Jurnal Matematika dan Statistika serta Aplikasinya, Vol. 13, No. 2, 134-143, 2025
DOI:https://doi.org/10.24252/msa.v13i2.61832

Application of ST-DBSCAN Algorithm in Clustering Earthquake Points in Sulawesi Region

Sutamrin[1], Irwan[2]*, Nur Insani Maiwa[3]
[1,2,3] *Universitas Negeri Makassar, Makassar, Indonesia*
***Corresponding author**: irwanthaha@unm.ac.id

***Submission date**: 08 October 2025, **Revision**: 11 October 2025, **Accepted**: 11 November 2025

ABSTRACT
Clustering is a method in data mining that aims to group data based on certain similarities or characteristics. One of the clustering methods or algorithms is the Spatial-Temporal Density-Based Spatial Clustering of Applications with Noise (ST-DBSCAN) algorithm. This algorithm was chosen because of its ability to analyse data based on spatial and temporal dimensions simultaneously, using the parameters of spatial distance (ε_1), temporal distance (ε_2), and minimum number of points (MinPts). This study aims to determine the results of the ST-DBSCAN algorithm in clustering earthquake points in the Sulawesi Region. The data analysed is secondary data obtained from the Meteorology, Climatology and Geophysics Agency (BMKG) for the period 2019-2023, covering 12109 earthquake points with magnitude ≥ 3 on the Richter scale. The results show that earthquake points in Sulawesi are concentrated in subduction zones and active faults. The most earthquake-prone areas include North Sulawesi and Gorontalo, which are affected by the subduction of the Pacific and Eurasian Plates. In addition, Central Sulawesi, West Sulawesi, South Sulawesi and Southeast Sulawesi are also at high risk due to the activity of the Palu-Koro Fault. Earthquake intensity around the Flores Sea and Banda Sea increases in 2021-2022 due to subduction of the Indo-Australian Plate. The optimal parameters for clustering varied every year during the study. The optimal parameters for clustering varied every year during the study period. This study provides new insights into seismic activity patterns in Sulawesi that can be utilised to support disaster mitigation and earthquake risk reduction policies

KEYWORDS
Clustering, ST-DBSCAN, Earthquakes, Spatial-Temporal.

1. INTRODUCTION

Earthquakes are one of the most serious natural disasters in Indonesia because their occurrence is difficult to predict, both in terms of time, location, and vibration strength [1]. Indonesia is one of the countries whose regions often experience high earthquake activity and volcanism, besides that Indonesia is located on the border of several large tectonic plates, such as the Eurasian Plate, Indo-Australian Plate, and Pacific Plate which often move and interact with each other. The interaction of these plates causes earthquakes that often occur in the Indonesian region [2].

The level of earthquake risk in an area not only depends on the frequency and intensity of earthquakes, but is also influenced by population density and existing infrastructure. Areas with high population and dense buildings are at risk of greater damage and an increase in the number of victims when an earthquake occurs. Therefore, one of the preventive efforts to minimize the negative impact is to map and group areas based on the pattern of frequent earthquake events [3].

The Sulawesi region is one of the seismically active regions in Indonesia [4]. This condition is due to the fact that this region is located at the meeting of three large plates which causes very complex tectonic conditions. The three plates are the Pacific plate, Eurasian plate, and Philippine plate. These plate movements form the four arms of Sulawesi with different tectonic

processes, creating a distinctive geological mosaic [5].

Clustering analysis is an effective approach to grouping earthquake regions, enabling more targeted mitigation measures and development planning in high-risk areas [6]. One of the clustering methods that can be applied is the ST-DBSCAN algorithm which is a development of the DBSCAN algorithm which as the name implies can handle spatial and temporal data at once [3].

Previous research by Fahira dan Nooraeni optimized the determination of ST-DBSCAN initial parameters using K-Nearest Neighbor and Genetic Algorithms on simulated data, applied to clustering natural disaster areas [7]. Johar, Vatresia dan Donny grouped hotspots based on distance and time [8], while Arafat, Hariyadi, Santoso dan Crysdian used DBSCAN for earthquake clustering without considering temporal distance [3]. In this study, ST-DBSCAN is applied by considering three main parameters, namely spatial distance (ε_1), temporal distance (ε_2), and the minimum number of cluster members (MinPts), with the support of earthquake data that includes complete spatial and temporal information. Based on this study, this research aims to analyze the results of the application of the ST-DBSCAN algorithm in clustering earthquake points in the Sulawesi Region. It is hoped that this research can make a significant contribution to the development of technology, knowledge in the field of seismology and mathematics, as well as support disaster mitigation efforts and earthquake risk reduction in the future.

