 https://doi.org/10.1051/e3sconf/202447902005

State transition matrix and Markov-chain diagram for frequent volcanic eruptions: Krakatoa, Indonesia

Danni Gathot Harbowo[1,2*] and *Triyana* Muliawati[3]

[1]Sedimentology, Stratigraphy, and Geodynamic Research Group, Department of Geology, Institut Teknologi Sumatera, 35365, Wayhui, Lampung, Indonesia
[2]Center for Sustainable Development Goals Studies (ITERA SDGs Center), Institut Teknologi Sumatera, 35365, Wayhui, Lampung, Indonesia.
[3]Department of Matematics, Institut Teknologi Sumatera, 35365, Wayhui, Lampung, Indonesia.

Abstract. Krakatoa has been a global attraction throughout history. Historical records of eruptions on this volcanic island complex provide thrill and excitement for visitors at once. They are often stunned by sudden eruptions, noticed by columns of ash billowing high into the sky and drawing attention to the scene. However, eruptions such as this also risk visitors of an improper radius. Therefore, in this study, we aimed to reveal the probability pattern of eruptions to sustain preparedness for the worst. Through the Anakrakatau eruption dataset that we collected from 2018 to 2023 [n=540], we propose a transition matrix diagram of eruption events generated from probability analysis. The approach of this method is based on a Markov-chain analysis. This study assessed the period between eruptions and the probabilities of observed column height. In this study, state determination refers to the *k*-means clustering (*k*=3) each variable. The results show that there are states that represent the variety of circumstances transitions in frequent small activity eruption. The highest probability achieved at eruption with maximum ash column below 800 m and time gaps between eruptions in less than two days. The results of this study provide new insights into the probability of annual eruptions and provide information for sustainable risk mitigation purposes. This report can be a reference for visitors to the Krakatau area, either for education or research.

1 Introduction

Krakatoa has attracted the attention of the world community for more than a century. Its position in the Sunda Strait (**Fig. 1**) has made this volcano known to travellers who have passed through it since time immemorial. Krakatau is known as an isolated volcanic island due to the frequency of its eruptive activity. Referring to its geohistorical, this volcanic activity can cause widespread disturbances and damage, both because of the volcanic material produced and the impact of eruptions that trigger tsunamis [1]. One historical record is that the Krakatoa eruption in 1883 caused tremendous damage, both from a tsunami as

* Corresponding author: danni.gathot@gl.itera.ac.id

high as 30 m and from instant weather changes worldwide [1,2]. This incident has attracted the attention of various groups worldwide regarding Krakatoa, from various public discussions to scientific studies [3-5].

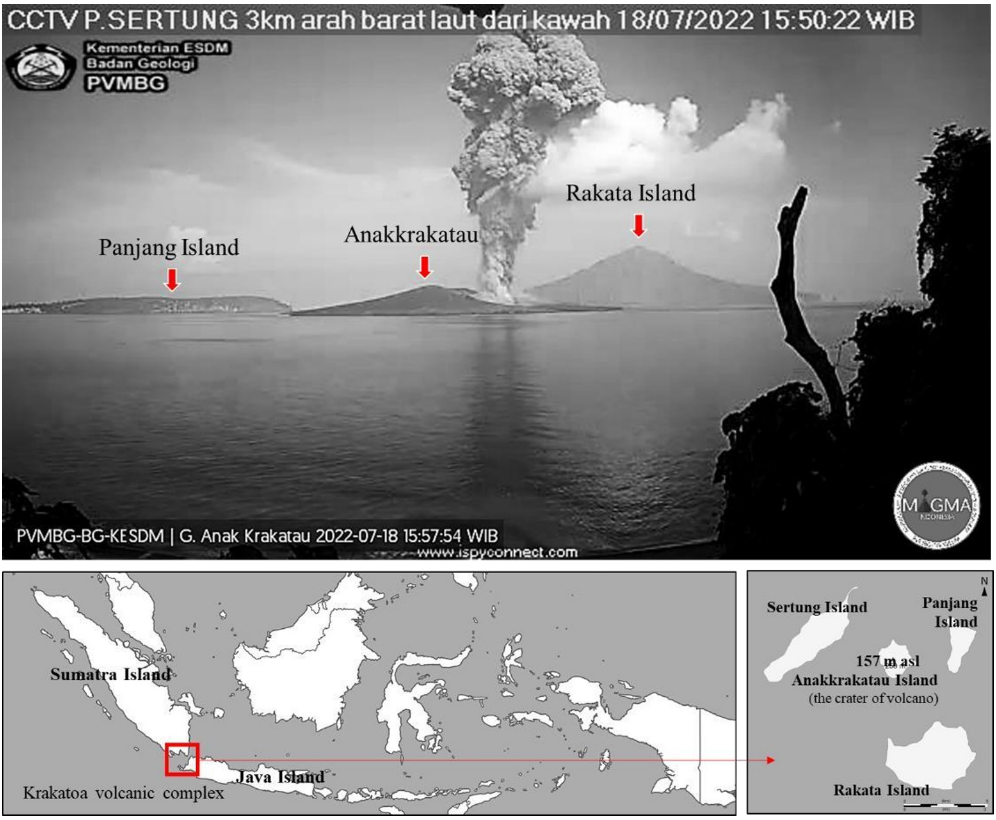

Fig. 1. Location of the Krakatoa volcanic complex and the appearance of Krakatoa from observation cameras on Sertung Island during the eruption on August 18, 2022, at 03.57 PM. (*magma.esdm.go.id*).

