullanılmasına rağmen, kendine özgü çeşitli sınırlamaları mevcuttur. İlk olarak, kümeleme işlemi doğası gereği denetimsiz bir öğrenme yöntemi olarak bilinmekte ve genellikle darboğaz bir problem oluşturmaktadır. Denetimli öğrenmeden farklı olarak, etiketsiz verileri anlamlı kümeler halinde gruplamaktadır ve dolayısıyla veriler hakkında öncül bir bilgi bulunmamaktadır ([Jain vd., 1999](#); [Alqurashi & Wang, 2019](#)).

İkinci olarak, [Kleinberg \(2002\)](#) tarafından ele alınan “Kümeleme için Olanaksızlık Teoremi” adlı çalışmadan yola çıkarak, farklı problemler için uygun ve optimum sonuçlar üretebilen tek bir kümeleme algoritmasının ve/veya belirli bir problem için herhangi bir kümeleme algoritmasının seçimi hakkında evrensel bir standart fikir birliği bulunmamaktadır.

Üçüncü ve son olarak, kümeleme çalışmalarında ele alınan çeşitli yöntemlerin birbirlerine göre bazı avantaj ve dezavantajları bulunmaktadır. Örneğin, konveks yapıda olmayan kümeleri bazı algoritmalar (örn. bölütleme odaklı algoritmalar) tespit edemezken, hiyerarşik ve graf (çizge) temelli çalışan bazı algoritmalar rastgele şekil gözlemlerine sahip veriler ile daha uyumlu çalışmaktadır. Dolayısıyla, çeşitli kümeleme algoritmalarından her biri, benzer bir veri kümesine örtük veya açık bir şekilde belirli bir yapı empoze ederek farklı sonuçlar üretebilmektedir. Bu noktada, birçok tekli kümeleme algoritmasının öznelliğinden kaçmak ve bir veri seti üzerine uygulanan çeşitli algoritmaların güçlü yönlerini birleştirmek için verimli yöntemlere ihtiyaç duyulmaktadır ([Ghaemi vd., 2009](#)). Bu nedenle, araştırmacılar tarafından bu türde problemlerin üstesinden gelebilmek ve tekli kümeleme algoritmalarının sonuçlarını iyileştirmek için topluluk kümeleme (konsensüs kümeleme) algoritmaları ortaya çıkarılmıştır ([Strehl & Ghosh, 2002](#)).

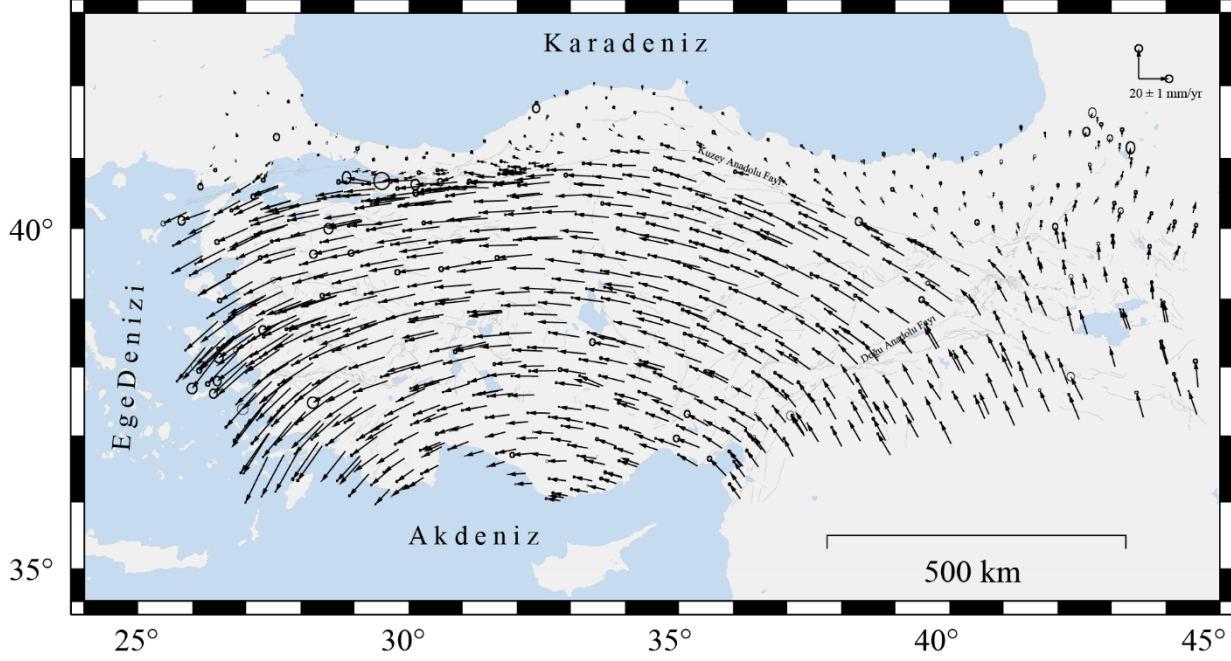
Bu çalışmada, [Kurt vd. \(2023\)](#)’de yayımlanmış Türkiye’nin en güncel ve kapsamlı GNSS verilerinden elde edilen yatay hızlar topluluk kümeleme algoritması ile kümelenmiş ve sonuçlar irdelenmiştir. Güncel hız alanı ile kümeleme işlemi öncesi kullanılacak yöntemlerinin geçerliliklerini test etmek ve veriyi en iyi temsil eden optimum küme sayısını belirlemek için Davies-Bouldin, Gap istatistik ve Silhouette olmak üzere üç farklı iç geçerlilik yöntemi kullanılmıştır. Optimum küme sayısının tespitinin gerçekleştirilmesi ile birlikte, k -ortalamalar, HAC ve spektral kümeleme ile yatay hızlar kümelenmiştir. Bu kümeleme işlemlerinin ardından, ayrı ayrı her bir tekli kümeleme yönteminin aksine, yöntemlerin güçlü yanlarını birleştirmeye dayalı bir topluluk kümeleme algoritması olan Meta Kümeleme Algoritması ile de kümelenmiştir. Meta Kümeleme yöntemi ile tekli kümeleme algoritmalarının sonuçları girdi olarak alınarak, son bir kümeleme işlemi gerçekleştirilmiş ve sonuçlar irdelenmiştir. Ayrıca, literatürde bulunan kümeleme çalışmaları, Türkiye’de gerçekleştirilen çalışmalar özelinde incelenecek olunursa, sadece sürekli verilerden oluşan ve yaklaşık 200 nokta kullanılarak ya da sürekli ve kampanya noktalarının farklı çalışmalardan birleştirilerek oluşturulduğu araştırmalar üzerinde yoğunlaşmıştır. Bu çalışmada kullanılan ve 836 noktadan oluşan hız alanı hem sürekli hem de kampanya noktalarının birlikte değerlendirilerek üretildiği, çalışma bölgesinin güncel hız alanıdır. Bu çalışma ile yoğunlaşan hız alanıyla elde edilen blok sınırlarının önceki kümeleme çalışmalarıyla oluşturulan blok sınırlarıyla farkları da vurgulanmaktadır. Bununla birlikte, literatürde var olan blok modelleme çalışmalarındaki blok sınırları ile de karşılaştırmalar yapılmıştır. Çalışmanın Türkçe literatüre katkısı dışında, buradan elde edilecek blok sınırlarının bundan sonra gerçekleştirilecek blok modelleme çalışmalarında altlık olarak kullanılabilir olması önemlidir.