2. LITERATURE REVIEW

2.1 Earthquakes

An earthquake is a phenomenon of sudden vibration due to the release of energy from the epicenter known as seismic waves [9]. The strength of an earthquake is measured using the Richter scale or moment magnitude scale, which measures the energy released. Earthquake impacts can be highly destructive, depending on depth, location and strength [10]. In Sulawesi, earthquakes are often destructive. With the high frequency and distribution of earthquakes, analyzing the timing and location of earthquakes is important [3].

2.2 Clustering

Clustering, or cluster analysis, is a statistical technique that is very useful in exploring and organizing data in a systematic way. Its main objective is to group objects or variables based on their similar characteristics into clusters, where similar objects tend to be grouped together in a single entity called a cluster. Conversely, objects that have significant differences are placed in different clusters, creating a structure where homogeneity within a cluster and heterogeneity between clusters are the main features [11].

2.3 Dissimilarity Matrix Euclidean Distance

This dissimilarity matrix records the degree of closeness or distance for each pair of n objects. Euclidean distance is the distance between two data objects (i,j) of n numeric-valued attributes, expressed as $i = x_{i1}, x_{i2}, \ldots, x_{in}$ and $j = x_{j1}, x_{j2}, \ldots, x_{jn}$ dimensional space $n(R^n)$ [6].

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

For the spatial aspect, the euclidean distance equation becomes [6]

$$d_s(i,j) = \sqrt{\left(X_{long\ i} - X_{long\ j}\right)^2 + \left(X_{lat\ i} - X_{lat\ j}\right)^2} \tag{2}$$

where:
$long\ i$ = longitude of $i-th$ data, $i = 1,2,\ldots,n$
$lat\ j$ = latitude of $j-th$ data, $j = 1,2,\ldots,n$
For the temporal aspect, the euclidean distance equation is modified into the equation [6]

$$d_t(i,j) = \left| X_{tanggal\ i} - X_{tanggal\ j} \right| \tag{3}$$

2.4 Parameter Determination

The determination of parameter values is done through repeated experiments using the K-Distance (k-dist) graph to measure and sort the distance between objects, with the Euclidean distance used to calculate the proximity between points. The point with the largest distance is considered as the point for elbow point detection [7]. One method for detecting elbow points in k-dist graphs is KneeLocator, which automatically identifies significant changes in the graph [12]. These changes are calculated by the vertical distance between the point (x_i, y_i) and a straight line representing the trend of the data. This vertical distance can be calculated by the formula [13]:

$$d_i = \frac{|m.x_i - y_i + c|}{\sqrt{m^2 + 1}} \tag{4}$$

where d_i is the distance from point (x_i, y_i) to the line $y = mx + c$, where m and c are the gradient and intercept of the line or the value at which the line intersects the y-axis on the cartesian graph estimated based on the existing data. Setting the values of ε_1 and ε_2 aims to minimize the total number of clusters, while setting the value of MinPts aims to reduce the amount of noise [8].

2.5 ST-DBSCAN Algorithm

The ST-DBSCAN algorithm has several stages in general, namely [6]:

1. Determine the parameters ε_1 and ε_2 and MinPts
2. Calculate all euclidean distances between objects based on spatial and temporal aspects
3. Form a distance matrix for all pairs of n objects based on spatial and temporal aspects
4. Start from the first point and then take all points in the spatial aspect and temporal aspect with conditions:
   $A = \{x | x \leq \varepsilon_1, x \in spatial\ distance\ matrix\}$
   $B = \{x | x \leq \varepsilon_1, x \in temporal\ distance\ matrix\}$
5. Take all slices of spatial aspect and temporal aspect with the condition: $A \cap B = \{x | x \in A \ \wedge \ x \in B\}$
6. If the number of objects in the slice is smaller than the MinPts value, then the point is considered as noise
7. Cluster is formed if the point meets the parameters ε_1, ε_2 and MinPts
8. If point p is a border point and there are no other points on the slice then proceed to the next point and a new cluster is formed
9. Repeat steps 4-8 until all points are processed.
10. If two clusters C1 and C2 are close to each other, a point q may belong to both clusters. However, this algorithm will declare point q as the cluster that finds it first.

2.6 Cluster Validation

In order for the clustering results to be appropriate and reflect the general population, validation is required. One of the validation methods used to evaluate cluster quality is the Silhouette Coefficient, which combines the cohesion and separation methods. The calculation stages are as follows [14]:

$$s(i) = \frac{b(i) - a(i)}{max\{a(i), a(i)\}} \tag{5}$$

where:
$s(i)$: Silhouette Coefficient value of i
$a(i)$: average distance of object i to all other objects in cluster a
$b(i)$: average minimum distance from object i to all other clusters (which are not clusters of object (i)

3. METHODOLOGY

3.1 Data and Research Variables

The data used in this study is secondary data which includes information on earthquake activity in the Sulawesi region for 5 years (2019-2023). The data was obtained through the official website of the Indonesian Meteorology, Climatology and Geophysics Agency (BMKG) at https://www.bmkg.go.id. The observation area in the study covers the range from 5° North latitude (LU) to $-7.7°$ South latitude (LS) and from 118.5° East longitude (BT) to 127.3° East longitude (BT) with a magnitude ≥ 3 on the Richter scale (SR). In the analysis, the variables taken involved important variables such as latitude, longitude, time of earthquake occurrence.