Until now, Krakatoa has shown frequent eruptions. Data collection on volcanic activity was recorded at observation posts in South Lampung and Banten using observation instruments located on site [6]. These data have been well collected and can be obtained publicly since 2018 via the *magma.esdm.go.id* webpage. On this page, information is available regarding the date and local time of the eruption, photos of the appearance when the eruption occurred, the height of the ash column formed, and a summary of the seismic record, which includes the maximum amplitude and seismic duration. Using these data, in this study, we describe the behaviour of Krakatoa activity based on a stochastic model, which refers to the probability of each eruption event depending only on the previous eruption event, also known as the Markov chain [7,8]. This study approaches the grouping of eruption events into states and calculates the transition between states. It is described as a transition matrix and Markov diagram. Through this study, it is hoped that the transition opportunities between eruptions can be described to estimate the probability of what will happen afterwards. This study was carried out to forecast the activity of the Krakatoa eruption in line with routine monitoring, which continues to this day.

2 Data and methods

This study uses a compilation of data from observations of Krakatoa activity collected at *magma.esdm.go.id*. Data were compiled from 01/29/2018 to 06/20/2023; the data within that range were only available in the database until this research was conducted (**Fig. 2**). From this information, information was compiled regarding the eruption date, ash column height, recorded eruptive seismic duration, and seismograph maximum amplitude.

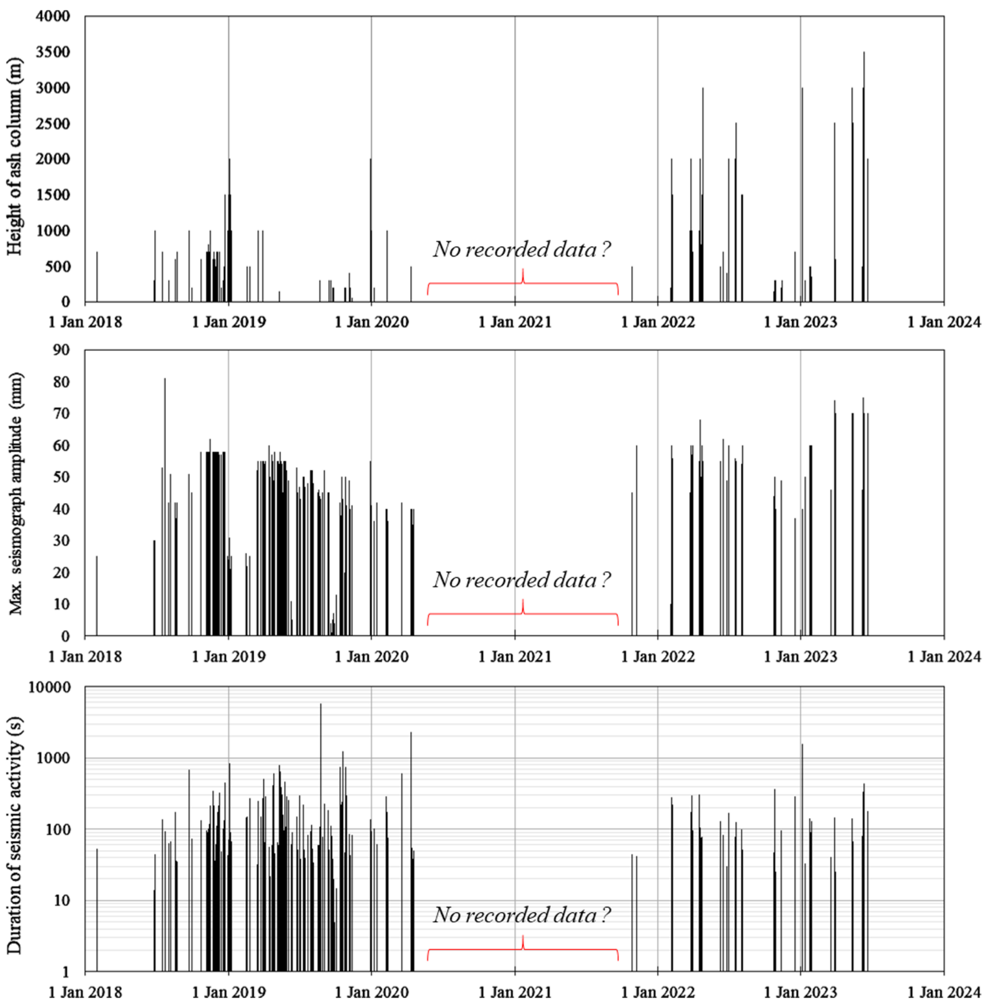

Fig 2. Distribution of data recordings of the Krakatoa eruption from January 1, 2018, to June 20, 2023.

