CAUCHY – Jurnal Matematika Murni dan Aplikasi
Volume 10 (2) (2025), Pages 916-928
p-ISSN: 2086-0382; e-ISSN: 2477-3344

Identification of Earthquake Prone Zones in Sumatra using Density Based Spatial Clustering of Applications with Noise

Dwi Agustin Nuriani Sirodj[1,2], Muhammad Nur Aidi[1*], Bagus Sartono[1], Utami Dyah Syafitri[1], and Bayu Pranata[3]

[1] *Statistics and Data Science Study Program, School of Data Science, Mathematics, and Informatics, IPB University, Bogor, Indonesia*
[2] *Statistics Study Program, Faculty of Mathematics and Natural Sciences, Universitas Islam Bandung, Bandung, Indonesia*
[3] *Earthquake and Tsunami Center, Agency for Meteorology, Climatology, and Geophysics (BMKG), Jakarta Pusat, Indonesia*

Abstract

This study investigates the spatial distribution of earthquakes in Sumatra using the DBSCAN clustering algorithm applied to seismic data spanning 1 January 2000 to 31 December 2023. The analysis identified two distinct seismic clusters: one in the northern region (Aceh and North Sumatra) and another in the southern region (Lampung, Bengkulu, and West Sumatra), while several events in central areas were classified as noise. Cluster validity assessment confirmed that the identified groups are compact and well separated, reflecting meaningful seismotectonic segmentation. Statistical testing further revealed significant differences in earthquake depth and magnitude between the clusters, supporting the robustness of the findings. Notably, the southern cluster corresponds to the Mentawai Fault system, whereas the northern cluster aligns with the subduction zone and the Sumatran Fault. DBSCAN proved particularly effective in this context as it can capture clusters of arbitrary shapes, consistent with the complex geological structures governing seismicity in Sumatra.

Keywords: DBSCAN, Earthquake Prone Zones, Spatial Clustering, Sumatra Seismic Risk.

1 Introduction

Spatial data refers to information associated with the geographic location or position of objects on the Earth's surface. Spatial clustering is widely applied to identify geographic distribution patterns and spatial relationships among locations or regions that share similar characteristics. According to Aldstadt [1], a spatial cluster is defined as a geographically bounded group of events with sufficient size and density such that their occurrence is unlikely to be due to random chance. Based on the techniques used to define clusters, spatial clustering algorithms can generally be categorized into four groups [2], [3]: hierarchical clustering methods [4], [5], [6], partitional clustering algorithms [7], [8], [9], density-based clustering algorithms [10], [11], [12], [13], and grid-based clustering algorithms [14].

*Corresponding author. E-mail: muhammadai@apps.ipb.ac.id

Submitted: August 16, 2025 Reviewed: September 3, 2025 Accepted: September 19, 2025
DOI: https://doi.org/10.18860/cauchy.v10i2.36120

Indonesia is located in one of the most tectonically active regions in the world [15], as it lies at the convergence of three major tectonic plates: the Eurasian, Indo-Australian, and Pacific Plates. According to data from Badan Pusat Statistik (BPS), in 2023 Indonesia experienced more than 11,000 earthquakes, 31 of which resulted in natural disasters [16]. Along the western margin of Sumatra, tectonic plate interactions have formed a major subduction zone, the Sumatran Fault System (SFS), and the Mentawai Fault. These geological structures make the Sumatra region highly vulnerable to seismic hazards [17].

Various clustering methods have been applied to seismic data. For instance, K-means clustering has been used to map earthquake vulnerability in Istanbul, Turkey [18], to analyze seismic events at the Yongshaba mine in China [19], and to identify epicenter clusters in Bengkulu Province, Indonesia, where five distinct clusters were found both on land and offshore [20]. Hierarchical clustering has also been employed to group earthquake events based on waveform similarity and travel-time differences, as shown in studies from Spanish Springs and Sheldon, Nevada [21]. While K-means and hierarchical clustering are often considered conventional methods due to their simplicity and ease of use, they tend to be less effective for datasets with noise or irregular spatial distributions [5].

Density-Based Spatial Clustering of Applications with Noise (DBSCAN) provides an alternative approach capable of detecting clusters of arbitrary shapes while effectively handling noise [22]. DBSCAN defines clusters as dense regions separated by areas of lower density, making it well-suited for large and complex spatial datasets. Unlike conventional clustering methods that rely mainly on distance metrics, DBSCAN employs a density-based approach, enabling the detection of complex and nonlinear spatial patterns that might otherwise be overlooked [23], [24].

Earthquake data are typically represented by geographic coordinates (longitude and latitude) together with attributes such as magnitude and depth. DBSCAN is particularly appropriate for analyzing such data, as it identifies density-based spatial patterns without requiring the number of clusters to be predefined. Unlike centroid-based methods such as K-means, DBSCAN can capture clusters of arbitrary shapes, which is critical since earthquake distributions often follow complex geological structures, including elongated subduction zones, active faults, or volcanic regions. Previous studies in Indonesia have employed DBSCAN, but most were conducted at the national scale, producing generalized patterns of earthquake distribution across the archipelago [11]. For example, research in West Java identified 12 earthquake clusters using DBSCAN [10]. However, despite the region's high seismic risk and tectonic complexity, applications of DBSCAN at the regional scale in Sumatra remain limited.

This study addresses this gap by applying DBSCAN to earthquake data from Sumatra, aiming to uncover spatial patterns of seismicity and contribute novel insights into the region's seismic segmentation. The findings are expected to strengthen the scientific basis for disaster risk reduction and mitigation strategies in one of Indonesia's most seismically active regions.

The remainder of this paper is organized as follows. Section 2 describes the data sources, magnitude harmonization, and declustering procedures, and details the methods employed, including Nearest-Neighbor Analysis, Moran's Index, and DBSCAN with its evaluation metrics. Section 3 presents the empirical findings, parameter sensitivity analysis, and spatial interpretation of the resulting clusters with respect to major tectonic structures. Finally, Section 4 summarizes the main conclusions, discusses implications for seismic hazard delineation, and outlines directions for future work.

