Earthquake behavior in Sumatra territory based on time and distance between consecutive large event

Samsul Anwar1*

1Universitas Syiah Kuala, Department of Statistics, Banda Aceh 23111, Indonesia

Abstract. The purpose of this study is to understand the behavior of earthquakes with a specific magnitude (M) in the Sumatra territory based on their duration of time and the distance between two earthquakes that happen consecutively. We explore 1,641 data sets of duration and distances with a specific magnitude that were generated from 458 earthquakes with M ≥ 5.0 during 2009-2017. This study revealed that the occurrence of the later (second) earthquake at any location (referred to in this paper as a recurring earthquake) with a bigger or equal magnitude as the previous one depends on duration especially for magnitude under 5.8. The comparison analyses showed that the recurring earthquake events tended to have similar average distances. However, the length of time between two earthquakes is different for lower magnitudes (under 6.0), whereas between an earthquake and a recurring earthquake with a magnitude above 6.0, the length of time tends to be similar.

1 Introduction

Indonesia, a country located in between two continents, Asia and Australia, is squeezed by four large tectonic plates of the Earth: Indies, Australia, Eurasian, and Pacific plates. The Indies-Australian plate moves to the north, pounding the Eurasian plate, moving at a speed of 50-70 mm/year [1]. The pounding zones between these two plates are along the sea trench of Sumatra, Java, Bali, and Lombok islands. The movement of collisions and the shifting of these large plates respond mechanically. These plate movements have caused many earthquakes to occur in the Indonesia archipelago, one of them in the fault of the Sumatra territory. According to NOAA’s National Geophysical Data Center, at least 235 tsunamis have occurred in the last four centuries and Indonesia is one of the countries that is most exposed to earthquakes in the world [2].

The oblique tugging between the Indies-Australian and Eurasian plates had a significant influence on tectonic and volcanic conditions in northern Sumatra [3]. It caused the Sumatra fracture zone to become reactive when the depth of the earthquake center was not more than 50 kilometers. In general, the Sumatra territory is divided into two sources of tectonic earthquakes, those that are in subduction zones along the sea, and those in the zones along the land with a fracture system of the Sumatra territory [4]. There are several massive earthquakes that have happened in this fractured area, including the earthquake in Padang in 2009, Mentawai in 2010 and the earthquake and tsunami Aceh in 2004. The earthquake is one of the tsunamigenic events that caused a tsunami. In Indonesia, tsunamis in the northern part of the equator tend to happen less frequently compared to the southern part of the equator [5].

The earthquake known as the Sumatra-Andaman earthquake happened on December 26, 2004 [6,7], and it is one of the earthquakes whose source is the subduction of oceanic plates into the Earth along the continental banks of western Sumatra, which extends to the south of Java, Bali, and Lombok islands [3]. This 2004 earthquake with a magnitude of 9.1 - 9.3 [6, 8–10], was also felt in Bangladesh, India, Malaysia, Myanmar, Thailand, Singapore, Sri Lanka and the Maldives [7] and caused a tsunami wave with a peak of 51 meters (167.3 ft) [11]. It destroyed everything not only in the southern coastal area of Sumatra island, but also in the other areas along the coastline of the Indian Ocean, and even reached some countries on the coast of the Africa continent. The number of victims caused by this earthquake and the tsunami was approximately 230,000 people in 14 countries [2]. The Sumatra-Andaman earthquake is the third biggest earthquake that has ever been recorded with a seismograph and had the longest duration in history, about 8.3 to 10 minutes [12]. The earthquake caused the entire planet to vibrate 1 centimeter (0.4 inches) [13]. In addition, the Sumatra-Andaman earthquake also triggered the formation of a new fault pattern, specifically around the Sumatra Island which could potentially cause other earthquakes with large magnitudes. It can be shown that an earthquake with a large magnitude (M > 5.5) might occur without being preceded by a smaller earthquake. Although the probability is relatively small, an earthquake with a magnitude > 7.6 can possibly occur before an earthquake with a magnitude between 5.0 and 7.6

* Corresponding author: samsul.anwar@unsyiah.ac.id
The 13th AIWEST-DR 2021

The 13th AIWEST-DR 2021

[14]. Another study showed that the Aceh region, the nearest area from the epicentre of the Sumatra-Andaman earthquake, has a high seismicity level [15].

