6th Intercontinental Geoinformation Days

igd.mersin.edu.tr

NMF-based ensemble clustering for GNSS velocity field: tectonic insights

Seda Özarpacı*[1], Batuhan Kılıç[1]

[1]Yildiz Technical University, Faculty of Civil Engineering, Department of Geomatic Engineering, Istanbul, Türkiye

Keywords
Block modeling
GNSS horizontal velocity field
Clustering
NMF
Ensemble clustering

Abstract

Block modeling is a frequently employed technique to yield the block motions such as block rotations and internal strains as well as to estimate fault slip rates via GNSS velocity data. The accuracy of the modeling outcomes heavily relies on the precise determination of the block boundaries. Typically, surface traces of mapped faults, seismological data, and/or GNSS velocity fields are utilized for this purpose. Nonetheless, the identification of suitable and accurate block boundaries for the velocity field is subject to interpretation and subjectivity. One method that can be used to determine block boundaries is the clustering method. Clustering analysis assigns data to similar groups based on similarities and differences in the data subject to clustering. Since the 2000s, this method has been used in the clustering of GNSS velocities and helps to determine block boundaries before block modeling. In this study, we utilized the Non-negative Matrix Factorization (NMF) method to cluster the current GNSS horizontal velocity field of Türkiye, and we compared the results with the literature. First, we clustered velocities from k = 2 to 9 using five distinct single clustering methods to determine the block boundaries. Then, we used the optimum number k = 5 to cluster the data using NMF-based ensemble clustering.

1. Introduction

The tectonics of Türkiye have a complex structure due to the collision of the Eurasian and Arabian plates, the escape resulting from this collision, and the subduction zones formed by the Anatolian and African plates along the Cyprus and Hellenic arcs. The North Anatolian Fault (NAF) and the East Anatolian Fault (EAF), which were formed as a result of these processes, are the most important strike-slip faults in the region (McClusky et al. 2000; Reilinger et al. 2006).

Although many models are used to understand and interpret these structures, block modeling produces the closest results to surface deformations. The accuracy of the modeling results depends on how well the block boundaries are defined (Thatcher 2009). Although many researchers have used mapped fault surface traces, seismology data, and/or GNSS velocity fields to determine block boundaries, the determination of block boundaries is subjective (Reilinger et al. 2006).

One method that can be used to determine block boundaries is the clustering method. Clustering analysis assigns data to similar groups based on similarities and differences in the data subject to clustering, without any prior information. This method has been used by different researches (Savage and Simpson 2013; Savage and Wells 2015; Özarpacı et al. 2023) in the clustering of GNSS velocities and helps to determine block boundaries before block modeling.

In previous studies, various clustering methods have been used to determine block boundaries from GNSS velocities, and their performances have been analyzed. However, it is also observed that different algorithms produce different results. In this study, an ensemble clustering algorithm is tested to eliminate the subjectivity arising from clustering methods. For this purpose, using the last velocity field obtained from updated, continuous, and campaign GNSS data for Türkiye (Kurt et al. 2023), first, individual clustering methods were used, and then the success of the Non-Negative Matrix Factorization (NMF) algorithm was tested based on the results obtained from these methods.

2. Method

In the clustering analysis, the horizontal velocities of 836 stations were evaluated by means of the current

* Corresponding Author

*(ozarpaci@yildiz.edu.tr) ORCID ID 0000-0002-1900-3725
(batuhank@yildiz.edu.tr) ORCID ID 0000-0002-0529-8569

Cite this study

Ozarpaci, S., & Kilic, B. (2023). NMF-based ensemble clustering for GNSS velocity field: tectonic insights. Intercontinental Geoinformation Days (IGD), 6, 207-210, Baku, Azerbaijan

GNSS velocity field (Kurt et al. 2023). Diverse clustering techniques were utilized to generate clustering models (k = 2 to 9) from multiple community members to form the NMF ensemble clustering approach. For each k value, the outputs of the members were subsequently combined using the NMF ensemble clustering method to derive the ultimate clustering results, and the conformity of the final solutions with the tectonic structure of the region was evaluated (Figure 1).

