УДК 550.3+519.2 DOI: https://doi.org/10.24028/gj.v47i2.322475

Application of cluster analysis to the study of spatial patterns of earthquakes in Azerbaijan and adjacent territories

S.I. Skurativskyi[1], S.V. Mykulyak[1], Yu.V. Semenova[1], K.S. Skurativska[2], 2025

[1]S. Subbotin Institute of Geophysics of National Academy of Sciences of Ukraine, Kyiv, Ukraine
[2]University of Padova, Padova, Italy

Based on the data of the earthquake catalog in Azerbaijan and adjacent areas from 2010 to 2023, a procedure for dividing earthquakes into clusters was developed using the DBSCAN algorithm for Python. In the process of dividing into clusters, a number

Citation: Skurativskyi, S.I., Mykulyak, S.V., Semenova, Yu.V., & Skurativska, K.S. (2025). Application of cluster analysis to the study of spatial patterns of earthquakes in Azerbaijan and adjacent territories. *Geofizychnyi Zhurnal*, *47*(2), 145—150. https://doi.org/10.24028/gj.v47i2.322475.

of indicators were used that allowed optimizing the number of clusters and the number of earthquakes in the clusters. A comparison of the location of earthquake clusters with tectonic faults in this region demonstrated their high correlation.

Key words: earthquake, Azerbaijan, clustering, DBSCAN algorithm, network of faults.

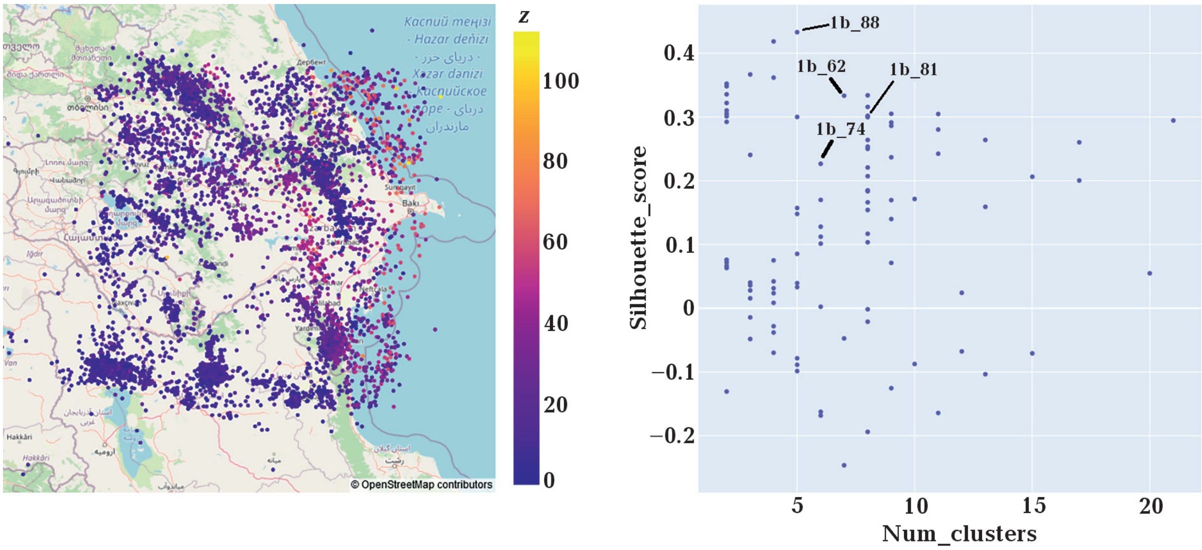
Introduction. Studying the patterns of earthquake location and their periodicity and intensity is an urgent task given the danger and consequences that earthquakes cause to people, buildings, structures, and various infrastructure facilities. This knowledge is important for understanding how earthquakes occur and is also an integral part of the theoretical basis for seismic hazard assessment methods [Scitovski, 2018]. It is well known that earthquakes happen in areas located on the boundaries of tectonic plates and faults [Scholz, 2019]. These areas should be considered as complex systems, and such an approach to their study has recently become quite widespread. The dynamic behavior of complex systems is described by statistical methods. Many models and approaches to studying dynamic processes in complex systems have been developed [Holovatch et al., 2017]. One of the most rapidly developing methods is the cluster analysis method, which allows one to group data of seismic observations algorithmically, reducing the procedure's subjective [Georgoulas et al., 2013] component.

