Kaunia : Integration and Interconnection of Islam and Science Journal
Vol. 21, No.1, April 2025, pp. 27-39
ISSN 1829-5266 (paper), E-ISSN 2962-2085 (online)
Available online at http://ejournal.uin-suka.ac.id/saintek/kaunia

Subduction and Local Fault Earthquake Analysis Using ST-DBSCAN Clustering Algorithm in The Special Region of Yogyakarta (DIY)

Wuri Handayani[1*]
[1] *Information Systems Program*
Mercubuana Yogyakarta University
*Corresponding author: *wurihandayanisutarjo@gmail.com*

Irfan Pratama[2]
[2] *Information Systems Program*
Mercubuana Yogyakarta University

Nugroho Budi Wibowo[3]
[3] *Sleman Geophysics Station*
Meteorology Klimatology and Geophysics Agency

Submitted: 8th July 2025 ; **Accepted**: 15th July 2025 ; **Published**: 27th August 2025

Abstract
This study aims to analyze the spatio-temporal patterns of subduction and local fault earthquakes in the Special Region of Yogyakarta using the ST-DBSCAN (Spatio-Temporal Density-Based Spatial Clustering of Applications with Noise) algorithm. A total of 5,403 earthquake events from 2019 to 2024 were clustered using spatial parameters (2–5 km) and a temporal window of 10 days. The results were evaluated using the Davies-Bouldin Index (DBI) and Silhouette Score. In the subduction zone, nine clusters were identified with a DBI of 3.23 and a Silhouette Score of 0.18, indicating moderate separation. Meanwhile, 25 clusters were detected in the local fault zone, particularly around the Opak and Oyo Faults, with a higher DBI of 3.82 and a negative Silhouette Score (-0.14), suggesting overlapping clusters and weak structure. The clustering outcomes correlate with geological features and offer insights for improving earthquake hazard assessment and early warning systems in Yogyakarta.
Keywords:earthquakes; subduction; local faults; Davies-Bouldin Index; Silhouette Score

---

INTRODUCTION

Earthquakes occur when energy stored in the Earth's crust is released suddenly, usually due to the breaking of rock layers or fault movement between tectonic plates (Shearer, 2019). Indonesia lies at the convergence of three major tectonic plates—Indo-Australian, Eurasian, and Pacific—which makes it highly prone to seismic and volcanic activity (Hutchings & Mooney, 2021). The Special Region of Yogyakarta (DIY), is one of the provinces in Indonesia located in the North of the Indo-Australian and Eurasian Plate Subduction Line. This condition makes the region have a history record of destructive and felt seismicity. BMKG's catalog of destructive and significant earthquakes, records that from 1800 to 2024 the DIY region experienced 4 destructive earthquakes and hundreds of earthquakes were felt (BMKG, 2024). In general, earthquakes in Yogyakarta originate from subduction activities and local faults. That earthquakes have thousands of casualties of fund for rehabilitation and reconstruction. The various potensial natural disaters we should have to know, so that the character of these natural hazard impacts can be minimized. Seismic sequences are not formatted randomly, but they follow a spatial pattern with consequent triggering of events or this event produces non-random grouping (Tanaka & Matsumoto, 2024). This grouping, well-known as clustering, allows retrieval of spatio-temporal patterns formed by earthquake events (Lee, 2023). The method that can be used to analyze earthquake data spatially and temporally is the ST-DBSCAN (Spatio-Temporal Density-Based Spatial Clustering of Applications with Noise) algorithm (Iswari, 2022).

The clustering process in this method is based on the distribution of the epicenter of the earthquake and the time of the earthquake, so that the clustering results can represent the source of the earthquake more accurately. In DBSCAN, the density associated with a point is obtained by counting the number of points in a region of specified radius around the point. Points with a density above aspecified threshold are constructed as clusters. Among the existing clustering algorithms, we have chosen DBSCAN because of its ability to discover clusters with arbitrary shapes—such as linear, concave, or oval. Furthermore, unlike other clustering algorithms, DBSCAN does not require a predefined number of clusters. DBSCAN has been proven in its ability of processing very large databases ((Kong, 2024); (Wang & Zhan, 2023); (Ester et al., 1996)).

This study examines the application of the ST-DBSCAN (Spatio-Temporal Density-Based Spatial Clustering of Applications with Noise) algorithm, which is a development of the DBSCAN algorithm with the addition of temporal dimensions to group earthquake data based on spatial (latitude-longitude coordinates) and temporal (date of occurrence) aspects. Based on a study (Gaonkar & Sawant, 2013), ST-DBSCAN proved to be effective in identifying earthquake clusters on the island of Java, including the Yogyakarta area, with optimal parameters of Eps1=20 (spatial distance), Eps2=5 (temporal distance), and MinPts=3, resulting in 13 clusters and 10 noise points with a Silhouette Coefficient value of 0.538, indicating good cluster quality. In Sulawesi, the study (Manalu et al., 2021) also utilized ST-DBSCAN to analyze spatial-temporal seismic patterns with parameters Eps1=0.45, Eps2=7, and MinPts=4, resulting in 60 clusters and 216 noise points. These two studies emphasize the importance of parameter adjustment through k-dist graph observation to optimize clustering results, especially in overcoming heterogeneous seismic data complexity. These results prove that ST-DBSCAN is able to uncover spatial-temporal patterns of earthquakes, although further validation is required to reduce cluster overlap and improve the accuracy of disaster risk prediction, so that the clustering results can represent the earthquake source more accurately. In recent years, this approach has been successfully applied in various seismic regions. For example, ST-DBSCAN-EV has been used to detect earthquake clusters in Chile (Nicolis et al., 2024), while similar studies in Greece have helped map active fault systems (Bountzis et al., 2022). These applications highlight how ST-DBSCAN can adapt to different geological contexts and contribute to a deeper understanding of seismic behavior.