2. Veri ve yöntem

2. Data and method

Bu çalışmada kullanılan veri seti [Kurt vd. \(2023\)](#)’den Türkiye’nin GNSS verilerinden elde edilen en güncel ve kapsamlı yatay hızlardır (Şekil 1). Bu hızların %78’i kampanya verilerinden %22’si ise sürekli istasyonlara ait

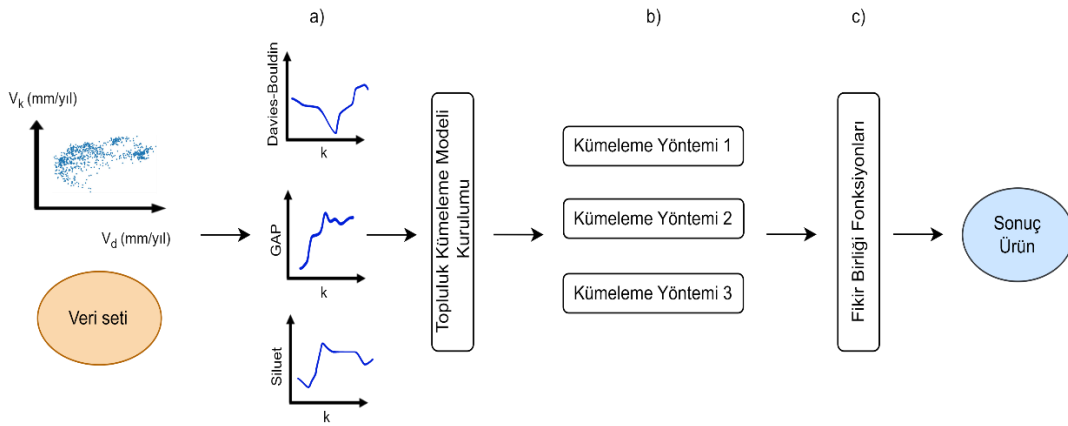
verilerin değerlendirilmesinden elde edilmiştir. Kampanya noktalarındaki ölçülerin 7-10 saat arasında yapıldığı ve 24 saatlik sürekli verilerin de dahil edildiği veri aralığı 1992 ile 2020 yılları arasını kapsamaktadır. Veriler genel olarak, Türkiye Ulusal Temel GPS Ağı, Türkiye Ulusal Sabit GNSS Ağı Aktif (TUSAGA-Aktif) ağlarından yararlanılarak oluşturulmuştur. Ayrıca, veri sıklığını arttırabilmek adına, Türkiye Marmara Araştırma Merkezi verileri (MAGNET) ve Bursa, Sakarya, İstanbul illerinin su ve kanalizasyon idarelerine ait bölgesel ağlar da veriye entegre edilmiştir.



Şekil 1. Türkiye'nin Avrasya referanslı güncel yatay hız alanı (Kurt vd., 2023). Gri çizgiler Emre vd. (2013)'den alınan diri fayları göstermektedir.

Figure 1. Eurasia-fixed current horizontal velocity field of Turkey (Kurt et al., 2023). The grey lines indicate the active faults from Emre et al. (2013).

Kümeleme analizi için güncel GNSS hız alanı kullanılarak, toplam 836 noktanın yatay hızları değerlendirilmiştir. İlk olarak, üç farklı iç geçerlilik indeksi kullanarak veriye en uygun optimum küme sayısı (k) belirlenmiştir (Şekil 2a). Üç yöntemden de en uyumlu sonuçlar elde edildikten sonra, topluluk kümeleme yaklaşımını oluşturmak adına farklı kümeleme tekniklerinden (topluluk üyeleri) kümeleme modelleri oluşturulmuştur (Şekil 2b). Son olarak, nihai kümeleme sonuçlarını elde etmek için topluluk kümeleme yöntemi kullanılarak üyelerin çıktıkları birleştirilmiştir (Şekil 2c) ve nihai çözümlerin blok sınırları ile uyumu incelenmiştir (Şekil 2).



Şekil 2. Akış diyagramı

Figure 2. The flowchart of this study

2.1. Küme geçerlilik analizleri

2.1. Cluster validity analyses

Kümeleme analizlerinde otomatik kümeleme problemi, hem veriye uygun küme sayısının belirlenmesi hem de bu duruma karşılık gelen doğal bölütleme işleminin incelenmesini gerektirir. Bir başka ifade ile, bir veri setinin analizinde ve yorumlanmasında kullanılacak parametreler için k küme sayısının otomatik olarak belirlenmesi, araştırmacılar, uygulayıcılar ve son kullanıcılar için temel bir sorun olmaktadır (Pakhira, 2012). Bununla birlikte, uygun k küme sayısının seçimi, bir veri setindeki parametrelerin dağılım şekline, yapısına ve ölçeğine bağlı olarak bir belirsizlik sorunu ortaya çıkarmaktadır. Bu türdeki problemleri ortadan kaldırmak için, bir kümeleme sonucunun ilgili verinin altında yatan yapıya hangi ölçüde iyi uyduğunu tahmin etme süreci olarak bilinen küme geçerlilik işlemi uygulanır. Literatürde, küme geçerlilik tespiti için çeşitli indeksler geliştirilmiştir. Küme geçerlilik indeksleri, bir kümeleme işleminde veriye uygun anlamlı gruplamayı oluşturabilmek için küme içi uyum (grup içi dağılım) ve kümeler arası ayırım (gruplar arası dağılım) arasındaki ilişkiyi tanımlar (Arbelaitz vd., 2013; Xu & Tian, 2015). Bu indeksler, kümeleme algoritmasının oluşturulması sürecinde test verileri açısından iç ve dış geçerlilik indeksleri olmak üzere iki kategoriye ayrılır. Dış geçerlilik indeksleri, bir veri kümesindeki veriye uygun küme sayısını tespit etmek için veri hakkında daha önceden belirlenen öncül bilgileri kullanır. Öte yandan, iç geçerlilik indeksleri, veri kümeleri için herhangi bir öncül bilgiye gerek duymamakta ve uyum iyiliğini veya geçerliliğini belirlemek için veriye özgü bilgileri kullanmaktadır (Rendón vd., 2011).

Bu çalışmada, veriye ait öncül bir bilgi olmaması nedeniyle Davies-Bouldin, Gap ve Silhouette gibi popüler üç farklı iç geçerlilik indeksi kullanılmıştır. Bu yöntemlerin her biri farklı hesaplama modellerine sahiptir (Tablo 1). Kümeleme çalışmalarında tek bir yöntem yerine farklı sayıda yöntemin incelenmesi, optimum küme sayısı tespitinin sağlıklı bir şekilde gerçekleştirilmesine katkı sağlamaktadır (Özarpacı vd., 2023).