3.2 Research Procedure

The research procedures carried out in this study are as follows:

1. Collecting earthquake data in the Sulawesi region
2. Preoricessing the data by reducing the data to focus on the variables used in this study using Microsoft Excel including latitude, longitude, time of earthquake occurrence.
3. Calculating the Euclidean distance between objects based on spatial aspects (longitude and latitude) using equation (2) and based on temporal aspects (date of earthquake occurrence) using equation (3) implemented in Python with the help of libraries such as NumPy, scikit-learn, matplotlib, and kneed to improve reproducibility.
4. Determining the parameter values ε_1, ε_2 and MinPts by plotting the K-Distance (k-dist) graph and using the KneeLocator method to find the elbow point to analyze the distance to the K-nearest neighbor using equation (4) with the help of Python software.
5. Performed cluster analysis iterations with the ST-DBSCAN algorithm based on the best parameters of ε_1 and ε_2 and MinPts from the k-dist graph with the help of Python software and saved the data in Microsoft Excel. Measuring cluster quality using Silhouette Coefficient using **equation** (5) with the help of Python software.
6. Visualizing and interpreting the results of the clusters from the ST-DBSCAN method using ArcGIS based on the best parameters of ε_1 and ε_2 and MinPts.
7. Analyzing the distribution pattern of dominant earthquakes in 2019-2023.

4. RESULT & DISCUSSION

4.1 Data and Data Preprocessing

Earthquake data from 2019-2023 totaling 12109 points were obtained, including earthquake data in 2019 totaling 3198 points, in 2020 totaling 2227 points, in 2021 totaling 2922 points, in 2022 totaling 1682 points, and in 2023 totaling 2071 points.

The data preprocessing stages carried out are as follows:

1. Changing the format of the date column

   The date column was converted from a custom format to a standard date format to generate a new column, namely "ordinal day," which facilitates temporal distance calculation in the program. To ensure comparability between spatial distance (ε_1) and temporal distance (ε_2), the temporal values were scaled to the same order of magnitude as the spatial distances using normalization, so that differences in units (days versus degrees) did not distort the ST-DBSCAN clustering results.

2. Separating data by year

   The purpose of separating earthquake data from 2019 to 2023 by year is to analyze temporal trends and identify seasonal patterns in seismic activity, so as to compare earthquake events from year to year.

4.2 Euclidean distance

In the process of cluster formation, calculating the distance between objects is an important step. The distance is calculated using the Euclidean method. The Euclidean distance calculation considers two main aspects, namely the spatial aspect (longitude and latitude) and the temporal aspect (date of earthquake occurrence) can be seen in **Table 1** for the 2019 earthquake data.

Table 1. Euclidean Distance Calculation for the 2019 Earthquake

(i,j)	$\mathbf{d(i,j)}$ spasial	$\mathbf{d(i,j)}$ temporal
(1,1)	0	0
(1,2)	4.809924	0
(1,3)	3.769898	0
(1,4)	3.089473	0
(1,5)	4.780677	0
(1,6)	3.044819	0
(1,7)	4.788179	0
(1,8)	4.554625	0
(1,9)	8.118105	1
(1,10)	4.030731	1
(1,11)	3.428740	1
⋮	⋮	⋮
(3198,3197)	3.349985	0
(3198,3198)	0	0

4.3 Parameter Determination

The selection of parameters ε_1 and ε_2 and MinPts in order to obtain the optimal value is selected based on the k-dist graph. Computation is performed to obtain the k-dist graph at values of $k = 3$, $k = 4$, $k = 5$, and $k = 7$, then the value of k will be the MinPts parameter. Furthermore, the determination of the elbow point with the KneeLocator method to analyze the distance to the nearest K-neighbor with the help of Python software to determine the parameter ε_1.