The implementation of ST-DBSCAN in Indonesia remains significantly constrained, especially for earthquake clusters (Ajitomo & Pratama, 2024). The earthquake clustering process needs to pay attention to the characteristics of clusters in subduction earthquakes and local fault earthquakes, so this study aims to model earthquake clustering in the DIY area using the ST-DBSCAN algorithm and test the quality of clustering results with validation metrics (Davies-Bouldin Index and Silhouette Score) (Wijaya et al., 2024).

METHOD

The methodology of this study consists of several stages to produce the best clustering of subduction earthquakes and local faults. The stages consist of data collection, data preprocessing and classification of tectonic zones, application of ST-DBSCAN algorithm, cluster validity, visualization and interpretation (Figure 1).

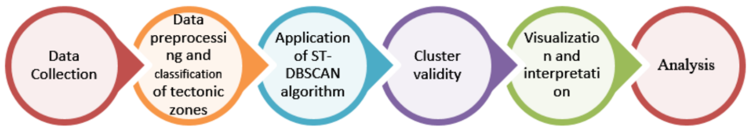

Figure 1: Stages of earthquake data clustering with ST-DBSCAN

Data Collection

The earthquake data used in this study was obtained from the Sleman Geophysics Station / Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) for the period 2019–2024, with a study area limit of $7.45530° - 9.66950°$ LS and $109.82080° - 110.96170°$ East (Figure 2). The earthquake data used was 5403 data in the form of point data. The point data consist of with earthquake magnitude varying between M 0.70 – 6.40 and earthquake depth varying between 1 – 266 km. The dataset in one earthquake data consists of the parameters of date time, epicenter, magnitude and depth (BMKG, 2025).

Data Preprocessing and Zone Classification

This stage is carried out to prepare data before the clustering process is carried out. Data preprocessing consists of the process of converting data from date format to date time, filtering of $M < 1$ magnitude parameters, unrealistic earthquake depth and classification of tectonic zones. The tectonic zones are grouped into two, namely the subduction zone and the local fault. The data group on the subduction zone used the latitude criteria $<$ 8,300 LS and the local fault zone $\geq$ 8.300 LS

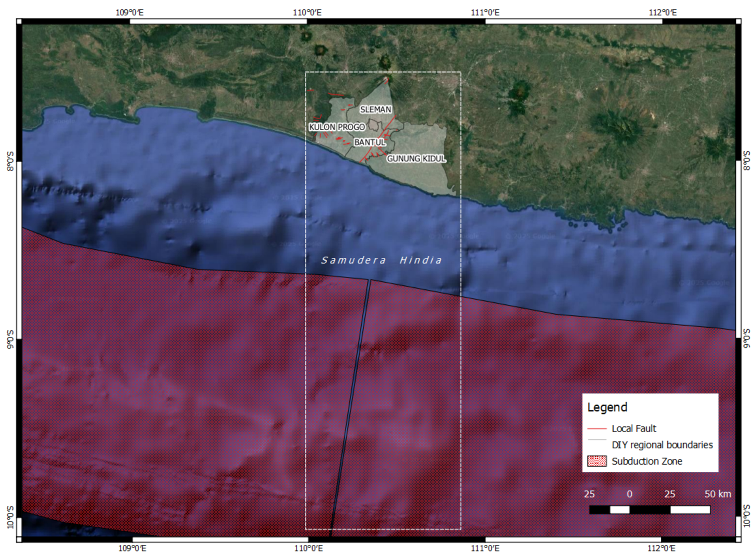

Figure 2: Map of the research area (white line) covering the subduction zone area and local fault in Yogyakarta

ST-DBSCAN Algorithm

The application of the ST-DBSCAN algorithm consists of the stages of deuclidean Distance Dissimilarity Matrix. The data input consists of the epicenter (latitude and longitude), time of occurrence (timestamp), depth, and magnitude (Figure 3).

The parameters of ST-DBSCAN consist of (Wang & Zhan, 2023):

1. EPS1, which is the maximum distance between points in the spatial dimension to be considered close to each other.

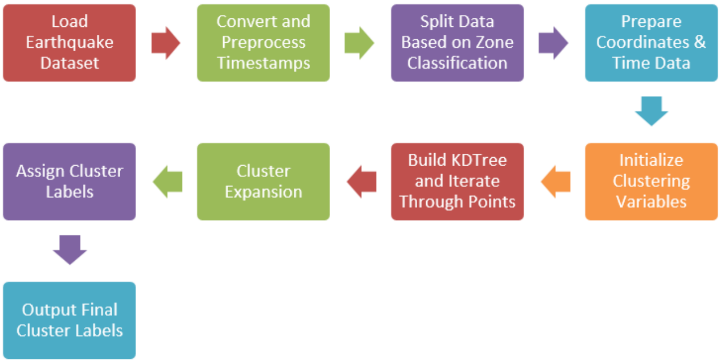

Figure 3: ST-DBSCAN algorithm flowchart

2. eps2: The maximum distance between points in the temporal dimension (time) to be considered close to each other.
3. min_samples: The minimum number of points required to form a cluster (similar to a regular DBSCAN).
4. metric: Distance measurement methods, such as 'euclidean', 'manhattan', or 'chebyshev'.
5. frame_size (optional): If used, the data will be divided into frames based on time.
6. frame_overlap (optional): The amount of overlap between frames to ensure cluster continuity.
7. n_jobs: The number of CPU cores used for parallel computing (e.g. -1 for all cores).
8. sparse_matrix_threshold: Limit the number of samples to switch to the sparsely matrix method for computational efficiency.