2 Methods

This study utilizes secondary data from an earthquake event catalog for the Sumatra region, obtained from two sources: the International Seismological Centre (ISC) Bulletin and the Indonesian Meteorology, Climatology, and Geophysics Agency (BMKG). The dataset covers the period from 1 January 2000 to 31 December 2023 and consists of 2,771 earthquake events with magnitudes greater than 5.0 Mw. According to [25], earthquakes with magnitudes between 5.0 and 5.9 have the potential to cause localized damage; therefore, this study focuses on events considered capable of generating such impacts. Geographically, the study area encompasses the entire island of Sumatra, bounded by coordinates ranging from 95° to 105° East longitude and 6° North to 6° South latitude. The earthquake catalog includes information such as epicenter coordinates (latitude and longitude), magnitude, and focal depth.

The seismic data initially comprised various types of magnitude scales, including Surface Wave Magnitude (M_s), Bodywave Magnitude (m_b), and Local Magnitude (M_L). To ensure consistency, all magnitudes were converted to Moment Magnitude (M_w) using established conversion formulas, as referenced in [26]. Prior to further analysis, the dataset was declustered to separate mainshock events from aftershocks. This process was carried out using the Uhrhammer algorithm [27].

2.1 Nearest-Neighbor Analysis and Moran's Index

Nearest-Neighbor Analysis (NNI) is a statistical method used to evaluate the presence and degree of spatial clustering or dispersion among geographic locations (geometric points) [28]. The procedure involves calculating the average distance from each point to its nearest neighbor. The NNI value is determined by comparing the observed mean nearest-neighbor distance with the expected mean distance under a random spatial distribution [28]. The NNI is interpreted relative to a reference value of 1. Values of NNI equal to 1 indicate a random spatial distribution. Values of NNI < 1 suggest a clustered pattern, with values approaching 0 indicating a stronger degree of clustering. Conversely, NNI > 1 suggests a dispersed spatial pattern [28]. To statistically evaluate the spatial distribution pattern, the following hypotheses are formulated [29].

Spatial autocorrelation refers to the correlation of a variable with itself across space. The presence of spatial autocorrelation indicates that the attribute value of a given region is associated with the attribute values of neighboring regions. One of the most commonly used statistical measures for quantifying spatial autocorrelation is Moran's Index. The Moran's Index can be expressed as in [30].

2.2 Density-Based Spatial Clustering of Applications with Noise

Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is a density-based clustering algorithm originally proposed by Ester et al. [24]. It is designed to detect clusters of arbitrary shapes in the presence of noise, particularly within high-dimensional spatial and non-spatial datasets. The core concept of DBSCAN is to identify clusters as regions with high point density, separated by areas of lower density. Unlike conventional clustering methods, DBSCAN does not require prior specification of the number of clusters and is capable of identifying noise or outlier points.

The algorithm relies on two main parameters: *epsilon* (Eps), which defines the radius of the neighborhood, and *minimum points* (MinPts), which specifies the minimum number of points required within that radius for a point to be classified as a core point. A neighborhood defined by radius Eps must contain at least MinPts points, implying that the local density exceeds a certain threshold. The ε-neighborhood of an arbitrary point p is formally defined as follows [22]:

$$N_\varepsilon(p) = \{q \in D \,|\, dist(p,q) < \varepsilon\} \tag{1}$$

where D is the object database. If the ε-neighborhood of a point p contains at least the minimum number of points (MinPts), then p is classified as a *core point*. A core point is formally defined as follows [22]:

$$|N_\varepsilon(p)| \geq \text{MinPts} \tag{2}$$

The following outlines the algorithm of the DBSCAN clustering method:

1. Define the values of ε (Eps) and MinPts to be used.

2. Randomly select an initial observation p.
3. Calculate the distance between observation p and all other observations using the Euclidean distance formula.
4. Identify all observations that are density-reachable from observation p.
5. If the number of observations within the ε-neighborhood of p is greater than or equal to MinPts, then p is categorized as a core point, and a cluster is formed.
6. If p is a border point with no density-reachable observations, then proceed to the next unvisited observation.
7. Repeat steps 3 to 6 until all observations have been processed.

2.3 Silhouette Coefficient

The silhouette coefficient is a commonly used metric for evaluating the quality and optimal number of clusters (k). It measures how similar an object is to other objects within the same cluster, denoted as $a(i)$, compared to the lowest average dissimilarity to objects in different clusters, denoted as $b(i)$. The silhouette coefficient for each data point is calculated using the following formula [31]:

$$s_i = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{3}$$

where $a(i)$ represents the average distance between a data point and all other points within the same cluster, and $b(i)$ denotes the lowest average distance between that point and all points in the nearest neighboring cluster. The interpretation of clustering quality based on silhouette coefficient values follows the proposal in [31].

2.4 Test of Normality and Equality of Means

To examine potential differences in the characteristics of the identified clusters, statistical tests were performed. The first step was to assess whether the data followed a normal distribution, which determined the use of either parametric or nonparametric inference. Normality was assessed using the Shapiro–Wilk test. When the data met the normality assumption, differences between clusters were tested using the parametric independent samples t-test; otherwise, the nonparametric Mann–Whitney test was applied.