Objects and Methodology. In this research, we deal with earthquakes in the Caucasus region characterized by strong seismicity [Tsereteli et al., 2016; Semenova et al., 2024]. In particular, we consider the spatial patterns of earthquakes in Azerbaijan and adjacent areas from 2010 to 2023. The earthquake catalog is clustered using the DBSCAN algorithm [Scitovski, 2018]. The Silhouette index is used for quality control as it provides the optimal number of clusters. The clustering result is not unique, so a similarity index, i.e. the Adjusted Rand Index (ARI), is used to compare different partitions.

Results and discussion. Using the catalog [ISC, 2024], the set of earthquakes containing 6201 readings is extracted. We will use a seismic event's latitude, longitude, and depth. To get better results, it is recommended to that the data be pre-processed. To do this, we scale not all variables but only the depth according to the relation $Z_j = z_j \Big/ \sum_j |z_j|$. The number of earthquakes with a magnitude greater than 5.5 was also extracted from the catalog. The epicenters of earthquakes are located between 38° and 42° N and 45° and 50° E. The spatial patterns of the selected earthquakes are shown in Fig. 1, *a*, where the earthquake depth is marked in color. For the cluster analysis, we use the DBSCAN algorithm. It is a density-based procedure governed by two intrinsic parameters — ε and *min_samples*. This algorithm does not require the pre-definition of several clusters and can identify clusters of arbitrary shapes. This is especially important for seismic analysis. The result of DBSCAN's work is the set of clusters accompanied by a set of unclassified points (noise points). The cluster testing procedure is implemented using the *sklearn.cluster* in Python.

Thus, selecting the parameter ε from the range (0.1; 1.1) with a step of 0.1 and the parameter *min_samples* from the range (25; 151) with a step of 10, we evaluate different partitions of the catalog into clusters. A validation index was applied to roughly evaluate the clustering results. Among many indices, we choose the Silhouette Index (*SI*), which approaches maximum when all clusters are well separated and dense enough. It varies from –1 (poorly classified elements) to 1 (perfect clustering). According to the dependence of *SI* on the number of clusters (Fig. 1, *b*), *SI* grows for the small number of clusters, achieves its local maximum at 5 clusters, and then decreases. Note that the same number of clusters can be evaluated for different parameters and *min_samples*.

Omitting the cases of the small number of clusters, let us consider the partitions with 5—8 clusters in more detail. The DBSCAN's parameters for themare written in Table.



a *b*

Fig. 1. The distribution of the earthquake epicenters in the selected region (*a*). The color bar marks the depth of earthquakes from 0.1 km (dark color) to 90 km (light color) (*b*). The dependence of *SI* on the number of clusters. Here, lb_88 and lb_81 mark the clustering cases.

The parameters of the DBSCAN algorithm

N	5 (lb_88)	6 (lb_74)	7 (lb_62)	8 (lb_81)
ε	0.4	0.3	0.3	0.4
min_samples	145	145	85	105

Like most indices, *SI* cannot conclusively confirm the correctness of clustering due to its limitations. Therefore, we deal not only with cases when *SI* reaches a maximum but also with cases close to it.

In particular, for *N*=6 and *N*=7, we only take cases with maximal *SI* because *SI* is not high enough (for *N*=6) or *SI* is positive for a single case (*N*=7). When we consider the partitions containing five clusters (*N*=5), we analyze the case labeled lb_88 and, in addition, cases that satisfy the constraint *SI*>0.1. There are four such cases, as shown in Fig. 1, *b*. Similarly, considering the partition consisting of *N*=8 clusters and the constraint *SI*>0.2, we obtain 9 cases.

To quantify the differences between the four partitions selected for *N*=5, we apply ARI, which measures similarity between two clusterings. This can help us to reduce the number of partitions that need to be separated. ARI ranges from –0.5 (especially discordant partitions) to 1 (identical partitions).

The results of ARI evaluation for a pair of clusterings are presented in Fig. 2, *a* in the matrix form. The captions of the rows and columns refer to the partitions for which ARI is evaluated. The evaluated ARI is in the corresponding cell. Thus, ARIs for identical partitions are equal to 1 and depicted in dark colors, while other combinations yield smaller ARIs and are represented in lighter. We see that for the pair lb_88 and lb_79, the ARI is 0.668; for the pair lb_75 and lb_73, the ARI is 0.981.Therefore, the selected partitions are highly similar. Other pairs are less similar since their ARIs are around 0.3.