2.3.1 Determination of Eps and MinPts

The determination of the value of these parameters is carried out by trial and error using the k-dist graph. The k-dist graph is formed by calculating the distances between objects and then sorting them in descending order and plotting the value of the smallest distance to k of each object (Figure 4). The x-axis expresses the object that has been sorted, while the y-axis expresses the distance value from the object. The point where there is a sharp change in k-dist will be used as Eps and the k value as MinPts (Ramadan & El-Bahnasy, 2023).

The Eps parameter consists of spatial Eps with a maximum distance in kilometers (2–5 km (adjusted for earthquake density) and temporal Eps with a maximum time interval (10 days (time range of correlated earthquake activity). MinPts parameters with a minimum number of neighbors to form clusters 5–10 (based on sensitivity analysis). To improve the reliability of this process, recent research has introduced heuristic methods for defining optimal Eps values, such as analyzing the curvature of the k-dist graph or evaluating clustering density trends ((Sharma & Nanda, 2024); (Cesca, 2020)). These approaches help reduce trial-and-error and make the parameter selection process more systematic.

2.3.2 Dissimilarity Matrix of Euclidean Distance

A dissimilarity matrix stores the proximity or distance for all pairs of n objects. Euclidean distance is the distance between two data objects (i, j) of n numerically valued attributes, which are expressed as $i = x_{i1}, x_{i2}, \ldots x_{in}$ and $j = x_{j1}, x_{j2}, \ldots x_{R_n}$ in space dimension $n(Rn)$(Han et al., 2012):

$$eq:1)d(i,j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2}^2 + \ldots + (x_{in} - x_{jn}^2} \tag{1}$$

For the spatial aspect of the Euclidean distance equation to be (Han et al., 2012):

$$d(i,j) = \sqrt{(x_{long\ i} - x_{long\ i})^2 + (x_{lat\ i} - x_{lat\ j})^2} \tag{2}$$

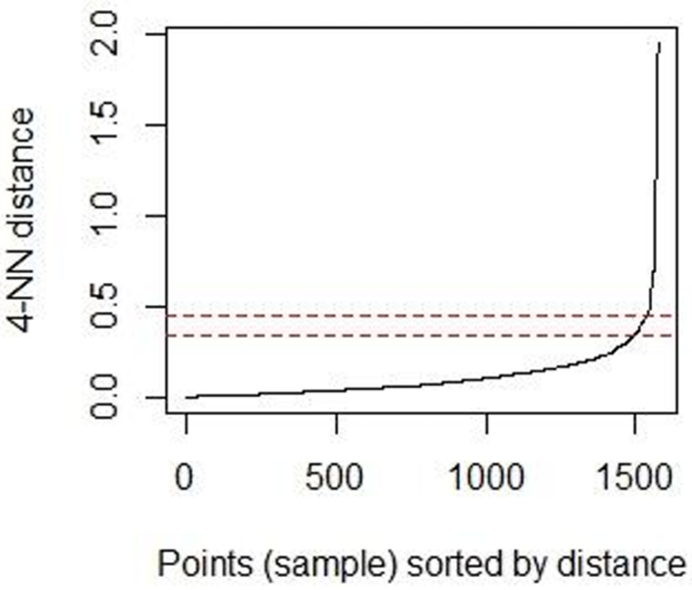

Figure 4: k-dist graph to define Eps and MinPts

where: long i = longitude data of-i, $i = 1, 2, \ldots, n$ lat j = latitude data of-j, $j = 1, 2, \ldots, n$ for the temporal aspect of the Euclidean distance equation is modified into an equation:

$$d(i, j) = |x_{tanggali} - x_{tanggalj} eq : 3) \tag{3}$$

where: $xtanggal$ is an object in the date column and is a one-dimensional object

2.4 Cluster validity

The cluster validity stage is carried out to validate that the resulting clusters have reflected the population in general. The cluster validity method used is the Davies-Bouldin Index (DBI) and Silhouette Coefficient. Complementary tools like Seiscloud have also been developed to assist in visualizing and validating clustering results in seismology (Cesca, 2020), making it easier to interpret complex spatial patterns in earthquake data.

2.4.1 Silhouette Coefficient

Silhouette Coefficient is a metric used to assess the quality of clusters in data analysis. This value measures how well a data point fits into its cluster compared to other clusters ((Rousseeuw, 1987); (Ramadan et al., 2022)). The Silhouette Coefficient equation is as follows:

$$s_{(i)} = \frac{b_i - a_i}{max(a_i, b_i)} eq : 4) \tag{4}$$

where,$s(i)$: value of Silhouette Coefficient from i, ai: Average distance of objects i to all other objects in the cluster a, bi: The average minimum distance from object i to all other clusters which are not clusters of the object (i).

Value of silhouette coefficient	***Structure***
0.7 <$SC \leq 1$	Strong Structure
0.5 <$SC \leq 0.7$	Medium Structure
0.25 <$SC \leq 0.5$	Weak Structure
$SC \leq 0.25$	Unstructured

Table 1: Category of Silhouette Coefficient

2.4.2 Davies-Bouldin Index (DBI)

Davies-Bouldin Index (DBI) is an evaluation metric used to measure the quality of clustering in data. This index combines a low intra-cluster size with a high inter-cluster size. The lower the DBI value, the better the quality of the clustering. DBI is calculated with the following equation (Davies & Bouldin, 1979):

$$DBI = \frac{1}{k}\sum_{i=1}^{k} max_{i \neq j}(R_{(i,j)}) \tag{5}$$

where, k is the number of clusters used. A DBI value of ≥ 0 indicates that the generated cluster is in the good category.

2.5 Visualization and interpretation

Validated subduction zone clusters and local faults can be interpreted after being presented in the form of maps and graphs. Mapping of subduction zone clusters and local faults using the Folium/QGIS application by overlaying megathrust subduction zones, local fault lines and administrative boundaries (shapefile format).