Davies-Bouldin (DB) indeksi, kompakt ve düzgün dağılımlı kümelerin belirlenmesi için sıklıkla kullanılan iç geçerlilik indekslerinden biridir. Bir kümede yer alan noktalardan merkeze olan uzaklığa göre küme içi uyum ve merkezler arasındaki uzaklığa bağlı olarak da kümeler arası ayırımı tahmin eder. $DB(k)$ değerini en küçük yapan k değeri, bu indekse göre ideal küme sayısını ifade eder (Tablo 1). Bir diğer yöntem olan Gap istatistiği, küme değerlendirmelerinin geçerliliğini ve uyum iyiliğini tespit eden bir veri madenciliği algoritmasıdır. Bu yöntemin diğer yöntemlere göre avantajı, herhangi bir kümeleme algoritmasına ve mesafe ölçüsüne uygulanabilecek şekilde tasarlanmış olmasıdır. Bu yöntem, k farklı sayıda kümedeki toplam küme içi varyasyonu, referans boş hipotez (W_k^{ref}) dağılımı (önemli kümelenecek yok) altında verilerin beklenen değerleriyle ($W_k^{göz}$) karşılaştırır (Tablo 1). Bu, hız uzayında test edilecek verilerle aynı uzamsal kapsamı işgal eden rastgele bir nokta dağılımı (Monte Carlo örnekleme yöntemi) oluşturularak elde edilir. Son olarak Silhouette indeksi, bir veri kümesindeki bir nesnenin diğer kümelere (ayırım) kıyasla kendi kümesine (uyum) ne kadar yakın olduğunu gösteren bir ölçüttür (Tablo 1). Uyumluluk, aynı kümedeki tüm noktalar arasındaki ortalama mesafeye göre ölçülür ve ayırım en yakın komşu mesafesine dayanır (Özarpacı vd., 2023). Burada, bir silhouette değeri -1 ve +1 aralığında değişim gösterir. Değer 1'e yaklaştığında, i . nesnenin atandığı küme daha uygun olmaktadır.

Tablo 1. Küme iç geçerlilik indeksleri

Table 1. Cluster internal validity indices

İsim	Formül	Açıklama	Referans
Davies-Bouldin	$R_i = \max_{i \neq j} \frac{e_i + e_j}{d_{ij}},$ $DB = \frac{1}{k} \sum_{i=1}^k R_i$	<p>DB değeri, hesaplanan R_i değerleri arasından en büyük değer seçilerek elde edilir.</p> <p>e_i kümenin ortalama hatası; d_{ij}, i ve j kümelerinin merkezleri arasındaki mesafeyi göstermektedir.</p>	Davies & Bouldin (1979)

Tablo 1. Devamı
Table 1. Continuing

GAP	$GAP_n(k) = E_n^* \{\log(W_k^{ref})\} - \log(W_k^{göz})$	E_n^* referans veri dağılımındaki n örneğin beklenen değerine karşılık gelir. En optimum küme sayısının seçimi, $\log(W_k)$ değerinin referans değerinin altında en uzun düşünüşe sahip olduğu k değeridir.	Tibshirani vd. (2001)
Silhouette	$s(i) = \frac{b(i) - a(i)}{\max\{b(i), a(i)\}}$	$a(i)$, i . nesne ile aynı kümedeki diğer tüm nesnelere arasındaki ortalama mesafeye; $b(i)$ ise i . nesne ile diğer kümedeki nesnelere arasındaki ortalama uzaklıkların minimum değerine karşılık gelir.	Rousseeuw (1987)

2.2. Topluluk kümeleme modelinin kurulumu

2.2. Generation of ensemble clustering model

Topluluk kümelemesi veya fikir birliği kümelemesi olarak da adlandırılan bir kümeleme yöntemi, mevcut birden fazla kümeleme algoritmasına ait sonuçları nihai bir kümede birleştirir. Bir başka ifade ile, kümeleme algoritmalarının sonuçlarının avantajlarını bir araya getirmek için alternatif bir yaklaşım sunar (Strehl & Ghosh, 2002; Vega-Pons & Ruiz-Shulcloper, 2011). Topluluk kümeleme yaklaşımları, birçok tekli kümeleme algoritmasının güçlü yönlerini birleştirerek bir sonuç ürün ortaya çıkarır (Ghaemi vd., 2009). Bu nedenle, tek bir kümeleme algoritmasına ait aşağıda çeşitli açılardan listelenen durumların ötesine geçebilmektedir:

- Sağlamlık: Birleştirme işlemi, tekli kümeleme algoritmalarına kıyasla daha iyi ortalama performansa sahip olmalıdır.
- Yenilik (Orijinallik): Herhangi bir kümeleme algoritması tarafından ulaşılamayan yeni bir çözüm bulma.
- Kararlılık: Gürültü ve aykırı değerlere karşı daha düşük hassasiyete sahip sonuçları üretme kabiliyeti.
- Ölçeklenebilirlik ve Parallelleştirme: Dağıtılmış verileri hesaplama ve birden fazla veri kaynağından gelen çözümleri entegre edebilme (paralel kümeleme) yeteneğine sahip olma (Topchy vd., 2005; Ghaemi vd., 2009; Ghosh & Acharya, 2011; Golalipour vd., 2021; Kılıç & Özarpacı, 2022).

Topluluk kümeleme işlemi iki temel aşamada gerçekleştirilmektedir. Bunlardan ilki, topluluk üyelerinin üretilmesi için üretim mekanizmasının oluşturulmasıdır. Diğer adım ise, üretilen topluluk üyelerini birleştirmek için bir uzlaşma ya da fikir birliği fonksiyonunun kullanılmasıdır. Üretim mekanizmalarının oluşturulması yani topluluk üyelerinin üretilmesi işleminde literatürde farklı kümeleme algoritmalarının kullanımı, tekli bir kümeleme algoritmasında farklı parametrelerin kullanılması, nesnelere farklı alt kümelerinin kullanımı (yeniden örnekleme) gibi çeşitli yaklaşımlar mevcuttur (Kılıç & Özarpacı, 2022). Buradaki temel amaç, topluluk kümeleme işleminin kalitesini artırmak için dağılımın (çeşitliliğin) oluşturulmasıdır. Bu çalışmada, her bir topluluk üyesinin oluşturulması için farklı kümeleme algoritmalarının kullanımı stratejisi dikkate alınmıştır.

Öte yandan, uzlaşma fonksiyonunun oluşturulması, herhangi bir kümeleme topluluğu algoritmasının temel işlem adımıdır ve nihai çözümleri elde etmek için topluluk üyelerinin sonuçlarını farklı yaklaşımlar kullanarak bir araya getirir. Bu yaklaşımlar, nesnelere birlikte oluşumuna dayalı ve medyan bölütlemeyle dayalı olarak iki sınıfta toplanmaktadır (Vega-Pons & Ruiz-Shulcloper, 2011). Bu çalışmada, nesnelere birlikte oluşumuna dayalı olarak çalışan ve popüler bir topluluk kümeleme yöntemi (Golalipour vd., 2021) olan Meta-Kümeleme Algoritması kullanılmıştır.

2.3. Topluluk kümeleme üyelerinin oluşturulması

2.3. Building of ensemble clustering members

Daha önceden belirtildiği gibi, her bir kümeleme algoritması veri kümesini özel bir bakış açısıyla (belirli kriterlere göre) bölümlere ayırır ve farklı algoritmalar ortak veri kümesi üzerinde farklı sonuçlar oluşturabilir.

Bu nedenle, verilerin altında yatan örüntüler hakkında daha fazla bilgi elde etmek için üye sonuçlarının mümkün olduğundan çeşitli olması gerekmektedir (Abbasi vd., 2019; Golalipour vd., 2021; Kılıç & Özarpacı, 2022). Bu çalışmada, topluluk kümeleme üye sonuçları için gerekli çeşitliliği oluşturmak için k -ortalamalar, HAC ve spektral kümeleme gibi üç farklı kümeleme tekniği kullanılmıştır. Bu yöntemler arasından, k -ortalamalar ve bölümlenme odaklı, HAC hiyerarşiye dayalı ve spektral kümeleme ise graf odaklı çalışma prensibine sahiptir.