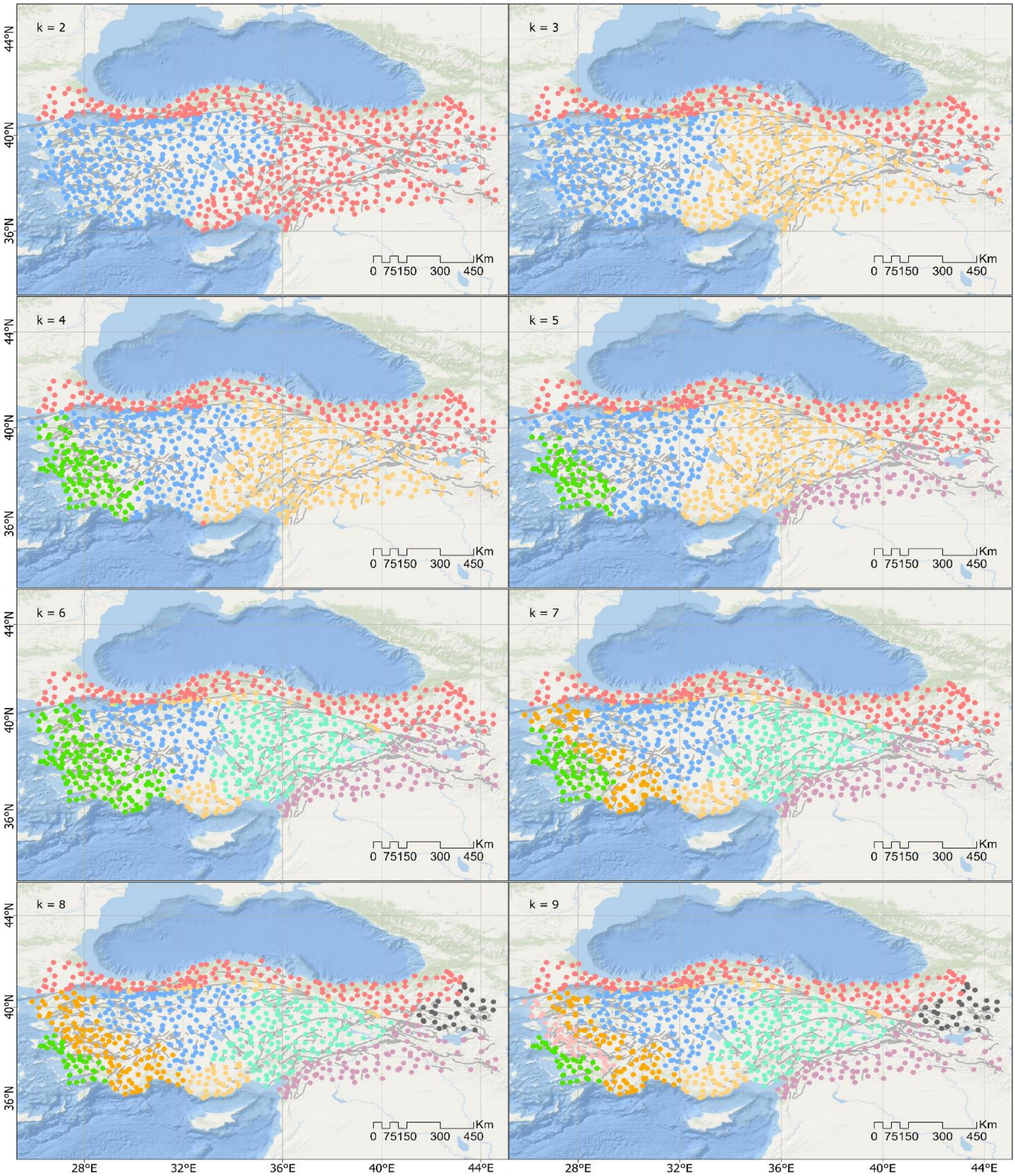

Figure 1. Ensemble clustering results with NMF (from k = 2 to k = 9). The gray thin lines indicate active faults (Emre et al. 2013)