Clustering was then carried out on the data using the *k*-means clustering algorithm with a value of k=3 for each variable. The variables in this study include the ash column height in meters, seismic duration in seconds, seismograph maximum amplitude in millimetres, and the time gap between eruption events in days. In this manner, we obtain three centroids for the three clusters for each variable.

Next, the cluster results are used as a state for each variable so that there are three states for each variable. Then, state transition calculations are carried out for each variable, which are presented in the state transition matrix and Markov chain diagram accompanied by the transition probability [3,4]. To explain the distribution and analytical description, correlation xy plots between variables, box plots for each state in each variable, and distribution of data for each variable in the Gregorian month were also performed.

3 Results

From the compilation of Anakkrakatau activity data, 540 data points were collected. As a note, according to the database in *magma.esdm.go.id*, there are no official data records for April to November 2021, resulting in a data gap. **Fig. 2** shows that the maximum ash column distribution recorded since 2018 can reach a height of 3500 m, which will occur in 2023. When recording seismographic activity, the maximum amplitude of Krakatoa can reach 80 mm. In addition, the duration of the seismograph activity for one activity can reach 3000 seconds or 50 min. Referring to these data, it seems that since 2018, there has been an increase in the height of the ash column in the Krakatau volcanic activity. The height of the ash column could indicate increasing pressure in the Krakatoa magma chamber [9], along with the growth of Krakatoa after its collapse in December 2018.

Next, correlation analysis was performed. In this correlation analysis, the only variables were the maximum seismograph amplitude, duration of seismic activity, and height of the ash column (**Table 1**). The variable time gaps between eruption events were included in k-means clustering and Markov chain analysis (**Table 2** and **Table 3**). The time gap between eruptions may be an indication of magma accumulation rate in the magma chamber [10].

Table 1. Correlation between variables

	Max. seismograph amplitude	**Duration of seismic activity**	**Height of ash column**
Max. seismograph amplitude		-0.02	0.28
Duration of seismic activity			0.14
Height of ash column			

The correlation results show a stronger correlation between the three variables at height of ash column *vs.* height of ash column *vs.* max. seismograph amplitude (0.28). Meanwhile, the correlation values between the other variables tended to be weak. This shows the characteristics of Krakatoa activity, in which some Krakatoa activities that produce eruption ash columns can be represented by the maximum amplitude that appears on the seismograph. To describe the form of this relationship, an xy plot was carried out between variables, as shown in **Fig. 3**. In this plot, regression was carried out on the highest variable correlation value.