In order to minimize the risk of disasters caused by earthquakes in the future, the mitigation effort needs to be prepared as much as possible. Knowledge of the past can help reduce losses in the future by identifying and defining earthquake and tsunami hazards [8]. This includes the study of earthquake behavior, which is the first step in understanding the phenomena of earthquakes. The behavior can be studied through the research of hazard seismicity. The base seismic hazard assessment is based on the statistical distribution of the recurrence interval [16]. It is assumed that the times between large consecutive earthquakes (inter-event times or recurrence intervals) follow a certain statistical distribution [17]. The duration and magnitude play an increasingly important role in earthquake prediction models [18].

The objective of this study is to understand the earthquake behavior in the Sumatra territory based on the duration and distance between two earthquake events with particular magnitudes. The duration data is analysed using a survival analysis approach in order to understand the hazard seismicity. Furthermore, for each set of earthquakes with a certain magnitude, the duration, as well as the distance between earthquakes was compared by using Analysis of Variance (Anova) or the Kruskal-Wallis test, when applicable. Together, these two analyses give more comprehensive information about the earthquake behavior in the Sumatra territory.

2 Materials and methods

2.1 Data and settings

This study analyses the earthquake dataset catalog of National Meteorology, Climatology, Geophysics Agency (BMKG) from 2009 to 2017 (http://dataonline.bmkg.go.id), with latitudes between -9.66 and 5.57 degrees, longitudes between 94.08 and 105.54 degrees, which is the area northern Sumatra, off the West Coast of northern Sumatra, southern Sumatra, and southwest of Sumatra, Indonesia. The data set used in this study is the length of time (in days) between two earthquakes, where the later (second) earthquake (termed as recurring earthquake) has a bigger or equal magnitude (M) to the previous one. A recurring earthquake happened for a similar duration and at a similar distance.

Comparing magnitudes evaluated to each other in a part of survival analysis. Additionally, Levene’s test is performed to check the homogeneity of variances assumption. As for the duration data, it is checked in a part of survival analysis. Analysis of data in this study uses software R version 3.5.0.

Performing the analysis in this study is divided into three steps:

1. Estimating the probability distribution of the duration of the recurring earthquake events with a specific magnitude using the Maximum Likelihood Estimator (MLE) method that is integrated in the GAMLSS package. The chosen probability distribution is the one having the smallest AIC value.

2. Analyzing the earthquake behavior through the hazard rate according to the probability distribution estimated in the first step.

3. Comparing magnitudes evaluated to each other in order to investigate whether or not the recurring earthquake happened for a similar duration and at a similar distance.

2.2 Methods

This duration data of a recurring earthquake event is a set of survival data. Hence, the data can be analysed by using the survival analysis method. Specifically, the hazard rate is used to help understand how the earthquake behavior in the Sumatra territory is related to the duration of the recurring earthquake events with the specific magnitudes, as stated previously. There are eight probability distributions used to test the duration data set namely the Exponential, Gamma, Gumbel, Logistic, Lognormal, Normal, Normal Family and Weibull distributions.

Furthermore, the length of time as well as the distance between two earthquakes of different magnitudes evaluated are compared using the Analysis of Variance (Anova) or Kruskal-Wallis test, depending on the normality and homogeneity variances assumption. A p-value less than 0.10 was considered significant. These analyses would give a better understanding of earthquake behavior, whether or not different magnitudes would require different durations and distances between the earthquake and the recurring earthquake events to happen. The Shapiro-Wilk test is employed to analyse the distance data set to test its normality assumption. As for the duration data, it is checked in a part of survival analysis. Additionally, Levene’s test is performed to check the homogeneity variances assumption. Analyzing data in this study used software R version 3.5.0.

Performing the analysis in this study is divided into three steps:

1. Estimating the probability distribution of the duration of the recurring earthquake events with a specific magnitude using the Maximum Likelihood Estimator (MLE) method that is integrated in the GAMLSS package. The chosen probability distribution is the one having the smallest AIC value.

2. Analyzing the earthquake behavior through the hazard rate according to the probability distribution estimated in the first step.

3. Comparing magnitudes evaluated to each other in order to investigate whether or not the recurring earthquake happened for a similar duration and at a similar distance.