Based on **Figure 1**, the optimal ε_1 value for each MinPts in earthquake data (2019-2023) is presented in **Table 2** as follows:

Table 2. Euclidean Distance Calculation for the 2019–2023 Earthquake

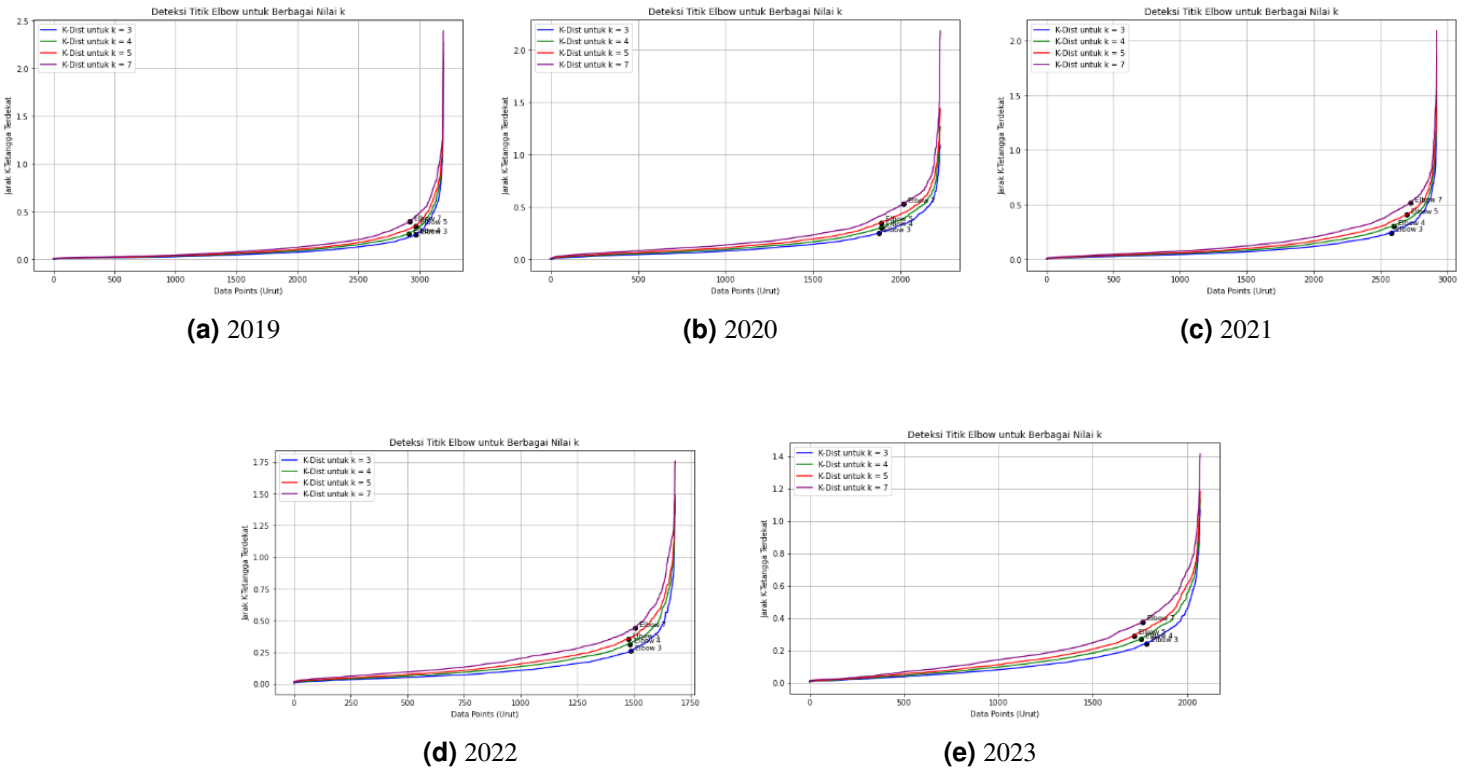
Earthquake Year	MinPts 3	MinPts 4	MinPts 5	MinPts 7
2019	0.26	0.27	0.35	0.39
2020	0.25	0.30	0.35	0.53
2021	0.25	0.30	0.41	0.51
2022	0.26	0.32	0.35	0.44
2023	0.24	0.27	0.29	0.38

Based on **Table 2**, a combination of ε_1 and MinPts values for each earthquake data was obtained. Then the selection of ε_2 values is 7 days, 14 days, 21 days, and 30 days, while ε_1 is determined based on four options taken from the k-disk graph. Thus, 16 combinations of each earthquake data (2019-2023) were formed with the help of Python software using the ST-DBSCAN algorithm.