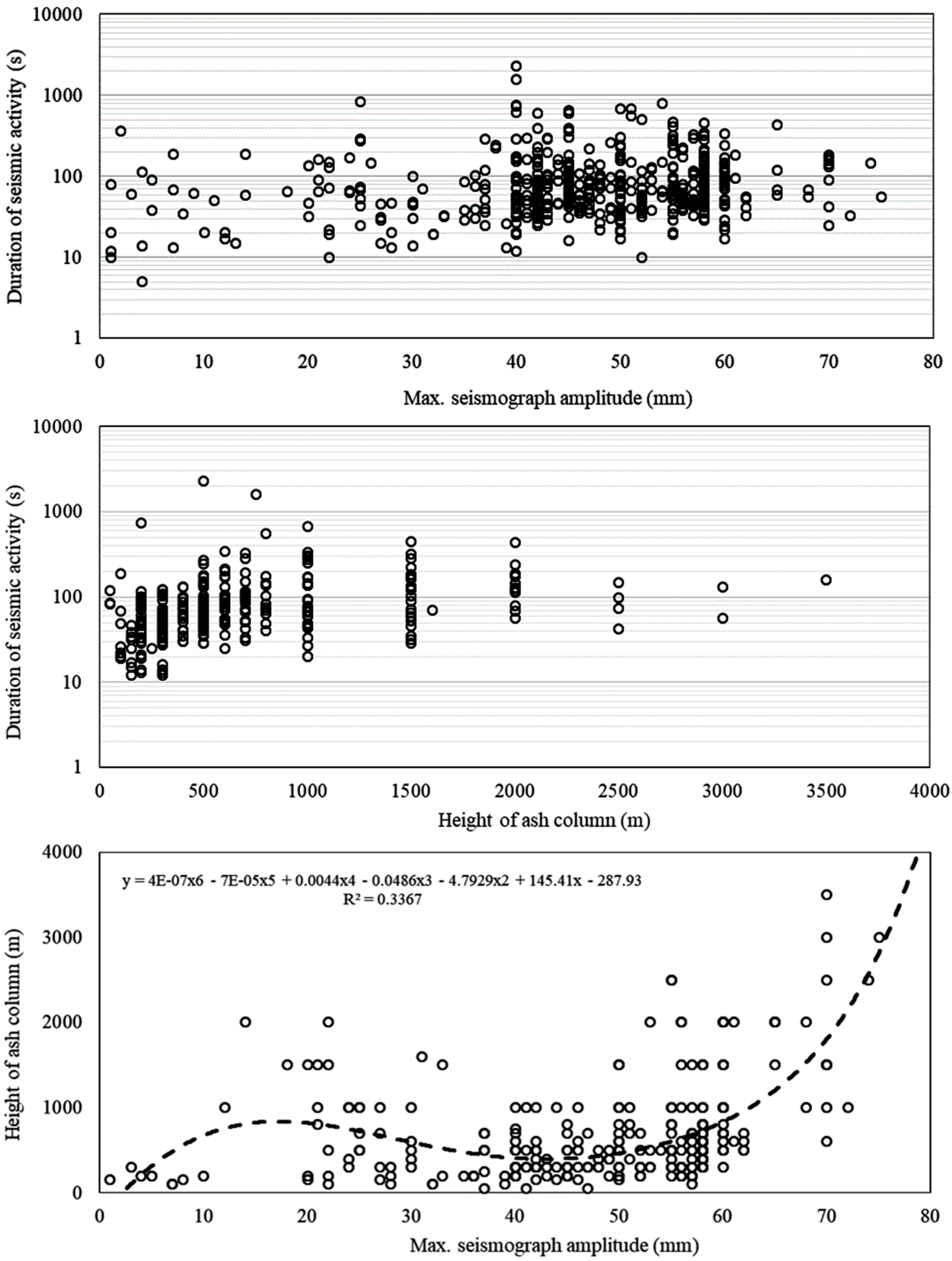

Fig. 3. xy plot between variables, on variable max. Seismograph amplitude vs. height of ash column (correlation coefficient = 0.28) polynomial regression analysis was carried out (R^2 = 0.3367).

In **Fig. 3**, it can be seen that for the variable duration of seismic activity vs. maximum seismograph amplitude, Krakatoa's activity pattern tends to be at maximum amplitude of 40-60 mm with a seismograph duration of 10-1000 seconds, and the resulting ash column ranged from less than 100-750 m. Interestingly, the data plot shows that the duration of seismic activity tends to decrease when the amplitude exceeds 60 mm, which may indicate a different behaviour of the different magmatic characteristics in eruptive events. The eruption behaviour of highly viscous magma (felsic magma) tends to significantly increase the seismograph amplitude but tends to occur in a shorter duration. On the other hand, less viscous magma (intermediate basaltic magma activity) tends to generate gentler eruptions over longer durations [11,12]. Additionally, the distribution of ash column values tended to be related to the pattern with the maximum amplitude on the seismograph. This positive correlation tendency was also reflected in the results of the x^6 polynomial regression, with an R^2 value exceeding 0.3. Next, a cluster analysis was performed on the image using the *k*-means clustering algorithm (*k*=3) of each variable. The results of the clustering are shown in **Fig. 4**.

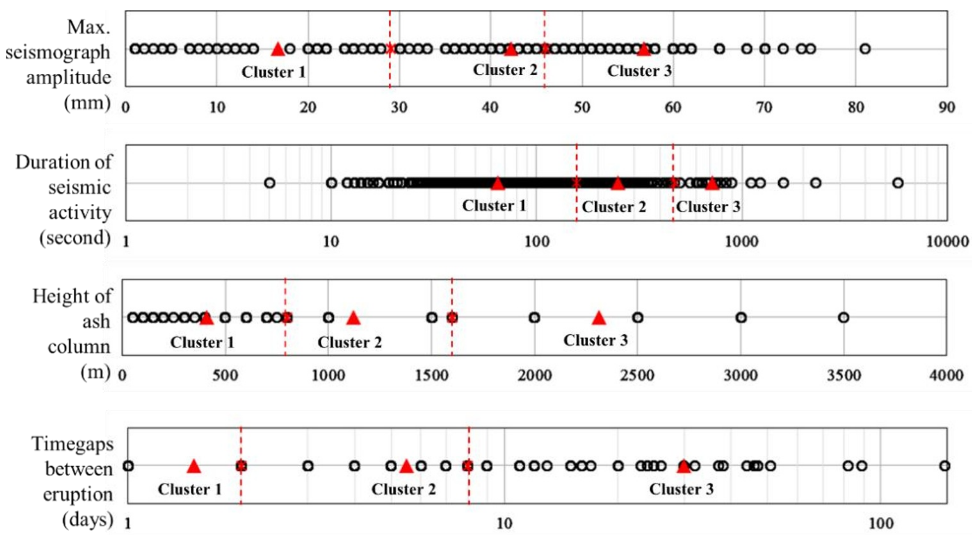

Fig. 4. *k*-means clustering for each variable. The circle indicates the data included in the cluster, the triangle is the centroid of the cluster, and the cluster boundary is marked with a dotted red line.

The cluster results serve as a reference for determining the states for preparing transition matrices and Markov chain diagrams. The cluster results show the concentration of data that can be grouped with the Euclidean value closest to the centroid. The centroid is obtained when it converges by rearranging the centroid position in the data. The cluster boundaries were determined from the smallest and highest values of the Cluster 2, respectively. Cluster 1 covered the lowest value of the Cluster 2 up to the minimum value in the data, and the second cluster covered the maximum value from Cluster 2 up to the maximum value. Centroid values and cluster boundaries of each variable is provided in **Table 2**.