RESULT AND DISCUSSION

This section presents the results of clustering of subduction earthquakes and local faults by applying the ST-DBSCAN algorithm. The algorithm is applied to earthquake data that has been analyzed spatially and temporally distributed. The spatial and temporal distribution shows that earthquake data for the period 2019 – 2024 is concentrated in the South and East regions of Yogyakarta. This study shows a pattern that is not entirely similar to previous research using the same method, namely ST-DBSCAN. In the study of (Manalu et al., 2021) in the Sulawesi region, they detected 60 clusters with 216 noise data, reflecting the high diversity of seismic activity in the region. In this study, 34 clusters were obtained, consisting of 9 clusters in the subduction zone and 25 in the local fault zone. The Silhouette Score values obtained (0.18 and -0.14) indicate a less compact cluster structure and indicate overlap between clusters in the fault zone. Although the quality of this metric is not optimal, the spatial distribution of these clusters is proven to be closely related to active fault lines, especially the Opak Fault and the Oyo Fault. This strengthens the relevance of the method used, especially due to its consistency with the results of other studies conducted in regions such as Greece ((Bountzis et al., 2022)) and Chile ((Nicolis et al., 2024)), which both demonstrated the ability of ST-DBSCAN to recognize active geological structures in various tectonic configurations. Thus, despite limitations in terms of cluster validity, this approach remains reliable for revealing seismic patterns in areas with high geological complexity such as Yogyakarta. The subduction earthquake and local fault data clusters in the initial plotting results did not show significant local subduction and fault clusters (Figure 5).

Subduction Zone Clustering

Clustering of earthquake data in the subduction zone using the ST-DBSCAN algorithm resulted in 9 main clusters (Figure 6). The results of the plotting of 9 clusters with subduction zones showed that six clusters were in the main zone of subduction (megathrust) or outer rise, two clusters were in the transition zone (the zone between the main and outside the subduction zone) or trench and two clusters were outside the main zone of subduction or accretionary prism (Figure 7). The optimal parameters used for clustering in the subduction zone are eps_spatial = 3 km, eps_temporal = 10 days, and minPts = 5.

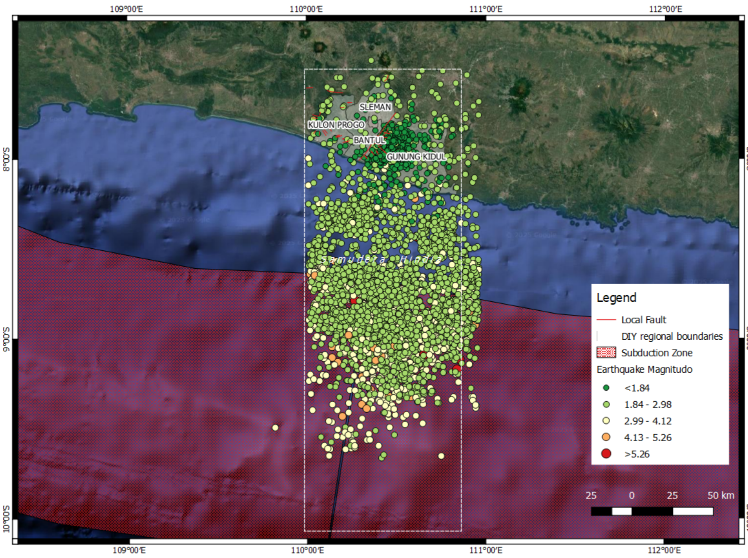

Figure 5: Distribution of earthquake magnitude data in the research area for the period 2019 – 2025

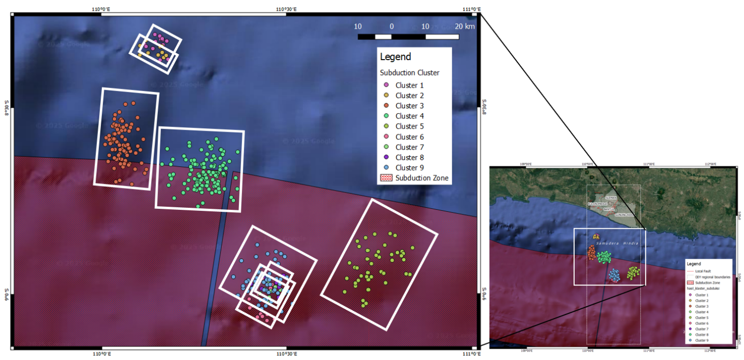

Figure 6: Map of earthquake clusters in the subduction zone as a result of the ST-DBSCAN algorithm (white line is the cluster boundary)

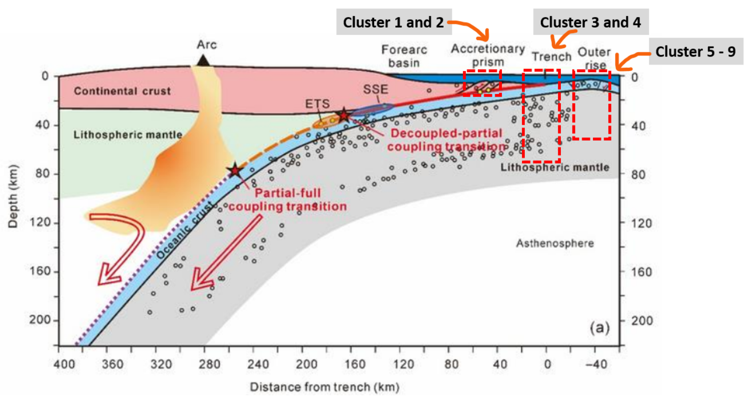

Figure 7: Distribution of subduction earthquake clusters in the subduction zone model ((Zhao et al., 2007) ; (Iwasaki et al., 2013); (Wang et al., 2022)). Clusters 1 and 2 are in the accretionary prism zone, clusters 3 and 4 are in the trench and clusters 5 – 9 are in the outer rise