4.4 Cluster formation

After the clusters are formed for each combination, the next step is to select the parameters with the highest Silhouette Coefficient value and stability between the number of clusters and the amount of noise from each MinPts presented in Table 4.3:

The negative Silhouette values obtained in both spatial and temporal dimensions indicate that some clusters overlap or that certain data points are located near the boundary between clusters. This suggests that the clustering structure is not well



(a) 2019 **(b)** 2020 **(c)** 2021

(d) 2022 **(e)** 2023

Figure 1. *k*-dist graph with KneeLocator Method

Table 3. Best Parameters for Each Year

Year	MinPts	ε_1	ε_2	Number of Clusters	Amount of Noise	Spatial Silhouette	Temporal Silhouette
2019	7	0.39	30	18	663	0.07	-0.72
2020	7	0.53	30	15	566	-0.11	-0.70
2021	7	0.51	30	23	611	-0.01	-0.60
2022	7	0.44	30	23	540	-0.09	-0.57
2023	7	0.38	30	23	764	0.02	-0.64

separated, possibly due to the high density and irregular spatial–temporal distribution of earthquake events. Despite these negative values, the clustering results still provide meaningful patterns by identifying regions and periods with relatively higher seismic activity

Next, visualize and interpret the results of the clusters from the ST-DBSCAN method on each earthquake data using ArcGIS as shown in **Figure 2**. Earthquake distribution patterns in Sulawesi from 2019 to 2023 show an increase in seismic activity with increasingly widespread and complex clusters. In 2019, earthquakes were widely distributed with major clusters in the Maluku Sea, west coast of Sulawesi and Flores Sea. In 2020, clusters in Bone Bay and Gorontalo began to develop, while the west coast of Sulawesi showed increased activity. In 2021, the Moluccas Sea remains the center of earthquake activity, while Mamuju and South Sulawesi experience an increase in earthquakes. In 2022, the cluster becomes clearer with an increase in earthquakes in Bone Bay and the west coast of Sulawesi. In 2023, the distribution pattern was similar to the previous year, but the clusters were more extensive and organized, especially in the Molucca Sea and Mamuju. Overall, earthquake activity continues to increase with increasingly complex clustering patterns, indicating changes in seismic dynamics in the Sulawesi region. As summarized in **Figure 1**, the number of clusters tends to increase after 2020, accompanied by relatively low or negative Silhouette values, which reflect overlapping spatial–temporal boundaries of seismic activity. The dominant clusters are concentrated in Central and North Sulawesi, regions that historically exhibit high tectonic activity due to the interaction of multiple microplates. These results confirm that the seismic patterns in Sulawesi have become more intricate over time, suggesting increased interaction

between fault segments.

4.5 Analysis of Dominant Earthquake Distribution Patterns

The **Figure 3** is a map of the distribution of dominant earthquakes in Sulawesi in 2019-2023 Earthquake points in Sulawesi are generally concentrated in areas near subduction zones and active faults. From 2019 to 2023, the most earthquake-prone areas include North Sulawesi and parts of Gorontalo, with the distribution of earthquake points from the North Sulawesi mainland to the Maluku Sea, influenced by the subduction of the Pacific and Eurasian Plates. In addition, Central Sulawesi, parts of West Sulawesi, South Sulawesi and Southeast Sulawesi are also at high risk, especially around Bone Bay and the Central Sulawesi Mountains, due to the activity of the Palu-Koro Fault.

In South Sulawesi, especially around the Flores Sea and Banda Sea near Selayar Regency, earthquake intensity increased significantly in 2021–2022. This seismic activity is caused by the subduction of the Indo-Australian Plate beneath the Eurasian Plate, triggering earthquakes in the Flores Sea, one of the most active seismic zones in eastern Indonesia.

The clustering results in this study are in line with the findings of [5], who identified high seismic hazard indices in Sulawesi based on historical earthquake data from 1905–2005, particularly around the Flores and Banda Seas. Similarly, [6] applied the ST-DBSCAN algorithm to Sulawesi earthquake data and found that spatial–temporal clustering tends to concentrate in subduction and fault boundary zones, which is consistent with the dominant clusters detected in this research. Moreover, [7] demonstrated that optimizing ST-DBSCAN parameters enhances clustering accuracy for disaster data, supporting the parameter tuning approach adopted in this study.

In addition, [8] and [9] also reported the effectiveness of spatio-temporal clustering methods in identifying regions of high hazard potential in Indonesia, confirming that clustering-based approaches can reveal meaningful spatial–temporal patterns in disaster events. Compared with those studies, this research provides a more detailed temporal evolution of earthquake activity in Sulawesi from 2019 to 2023, showing an increasing trend and more complex clustering structure in recent years.

5. CONCLUSION

The results of the application of the ST-DBSCAN algorithm in clustering earthquake points in the Sulawesi region show that in the period 2019 to 2023, the areas with the highest earthquake density are in North Sulawesi Province and parts of Gorontalo, with the distribution of earthquake points stretching from the mainland of North Sulawesi to the Maluku Sea. In addition, significant earthquake density was also detected in Central Sulawesi, parts of West Sulawesi, South Sulawesi and Southeast Sulawesi, covering the area around Bone Bay and extending into the mountainous and lowland parts of Central Sulawesi. A spatially and temporally high increase in earthquake intensity was also observed around the Flores Sea and Banda Sea, particularly around the land of Selayar Regency in 2021 to 2022. These findings indicate that these areas have a high level of seismic activity, so further monitoring and mitigation efforts are needed to reduce the risk of future disasters.