2.3 Literature background

In order to support the methodology used in this study, some brief background theories are provided.

2.3.1 Survival analysis

The lifetime of a unit is the length of time, $T$, from the first time the unit works until the time it fails. The time variable $T$ is defined as a random variable because it cannot be predicted precisely. There are four important functions in the survival analysis, namely: probability
density function (p.d.f.), cumulative density function (c.d.f.), survival function, and hazard function. One of the most important functions in the survival analysis is a hazard function \( h(t) \). \( h(t) \) is a function that provides the instantaneous failed rate of a unit that already has lifetime at time \( t \) \( (0 < t < \infty) \), or could be written as follows [19]:

\[
h(t) = \lim_{\Delta \to 0} \frac{P(t < T < t + \Delta | T > t)}{\Delta}
\]

where the \( f(t) \) is a p.d.f., \( F(t) \) is a c.d.f. and \( S(t) \) is a survival function with the following forms:

\[
f(t) = P\{T = t\}
\]

\[
F(t) = P\{T \leq t\} = \int_0^t f(x) dx
\]

\[
S(t) = 1 - \{T < t\} = P\{T > t\} = \int_t^\infty f(x) dx
\]

Accordingly, knowing the p.d.f. and c.d.f. would lead to the hazard function based on Equation (1) which is also related to Equation (2) and (4).

### 2.3.2 Maximum likelihood estimator (MLE)

Let \( \theta \) denote the parameter set of the distribution function \( F(t) \). Since the likelihood function for a failure observation \( t \) is \( L(t) = f(t; \theta) \), then the overall likelihood function is written as:

\[
L(\theta) = \prod_{i=1}^{n} L_i(\theta)
\]

The parameter is determined by maximizing the \( L(\theta) \) or easier in logarithmic form, i.e.

\[
\ln L(\theta) = \sum_{i=1}^{n} \ln L_i(\theta)
\]

The maximizing problem can be solved by doing differentiation on \( \ln L(\theta) \) against each parameter in the probability distribution [20].

### 2.3.3 Akaike information criterion (AIC)

The information criterion is appropriate for model selection when the candidate models are either nested or non-nested [20]. The AIC is formulated as:

\[
AIC = -2 \ln(L) + 2m
\]

where \( L \) is likelihood function and \( m \) is the number of parameters in the model. Model with the smallest AIC value is better than others [21].

### 2.3.4 Probability distributions in survival analysis

There are eight probability distributions tested in this study, namely: Exponential, Gamma, Gumbel, Logistic, Lognormal, Normal, Normal Family, and Weibull distributions. Table 1 provides the p.d.f. and c.d.f. of each probability distribution tested [22].

| Probability Distribution | p.d.f. | c.d.f. |
|-------------------------|--------|-------|
| Exponential, EX(\( \mu \)) | \( \frac{1}{\mu} e^{\left(-\frac{y}{\mu}\right)} \) | \( 1 - \exp \left(\frac{-y}{\mu}\right) \) |
| Gamma, GA(\( \mu, \sigma \)) | \( \frac{\frac{1}{\sigma} \gamma \left(\frac{\gamma - \mu}{\sigma}\right)}{\Gamma(\gamma)} e^{\left(\frac{\gamma - \mu}{\sigma}\right)} \) | \( \frac{\gamma(\gamma^{-1}, \mu^{-1}, \sigma^{-1})}{\Gamma(\gamma^{-1})}, \text{where} \ \gamma(\alpha, \chi) = \int_0^\infty t^{\alpha-1} e^{-\chi} dt \) |
| Gumbel, GU(\( \mu, \sigma \)) | \( \frac{1}{\sigma} \left(\frac{1}{\sigma} e^{\left(-\frac{y-\mu}{\sigma}\right)} - e^{\left(-\frac{y}{\sigma}\right)}\right) \) | \( 1 - e^{\left(-\frac{y}{\sigma}\right)} \) |
| Logistics, LG(\( \mu, \sigma \)) | \( \frac{1}{\sigma} \left(1 + e^{\left(-\frac{y-\mu}{\sigma}\right)}\right)^{-1} \) | \( \Phi \left(\frac{\log(y) - \mu}{\sigma}\right) \) |
| Log Normal, LOGNO (\( \mu, \sigma \)) | \( \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left[-\frac{\log^2(y) - \mu^2}{2\sigma^2}\right]} \) | \( \Phi \left(\frac{\log(y) - \mu}{\sigma}\right) \) |
| Normal, NO (\( \mu, \sigma \)) | \( \frac{1}{\sqrt{2\pi\sigma}} e^{\left[-\frac{(y-\mu)^2}{2\sigma^2}\right]} \) | \( \Phi \left(\frac{y - \mu}{\sigma}\right) \) |
| Normal Family, NOF(\( \mu, \sigma, \nu \)) | \( \frac{1}{\sqrt{2\pi\sigma^2\nu}} e^{\left[-\frac{(y-\mu)^2}{2\sigma^2\nu}\right]} \) | \( \Phi \left(\frac{y - \mu}{\sigma\nu}\right) \) |
| Weibull, WEI(\( \mu, \sigma \)) | \( \frac{\sigma y^{\sigma - 1} e^{\left(-\frac{y}{\mu}\right)^\sigma}}{\mu^\sigma} \) | \( 1 - e^{\left(-\frac{y}{\mu}\right)^\sigma} \) |
2.3.5 Kruskal-wallis test