Table 2. Centroid and range of the cluster each variable

Variable	Max. seismograph amplitude (mm)		Duration of seismic activity (s)		Height of ash column (m)		Gaps between eruption (days)	
	Centroid	Range	Centroid	Range	Centroid	Range	Centroid	Range
Cluster 1	16.7	< 29	64.5	<157	406	<800	1.5	<2
Cluster 2	42.2	29 to 46	249.1	157 to 467	1120	800 to 1600	5.5	2 to 8

Cluster 3	65.8	>46	714.5	>467	2312	>1600	30	>8

Next, the cluster results were used as a state reference for the sequence of events from 2018 to 2023, as described in **Fig. 2**. Therefore, based on this series of events, the number of state transitions and the probability of transition from one state to another can be calculated. The results of this analysis are presented in the state transition matrix (**Table 3** and **Table 4**) and Markov chain diagram (**Fig. 5** and **Fig. 6**).

Table 3. State (S) transition matrix diagram for variable max. Seismograph amplitude and duration of seismic activity.

Max. seismograph amplitude			
State	S1	S2	S3
S1	0.55	0.32	0.13
S2	0.09	0.68	0.24
S3	0.03	0.19	0.78

Duration of seismic activity			
State	S1	S2	S3
S1	0.88	0.11	0.02
S2	0.66	0.25	0.09
S3	0.56	0.22	0.22

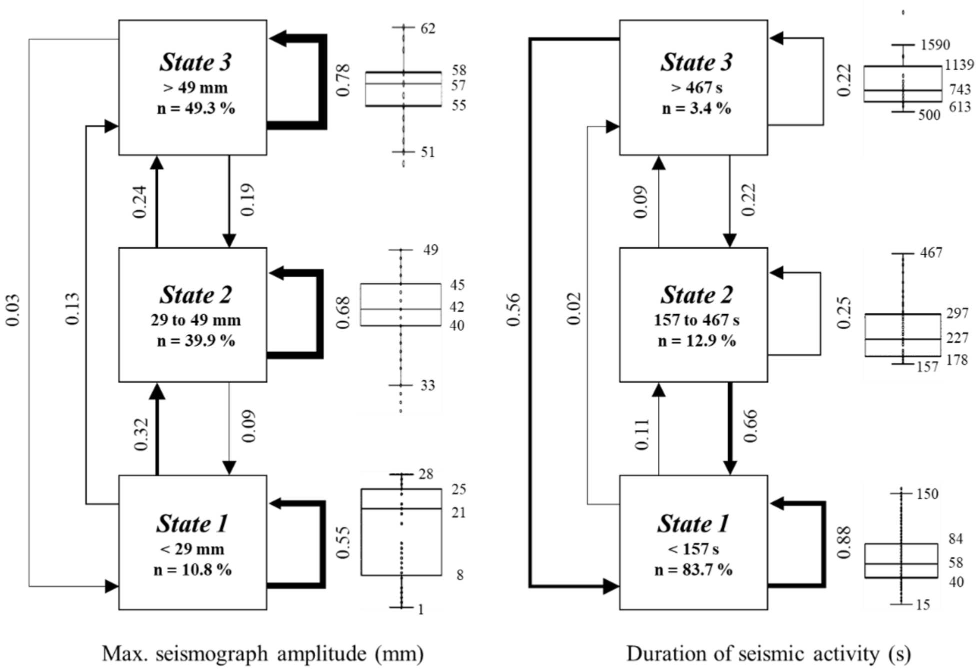

Fig. 5. Markov-chain diagram on variable max. Seismograph amplitude and duration of seismic activity. In the attached figure, the percentage of the number and range of data for each state, the relative transition probability values are also represented by the thickness of the transition line, and the distribution of data is represented by the boxplot.

In the Markov chain diagram, the transition trend for each variable can be observed. The maximum seismograph amplitude tends to repeat states compared with transitions, and the tendency for transition shifts to occur when changing states, both from State 1 (S1) to State 2 (S2) and from State 2 (S2) to State 3 (S3). Meanwhile, a variable duration of seismic activity tends to occur in State 1, that is, a shorter seismic duration. than 157 seconds. Increasing duration tends to occur between eruptions very rarely, after a long duration of activity, anakkrakatoa tends to be followed by activity of a shorter duration (to State 1), both from State 2 to State 1 (0.66) and from State 3 directly to State 1 (0.56). Then, the variable ash column and repetition of ash column ejection events tended to occur at heights <800 m (0.88). There were no frequent events where the ash emissions were higher than this value. Ash ejections up to 1600 m can occur frequently (19.1%) with a relatively frequent chance of repetition (0.48). The ash plume tended to fall to lower states, from State 3 to State 2, and from State 2 to State 1. According to the variable time gaps between eruptions, eruptions within <2 days occur more frequently (State 1). If an eruption does not occur for more than two days, it is very likely that subsequent eruptions will occur more frequently.