In the same way, the 9×9 ARI matrix is calculated for partitions of eight clusters (*N*=8) (Fig. 2, *b*). Note that all ARIs are greater than 0.54 (thus, the selected partitions are highly similar).

Thus, we select the partition lb_88 for *N*=5, the partition lb_62 for *N*=7, and the partition lb_81 for *N*=8. We map these cases (Fig. 3), where the noise points are eliminated. In addition, the map also shows the network of faults and locations of the strongest earthquakes (*M*>5.5). Analyzing Fig. 3, we can conclude that the sections correlate well with the

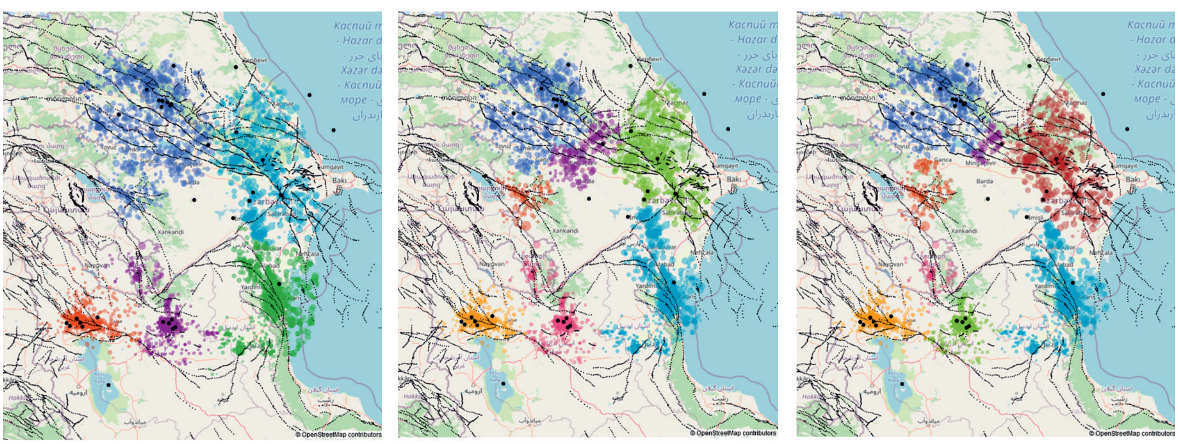
	Labels_88	Labels_79	Labels_73	Labels_75
Labels_88	1	0.6684307	0.3287655	0.3204467
Labels_79	0.6684307	1	0.3140118	0.309387
Labels_73	0.3287655	0.3140118	1	0.980879
Labels_75	0.3204467	0.309387	0.980879	1

a

	Labels_83	Labels_66	Labels_81	Labels_64	Labels_70	Labels_72	Labels_89	Labels_65	Labels_67
Labels_83	1	0.9116091	0.9068556	0.8381935	0.7891408	0.7646648	0.8037293	0.6517669	0.6204884
Labels_66	0.9116091	1	0.8284298	0.832619	0.8659147	0.8399748	0.8414726	0.7200661	0.6866225
Labels_81	0.9068556	0.8284298	1	0.8931901	0.7076072	0.6848968	0.7214077	0.5769968	0.5479461
Labels_64	0.8381935	0.832619	0.8931901	1	0.7079607	0.6848253	0.705369	0.574715	0.5449491
Labels_70	0.7891408	0.8659147	0.7076072	0.7079607	1	0.9724405	0.9228662	0.8458201	0.8109603
Labels_72	0.7646648	0.8399748	0.6848968	0.6848253	0.9724405	1	0.9166669	0.8607558	0.8327758
Labels_89	0.8037293	0.8414726	0.7214077	0.705369	0.9228662	0.9166669	1	0.7989854	0.7686968
Labels_65	0.6517669	0.7200661	0.5769968	0.574715	0.8458201	0.8607558	0.7989854	1	0.9609688
Labels_67	0.6204884	0.6866225	0.5479461	0.5449491	0.8109603	0.8327758	0.7686968	0.9609688	1

b

Fig. 2. Adjusted Rand Index for the 5-cluster partition (*a*) and 8-cluster partition (*b*).



a *b* *c*

Fig. 3. The partitions containing 5 (*a*), 7 (*b*), and 8 (*c*) clusters. The solid curves mark faults. Black bullets correspond to the region's strongest earthquakes (*M*>5.5).

fault network and the distribution of strong seismic events.