Bölümlenme odaklı çalışan kümeleme algoritmalarındaki temel düşünce, veri noktalarının merkezini ilgili kümenin merkezi olarak kabul etmektir. Bu yöntemlerden biri olan k -ortalamalar yöntemi, ilk olarak MacQueen (1967) tarafından kullanılan denetimsiz bir öğrenme algoritmasıdır. Yöntemin altında yatan ana fikir, n veriden oluşan bir veri tabanı içerisinde her nesnenin en yakın ortalamaya sahip kümeye ait olduğu k adet kümeyi bölümlenmektir. Bu sayede, küme içi varyanslar minimize edilmekte ve küme içi benzerlikler homojen halde olmaktadır.

Hiyerarşiye dayalı olarak çalışan algoritmalarındaki temel fikir, kümeleme işleminde veri grupları arasında hiyerarşik bir ilişki oluşturmaktır. İki türlü çalışma prensibine sahiptir. Veri tabanı içerisinde yer alan her bir veri noktası başlangıçta ayrı bir küme olarak temsil edilir ve daha sonrasında en yakın komşu iki küme birleştirilir ve yeni bir kümede toplanır. Buradaki işlem $n - 1$ defa tekrarlanır ve son olarak iki küme birleştirilerek tek bir küme oluşturulur. Kullanılan algoritmanın çalışma prensibine göre gerçekleştirilen bu işlem adımlarının tersi de uygulanabilmektedir. HAC yöntemi, bir veri setindeki belirli seviyelerdeki benzerlikleri (uzaklık ölçüleri) dikkate alarak nesnelere birleştirmeyi amaçlayan bir algoritmadır. Her yinelemede HAC kümeleme yöntemi en yakın küme çiftini birleştirmektedir. Bir başka ifade ile veri setindeki her bir nesneden başlayarak aşağıdan yukarıya doğru bir hareket izler. Tüm bu işlemler, nesnelere dendrogram olarak bilinen ağaç yapıları halinde gruplandırmaya dayanır (Kaufman & Rousseeuw, 1990). Hangi nesnelere/kümelere birleştirilmesi gerektiğine karar vermek açısından karesel Öklid, Manhattan, Kosinüs ve Minkowski gibi yöntemler aralarındaki benzerliği ölçmek için kullanılır. Öte yandan, nesne çiftlerinin benzerliklerine dayalı olarak kümelere olan yakınlıkları tanımlamak için ise: (1) Tek ya da minimum bağlantı (en yakın komşu); (2) Ortalama bağlantı (aritmetik ortalama ile ağırlıklandırılmamış grup çifti yöntemi); (3) Tam bağlantı; (4) Ağırlıklı bağlantı (aritmetik ortalama ile ağırlıklandırılmış grup çifti yöntemi); (5) Merkezi bağlantı; (6) Medyan bağlantı ve (7) Ward bağlantı kümeleme yöntemleri kullanılmaktadır (Ward, 1963).

Graf ya da spektral graf temelli çalışan algoritmalarındaki temel fikir, kümeleme problemini bir graf bölümlenme problemine dönüştürmektir. Bunun için verilen bir veri kümesinde yer alan her nesneyi bir düğüm olarak ele alarak graf oluşturur ve nesnelere arasındaki benzerliği ağırlıklı kenar olarak kabul eder (Xu & Tian, 2015). Farklı gruplar arasındaki bağlantı ağırlığını mümkün olduğunca küçük ve aynı grup içindeki kenarlar arasındaki toplam bağlantı ağırlığını mümkün olduğunca yüksek yapan bir graf bölümlenme işlemi uygular. Bu yöntem, parametrik olmayan bir model ve daha geniş bir geometri yelpazesini yakalama yeteneğine sahiptir ve güçlü bir denetimsiz öğrenme algoritmasıdır (Von Luxburg, 2007; Yan vd., 2009).

Çalışmada ele alınan her bir topluluk kümeleme üyesi verinin farklı yönlerini ortaya çıkarsa da kendi aralarında bazı avantaj ve dezavantajları da bulunmaktadır (Tablo 2).

Tablo 2. Kümeleme yöntemlerinin avantaj ve dezavantajları
Table 2. Advantages and disadvantages of the clustering methods

Yöntem	Çalışma prensibi	Avantaj	Dezavantaj
HAC	Hiyerarşi odaklı	<ul style="list-style-type: none"> Keyfi (rastgele) türde şekillere ve özniteliklere sahip veri setleri için uygun olma, Yüksek ölçeklenebilirlik 	<ul style="list-style-type: none"> Genel anlamda, zaman karmaşıklığının yüksek olması
K-ortalamar	Bölümleme odaklı	<ul style="list-style-type: none"> Genel olarak, düşük zaman karmaşıklığına ve yüksek hesaplama verimliliğine sahip olma 	<ul style="list-style-type: none"> Konveks olmayan verilere uygun olmama, Gürültülü veriye duyarlılığının olması, Kümeleme sonucunun küme sayısına duyarlı olması
Spektral kümeleme	Graf odaklı	<ul style="list-style-type: none"> Keyfi (rastgele) türde şekillere ve yüksek boyutlu veri setlerine uygun olma, Girdi olarak yalnızca benzerlik matrisi kullanma, Aykırı değerlere duyarlı olmama 	<ul style="list-style-type: none"> Kümeleme sonucunun ölçekleme parametresine duyarlı olması, Zaman karmaşıklığının yüksek olması, Öz vektör seçiminin optimize edilememesi

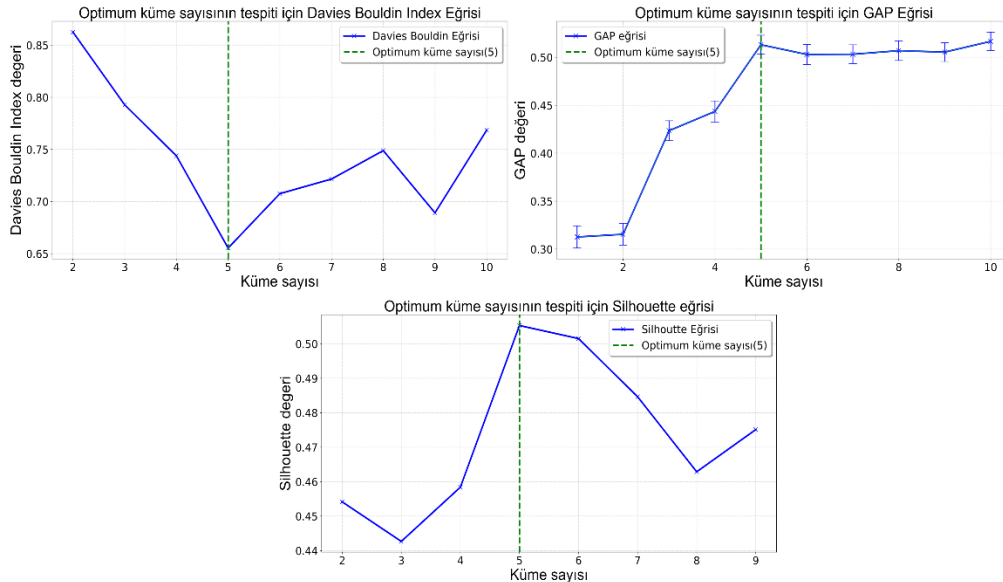
3. Bulgular ve tartışma

3. Results and discussion

3.1. Optimum küme sayısının belirlenmesi

3.1. Determining the optimum number of clusters

Kavramsal olarak, en uygun küme sayısının belirlenmesi genellikle öznel olarak gerçekleştirilir. Ancak veri setindeki parametrelerin dağılım şekline, yapısına ve ölçeğine bağlı olarak, veriye en uyumlu olan küme sayısının seçimi için Davies-Bouldin, Gap istatistik ve Silhouette küme geçerlilik indeksleri kullanılmıştır (Şekil 3).