Given data set \( t_{11}, t_{12}, t_{13}, ..., t_{1n_1} \) are samples from population 1 and \( t_{k1}, t_{k2}, t_{k3}, ..., t_{kn_k} \) are samples from population \( k \). Let \( R_{ij} \) be the rank of the data point \( t_{ij} \) and \( R_i = \sum_{j=1}^{n_i} R_{ij} \) be the sum of ranks in the \( i \)th sample and \( N = \sum_{i=1}^{k} n_i \). Thus, the statistics \( H \) of Kruskal-Wallis test is defined as:

\[
H = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(N + 1)
\]

The distribution of statistics \( H \) is approximated by Chi-square distribution with \( k - 1 \) degree of freedom [23].

2.3.6 Mann-whitney u test

Given data sets from two groups, say \( X \) and \( Y \), with \( n_X \) and \( n_Y \) as the number of samples and the sum of ranks assigned to the first group respectively, then \( n_Y \) and \( R_Y \) are the number of samples and the sum of ranks assigned to the second group respectively, the Mann-Whitney U test for each group is defined as:

\[
U_X = n_X n_Y + \left( (n_X(n_X + 1)/2) - R_X \right) - U_Y
\]

In terms of median, two groups are considered to have different medians if, by consulting the Mann and Whitney tables, the probability corresponding to the min\( (U_X, U_Y) \) is smaller than the probability of the predetermined \( \alpha \) threshold [24].

3 Results and discussion

3.1 Data characteristics

During 9 years (2009 - 2017), there were 46,335 earthquakes that occurred in Indonesia with a magnitude of between 1.5 and 7.8 and with a depth of hypocenters of between 1 to 750 meters. 7,970 (17.20%) of those earthquakes happened in the Sumatra territory between the latitudes of -9.66 to 5.57, and longitudes of 94.08 to 105.54. In agreement with a previous study [25], this paper shows that almost all segments in the northern part of Sumatra are in active condition. Among those 7,970 earthquakes, 503 (6.31%) of them have \( M \geq 5.0 \). Table 2 provides the number of earthquake events in Indonesia and Sumatra territory for \( M \geq 1 \) and \( M \geq 5 \).

| Area (Magnitude) | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 |
|------------------|------|------|------|------|------|------|------|------|------|
| Indonesia ( \( M \geq 1.0 \)) | 3,676 | 5,999 | 4,451 | 6,624 | 4,130 | 4,668 | 5,073 | 5,625 | 6,089 |
| Sumatra ( \( M \geq 1.0 \)) | 827 | 1363 | 944 | 1085 | 814 | 751 | 696 | 699 | 791 |
| Sumatra ( \( M \geq 5.0 \)) | 107 | 82 | 69 | 48 | 45 | 38 | 34 | 36 | 44 |

Table 2. Number of earthquake events in Indonesia (2009 – 2017)