Conclusion and perspective. In this research, the earthquake clustering procedure is developed using cluster analysis tools. In particular, earthquakes in Azerbaijan and adjacent areas from 2010 to 2023 were considered. The catalog of earthquakes as a dataset of elements characterized by longitude, latitude, and depth was divided into clusters via the DBSCAN algorithm. The application of the algorithm provides a variety of partitions into clusters depending on the intrinsic algorithm's parameters. To validate the DBSCAN results, the Silhouette Index was used to identify the appropriate clusterings with maximal Silhouette Indices. After this, the final clustering of the earthquake catalog is still not unique. The further application of the Adjusted Rand Index showed that some partitions are similar and almost indistinguishable. Another auxiliary criterion for determining the correct clustering of earthquakes is based on comparing the location of the clusters with the structure of faults and the locations of strong earthquakes with a magnitude of more than 5.5. In consequence, a reduced number of earthquake clusterings was obtained.

The results can be used to identify particularly dangerous seismogenic zones and determine the type of seismicity [Taroni et al., 2024] and the geometry of faults in cluster areas. They can also be used to perform risk assessments and clarify the statistical regularities of earthquake sequences.

References

Georgoulas, G., Konstantaras, A., Katsifarakis, E., Stylios, C.D., Maravelakis, E., & Vachtsevanos, V. (2013). «Seismic-mass» density-based algorithm for spatio-temporal clustering. *Expert Systems with Applications, 40*, 4183—4189. https://doi.org/10.1016/j.eswa.2013.01.028.

Holovatch, Yu., Kenna, R., & Thurner, S. (2017). Complex systems: physics beyond physics. *European Journal of Physics, 38*, 02300. https://doi.org/10.1088/1361-6404/aa5a87.

ISC: International Seismological Centre. (2024). Online Bulletin.Retrieved from http://www.isc.ac.uk/iscbulletin/search/.

Semenova, Yu., Kendzera, O., Skurativskyi, S., Mykulyak, S., Skurativska, I., & Topoliuk, O. (2024). Seismic hazard assessment in the Shamkir-Mingachevir reservoir region through ground response analysis. *IOP Conference Series: Earth and Environmental Science (EES), 1415*(1), 012007. https://doi.org/10.1088/1755-1315/1415/1/012007.

Scholz, C.H. (2019). *The mechanics of earthquakes and faulting* (3rd ed.) Cambridge, UK: Cambridge University. https://doi.org/10.1017/978 1316681473.

Scitovski, S. (2018). A density-based clustering algorithm for earthquake zoning.*Computers and Geosciences, 110*, 90—95. https://doi.org/ 10. 1016/j.cageo.2017.08.014.

Taroni, M., Consoler, R., Montuori, C., Murru, M., Falcone, G., Chiaraluce, L., & Pastoressa, A.L. (2024). Statistically significant difference between earthquake size distributions of independent and triggered seismicity. *Communications Earth & Environment, 5*, 193 https://doi.org/10.1038/s43247-024-01367-x.

Tsereteli, N., Tibaldi, A., Alania, V., Gventsadse, A., Enukidze, O., Varazanashvili, O., & Muller, B. (2016). Active tectonics of central-western Caucasus, Georgia. *Tectonophysics, 691*, 328—344. https://doi.org/10.1016/j.tecto.2016.10.025.

Застосування кластерного аналізу до вивчення просторових патернів землетрусів Азербайджану та прилеглих територій

С.І. Скуратівський[1], С.В. Микуляк[1], Ю.В. Семенова[1], К.С. Скуратівська[2], 2025

[1]Інститут геофізики ім. С.І. Субботіна НАН України, Київ, Україна
[2]Університет Падови, Падова, Італія

На підставі даних каталогу землетрусів в Азербайджані та прилеглих до нього районів за період з 2010 по 2023 р. розроблено процедуру розбиття землетрусів на кластери з використанням алгоритму DBSCAN для Python. У процесі розбиття на кластери використовувався ряд індикаторів, які дали змогу оптимізувати кількість кластерів і кількість землетрусів у кластерах. Порівняння розташування кластерів землетрусів з тектонічними розломами в цьому регіоні продемонструвало їх високу кореляцію.

Ключові слова: землетруси, Азербайджан, кластеризація, алгоритм DBSCAN, мережа розломів.