The clusters in the accretionary prism zone (clusters 1 and 2) have the character of an earthquake with a depth of 6 – 25 km with a magnitude varying between $2.2 \leq M \leq 2.9$. Earthquakes in this zone are included in the category of small earthquakes with shallow depths that can only be recorded by seismographs. The clusters in the trenches (clusters 3 and 4) have the character of an earthquake with a depth of 12 – 79 km with a magnitude varying between $2.2 \leq M \leq 6.4$. Earthquakes in this zone are included in the category of small to strong earthquakes with shallow to medium depths. Earthquakes in this cluster allow it to be felt or even have an impact on the surface. Meanwhile, the outer rise zone (cluster 5 – 9) has the character of an earthquake with a depth of 4 – 50 km with a magnitude varying between $2.2 \leq M \leq 4.4$. Earthquakes in this zone are included in the category of small to light earthquakes with shallow depths. Earthquakes in this cluster were mostly recorded only by seimographs. The resulting cluster pattern indicates that the subduction zone is not homogeneous, the presence of separate active clusters indicates the complexity of plate interactions. These patterns are not unique to Yogyakarta. Similar segmentation of fault systems has been identified in Mount Etna and the Ionian Islands through detailed cluster analysis, underscoring the role of advanced methods in unveiling subsurface geological structures.

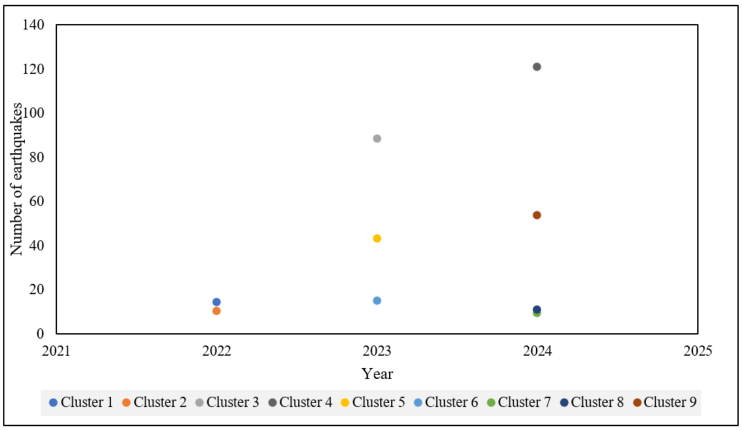

Figure 8: Temporal trends (Year 2022 – 2024) of 9 clusters in the subduction zone

The temporal tendency of earthquakes in the subduction zone shows that each cluster has a different period of activity (Figure 8). Clusters 1 and 2 located in the accretionary prism zone were predominantly active in 2022 despite the number of earthquake events < 20 events. In 2023, the trench and outer rise zones will begin to be active, as shown by seismic activity in clusters 3, 5 and 6. The highest number of earthquake events was in cluster 3 in the trench zone, with the number of earthquake events being 88 events. The temporal trend for 2024 shows an increase in the number of earthquake events in clusters 4, 7, 8 and 9. The four clusters are located in the trench and outer rise zones. The highest number of incidents was found in cluster 4 with 121 events in the trench zone. In general, the temporal trend shows an increase from 2022 – 2024, which follows the pattern of borrowing zones. The zones that are activated due to subduction are the accretionary prism zone, followed by the trench and outer rise zones. The accretionary prism zone is made up of deep-sea sediments, volcanic rocks, and material from the oceanic crust that has been released. This zone has a complex structure and is located at the boundary of a convergent plate ((Bountzis et al., 2022)).

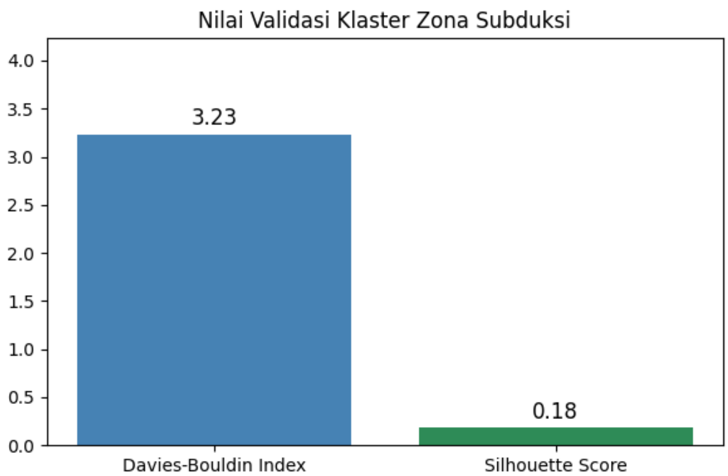

Figure 9: DBI index validation metrics and Silhouette Score in subduction zones

Subduction zone cluster validation metrics were performed using the DBI index and Silhouette Score. The DBI in the subduction zone is 3.23 and be in a good category, indicating significant overlap between clusters, especially in the transition area between clusters 6 - 9. However, the Silhouette Score (0.18) was positive despite being low and be in a weak structure, indicating that the cluster had a less dense internal structure (Figure 9). This may be due to variations in earthquake depth or rock heterogeneity in the subduction zone.