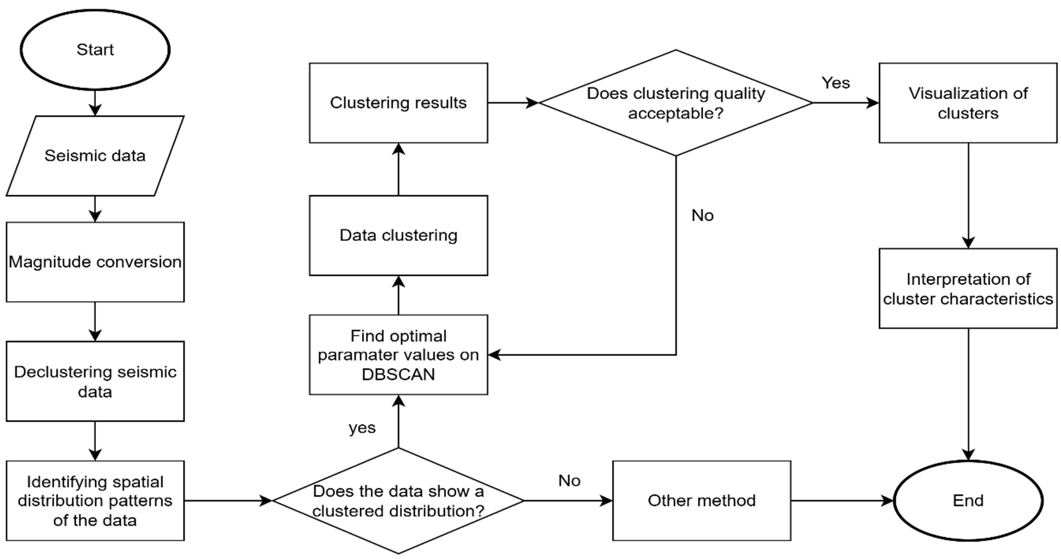


Figure 1: Research Flow Chart

3 Results and Discussion

The earthquake event with the highest magnitude in a given sequence is generally categorized as the mainshock. Within a specific time window and spatial extent, subsequent earthquake events are classified as aftershocks for each event recorded in the earthquake catalog [32]. The process of separating mainshocks from aftershocks in the catalog is referred to as declustering. After the declustering process, the dataset consisted of 1,479 mainshock events, with the maximum magnitude reaching 7.9 Mw. Earthquake activity in the Sumatra region between 1 January 2000 and 31 December 2023 was predominantly distributed along the western coast of Sumatra. The distribution of distances between earthquake events is illustrated in Figure 2.

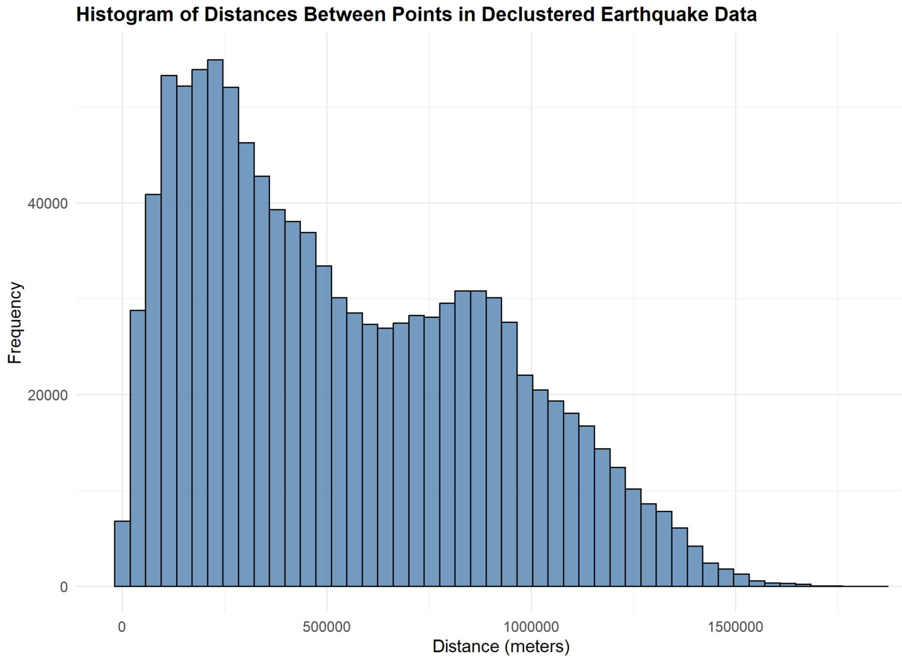


Figure 2: Spatial Distance Distribution of Seismic Events in Sumatra

Many earthquakes in Sumatra are concentrated along the island's western coast. The distribution of distances between earthquake events, as shown in Figure 2, exhibits a positively skewed pattern, indicating that most inter-event distances are relatively short, while a few point pairs are separated by large distances exceeding 1,500 km. Additionally, an increase in frequency is observed in the 800–1,000 km range, which may suggest the presence of a regional spatial structure or multiple earthquake concentrations located in distinct and widely separated zones. To assess whether the seismic data exhibits a non-random spatial pattern, a preliminary analysis using the Nearest-Neighbor Index (NNI) was conducted. The results of this analysis are presented in Table 1.

Table 1: NNI Results

Nearest Neighbor Index	Z-score	P-value
0.682	-23.383	6.25×10^{-121}

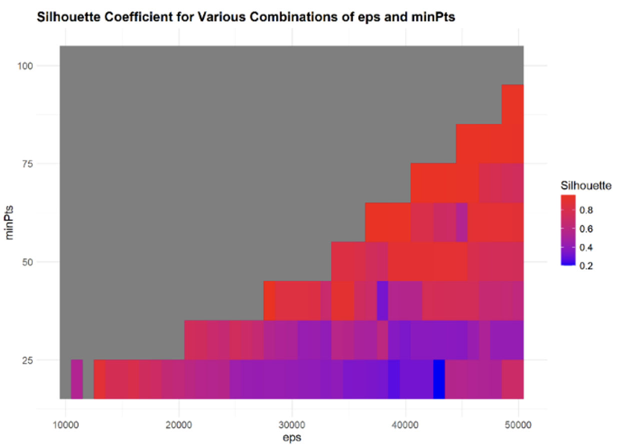
Table 2: Moran's I Statistic

Variable	Moran I	P-value
Magnitude	0.044	0.004
Depth	0.468	$< 2.2 \times 10^{-16}$