Şekil 3. Optimum küme sayısının tespiti için küme geçerlilik indeksi sonuçları

Figure 3. Cluster validity index results for determining the optimum number of clusters

Şekil 3'de, $k = 2$ 'den 10'a kadar her bir küme sayısına ait DB değerlerinin değişkenlik gösterdiği görülmektedir. Küme sayısının 5 ve 9 olduğu durumlarda DB değerlerinin birbirine yakın en düşük sonuçlar elde ettiği

söylenbilir (sırasıyla $DB(5)=0.655$ ve $DB(9)=0.689$). DB sonuçları arasından, bu iki değer en iyi optimum küme sayısını temsil edebileceği görülmektedir. Bununla birlikte, $DB(k)$ değerini en küçük yapan k değerinin, bu indekse göre ideal küme sayısını ifade ettiği düşünüldüğünde, bu veri seti için en uygun değer $k=5$ için olduğu sonucuna varılmaktadır.

Gap istatistiği algoritması, daha önceden belirtildiği gibi $\log(W_k)$ değerinin, veri setine ait boş referans dağılımı ile karşılaştırılmasını sağlamaktadır. Gap istatistik grafiği veriyi standart sapmaları ile birlikte temsil etmektedir. Eğriye ait sonuçlar incelendiğinde k değerinin 1'den 5'e kadar artış gösterdiği, sonrasında ise önemli bir artışın olmadığı görülmektedir. Gap istatistiği ($GAP_n(k)$) ile en uygun küme sayısının seçimi, $\log(W_k)$ değerinin referans değerinin altında en uzun düşüşe sahip olduğu k değeridir. Bu nedenle, $k=5$ 'in optimum k seçimi olduğu söylenebilir.

Bir diğer yöntem olan Silhouette indeksine ait sonuçlar incelendiğinde (Şekil 3), k küme sayılarına ait Silhouette değerinin $k=5$ ve 6 için en uygun sonuçları elde ettiği söylenebilir (sırasıyla $s(5)=0.505$, $s(6)=0.502$). Bununla birlikte, Silhouette değerinin 1'e daha yakın olduğu ve buna bağlı olarak en iyi sonuçları verdiği göz önüne alındığında, $s(i)=5$ 'in en uygun k seçimi olduğu sonucuna varılabilir. Bu durumlar, GNSS hızlarının kümelenmesi için en uygun küme sayısının 5 olduğunu göstermektedir.

Literatür incelendiğinde, bu bölgede yapılan çalışmalar için optimum küme sayısının beş olduğunu (Özdemir & Karslıoğlu, 2019; Özarpacı vd., 2023) ve veri yoğunluğunun ya da verilerin kampanya / sürekli GNSS verilerinden elde edilmiş olmasının, optimum küme sayısını değiştirmedeği görülmektedir.

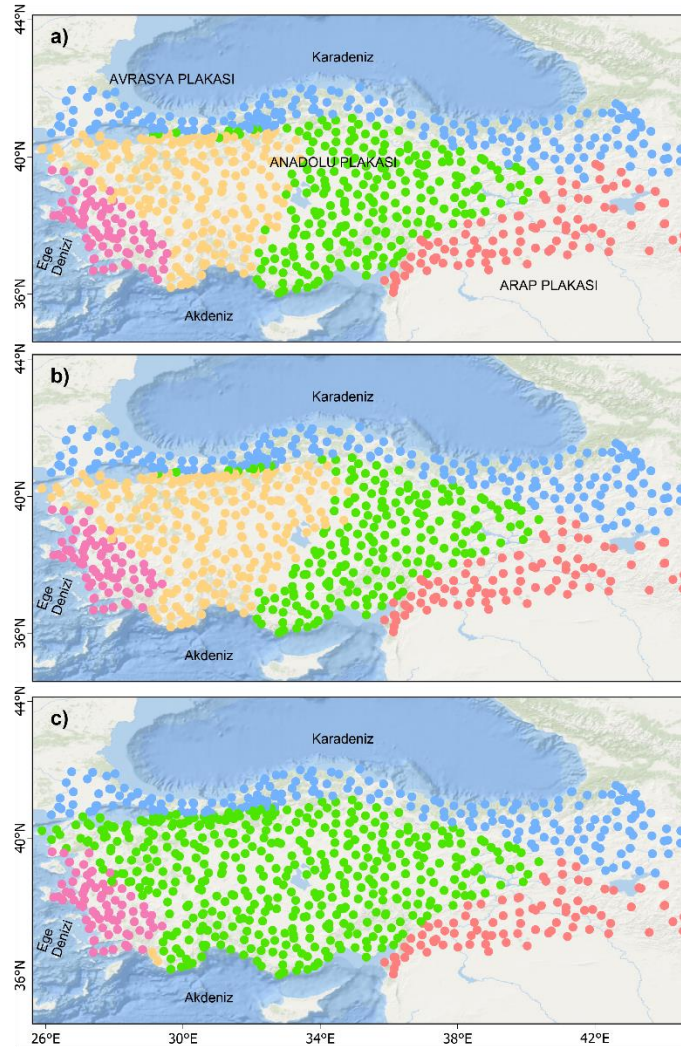
3.2. Tekli kümeleme algoritma sonuçları

3.2. Results of the single clustering algorithm

Tekli kümeleme yöntemleri ile kümelere ayrılan güncel hız alanı Şekil 4'te paylaşılmaktadır. Kümeleme sonuçları şekillendirilirken, aynı bölgeye gelen kümeler benzer renklerde gösterilmiştir. Genel olarak, mavi daireler Avrasya levhasını, pembe daireler Arap levhasını, mor daireler Ege bloğunu, sarı ve yeşil daireler Anadolu bloğunda kalan noktaların renklendirilmesi için kullanılmıştır. Şekil 4a, k -ortalamlar, 4b HAC ve 4c spektral kümeleme yöntemlerinin sonuçlarını göstermektedir. Sonuçlar incelendiğinde (Şekil 4), Anadolu bloğunu Avrasya'dan ayıran KAF ve Arap levhasından ayıran DAF, bütün yöntemlerde hemen fark edilmektedir. Ayrıca Ege bloğu dikkat çekmektedir. Rijit bir yapısı olduğu bilinen Anadolu bloğunun Şekil 4a ve b'de doğu ve batı şeklinde ikiye ayrılması bu bölgedeki hız değişiminin etkisinden kaynaklanmaktadır. Özarpacı vd. (2023), çalışmalarında k -ortalamlar yöntemini kullanarak ve optimum küme sayısını $k=2-10$ arasında kümeleme yaparak değişimleri incelemiş ve küme sayısının artmasıyla değişmeyen sınırların blok sınırı olarak kabul edilebileceğini vurgulamıştır (Savage & Wells, 2015). Çalışmada artan k optimum küme sayısı ile Anadolu bloğu sürekli daha küçük kümelere ayrılarak bu bölgede gerçekten bir blok sınırının olmadığı gösterilmiştir.