Local Fault Zone Clustering

Clustering of earthquake data in the local fault zone using the ST-DBSCAN algorithm resulted in 25 main clusters (Figure 10). The results of the plotting of 25 clusters with local fault zone showed that one clusters were in the main zone of the opak fault, one clusters were in the outside of the opak fault and 23 overlapping clusters in the Oyo fault area. The optimal parameters used for clustering in the local fault zone are eps_spatial = 2 km, eps_temporal = 10 days, and minPts = 8. A map of earthquake clusters in the local fault zone shows that the center of seismic activity is on the east side of the main path of the Opak fault. The center of activity is correlated with the existence of a fault that is suspected to be an oyo fault. Overlapping between clusters occurs in these areas, except for clusters 4 and 22. Cluster 22 is related to the activities of the opak fault and cluster 4 is related to the new fault system outside the opak and oyo faults. This type of overlapping activity also appears in other tectonically active regions, such as volcano-tectonic zones, where unsupervised learning techniques like clustering algorithms are essential to distinguish between closely spaced seismic events ((Piegari et al., 2024)). The overlap that occurred in the results of the earthquake data clustering showed the complexity of the fault system contained in the local fault ((Bountzis et al., 2022)). These clusters have a magnitude variation of $0.1 \leq M \leq 3.7$ and an earthquake depth of 4 – 127 km.

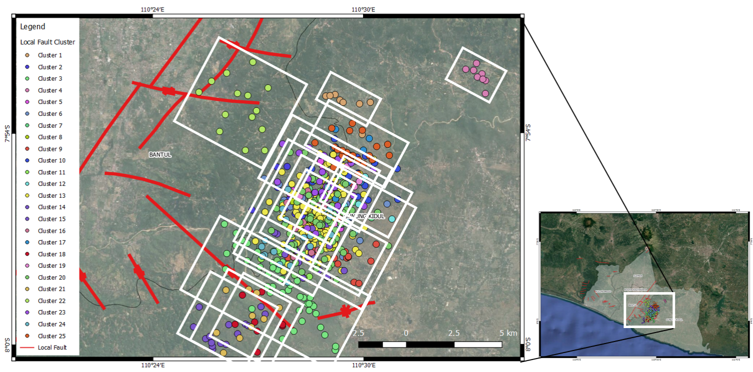

Figure 10: Map of earthquake clusters in the local fault zone as a result of the ST-DBSCAN algorithm (white line is the cluster boundary).

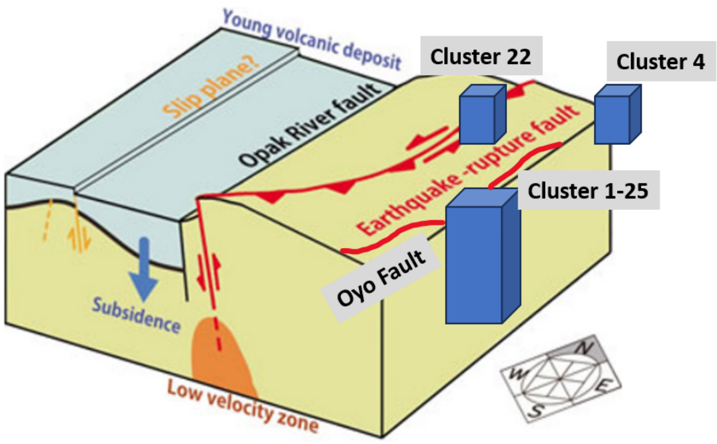

Figure 11: Geological models of the Opak Fault and the Oyo Fault and earthquake clusters in the local fault zone (modification from (Wagner et al., 2007))

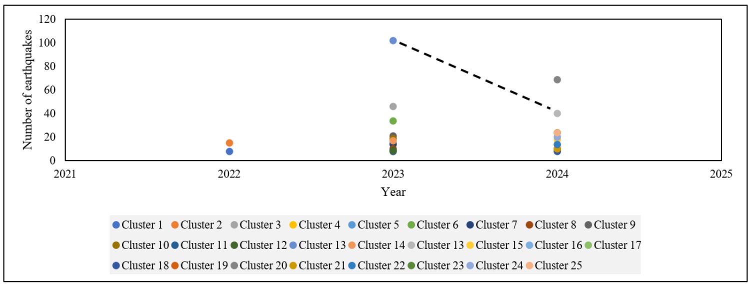

Figure 12: Annotate the dotted line is temporal trends (Year 2022 – 2024) of 25 clusters in the local fault zone

The temporal trend of earthquakes in the fault zone shows the dominant level of seismic activity in 2023 (Figure 12). These temporal trends show very heterogeneous dynamics. Cluster 13 dominates with the number of earthquakes reaching ~ 100 at the beginning of 2023, but experiencing a drastic decrease to ~ 40 by the end of 2024. This condition indicates that the energy release in the fault structure decreases and there is a redistribution of tectonic energy after a period of intensive activity.

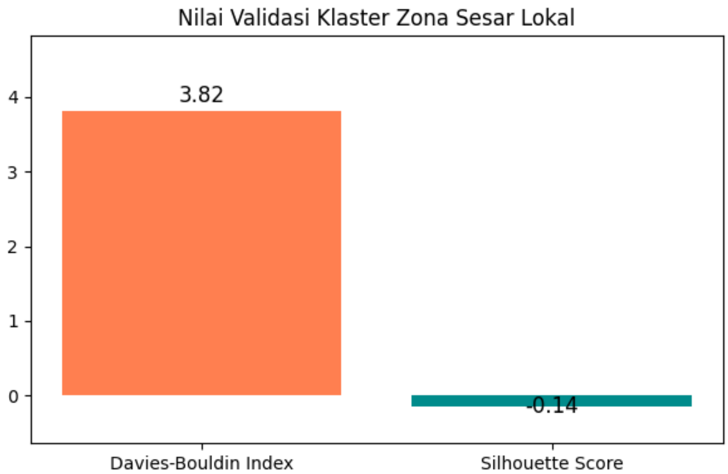

Figure 13: DBI index validation metrics and Silhouette Score in local fault zone

Validation metrics show suboptimal results. The Davies-Bouldin Index (3.82) have a good category indicates significant overlap between clusters, especially in the transition area between clusters 2/22 and 6/23. A negative Silhouette Score (-0.14) show unstructure and indicates a cluster is not dense and the point tends to be closer to another cluster, undermining the validity of the cluster. This may be due to the diversity of earthquake depth or the heterogeneity of the rocks in the fault zone.