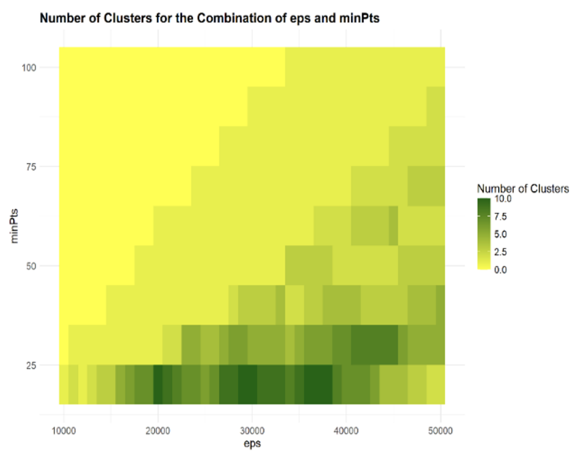
Based on the nearest-neighbor analysis presented in Table 1, an NNI value of 0.682 was obtained, indicating the presence of a clustered spatial pattern. Furthermore, using a significance level of $\alpha = 0.05$, the hypothesis test results in rejection of the null hypothesis (H_0), as the p-value (6.25×10^{-121}) is significantly less than α. This suggests that the spatial distribution of earthquake events in Sumatra from 1 January 2000 to 31 December 2023 exhibits a non-random, clustered pattern.

Table 2 presents the results of Moran's I statistic used to assess the spatial autocorrelation of earthquake magnitude and depth. The Moran's I value for magnitude is relatively low ($I = 0.044$), but with a significant p-value (0.004), indicating the presence of a weak yet statistically significant positive spatial autocorrelation. In contrast, earthquake depth exhibits a much stronger spatial dependence, with a Moran's I value of 0.468 and an extremely significant p-value ($< 2.2 \times 10^{-16}$). These results suggest that while earthquake magnitudes show only a slight tendency to cluster spatially, earthquake depths display a pronounced spatial clustering pattern, confirming that depth is a key variable underlying the spatial structure of seismicity in the study area.

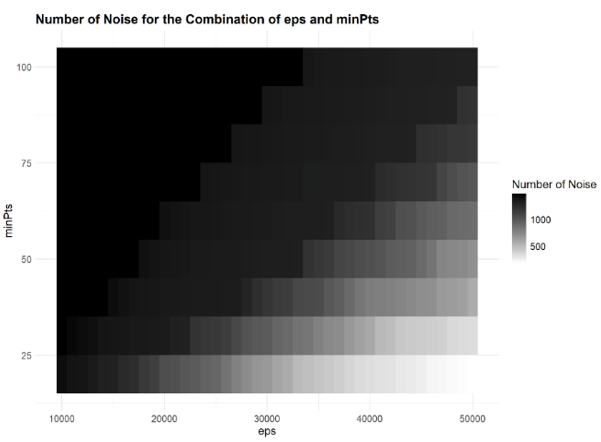
To further delineate regional groupings, clustering was continued using the DBSCAN algorithm. DBSCAN requires two parameters before clustering: Eps and MinPts. To determine the optimal combination of these parameters, various combinations of Eps and MinPts were tested. The evaluation was based on four main criteria: the silhouette coefficient, the number of clusters formed, the number of points classified as noise, and the number of points successfully clustered. The results of this evaluation are visualized in Figure 3.



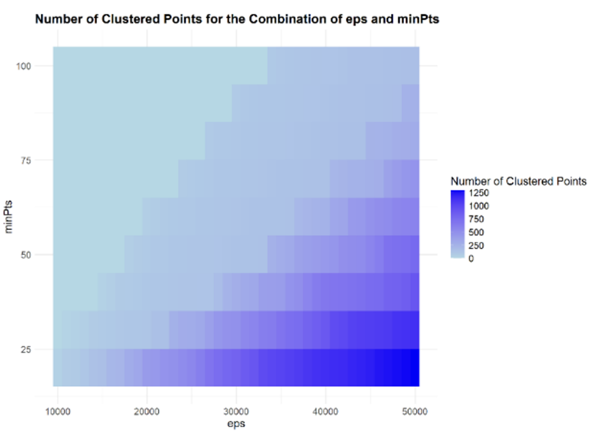
(a) Silhouette Coefficient



(b) Number of Clusters



(c) Number of Noises



(d) Number of Clustered Points

Figure 3: Results of clustering with various Eps and MinPts combinations.

The silhouette coefficient values for different parameter combinations are shown in Figure 3a.

High silhouette scores (highlighted in red) indicate well-separated clusters with strong internal cohesion. The parameter ranges of *Eps* between 30,000 and 45,000 and *MinPts* between 25 and 75 produced the highest silhouette values, suggesting the formation of well-structured clusters. Grey areas represent parameter combinations that failed to produce any clusters.

As illustrated in Figure 3b, the number of clusters varies across parameter settings. Darker green regions correspond to a larger number of clusters, typically observed at moderate to high *Eps* values and low *MinPts*. However, too many clusters may reflect over-segmentation of the data rather than meaningful seismic groupings.

The distribution of noise points is presented in Figure 3c. Black areas indicate parameter settings that produced a very high number of noise points, usually occurring when *Eps* is too small or *MinPts* is too large. In contrast, light grey regions correspond to settings with minimal noise, highlighting more suitable parameter choices.

Finally, the number of points successfully assigned to clusters is depicted in Figure 3d. Darker blue areas indicate more clustered points, whereas lighter shades correspond to fewer clustered points. These results are consistent with Figure 3c, as parameter settings yielding fewer noise points also lead to a greater number of clustered points.