REFERENCES

[1] N. H. Qothrunnada, R. Y. Utami, and S. A. Rizky, "Menganalisis bencana alam gempa bumi dalam perspektif al-quran," *Jurnal Konferensi Integrasi Interkoneksi Islam dan Sains*, vol. 4, pp. 257–260, 2022, [Online]. [Online]. Available: https://ejournal.uin-suka.ac.id/saintek/kiiis/article/view/3225

[2] I. N. Setiawan, D. Krismawati, S. Pramana, and E. Tanur, "Klasterisasi wilayah rentan bencana alam berupa gerakan tanah dan gempa bumi di indonesia," in *Seminar Nasional Official Statistics*, vol. 2022, no. 1, 2022, pp. 669–676.

[3] I. B. F. Arafat, M. A. Hariyadi, I. B. Santoso, and C. Crysdian, "Clustering gempabumi di wilayah regional vii menggunakan pendekatan dbscan," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 10, no. 4, pp. 823–830, 2023.

[4] N. E. Dwiyanti *et al.*, "Analisi hubungan magnitudo gempa bumi terhadap hasil frekuensi dominan pada rangkaian gempa aceh 2004, yogyakarta 2006, palu dan lombok 2018 sebagai upaya mitigasi bencana," *Jurnal Meteorologi Klimatologi dan Geofisika*, vol. 7, no. 3, pp. 44–50, 2020, [Online]. [Online]. Available: https://jurnal.stmkg.ac.id/index.php/jmkg/article/view/203

[5] I. I. Alfiansyah, B. Rahmadhaniyah, N. Khikmah, F. Roshmiasih, and W. Praditya, "Kajian indeks bahaya seismik regional menggunakan data seismik pulau sulawesi tahun 1905–2005," in *Prosiding Seminar Nasional Fisika Festival*, vol. 01, no. 01, 2020, pp. 77–82.

[6] D. J. Manalu, R. Rahmawati, and T. Widiharih, "Pengelompokan titik gempa di pulau sulawesi menggunakan algoritma st-dbscan (spatio temporal-density based spatial clustering application with noise)," *Jurnal Gaussian*, vol. 10, no. 4, pp. 554–561, 2021.

[7] A. N. Fahira and R. Nooraeni, "Optimasi parameter st-dbscan dengan knn dan algoritma genetika studi kasus: Data bencana alam di pulau jawa 2021," *Jurnal Komputasi*, vol. 11, no. 1, pp. 24–33, 2023.

[8] A. Johar, A. Vatresia, and I. A. Donny, "Implementasi metode spatio-temporal clustering dengan algoritma st-dbscan pada titik api kebakaran hutan indonesia (2015–2020)," *Jurnal Rekursif*, vol. 11, no. 1, pp. 1–9, 2023.

[9] S. Harini, H. Fahmi, A. D. Mulyanto, and M. Khudzaifah, "The earthquake events and impacts mapping in bali and nusa tenggara using a clustering method," in *IOP Conference Series: Earth and Environmental Science*, vol. 456, no. 1, 2020.

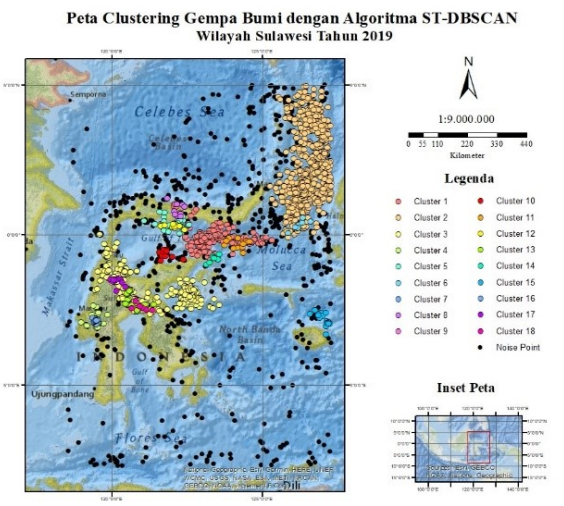
[10] M. A. P. Saputri and D. Sunarya, "Analisis perbandingan energi gempabumi utama dengan gempabumi susulan: Studi kasus gempabumi cianjur 21 november 2022," vol. 4, no. 3, pp. 1–7, 2023.