Identification of Potential Earthquake Hazards in DIY

The distribution of local fault earthquake clusters is predominantly in Gunungkidul and Bantul Regencies (Figure 14). These clusters are correlated with the existence of the Opak and Oyo Faults in the region. The clustering results showed that the earthquake from the Opak and Oyo fault activities had a maximum magnitude of M3.7 with a shallow depth of 4 km. The maximum magnitude, if converted into an earthquake intensity value, the potential earthquake hazard produced is equivalent to III – IV MMI. The earthquakes that occur are predominantly felt by humans, but do not cause significant damage. The areas that have the potential to be affected by the earthquake model are Gunungkidul and Bantul Regencies, Sleman and Yogyakarta City. The potential hazards produced in this study are based on earthquake data for the 2019 – 2024 period, so it is very possible that there is a greater potential for earthquake danger if a longer data period is used.

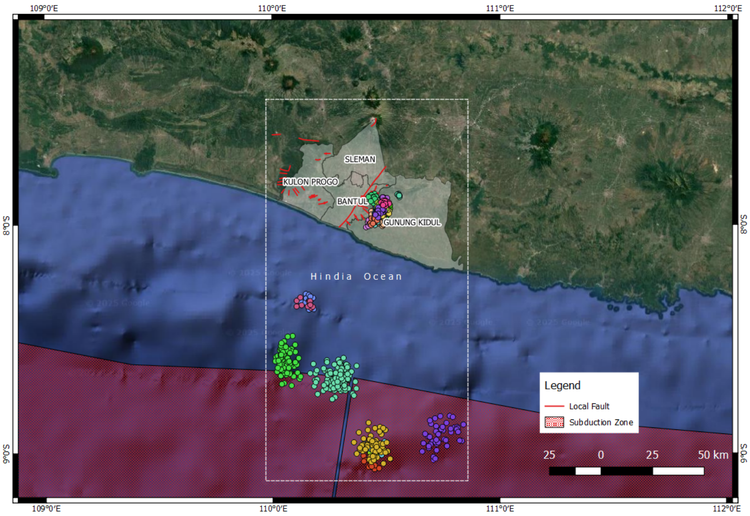

Figure 14: Map of local subduction zone and fault earthquake clusters in the DIY area

The potential hazard of earthquakes resulting from the subduction zone cluster is the area on the South coast of Yogyakarta, namely Kulonprogo, Bantul and Gunungkidul Regencies. The maximum magnitude resulting from the cluster process is M6.4 with a shallow depth of 12 km, which is found in the trench zone. Subduction earthquakes with a magnitude of M<7 are not yet included in the category of tsunami-generating earthquakes, but earthquakes with this category are possible to detect underwater landslides. The phenomenon of underwater landslides can trigger a tsunami with the category of non-tectonic tsunami generating earthquakes.

The results of the clustering of local subduction zones and fault earthquakes with the 2019 – 2024 data period show that the DIY region has the potential for earthquake hazards derived from fault activities on land and at sea

One of the limitations of this study lies in how earthquake events are analyzed over time. The ST-DBSCAN method used in this research groups earthquakes based on their location and time of occurrence, which works well for identifying patterns like foreshocks, mainshocks, and aftershocks ((Wijaya et al., 2024)). However, not all earthquakes happen in such sequences (Sharma et al., 2023). In reality, some earthquakes occur as standalone events — they strike once and are not followed by any significant aftershocks or preceded by foreshocks.

These single events do not form clusters in space or time, and therefore, cannot be grouped effectively using a spatio-temporal method like ST-DBSCAN. This means that some earthquake events may go unclustered, even though they are geologically important. It highlights that while clustering methods are powerful for spotting patterns, they also have blind spots — especially for isolated earthquakes that don't fit the typical seismic sequence.

This limitation suggests the need for complementary analysis techniques in future studies to better capture the full range of earthquake behaviors, particularly in seismically active and complex areas like the Special Region of Yogyakarta.

CONCLUSION

The clustering results revealed that, of the nine earthquake clusters identified in the subduction zone, six were located within the main subduction area (megathrust or outer rise), two in the transitional trench zone, and two in the accretionary prism outside the primary subduction zone. The Davies-Bouldin Index (DBI) for these subduction clusters was 3.23, indicating notable overlap between several clusters, particularly between clusters 6 and 9 in the transition zone. Despite this, the Silhouette Score was slightly positive at 0.18, suggesting the clusters had some internal consistency, though they were not particularly compact. In contrast, the clustering of 25 local fault-related earthquakes showed one cluster aligned with the main Opak Fault zone, one outside of it, and the remaining 23 overlapping in the Oyo Fault region. A high DBI of 3.82 points to substantial overlap between these clusters, especially between clusters 2 and 22, and clusters 6 and 23. The negative Silhouette Score (-0.14) indicates poor cluster cohesion, with many points positioned closer to neighboring clusters than their own, thus weakening the validity of the cluster structure. In terms of hazard potential, subduction zone clusters pose a risk to the southern coastal areas of Yogyakarta, including Kulonprogo, Bantul, and Gunungkidul Regencies, with a maximum observed magnitude of M6.4 at a shallow depth of 12 km in the trench area. Meanwhile, earthquakes linked to the Opak and Oyo faults had a lower maximum magnitude of M3.7 at 4 km depth, which corresponds to an intensity level of III–IV MMI — generally felt by people but not destructive. Moving forward, future studies should consider expanding the scope of analysis to broader subduction zones, as seismic activity is governed more by geological fault dynamics than by administrative boundaries.