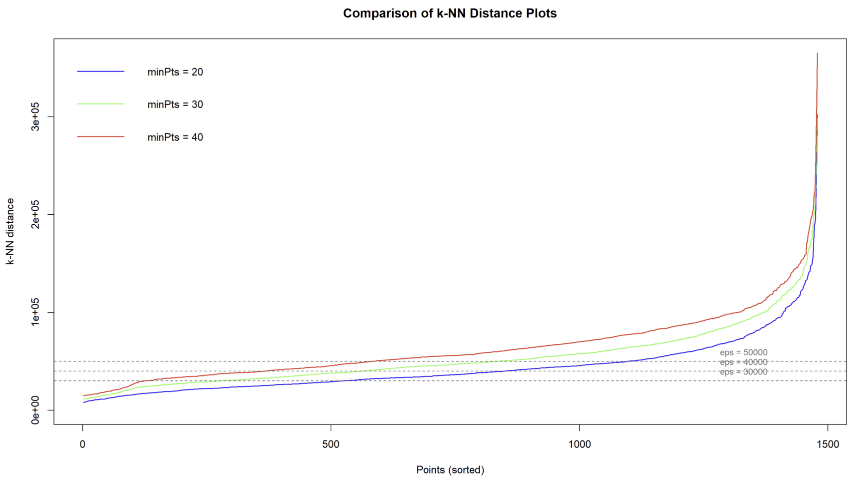


Figure 4: k-NN Distance Plots

To determine the optimal values of Eps and MinPts (Figure 3), we employed k-NN distance plots for several candidate parameters (MinPts = 20, 30, 40; Eps = 30,000; 40,000; 50,000). As shown in Figure 4, the k-NN distance curve remains relatively flat at the beginning and increases sharply toward the end. Larger MinPts values shift the curve upward, implying that higher Eps thresholds are required to retain points within clusters. An Eps of 30,000 appears too restrictive, leading to excessive noise and fragmentation of large clusters. In contrast, the combination of Eps $\approx$ 50,000 with MinPts = 20 is positioned near the elbow of the curve, representing a balance between overly compact and overly diffuse clustering. This conclusion is supported by the silhouette coefficient, the number of clusters, noise points, and clustered points across parameter combinations in Table 3.

Overall, the optimal parameter region is defined by a high silhouette coefficient (indicating good clustering quality), a reasonable number of clusters, minimal noise, and a high proportion of points successfully assigned to clusters. In DBSCAN, noise refers to points that are neither core points nor density-reachable from any other point, typically located in low-density regions and therefore excluded from any cluster [33]. Several candidate parameter combinations were examined to obtain the most reliable results, which are summarized in Table 3.

Table 3 presents the parameter sensitivity analysis for the DBSCAN clustering algorithm. The combination of Eps = 40,000 with MinPts = 30 produced the highest silhouette coefficient; however, it also yielded the largest number of noise points, indicating that this configuration does

Table 3: Optimal Parameter Candidates (Eps and MinPts) for DBSCAN

Eps	MinPts	Silhouette	Clusters	Noise	Clustered Points
30,000	20	0.420	10	662	817
40,000	20	0.320	7	338	1141
50,000	20	0.693	2	190	1289
30,000	30	0.515	5	980	499
40,000	30	0.340	7	589	890
50,000	30	0.415	5	342	1137
30,000	40	0.840	3	1191	288
40,000	40	0.556	4	806	673
50,000	40	0.615	5	571	908

not provide a robust clustering solution. In contrast, the combination of Eps = 50,000 and MinPts = 20 resulted in a Silhouette Coefficient of 0.693, reflecting a good level of cluster separation quality. With this configuration, DBSCAN successfully identified two clusters, assigning 1,289 points to clusters and classifying 190 points as noise. These results demonstrate that the selected parameter setting offers the most reliable balance between cluster compactness and noise reduction, effectively distinguishing high-density regions from low-density areas or outliers. The visual representation of the clustering results is provided in Figure 5.

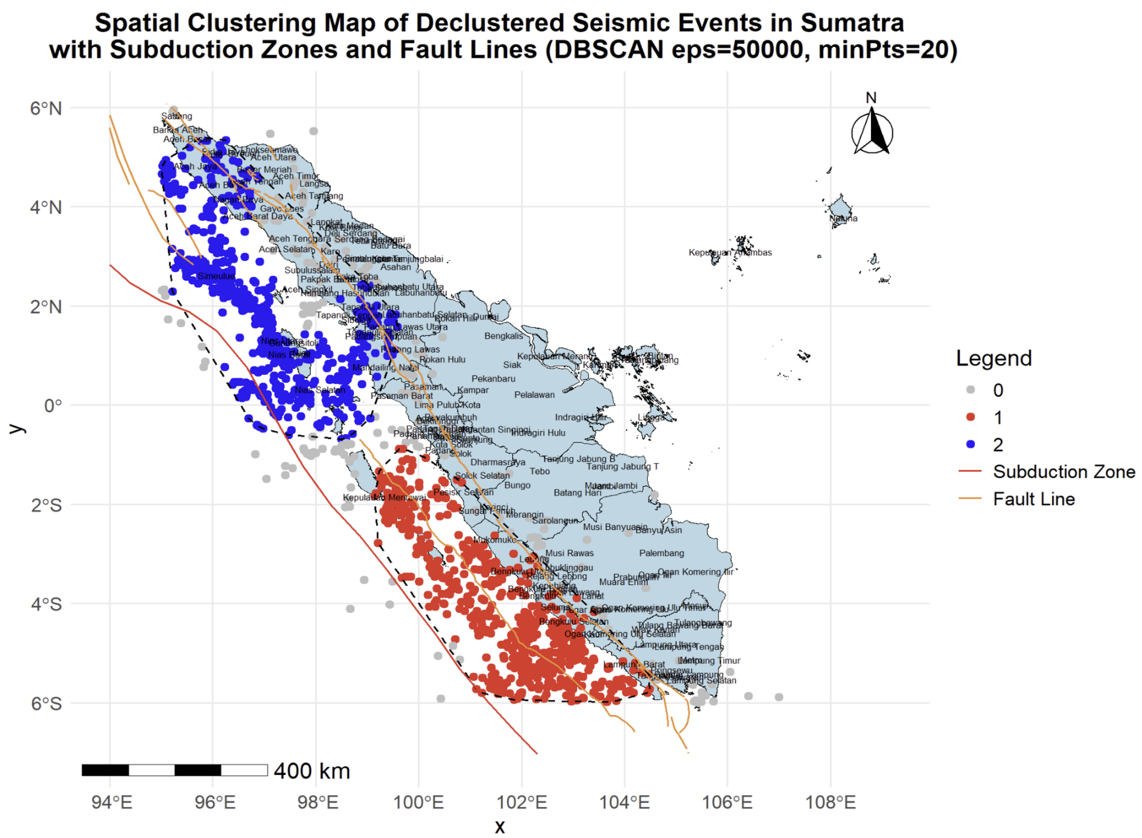


Figure 5: Spatial Clustering Map of Seismic Data in Sumatra using DBSCAN

Two main clusters were identified based on the clustering results shown in Figure 5. Cluster 1 (shown in red) is concentrated in southern to central Sumatra, encompassing regions such as Lampung, Bengkulu, and southern West Sumatra. This area lies near the southern segment of the Mentawai Fault. The Mentawai segment, located along the western coast of Sumatra, Indonesia, is characterized by the convergence of the Indo-Australian Plate and the Sunda Plate. The complex interactions between these tectonic plates have historically generated numerous significant earthquakes in the region [34].