[11] Irwan, A. Y. Hashari, H. Ihsan, and A. Zaky, "Penggunaan self organizing map dalam pengelompokan tingkat kesejahteraan masyarakat," *Jambura Journal of Probability and Statistics*, vol. 1, no. 2, pp. 57–68, 2020.

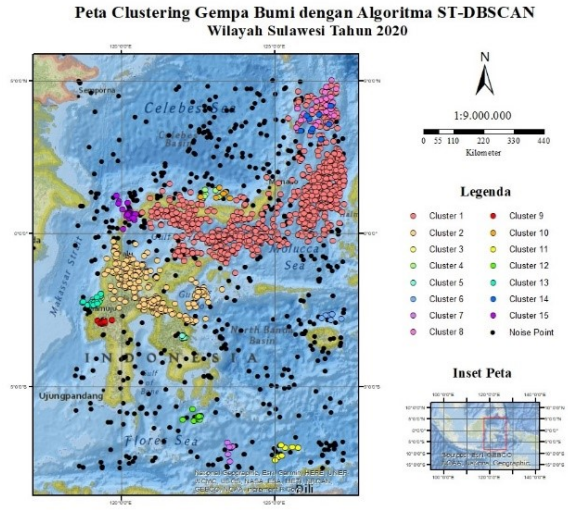
[12] V. Satopää, J. Albrecht, D. Irwin, and B. Raghavan, "Finding a 'kneedle' in a haystack: Detecting knee points in system behavior," in *Proceedings of the International Conference on Distributed Computing Systems Workshops*, 2011, pp. 166–171.

[13] H. Anton and C. Rorres, *Elementary Linear Algebra*, 2014, vol. 11.

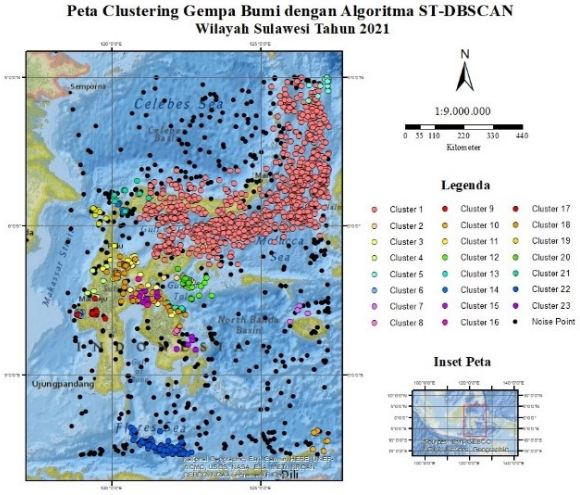
[14] R. Handoyo, R. R. M, and S. M. Nasution, "Perbandingan metode clustering menggunakan metode single linkage dan k-means pada pengelompokkan dokumen," *JSM STMIK Mikroskil*, vol. 15, no. 2, pp. 73–82, 2014.