2.1. Single clustering techniques

Each clustering technique partitions the dataset based on certain criteria and, as a result, generates different results on the same dataset from a specific perspective. Hence, in order to obtain more profound insights into the underlying patterns in the data, it is crucial to maintain a diverse range of member results (Golalipour et al. 2021; Kılıç and Özarpacı 2022). This study utilized various clustering algorithms, including k-means, Balanced Iterative Reducing and Clustering using Hierarchies (BIRCH), Hierarchical Agglomerative Clustering (HAC), mini-batch k-means, and spectral clustering techniques, to generate diversity in the ensemble clustering member results. Among these methods, k-means and mini-batch k-means are partitioning-based, BIRCH and HAC are hierarchy-based, and spectral clustering is graph-based.

2.2. NMF ensemble clustering technique

NMF is a clustering ensemble method that contains a non-negative matrix factorization process. This approach was initially introduced by Paatero and Tapper (1994). The NMF method is characterized by factorizing a non-negative data matrix V into two matrix factors, W and H, where V ≈ WH and both W and H are constrained to be non-negative. The matrix W represents the dimensionally reduced data matrix, whereas the matrix H corresponds to the associated coefficient matrix, i.e., the weights associated with W. Despite the emergence of various extensions that incorporate the clustering feature, the initial model formulation of NMF did not explicitly include a clustering objective, as it was primarily introduced as a dimensionality reduction algorithm (Kılıç and Özarpacı 2022).

3. Results and Discussion

For each value of k, the outcomes of the NMF ensemble clustering method produced over the results obtained from single clustering techniques are shown in Figure 1. As one can see that the NAF and EAF exhibit distinct separation of clusters after k = 5. Also, after k = 4 clusters, the Aegean coast obviously creates a block with a velocity gradient. In our clustering results from k = 2 to 9, we see that Anatolia is being divided by different clusters. However, Anatolia's geological structure is widely believed to be tectonically rigid and characterized by uniform motion (McClusky et al. 2000; Reilinger et al. 2006). As pointed out by Savage and Wells (2015), "cluster boundaries that remain relatively stable as k is increased are tentatively identified as block boundaries" in geographical space. Therefore, our results demonstrate that Anatolia is a rigid block because the pattern of clusters does not remain the same as it did along NAF or EAF block boundaries (Özarpacı et al. 2023).

We take the optimal number of clusters for Türkiye as five in terms of the literature (Özdemir and Karslıoğlu 2019; Kılıç and Özarpacı, 2022; Özarpacı et al. 2023), as it is the minimum number required to adequately explain the data without over-parametrization. Furthermore, we see some stations around the 1999 Izmit and Duzce earthquakes behave differently. The underlying reason for this phenomenon can be interpreted as the block boundary that exists along the NAF being more influential as a distinctive characteristic, relative to the clustering approach employed (Savage and Simpson 2013).

Figure 2 illustrates the outcomes of the NMF ensemble clustering approach with 5 clusters applied to the GNSS velocity data of Türkiye, along with the block boundaries derived from Reilinger et al. (2006) and NMF-based analyses.