Cluster 2 (shown in blue) is located in northern Sumatra, covering Aceh, North Sumatra, and surrounding areas. This region is part of the subduction zone, where the Indo-Australian Plate subducts beneath the Eurasian Plate, and is also influenced by the activity of the northern segment of the Sumatran Fault. Meanwhile, grey dots represent earthquake events that were not assigned to any cluster and are therefore classified as noise by the DBSCAN algorithm.

The overlay with tectonic features shows that the identified clusters correspond closely with major geological structures. The red line represents the subduction zone, which runs parallel to the western coast of Sumatra, while the orange lines denote fault systems that cut across the island. The alignment of seismic clusters with these structures suggests that DBSCAN effectively delineates seismotectonic patterns, with distinct clustering in the northern and southern segments of Sumatra.

Several fault segments of the Sumatran Fault System, as defined by the National Center for Earthquake Studies, include Nicobar, Seulimeum-North, Aceh-North, Seulimeum-South, Aceh-Central, and Aceh-South in the northern part of Sumatra. Further south, the identified segments comprise Lok Tawar, Peusangan, Tripal-thrust, Tripa2, Tripa3, Tripa4, Tripa5, Batee-A, Batee-B, Batee-C, Renun-A, Renun-B, Renun-C, Toru, Angkola, Barumun, Sumpur, Sianok, Suliti, Siulak, Dikit, Ketaun, Musi, Manna, Kumering-North, Kumering-South, Semangko Barat-A, Semangko Barat-B, Semangko Barat-C, Semangko Timur-A, Semangko Timur-B, and the Semangko Graben. A detailed overview of these fault segments is provided in [35].

In addition, two major offshore fault systems are recognized: the West Andaman Fault (WAF) in northern Sumatra, and the Mentawai Fault (MF) in southern Sumatra, the latter of which is characterized as a backthrust delineating the boundary between the accretionary prism and the forearc basin [35], [36].

This spatial clustering pattern aligns with the findings of [37], who analyzed the seismicity and seismic hazard potential of the Sumatran region by evaluating a-values, b-values, and anomalous b-values. Their study found that the highest seismicity levels are concentrated in the southern Sumatra zone, while the subsurface rock structure in northern Sumatra exhibits the highest stress levels, corresponding to the frequent occurrence of large-magnitude earthquakes in that area. The detailed characteristics of each identified cluster are presented in Table 4.

Table 4: Earthquake Characteristics by Cluster

Cluster	**Magnitude (Mw)**		**Depth (km)**	
	1	2	1	2
Number of Events	622	667	622	667
Median	5.30	5.24	35.25	29.00
Mean	5.40	5.35	39.77	36.34
Std. Deviation	0.36	0.32	19.83	26.08
Maximum	7.90	7.70	122.60	183.00

Table 5: Hypothesis Testing on the Characteristics of Clustering Results

Variable	**Test of Normality (Shapiro–Wilk)**		**Mann–Whitney Test**	
	W	P-value	Statistic	P-value
Depth	0.812	< 0.001	252591	< 0.001
Magnitude	0.765	< 0.001	225807.5	0.005

Table 4 presents the descriptive statistics of earthquake events based on the two main clusters. The average magnitudes in both clusters are relatively similar: 5.40 Mw in Cluster 1 and 5.35 Mw in Cluster 2, with maximum magnitudes reaching 7.90 Mw and 7.70 Mw, respectively. However, a notable difference is observed in the depth variable. Cluster 2 exhibits a wider variation in depth, with a standard deviation of 26.08 km and a maximum depth of 183.00 km, compared to 122.60 km in Cluster 1.

To further clarify the differences in characteristics between the two clusters, an inferential analysis was conducted. Table 5 presents the results of the hypothesis testing on the clustering outcomes. The Shapiro–Wilk test indicated that both depth ($W = 0.812$, $p < 0.001$) and magnitude ($W = 0.765$, $p < 0.001$) did not follow a normal distribution. Consequently, the nonparametric Mann–Whitney test was employed to examine mean differences between clusters. The results revealed statistically significant differences in both depth (Statistic = 252,591, $p < 0.001$) and magnitude (Statistic = 225,807.5, $p = 0.005$) across clusters. These findings suggest that the clusters identified by the DBSCAN algorithm are characterized by distinct seismic properties, with significant variations in earthquake depth and magnitude, thereby supporting the robustness of the clustering results in distinguishing different seismic regimes.

Regarding average focal depth, both clusters are dominated by shallow earthquake sources. This is expected, as seismic activity in the Sumatra region is typically generated by interplate earthquakes originating from tectonic processes along the Sumatra subduction zone, which marks the boundary between the Indo-Australian Plate and the Eurasian Plate, as well as from the Mentawai Fault and intraplate earthquakes that occur along the Great Sumatran Fault system on land [37].

4 Conclusion

The spatial clustering of seismic data using the DBSCAN algorithm revealed two main earthquake clusters in Sumatra: Cluster 1 in the south–central region near the Mentawai Fault, and Cluster 2 in the north, shaped by the subduction zone and the northern segment of the Sumatran Fault. The close alignment of these clusters with major tectonic structures highlights DBSCAN's effectiveness in capturing seismotectonic patterns and distinguishing regimes with different depth and magnitude characteristics.