Üç yöntemin de benzer sonuçları olmasına rağmen, tekli kümeleme yöntemlerinden elde edilen sonuçlar incelendiğinde aradaki farklar hemen göze çarpmaktadır. K -ortalamlar yönteminin Ege bloğu ve Arap levha sınırları diğer yöntemlere göre daha genişken, spektral kümeleme algoritması Anadolu levhasının neredeyse tamamını tek bir kümeye atamıştır. K -ortalamlar ve HAC yöntemlerinde KAF boyunca özellikle 1999 depremlerinin yüzey kırığına yakın alanda bulunan noktaların fayın sınırından etkilendiği ve sınırın kümeleme yönteminden baskın çıktığı (Savage & Simpson, 2013a) görülürken, spektral kümeleme yönteminde Anadolu bloğunun tamamı aynı kümede olduğundan bu etki görülmemektedir.

Tekli kümeleme algoritmalarının dezavantajı olan bu değişken sonuçlar, Meta-Kümeleme yöntemi için girdi olarak kullanılmış ve Meta-Kümeleme yardımıyla tek bir kümeleme sonucu üretilmiştir. Konsensüs kümeleme yöntemi olan Meta-kümeleme ile GNSS verilerinden elde edilen yatay hız alanı kümelenecek sonuçlar irdelenmiştir. Literatürde oylama yöntemi ile GNSS verilerinden elde edilen hızlar Kılıç ve Özarpacı (2022) ile kümelenecek ancak bu çalışmada, Kılıç ve Özarpacı (2022)'de kullanılan hız alanından daha yoğun bir veri seti ele alınmış ve sonuçları aşağıda paylaşılmıştır.



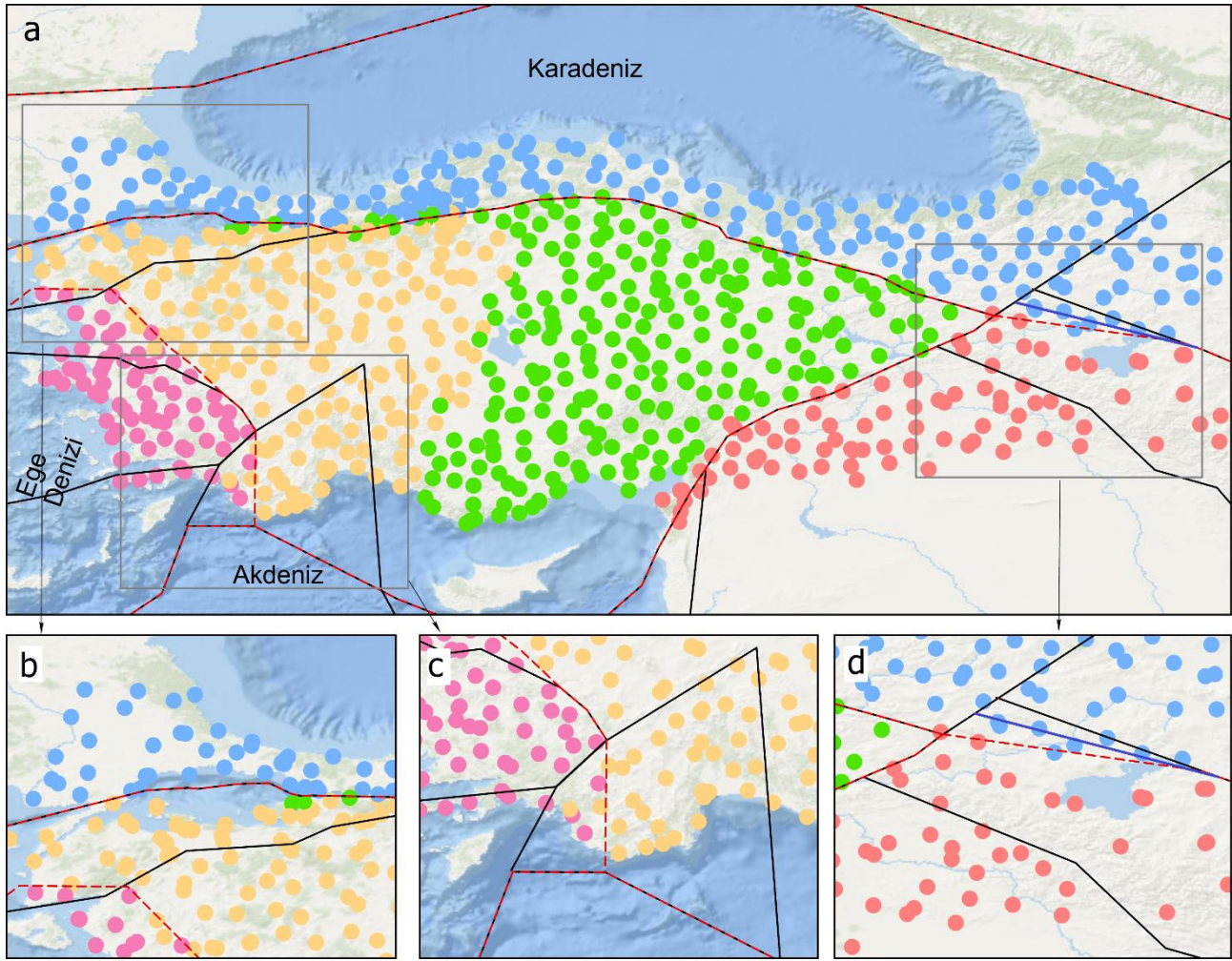
Şekil 4. Tekli kümeleme algoritmaları ile kümeleme sonuçları. a) *k*-ortalamalar b) HAC c) spektral kümeleme yöntemlerinin sonuçlarını göstermektedir. Renkli daireler her bir kümeye denk gelen GNSS gözlem noktasının yerini ifade etmektedir.
Figure 4. Clustering results with single clustering algorithms. a) *k*-means b) HAC c) spectral clustering methods. Coloured circles indicate the location of the GNSS observation point corresponding to each cluster.

3.3. Meta-Kümeleme yöntemi sonuçları

3.3. Meta-Clustering method results

Tekli kümeleme sonuçlarının oylanması işleminin gerçekleştirildiği Meta-Kümeleme topluluk kümelemesi ile elde edilen sonuçlar (Şekil 5) incelendiğinde, tekli kümeleme yöntemlerinin sonuçlarında ortak olarak görülen, KAF ve DAF hemen fark edilmektedir. Ayrıca Ege bloğu göze çarpmakta ve Anadolu bloğunun ikiye ayrılarak beş küme oluşturulduğu görülmektedir.

Literatürde yapılan oylama kümeleme sonuçları ile (Kılıç & Özarpacı, 2022) Doğu Anadolu bölgesindeki kümeleme sonucunda oluşan blok sınırı karşılaştırıldığında, hız alanındaki yoğunluğun artmasının buradan geçen blok sınırının değişmesine neden olduğu görülmektedir. Literatürde belirlenen blok sınırı mavi düz çizgi ile gösterilirken, yeni hız alanının kümelenmesi sonucunda elde edilen blok sınırı kırmızı kesikli çizgi olarak gösterilmektedir. Yeni blok sınırının KAF'ın Karlıova'da DAF ile birleştiği üçlü eklem bölgesinden başladığı göze çarpmaktadır.



Şekil 5. Meta-Kümeleme ile yapılan topluluk kümeleme sonuçları a) Türkiye'nin tamamını b) Marmara Bölgesi'ni c) Güney Ege ve Batı Akdeniz civarını d) Türkiye'nin doğusunu göstermektedir. Siyah sürekli çizgiler Reilinger vd. (2006) çalışmasından alınan blok sınırlarını, mavi sürekli çizgi Kılıç ve Özarpacı (2022) çalışmasından alınan sınır çizgisini, kırmızı kesikli çizgiler ise bu çalışma sonucunda elde edilen yenilenmiş blok sınırlarını göstermektedir.