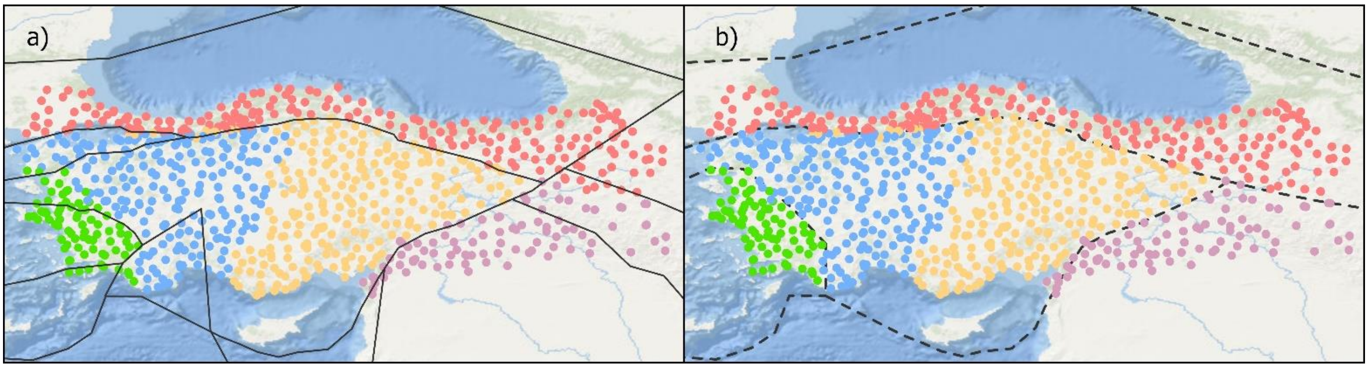

Figure 2. Horizontal velocity clustering results for k = 5 clusters with NMF: (a) black solid lines indicate block boundaries from Reilinger et al. (2006); (b) black dashed lines indicate the revised block boundaries obtained by the NMF ensemble clustering technique

Observations indicate that NAF is partitioned into two branches, namely, the southern and the northern branches, in the Marmara region. The velocity changes along the northern branch, which bears the majority of the velocity of the GNSS stations along NAF, manifest in the clustering results, and the resultant clusters are observed to be disjoined along the said branch (as depicted in Figure 2). Reilinger et al. (2006) established a block boundary along the southern branch based on their study. However, since the velocity gradients along this branch were not significant, no boundary was determined in the given study (as illustrated in Figure 2b). The present study reveals differences in the boundaries of the Aegean block compared to Reilinger et al.'s (2006) investigation. Specifically, the boundaries of the Aegean block appear to have undergone changes, and no significant velocity change was observed in the region along the triangular-shaped block boundary to the west of the Mediterranean Region (Figure 2b). Finally, upon analyzing the block boundaries of the eastern part of Türkiye in Figure 2b, it is observed that the velocity gradient in this area results in two distinct blocks, which are distinct from the findings of Reilinger et al. (2006).

4. Conclusion

In this study, the effectiveness of non-negative matrix factorization (NMF), an ensemble clustering method, in clustering the horizontal velocities obtained from the latest and most comprehensive GNSS velocity field of Türkiye and its compatibility with the tectonic structure of the region were investigated. GNSS velocities were subjected to clustering analysis ranging from 2 to 9 using five distinct single clustering methods to identify the most appropriate number of clusters for the GNSS horizontal velocity field. The outcomes yielded from each of the clustering methods were then incorporated into the NMF approach to assess the consistency with the study area. Beyond five clusters, it can be observed that the Eurasian and Arabian plates maintain their distinguishability, the Southern Aegean region exhibits consistency, and the Anatolian plate becomes increasingly partitioned into additional clusters. Furthermore, it is important to note that no block boundary exists along the southern branch of the NAF in the Marmara Region, and the velocity gradient in the eastern part appears as a single branch rather than being separated into two distinct branches.

References

Emre, Ö., Duman, T. Y., Özalp, S., Elmacı, H., Olgun, S., & 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.

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.

Kılıç, B., & Özarpacı, S. (2022). Ensemble Clustering in GPS Velocities: A Case Study of Turkey. Applied Sciences, 12(24), 12636.

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(3), 275-293.

Mcclusky, S., Balassanian, S., Barka, A., Demir, C., Ergintav, S., Georgiev, I., ... & 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.

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.

Özdemir, S., & Karslıoğlu, M. O. (2019). Soft clustering of GPS velocities from a homogeneous permanent network in Turkey. Journal of Geodesy, 93(8), 1171-1195.

Paatero, P., & Tapper, U. (1994). Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. Environmetrics, 5(2), 111-126.

Reilinger, R., McClusky, S., Vernant, P., Lawrence, S., Ergintav, S., Cakmak, R., ... & 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).

Savage, J. C., & Simpson, R. W. (2013). Clustering of GPS velocities in the Mojave Block, southeastern California. Journal of Geophysical Research: Solid Earth, 118(4), 1747-1759.

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.

Thatcher, W. (2009). How the continents deform: The evidence from tectonic geodesy. Annual Review of Earth and Planetary Sciences, 37, 237-262.