This study contributes to advancing the use of density-based clustering in seismology by offering a data-driven framework for delineating earthquake-prone zones. Future research should integrate temporal seismicity patterns and additional geophysical parameters, alongside comparative evaluations with alternative clustering methods, to further strengthen predictive accuracy and regional hazard mapping.

Credit Authorship Contribution Statement

Dwi Agustin Nuriani Sirodj: Conceptualization, Writing – Original Draft, Visualization, Formal analysis. **Muhammad Nur Aidi**: Writing – Review & Editing, Supervision, Methodology. **Bagus Sartono**: Writing – Review & Editing, Supervision. **Utami Dyah Syafitri**: Writing – Review, Supervision. **Bayu Pranata**: Writing – Review, Supervision, Data Curation.

Declaration of Generative AI and AI-assisted technologies

Generative AI (ChatGPT, OpenAI) was used solely for minor language editing and grammar correction. All aspects of the research design, data analysis, interpretation of results, and conclusions were entirely performed by the authors.

Declaration of Competing Interest

The authors declare no competing interests.

Funding and Acknowledgements

This research was supported by the Center for Higher Education Funding and Assessment (PPAPT) and the Indonesian Endowment Fund for Education (LPDP), Ministry of Finance of The Republic of Indonesia through the Indonesian Education Scholarship (BPI ID: 202231103733). The authors sincerely thank these institutions for their valuable support.

Data Availability

The dataset analyzed in this study is publicly available from the International Seismological Centre (ISC) Bulletin[1] and the Indonesian Meteorology, Climatology, and Geophysics Agency (BMKG)[2].

References

[1] J. Aldstadt, "Spatial clustering," in *Handbook of Applied Spatial Analysis*, 2010, pp. 279–300. DOI: 10.1007/978-3-642-03647-7_15.

[2] C. Neethu and S. Surendran, "Review of spatial clustering methods," *International Journal of Information Technology Infrastructure*, vol. 2, 2013.

[3] S. Chawla, S. Shekhar, W. Wu, and U. Ozesmi, "Modeling spatial dependencies for mining geospatial data," in *Proceedings of the Western Marketing Education Association Conference*, 2001, pp. 1–17. DOI: 10.1137/1.9781611972719.27.

[4] B. Czecze and I. Bondár, "Hierarchical cluster analysis and multiple event relocation of seismic event clusters in hungary between 2000 and 2016," *Journal of Seismology*, vol. 23, pp. 1313–1326, 2019. DOI: 10.1007/s10950-019-09868-5.

[5] A. Gupta, H. Sharma, and A. Akhtar, "A comparative analysis of k-means and hierarchical clustering," *EPRA International Journal of Multidisciplinary Research (IJMR)*, 2021. DOI: 10.36713/epra2013.

[6] H. Nashir, A. Kurnia, and A. Fitrianto, "Subdistrict clustering in west java province based on disease incidence of jkn participants primary services," *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 17, pp. 295–304, 2023. DOI: 10.30598/BAREKENGVOL17ISS1PP0295-0304.

[7] A. Famalika and P. R. Sihombing, "Implementation of k-means and k-medians clustering in several countries based on global innovation index (gii) 2018," *Advance Sustainable Science, Engineering and Technology*, vol. 3, p. 0 210 107, 2021. DOI: 10.26877/asset.v3i1.8461.

[8] R. A. Ramadhan, D. Swanjaya, and R. Helilintar, "Optimizing predictive accuracy: A study of k-medoids and backpropagation for mpx2 oil sales forecasting," *Advance Sustainable Science, Engineering and Technology*, vol. 6, p. 02 401 010, 2024. DOI: 10.26877/asset.v6i1.17665.

[9] M. N. Aidi, T. R. Aditra, F. Ernawati, N. Nurjanah, E. Efriwati, E. D. Julianti, et al., "Clustering of communicable diseases in indonesia and the factors that affect them: 2018 basic health research data statistical review," *Journal of Population and Social Studies*, vol. 33, pp. 88–109, 2024. DOI: 10.25133/JPSSV332025.005.

[10] M. Bariklana and A. Fauzan, "Implementation of the dbscan method for cluster mapping of earthquake spread location," *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 17, pp. 867–878, 2023. DOI: 10.30598/BAREKENGVOL17ISS2PP0867-0878.

[1] https://www.isc.ac.uk/iscbulletin/search/catalogue
[2] https://repogempa.bmkg.go.id/eventcatalog

[11] A. Amalia, U. Harmoko, and G. Yuliyanto, "Clustering of seismicity in the indonesian region for the 2018–2020 period using the dbscan algorithm," *Journal of Physics and Its Applications*, vol. 4, pp. 1–6, 2021. DOI: 10.14710/JPA.V4I1.11884.

[12] M. T. Anwar, W. Hadikurniawati, E. Winarno, and A. Supriyanto, "Wildfire risk map based on dbscan clustering and cluster density evaluation," *Advance Sustainable Science, Engineering and Technology*, vol. 1, p. 0 190 102, 2019. DOI: 10.26877/asset.v1i1.4876.

[13] R. J. G. B. Campello, P. Kröger, J. Sander, and A. Zimek, "Density-based clustering," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 10, no. 2, e1343, 2020. DOI: 10.1002/widm.1343.

[14] A. Starczewski, M. M. Scherer, W. Ksiek, M. Dębski, and L. Wang, "A novel grid-based clustering algorithm," *Journal of Artificial Intelligence and Soft Computing Research*, vol. 11, pp. 319–330, 2021. DOI: 10.2478/JAISCR-2021-0019.

[15] P. R. Cummins, "Geohazards in indonesia: Earth science for disaster risk reduction - introduction," in *Geological Society Special Publications*, vol. 441, 2017, pp. 1–7. DOI: 10.1144/SP441.1.

[16] A. B. Jatmiko, R. Ghaniswati, F. V. P. E. Utami, M. Burhan, C. A. Ardania, and A. K. Wulandari, *Statistik Indonesia dalam Infografis*, 8th. Badan Pusat Statistik, 2024.