Table 3. State (S) transition matrix diagram for variable height of ash column and time gaps between eruptions.

Height of the ash column			
State	S1	S2	S3
S1	0.88	0.09	0.02
S2	0.35	0.48	0.17
S3	0.25	0.46	0.29

Time gaps between eruptions			
State	S1	S2	S3
S1	0.83	0.12	0.06
S2	0.63	0.24	0.14
S3	0.77	0.13	0.10

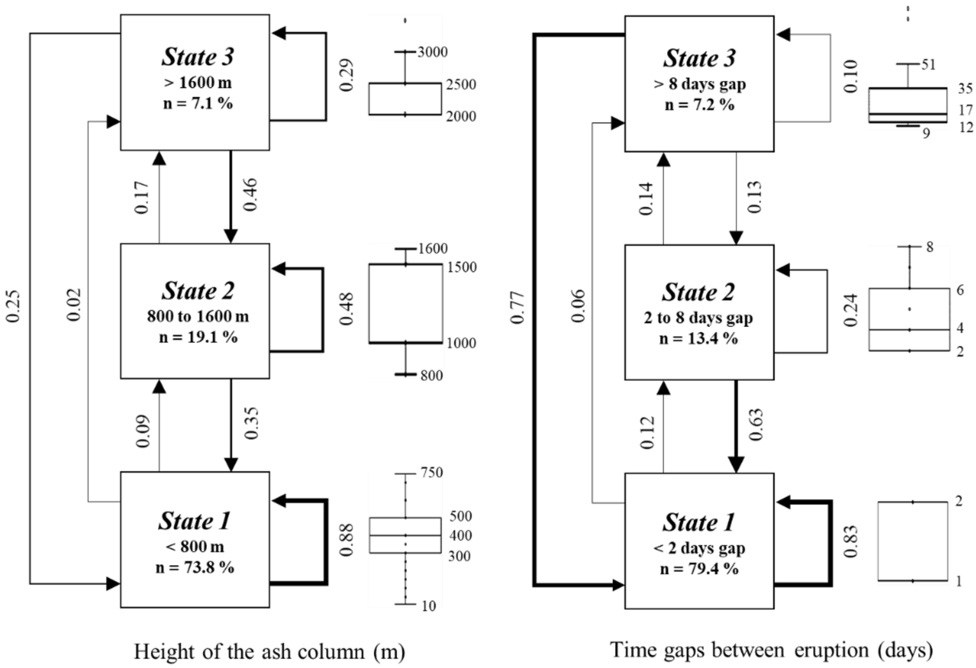

Fig. 6. Markov-chain diagram on variable height of the ash column and time gaps between eruptions. In the attached figure, the percentage of the number and range of data for each state, the relative transition probability values are also represented by the thickness of the transition line, as well as the distribution of data represented by the boxplot.

In this study, we reviewed the activity characteristics of each Gregorian month (**Fig 7**). The Gregorian calendar represents the position of the Earth relative to the sun. This position can have an impact on gravitational attraction, which affects fluid dynamics in the magma chamber, as well as the surface hydrostatic pressure caused by tidal changes. A similar phenomenon can also be found in several volcanoes, such as Mount Etna in Italy, Mount Lewotolo in Indonesia, and other frequently erupting volcanoes that appear to interact with annual planetary gravity [13-17].