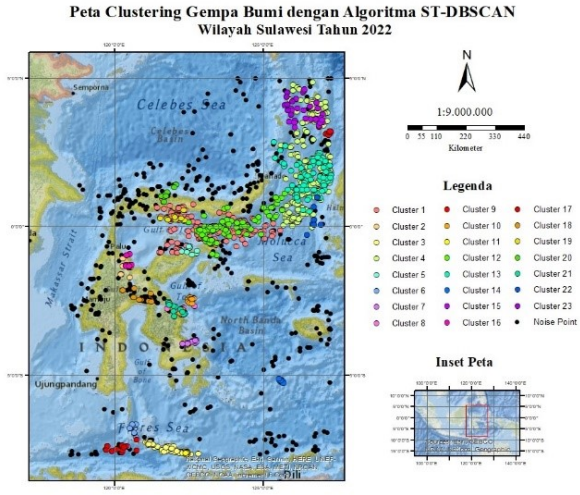
(a) 2019



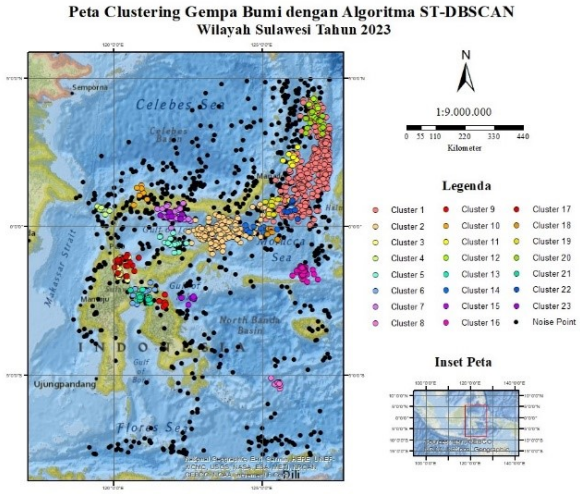
(b) 2020



(c) 2021

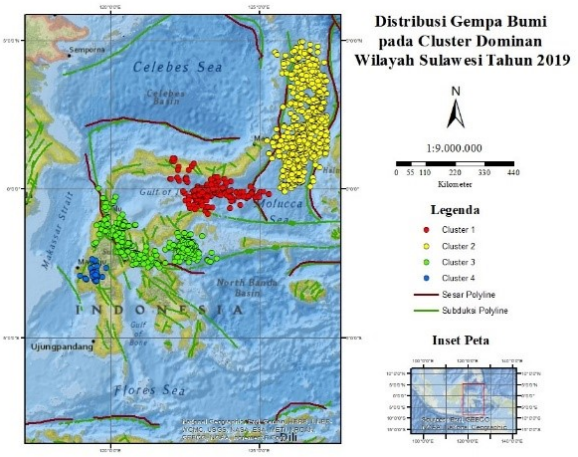


(d) 2022

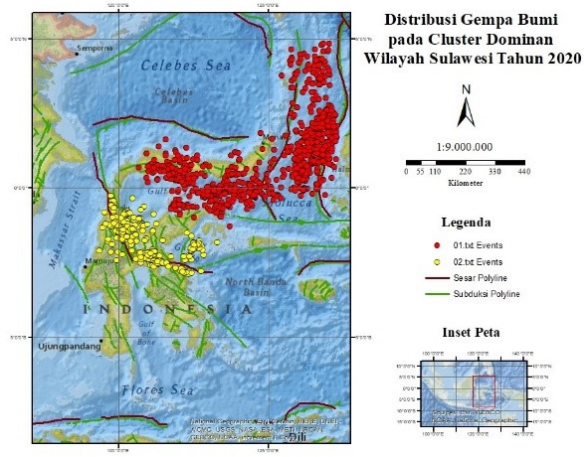


(e) 2023

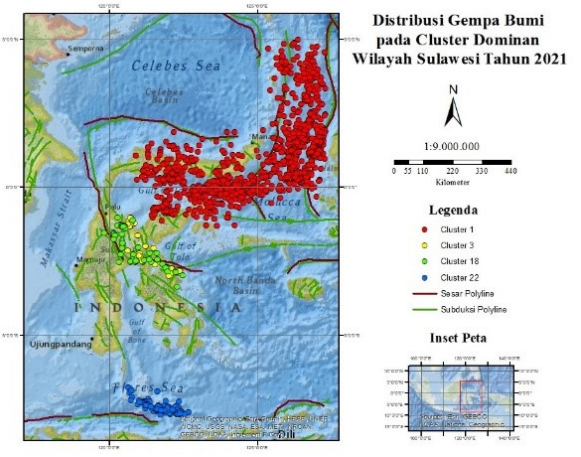
Figure 2. Earthquake Distribution Map



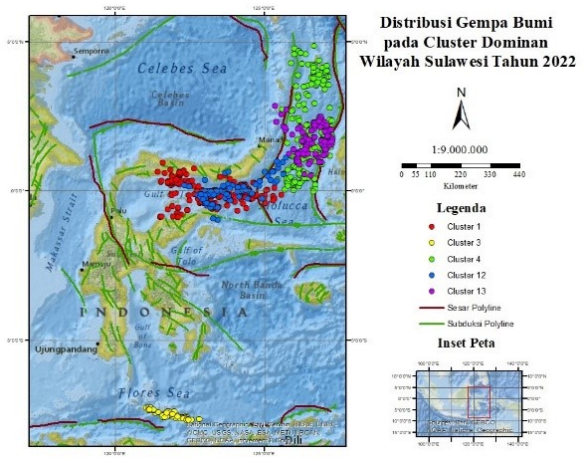
(a) 2019



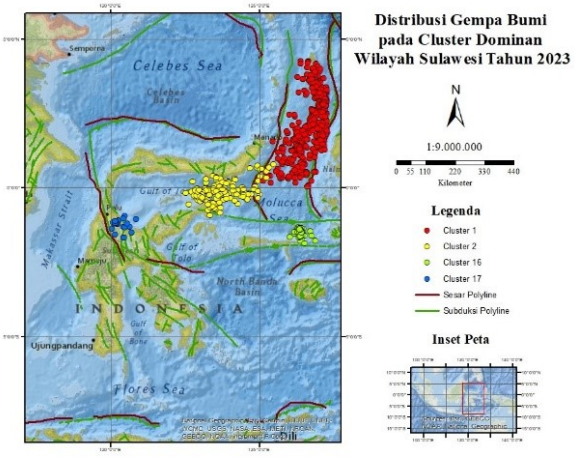
(b) 2020



(c) 2021



(d) 2022



(e) 2023

Figure 3. Distribution map of dominant earthquakes