ACKNOWLEDGMENT

We sincerely thank the Sleman Geophysics Station / Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) for generously providing access to detailed earthquake data from 2019 to 2024. We also appreciate the support and input from colleagues at Universitas Mercu Buana Yogyakarta and BMKG Yogyakarta, whose expertise helped shape the quality of this research.

References

Ajitomo, W. & Pratama, I. (2024). Application of the dbscan method for the identification of earthquake clusters in the yogyakarta region. *Journal of Informatics and Software Engineering*, 6(1).

BMKG (2024). Katalog gempa bumi signifikan dan merusak (tahun 1821–2023). https://yogyakarta.bmkg.go.id/buletin-mkg/katalog-gempa-bumi-signifikan-merusak-tahun-1821-2023/ [Accessed: 2025-08-13].

BMKG (2025). Katalog gempabumi dan tsunami untuk mitigasi. https://www.bmkg.go.id/gempabumi/mitigasi/katalog-gempabumi-tsunami [Accessed: 2025-08-13].

Bountzis, P., Papadimitriou, E., & Tsaklidis, G. (2022). Identification and temporal characteristics of earthquake clusters in selected areas in greece. *Applied Sciences*, 12(4):1908.

Cesca, S. (2020). Seiscloud, a tool for density-based seismicity clustering and visualization. *Journal of Seismology*, 24(3):443–457.

Davies, L. & Bouldin, D. W. (1979). A cluster separation measure. *IEEE Trans Pattern Anal Mach Intell*, PAMI1.

Ester, M., Kriegel, H., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*.

Gaonkar, M. N. & Sawant, K. (2013). Autoepsdbscan: Dbscan with eps automatic for large dataset. *International Journal on Advanced Computer Theory and Engineering*, 2(2):11–16.

Han, J. W., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques (3rd ed.)*. Morgan Kaufmann Publishers.

Hutchings, S. J. & Mooney, W. D. (2021). The seismicity of indonesia and tectonic implications. *Geochemistry, Geophysics, Geosystems*, 22(9):e2021GC009812.

Iswari, L. (2022). Profiling the spatial and temporal properties of earthquake occurrences using st-dbscan algorithm. In *2022 IEEE 7th International Conference on Information Technology and Digital Applications (ICITDA)*, pages 1–8. IEEE.

Iwasaki, T., Levin, V., Nikulin, A., & Iidaka, T. (2013). Constraints on the moho in japan and kamchatka. *Tectonophysics*, 609:184–201.

Kong, Q., e. a. (2024). Spatiotemporal clustering of seismic events using modified dbscan. *Earthquake Science*.

Lee, C. J., e. a. (2023). Earthquake segmentation using spatiotemporal clustering. *Bulletin of the Seismological Society of America*.

Manalu, D. J., Rahmawati, R., & Widiharih, T. (2021). Pengelompokan titik gempa di pulau sulawesi menggunakan algoritma st-dbscan (spatio temporal-density based spatial clustering application with noise). *Jurnal Gaussian*, 10(4):554–561.

Nicolis, O., Delgado, L., Peralta, B., Díaz, M., & Chiodi, M. (2024). Space-time clustering of seismic events in chile using st-dbscan-ev algorithm. *Environmental and Ecological Statistics*, 31(2):509–536.

Piegari, E., Camanni, G., Mercurio, M., & Marzocchi, W. (2024). Illuminating the hierarchical segmentation of faults through an unsupervised learning approach applied to clouds of earthquake hypocenters. *Earth and Space Science*, 11(10):e2023EA003267.

Ramadan, H. S. & El-Bahnasy, K. (2023). Enhanced st-dbscan with auto-parameter estimation. *Journal of Applied Geophysics*.

Ramadan, H. S., Maghawry, H. A., El-Eleamy, M., & El-Bahnasy, K. (2022). A heuristic novel approach for determination of optimal epsilon for dbscan clustering algorithm. *Journal of Theoretical and Applied Information Technology*, 100(7).

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.

Sharma, A. & Nanda, S. J. (2024). A many objective chimp optimization algorithm to de-cluster earthquake catalogs in space time domain. *Soft Computing*, 28(6):5287–5320.

Sharma, A., Vijay, R. K., & Nanda, S. J. (2023). Identification and spatio-temporal analysis of earthquake clusters using som–dbscan. *Neural Computing and Applications*.

Shearer, P. M. (2019). *Introduction to seismology*. Cambridge university press.

Tanaka, T. & Matsumoto, H. (2024). Adaptive spatiotemporal clustering of japanese seismicity. *Geophysical Journal International*.

Wagner, D., Koulakov, I., Rabbel, W., Luehr, B.-G., Wittwer, A., Kopp, H., Bohm, M., Asch, G., & scientists, M. (2007). Joint inversion of active and passive seismic data in central java. *Geophysical Journal International*, 170(2):923–932.

Wang, S. & Zhan, Z. (2023). Clustering of earthquakes using st-dbscan on large datasets. *Seismological Research Letters*.

Wang, Z., Jin, Z., & Lin, J. (2022). Slab melting and arc magmatism behind the japan trench: Evidence from seismic and thermal structure imaging. *Tectonophysics*, 833:229340.

Wijaya, O. O. et al. (2024). Analysis of sulawesi earthquake data from 2019 to 2023 using dbscan clustering. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 8(4):454–465.

Zhao, D., Wang, Z., Umino, N., & Hasegawa, A. (2007). Tomographic imaging outside a seismic network: Application to the northeast japan arc. *Bulletin of the Seismological Society of America*, 97(4):1121–1132.

THIS PAGE INTENTIONALLY LEFT BLANK