[17] I. N. Sari and T. Prastowo, "Analisis seismisitas dan potensi bahaya bencana seismik di wilayah selatan pulau sumatera," *Inovasi Fisika Indonesia*, vol. 11, pp. 12–19, 2022. DOI: 10.26740/IFI.V11N2.P12-19.

[18] M. Shafapourtehrany, P. Yariyan, H. Özener, B. Pradhan, and F. Shabani, "Evaluating the application of k-mean clustering in earthquake vulnerability mapping of istanbul, turkey," *International Journal of Disaster Risk Reduction*, vol. 79, p. 103 154, 2022. DOI: 10.1016/j.ijdrr.2022.103154.

[19] X. Shang, X. Li, A. Morales-Esteban, G. Asencio-Cortés, and Z. Wang, "Data field-based k-means clustering for spatio-temporal seismicity analysis and hazard assessment," *Remote Sensing*, vol. 10, p. 461, 2018. DOI: 10.3390/rs10030461.

[20] P. Novianti, D. Setyorini, and U. Rafflesia, "K-means cluster analysis in earthquake epicenter clustering," *International Journal of Advances in Intelligent Informatics*, vol. 3, pp. 81–89, 2017. DOI: 10.26555/ijain.v3i2.100.

[21] D. T. Trugman and P. M. Shearer, "Growclust: A hierarchical clustering algorithm for relative earthquake relocation, with application to the spanish springs and sheldon, nevada, earthquake sequences," *Seismological Research Letters*, vol. 88, pp. 379–391, 2017. DOI: 10.1785/0220160188.

[22] K. Khan, S. U. Rehman, K. Aziz, S. Fong, S. Sarasvady, and A. Vishwa, "Dbscan: Past, present and future," in *5th International Conference on the Applications of Digital Information and Web Technologies (ICADIWT)*, 2014, pp. 232–238. DOI: 10.1109/ICADIWT.2014.6814687.

[23] T. Ali, S. Asghar, and N. A. Sajid, "Critical analysis of dbscan variations," in *International Conference on Information and Emerging Technologies (ICIET)*, 2010. DOI: 10.1109/ICIET.2010.5625720.

[24] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining*, 1996, pp. 226–231. Available online.

[25] L. Irawan, L. H. Hasibuan, and F. Fauzi, "Analisa prediksi efek kerusakan gempa dari magnitudo (skala richter) dengan metode algoritma id3 menggunakan aplikasi data mining orange," *Jurnal Teknologi Informasi: Jurnal Keilmuan dan Aplikasi Bidang Teknik Informatika*, vol. 14, pp. 189–201, 2020. DOI: 10.47111/jti.v14i2.1079.

[26] A. Arimuko, S. Rohadi, and A. S. Rahman, "Seismotectonic studies to determine the recurrence of earthquakes mw > 7 using a statistical approach and plate motion in the megathrust western part of java," *Geotechnical and Geological Engineering*, vol. 41, pp. 1397–1406, 2023. DOI: 10.1007/s10706-022-02342-z.

[27] N. A. Galina, V. V. Bykova, R. N. Vakarchuk, and R. E. Tatevosian, "Effect of earthquake catalog declustering on seismic hazard assessment," *Seismic Instruments*, vol. 55, pp. 59–69, 2019. DOI: 10.3103/S0747923919010079.

[28] R. Wilson and A. Din, "Calculating varying scales of clustering among locations," *Cityscape: A Journal of Policy Development and Research*, pp. 215–231, 2018.

[29] B. N. Boots and A. Getls, *Point Pattern Analysis*. Regional Research Institute, West Virginia University, 2020.

[30] Y. Chen, "An analytical process of spatial autocorrelation functions based on moran's index," *PLoS One*, vol. 16, no. 3, e0249589, 2021. DOI: 10.1371/journal.pone.0249589.

[31] L. Kaufman and P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley & Sons, Inc, 1990. DOI: 10.1002/9780470316801.

[32] Juellyan, B. Setiawan, M. Hasan, H. Yunita, M. Sungkar, and T. Saidi, "Comparing gardner-knopoff, gruenthal, and uhrhammer earthquake declustering methods in aceh, indonesia," in *IOP Conference Series: Earth and Environmental Science*, vol. 1245, 2023, p. 012 010. DOI: 10.1088/1755-1315/1245/1/012010.

[33] A. Sharma, R. K. Gupta, and A. Tiwari, "Improved density based spatial clustering of applications of noise clustering algorithm for knowledge discovery in spatial data," *Mathematical Problems in Engineering*, vol. 2016, p. 1 564 516, 2016. DOI: 10.1155/2016/1564516.

[34] A. N. Aulia, J. Jatnika, D. Arisa, M. Ramdhan, A. Patria, and L. Handayani, "Earthquake relocation and deformation analysis on the mentawai segment of the sumatra subduction zone," in *IOP Conference Series: Earth and Environmental Science*, vol. 1437, 2024, p. 012 023. DOI: 10.1088/1755-1315/1437/1/012023.

[35] Tim Pusat Studi Gempa Nasional, *Peta sumber dan bahaya gempa Indonesia tahun 2017*. Kabupaten Bandung: Pusat Penelitian dan Pengembangan Perumahan dan Permukiman, Kementerian PUPR, 2017, vol. 1.

[36] M. M. Mukti, S. C. Singh, I. Deighton, N. D. Hananto, R. Moeremans, and H. Permana, "Structural evolution of backthrusting in the mentawai fault zone, offshore sumatran forearc," *Geochemistry, Geophysics, Geosystems*, vol. 13, p. 12 006, 2012. DOI: 10.1029/2012GC004199.

[37] M. Madlazim and N. I. D. Lestari, "Analisis seismisitas dan potensi bahaya bencana seismik pulau sumatera berdasarkan data gempa 1970–2020," *Inovasi Fisika Indonesia*, vol. 11, pp. 1–11, 2022. DOI: 10.26740/IFI.V11N02.P1-11.