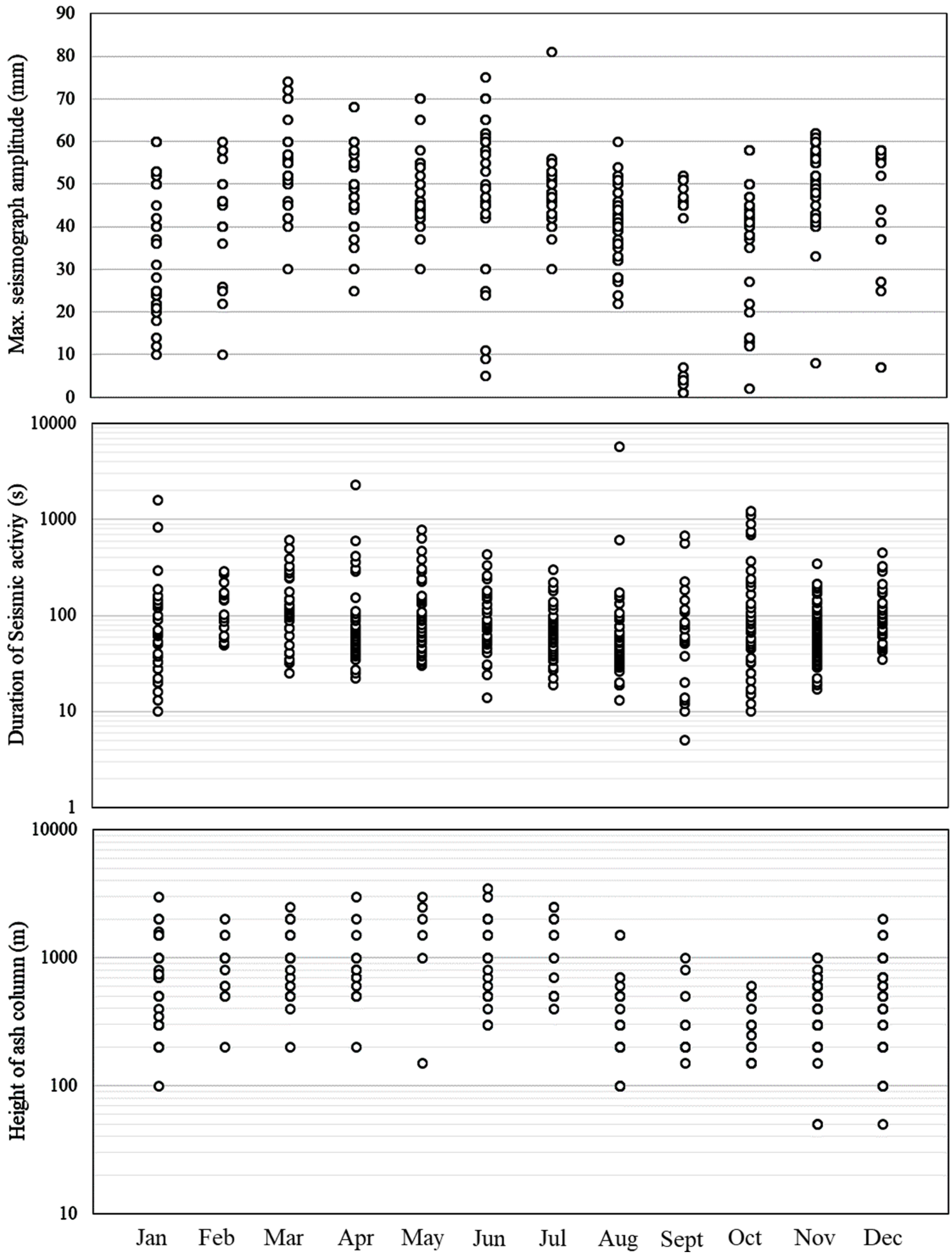

Fig. 7. Data plot of Krakatoa activity for each variable in the Gregorian month (calendar).

The plot results showed a relationship between increased activity and dates in the Gregorian calendar. As shown in **Fig. 7**, a recurring sinusoidal pattern was noticeable in the annual cycle. An increasing trend was observed from October to June, whereas a decreasing trend was observed from June to August. This phenomenon is intriguing and warrants further elaboration in the following research. This appears to be related to volcanic magmatism and gravitational tide pressure resulting from the gravitational interaction between the Earth and the Sun during specific cycle periods.

4 Discussion

This study provides the results of the Krakatoa volcano activity data analysis along with the relationships between its variables, including data distribution, correlation, clustering, state transition, and activity trends on the Gregorian calendar. The data used were compiled from eruptive event records in the magma.esdm.go.id database, covering the period 2018 to 2023. In total, 540 data points were used in this study.

Interestingly, there are indications of increased activity from 2018 to 2023. The ash column height in 2023 tends to be higher compared to previous years, reaching up to 3500

m. This may indicate more intense volcanic activity than in previous years. The increase of the volcanic activity can be triggered by several factors: (1) magma mixing or assimilation of the magma in the magma chamber, (2) dynamic of the pressure and temperature, (3) regional seismic activity (by tectonic), and (4) the seawater intrusion to the magma chamber, which could trigger the formation of high-pressure gas in magma chamber.

The purpose of the correlation analysis in this study was to examine the connection between three main variables: (1) maximum seismograph amplitude, (2) duration of seismic activity, and (3) height of the ash column. The analysis outcomes revealed a positive correlation between the ash column height and maximum amplitude (*correlation coefficient* = 0.28). This correlation coefficient was relatively high compared with the other variables, further confirmed by polynomial regression analysis ($R^2 = 0.3367$). This pattern also suggests the presence of both effusive and explosive characteristics of frequent Krakatoa eruptions.

The intensity of the pendulum oscillations in seismograph would demonstrate the amplitude on the seismograph. The value of this variable is an expression of seismic activity such as magma, rock fractures, or tectonic interactions. When seismic activity increases, indicating increased dynamics within the magma pocket, pressure within the magma pocket can also increase. When this dynamic trigger an increase in pressure within the magma pocket under certain conditions, this pressure increase will result in an observable height of the ash column on the Earth's surface. The positive correlation between the maximum seismograph amplitude and the height of the ash column indicates that changes in the intensity of seismic activity can provide clues regarding the potential for volcanic eruptions and the resulting height of the ash column. Polynomial regression results confirming this relationship show a stronger correlation between these variables, indicating an increased potential for eruption.

Furthermore, by this study, Krakatau activity data was grouped based on cluster analysis using the k-means clustering algorithm with k=3 for each variable. The clustering results revealed three typical data groups differentiated by activity levels defined as states. These results were used as a reference for constructing the transition matrix and Markov chain diagrams. Based on the clustering results, the state transitions and transition probabilities from one state to another were calculated to create a state transition matrix. A Markov chain diagram was created to describe the transition trends of the events in each variable. This analysis provides a better understanding of how the Krakatoa Volcano activity developed from one event to the next.