Figure 5. Ensemble clustering results with Meta-Clustering. a) All Türkiye b) Marmara Region c) South Aegean and Western Mediterranean region d) Eastern Turkey. Black solid lines indicate block boundaries from Reilinger et al. (2006), blue solid line show boundary line from Kılıç and Özarpacı (2022) and red dashed lines indicate the revised block boundaries obtained in this study.

Bu çalışmada oluşturulan blok sınırları (Şekil 5), literatürdeki blok modelleme çalışmalarında kullanılan blok sınırları (Reilinger vd., 2006) referans alınarak incelenmiştir. Şekil 5b, Marmara Bölgesi'nin kümeleme sonuçlarını göstermektedir. KAF, Marmara bölgesinde güney ve kuzey kollara ayrılmaktadır (Ergintav vd., 2014). KAF'ın kayma hızının büyük bir bölümünü üzerine alan kuzey kol (Meade vd., 2002) boyunca hızdaki değişim kümeleme sonuçlarına yansımış ve kümeler kuzey kol boyunca birbirinden ayrılmıştır (Şekil 5b). Reilinger vd. (2006) çalışmasında, güney kol boyunca da bir blok sınırı belirlerken, burada önemli bir hız değişimi olmadığı görüldüğünden, bir sınır oluşturulmamıştır (Şekil 5b).

Ege bloğu da Reilinger vd. (2006) çalışmasından farklılaşmakta ve sınırlarının değiştiği görülmekte olup, Akdeniz Bölgesinin batısında bulunan üçgen benzeri blok sınırı boyunca çalışmamızda herhangi bir hız değişimi görülmemiştir (Şekil 5c).

Son olarak, Şekil 5d'de Türkiye'nin doğusu, Arap levhası ile Avrasya levhasının birleştiği bölgenin blok sınırları incelendiğinde, Reilinger vd. (2006)'den farklı olarak buradaki hız değişiminin Karlıova'dan başlayarak iki ayrı blok oluşturduğu görülmektedir.

4. Sonuçlar

4. Conclusions

Bu çalışma, topluluk kümeleme yöntemlerinden biri olan Meta-Kümeleme yönteminin, Türkiye'nin yeni yayımlanmış, en güncel ve kapsamlı GNSS hız alanından elde edilen yatay hızlar üzerindeki performansı incelenmiştir. Bu kapsamda, GNSS yatay hız alanına en uygun optimum küme sayısı tespit edilmiş ve farklı türde çalışma prensibine sahip üç tekli kümeleme yöntemi ile GNSS hızları kümelendi. Kümeleme yöntemleri ile elde edilen değerler, Meta-Kümeleme yöntemine entegre edilerek çalışma bölgesi ile olan uyumluluk düzeyi irdelenmiştir.

Sonuç olarak yeni bulunan blok sınırları; literatürle benzerlik ve farklılıklar göstermektedir. Kümeleme sonucunda, KAF ve DAF net bir şekilde görülürken, Ege bloğu da Anadolu bloğundan ayrılmaktadır. Bunlarla birlikte, KAF'ın Marmara Bölgesi'ndeki güney kolu boyunca bir blok sınırı oluşmamakta ve çarpışma kuşağındaki hız değişimi ise, iki ayrı değil tek bir çizgi şeklinde kendini göstermektedir.

Çalışmamız, topluluk kümelemesi yöntemlerinden biri olan Meta-Kümeleme algoritmasının GNSS verilerinden elde edilen yatay hız alanında kullanılabildiğini göstermektedir. Ancak veri yoğunluğu bu kümeleme yönteminin sonuçlarını etkilemektedir.

Güncel hız alanı ile elde edilen sonuçlar, bu bölgede yapılmış önceki çalışmaları geliştirir niteliktedir.

Yazar katkısı

Author contribution

Bu çalışmada Özarpacı ve Kılıç yazım ve içeriğin oluşturulmasından, Köküm ve Doğan okuma ve düzenlemeden sorumludur.

Etik beyanı

Declaration of ethical code

Bu makalenin yazarları, bu çalışmada kullanılan materyal ve yöntemlerin etik kurul izni ve / veya yasal-özel izin gerektirmediğini beyan etmektedir.

Çıkar çatışması beyanı

Conflicts of interest

Yazarlar herhangi bir çıkar çatışması olmadığını beyan eder.

Kaynaklar

References

- Abbasi, S. O., Nejatian, S., Parvin, H., Rezaie, V., & Bagherifard, K. (2019). Clustering ensemble selection considering quality and diversity. *Artificial Intelligence Review*, 52(2), 1311-1340. <https://doi.org/10.1007/s10462-018-9642-2>
- Alqurashi, T., & Wang, W. (2019). Clustering ensemble method. *International Journal of Machine Learning and Cybernetics*, 10, 1227-1246. <https://doi.org/10.1007/s13042-017-0756-7>
- Arbelaitz, O., Gurrutxaga, I., Muguerza, J., Pérez, J. M., & Perona, I. (2013). An extensive comparative study of cluster validity indices. *Pattern Recognition*, 46(1), 243-256. <https://doi.org/10.1016/j.patcog.2012.07.021>
- Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2), 224-227. <https://doi.org/10.1109/TPAMI.1979.4766909>
- Driver, H. E., & Kroeber, A. L. (1932). Quantitative expression of cultural relationships. *University of California Publications in American Archaeology and Ethnology*, 31(4), 211-256.
- Emre, Ö., Duman, T.Y., Özalp, S., Elmacı, H., Olgun, Ş., & Şaroglu, F. (2013). Açıklamalı Türkiye Diri Fay Haritası, Ölçek 1:1.250.000. *Maden Tetkik ve Arama Genel Müdürlüğü, Özel Yayın Serisi*, 30, Ankara, Türkiye. ISBN: 978-605-5310-56-1

- Ergintav, S., Reilinger, R. E., Çakmak, R., Floyd, M., Cakir, Z., Doğan, U., King, R. W., McClusky, S., & Özener, H. (2014). Istanbul's earthquake hot spots: Geodetic constraints on strain accumulation along faults in the Marmara seismic gap. *Geophysical Research Letters*, 41(16), 5783-5788. <https://doi.org/10.1002/2014GL060985>
- Ghaemi, R., Sulaiman, M. N., Ibrahim, H., & Mustapha, N. (2009). A survey: clustering ensembles techniques. *International Journal of Computer and Information Engineering*, 3(2), 365-374. <https://doi.org/10.5281/zenodo.1329276>
- Ghosh, J., & Acharya, A. (2011). Cluster ensembles. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1(4), 305-315. <https://doi.org/10.1002/widm.32>
- Golalipour, K., Akbari, E., Hamidi, S. S., Lee, M., & Enayatifar, R. (2021). From clustering to clustering ensemble selection: A review. *Engineering Applications of Artificial Intelligence*, 104, 104388. <https://doi.org/10.1016/j.engappai.2021.104388>
- Granat, R., Donnellan, A., Heflin, M., Lyzenga, G., Glasscoe, M., Parker, J., Pierce, M., Wang, J., Rundle, J., & Ludwig, L. G. (2021). Clustering analysis methods for GNSS observations: A data-driven approach to identifying California's major faults. *Earth and Space Science*, 8(11), e2021EA001680. <https://doi.org/10.1029/2021EA001680>
- Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: a review. *ACM Computing Surveys (CSUR)*, 31(3), 264-323. <https://doi.org/10.1145/331499.331504>
- Karypis, G., & Kumar, V. (1998). A fast and high quality multilevel scheme for partitioning irregular graphs. *SIAM Journal on Scientific Computing*, 20(1), 359-392. <https://doi.org/10.1137/S106482759528799>
- Kaufman, L., & Rousseeuw, P. J. (1990). *Finding groups in data: An introduction to cluster analysis*. Wiley: New York.
- Kılıç, B., & Özarpacı, S. (2022). Ensemble clustering in GPS velocities: A case study of Turkey. *Applied Sciences*, 12(24), 12636. <https://doi.org/10.3390/app122412636>
- Kleinberg, J. (2002). An impossibility theorem for clustering. In S. Becker, S. Thrun, & K. Obermayer (Eds.), *Advances in Neural Information Processing Systems* (ss. 446-453.). MIT Press.
- Kurt, A. İ., Özbakir, A. D., Cingöz, A., Ergintav, S., Doğan, U., & Özarpacı, S. (2023). Contemporary velocity field for Turkey inferred from combination of a dense network of long term GNSS observations. *Turkish Journal of Earth Sciences*, 32(SI-3), 275-293. <https://doi.org/10.55730/1300-0985.1844>
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In L.M. Le Cam, & J. Neyman (Eds.), *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (ss. 281-297). University of California Press.
- Mcclusky, S., Balassanian, S., Barka, A., Demir, C., Ergintav, S., Georgiev, I., Gurkan, O., Hamburger, M., Hurst, K., Kahle, H., Kastens, K., Kekelidze, G., King, R., Kotzev, V., Lenk, O., Mahmoud, S., Mishin, A., Nadariya, M., Ouzounis, A., ... Veis, G. (2000). Global Positioning System constraints on plate kinematics and dynamics in the eastern Mediterranean and Caucasus. *Journal of Geophysical Research: Solid Earth*, 105(B3), 5695-5719. <https://doi.org/10.1029/1999JB900351>
- Meade, B. J., Hager, B. H., McClusky, S. C., Reilinger, R. E., Ergintav, S., Lenk, O., Barka, A., & Ozener, H. (2002). Estimates of seismic potential in the Marmara Sea region from block models of secular deformation constrained by Global Positioning System measurements. *Bulletin of the Seismological Society of America*, 92(1), 208-215. <https://doi.org/10.1785/0120000837>
- Özarpacı, S., Kılıç, B., Bayrak, O. C., Özdemir, A., Yılmaz, Y., & Floyd, M. (2023). Comparative analysis of the optimum cluster number determination algorithms in clustering GPS velocities. *Geophysical Journal International*, 232(1), 70-80. <https://doi.org/10.1093/gji/ggac326>
- Özdemir, S., & Karshoğlu, M. O. (2019). Soft clustering of GPS velocities from a homogeneous permanent network in Turkey. *Journal of Geodesy*, 93(8), 1171-1195. <https://doi.org/10.1007/s00190-019-01235-z>
- Pakhira, M. K. (2012). Finding number of clusters before finding clusters. *Procedia Technology*, 4, 27-37. <https://doi.org/10.1016/j.protey.2012.05.004>

- Reilinger, R., McClusky, S., Vernant, P., Lawrence, S., Ergintav, S., Cakmak, R., Ozener, H., Kadirov, F., Guliev, I., Stepanyan, R., Nadariya, M., Hahubia, G., Mahmoud, S., Sakr, K., ArRajehi, A., Paradissis, D., Al-Aydrus, A., Prilepin, M., Guseva, T., ... Karam, G. (2006). GPS constraints on continental deformation in the Africa-Arabia-Eurasia continental collision zone and implications for the dynamics of plate interactions. *Journal of Geophysical Research: Solid Earth*, 111(B5). <https://doi.org/10.1029/2005JB004051>
- Rendón, E., Abundez, I., Arizmendi, A., & Quiroz, E. M. (2011). Internal versus external cluster validation indexes. *International Journal of Computers and Communications*, 5(1), 27-34.
- Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53-65.
- Savage, J. C., & Simpson, R. W. (2013a). Clustering of GPS velocities in the Mojave Block, southeastern California. *Journal of Geophysical Research: Solid Earth*, 118(4), 1747-1759. <https://doi.org/10.1029/2012JB009699>
- Savage, J. C., & Simpson, R. W. (2013b). Clustering of velocities in a GPS network spanning the Sierra Nevada Block, the northern Walker Lane Belt, and the central Nevada Seismic Belt, California-Nevada. *Journal of Geophysical Research: Solid Earth*, 118(9), 4937-4947. <https://doi.org/10.1002/jgrb.50340>
- Savage, J. C., & Wells, R. E. (2015). Identifying block structure in the Pacific Northwest, USA. *Journal of Geophysical Research: Solid Earth*, 120(11), 7905-7916. <https://doi.org/10.1002/2015JB012277>
- Savage, J. C. (2018). Euler-vector clustering of GPS velocities defines microplate geometry in southwest Japan. *Journal of Geophysical Research: Solid Earth*, 123(2), 1954-1968. <https://doi.org/10.1002/2017JB014874>
- Simpson, R. W., Thatcher, W., & Savage, J. C. (2012). Using cluster analysis to organize and explore regional GPS velocities. *Geophysical Research Letters*, 39(18). <https://doi.org/10.1029/2012GL052755>
- Strehl, A., & Ghosh, J. (2000). Value-based customer grouping from large retail data sets. *Proceedings SPIE 4057, Data Mining and Knowledge Discovery: Theory, Tools, and Technology II* (pp. 33-42), Orlando, FL, United States.
- Strehl, A., & Ghosh, J. (2002). Cluster ensembles – A knowledge reuse framework for combining multiple partitions. *Journal of Machine Learning Research*, 3, 583-617. <https://doi.org/10.1162/153244303321897735>
- Takahashi, A., Hashimoto, M., Hu, J. C., Takeuchi, K., Tsai, M. C., & Fukahata, Y. (2019). Hierarchical cluster analysis of dense GPS data and examination of the nature of the clusters associated with regional tectonics in Taiwan. *Journal of Geophysical Research: Solid Earth*, 124(5), 5174-5191. <https://doi.org/10.1029/2018JB016995>
- Thatcher, W. (2009). How the continents deform: The evidence from tectonic geodesy. *Annual Review of Earth and Planetary Sciences*, 37, 237-262. <https://doi.org/10.1146/annurev.earth.031208.100035>
- Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2), 411-423. <https://doi.org/10.1111/1467-9868.00293>
- Topchy, A., Jain, A. K., & Punch, W. (2005). Clustering ensembles: Models of consensus and weak partitions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(12), 1866-1881. <https://doi.org/10.1109/TPAMI.2005.237>
- Xu, D., & Tian, Y. (2015). A comprehensive survey of clustering algorithms. *Annals of Data Science*, 2, 165-193. <https://doi.org/10.1007/s40745-015-0040-1>
- Vega-Pons, S., & Ruiz-Shulcloper, J. (2011). A survey of clustering ensemble algorithms. *International Journal of Pattern Recognition and Artificial Intelligence*, 25(03), 337-372. <https://doi.org/10.1142/S0218001411008683>
- Von Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and computing*, 17, 395-416. <https://doi.org/10.48550/arXiv.0711.0189>
- Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301), 236-244. <https://doi.org/10.1080/01621459.1963.10500845>
- Yan, D., Huang, L., & Jordan, M. I. (2009). Fast approximate spectral clustering. *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 907-916). Paris, France.