The activity transition trend in the Markov chain diagram observed in this study reflects the complexity of Krakatau volcanic activity and the factors that influence it. First, state repetition occurs more often in variable maximum amplitude seismographs than in the transitions between other states. This indicates that strong seismic activity tends to last for a certain period before entering the next phase. Transitions occurred during significant changes in seismic activity, which can indicate potential eruptions. The duration of the seismic activity variable tends to occur in State 1 (lower state), the short duration of seismic activity. This indicates that brief seismic activity is most common, and significant increases in duration are rare. After a longer period of seismic activity (long duration of activity), volcanoes tend to return to shorter periods of activity, indicating that they are experiencing a cycle of activity.

The ash column variable indicates that ash column ejection events tend to occur at heights <800 m and have relatively frequent recurrence opportunities. Ash ejections higher than 1600 m occur less frequently, and when they occur, there is a lower chance of recurrence. This may reflect the characteristics of Krakatau's eruptive activity, which tends to produce ash columns that are not consistently high.

Finally, the time gaps between eruption variables show that eruptions less than two days apart occur more frequently (State 1), indicating more frequent eruptive activity or repeated eruption events within a short period. If an eruption does not occur for more than two days, another eruption will likely occur. This highlights the importance of constant monitoring and surveillance of Krakatau Volcano, especially if there are long periods of no eruption, as this may indicate an increase in the potential danger of subsequent higher eruptions.

5 Conclusion

According to Krakatau activity data from 2018 to 2023, there is an indication of a correlation between variable of the seismograph amplitude expression and the resulting eruption ash column height. The Markov chain diagram also shows a higher frequency tendency in the lower state, specifically for a maximum amplitude expression of <29 mm, seismic duration of <157 s, and height of ash column <800 m, with eruptions occurring relatively on a daily scale (the time gap between eruptions is no more than two days). Additionally, there is an indication of a sinusoidal relationship between the Krakatau activity levels in an annual cycle, with an upward trend from October to June and a downward trend from June to August.

We express our appreciation to PVMBG for providing data for this research via *magma.esdm.go.id*, which was fundamental in our study. We thank Tri Lestari (ITERA) for her valuable contribution to collecting data. We express our gratitude to Ir. Imam Santosa, M.Sc, from the Indonesian Geological Agency, E3S editor teams, and Committee of the International Seminar of Science and Applied Technology (ISSAT) for their support and recommendations in this study. These contributions enriched the analysis. We also thank our colleagues for their fruitful discussion in bringing this research.

References

1. D.G. Harbowo. *Journal of Applied Geoscience and Engineering* **2**, 11-25 (2023).
2. M.A. Obregón, M.C. Gallego, M. Antón, J.M. Vaquero. *Geoscience Data Journal* **7**, 185-191 (2020).
3. Z. Hidayat, R. Indra, W.E. Lidya, S.S. Tukina. *Journal of Theoretical and Applied Information Technology*, **100**,192-208 (2022).
4. H.C. Natalia, D.G. Harbowo, R. Ikhram. *Journal of Science and Applicative Technology* **5**, 47-57 (2021).

5. C. von Savigny, A. Lange, C. Hoffmann, A. Rozanov. *EGUsphere*, 1-12 (2023)
6. Pusat Vulkanologi dan Mitigasi Bencana Geologi. *Magma Indonesia,* URL: https://magma.esdm.go.id/ (2023)
7. M.S. Bebbington, S.F. Jenkins. *Bulletin of Volcanology* **81**, 1-17 (2019).
8. H.S. Kuyuk, E. Yildirim, E. Dogan, G. Horasan. *Nonlinear Processes in Geophysics* **19**, 411-419 (2012).
9. N. Indrastuti, A. Basuki, H.D. Purnamasari, S. Adi, C. Patria, N. Haerani. *IOP Conference Series: Earth and Environmental Science* **873**, 012021 (2021).
10. M.R. Reid. *Elements*, **4**, 23-28 (2008).
11. B. Dahren, V.R. Troll, U.B. Andersson, J.P. Chadwick, M.F. Gardner, K. Jaxybulatov,
12. Koulakov. *Contributions to Mineralogy and Petrology*, **163**, 631-651 (2012).
13. R. Scandone, K.V. Cashman, S.D. Malone. *Earth and Planetary Science Letters* **253**, 513-529 (2007).
14. H. Munekane, *Earth, Planets and Space* **73**, 192 (2021).
15. G. Sottili, S. Martino, D.M. Palladino, A. Paciello, F. Bozzano. 2007. *Geophysical research letters*, *34*(1).
16. T. Muliawati, D.G. Harbowo. *IOP Conference Series: Earth and Environmental Science* **1245**, 012006 (2023).
17. S. Dumont, S. Custódio, S. Petrosino, A.M. Thomas, G. Sottili. 2023. *A Journey through Tides*, 333-364.