Table 3. Descriptive statistics of the recurring earthquake events at all evaluated magnitudes

| Magnitude (\( M \)) | Frequency | Duration time (days) | Distance (km) |
|---------------------|-----------|----------------------|---------------|
|                     |           | Min | Median | Mean | Max | Min | Median | Mean | Max |
| 5.0                 | 458       | 0.50 | 3.00 | 6.82 | 70.00 | 2.22 | 434.20 | 529.70 | 1775.00 |
| 5.1                 | 331       | 0.50 | 4.50 | 9.21 | 77.00 | 2.22 | 448.50 | 542.50 | 1775.00 |
| 5.2                 | 246       | 0.50 | 6.00 | 11.69 | 90.00 | 2.22 | 470.80 | 535.60 | 1634.00 |
| 5.3                 | 178       | 0.50 | 8.00 | 15.76 | 145.00 | 2.23 | 397.50 | 492.60 | 1501.00 |
| 5.4                 | 122       | 0.50 | 11.00 | 20.58 | 165.00 | 2.23 | 355.10 | 459.30 | 1494.00 |
| 5.5                 | 86        | 0.50 | 13.00 | 24.44 | 176.00 | 2.23 | 432.10 | 517.30 | 1494.00 |
| 5.6                 | 71        | 0.50 | 16.00 | 29.66 | 176.00 | 2.23 | 490.90 | 524.30 | 1408.00 |
| 5.7                 | 43        | 0.50 | 27.00 | 36.29 | 162.00 | 2.23 | 490.90 | 515.80 | 1228.00 |
| 5.8                 | 28        | 0.50 | 32.00 | 50.04 | 165.00 | 2.23 | 329.00 | 444.60 | 1033.00 |
| 5.9                 | 24        | 0.50 | 33.00 | 58.09 | 165.00 | 2.23 | 329.00 | 434.00 | 1045.00 |
| 6.0                 | 19        | 0.50 | 48.50 | 61.39 | 173.00 | 2.23 | 232.00 | 415.00 | 1045.00 |
| 6.1                 | 15        | 0.50 | 32.50 | 55.75 | 169.00 | 2.23 | 518.50 | 509.00 | 1033.00 |
| 6.2                 | 11        | 0.50 | 32.50 | 58.95 | 231.00 | 2.23 | 774.00 | 583.90 | 1033.00 |
| 6.3                 | 9         | 0.50 | 33.00 | 44.81 | 123.00 | 2.23 | 242.70 | 390.30 | 803.30 |

However, this study only analysed data from 458 (91.05%) earthquake events with \( M \geq 5.0 \). There was too small a sample size for events with magnitudes above 6.3, so they were not analysed. Note that the duration and distance between epicenters can only be calculated between two events. Accordingly, the number of data...
points in the data set of duration and distance can be formulated as \( n - 1 \), where \( n \) is the number of earthquake events. Furthermore, earthquakes with \( M \geq 5.0 \) can be reclassified into several new categories with bigger magnitude. For example, from this data set, earthquake events can be subdivided into specific \( M \geq 5.1 \) intervals and then the duration and distance between the first to the next earthquake with equal or bigger magnitude can be calculated. As mentioned before, in this study earthquake events are evaluated at \( M \geq 5.0, M \geq 5.1, \ldots \) and \( M \geq 6.3 \). Using this logic, a total of 1,641 new data points of duration and distances are generated from 458 earthquakes with \( M \geq 5.0 \). Some descriptive statistics of duration and distance between recurring earthquake events are provided in Table 3.

According to Table 3, we can see the most frequent recurring earthquakes that happened in the Sumatra territory have a magnitude of \( M \geq 5.0 \), and this magnitude earthquake has happened 458 times between 2009 - 2017. Consequently, there are 457 durations and distances or 27.85\% of total data used in this study. The least frequent magnitude is nine times for \( M \geq 6.3 \). In general, it could also be seen that the intensity of the recurring earthquake tends to be smaller exponentially along with the growth of the earthquake magnitude. Fig. 1 shows the distribution of duration for all recurring earthquake events in time sequence order of each magnitude evaluated during 2009 - 2017.

According to Figure 1, it is known that the duration of the recurring earthquake events with an increasing magnitude measured tends to be longer. It indicates that earthquakes with a bigger magnitude would occur less often when compared to earthquakes with smaller magnitude. Moreover, the larger the magnitude of earthquakes, the less frequently they occur. Note that, in the survival analysis, data points with zero value could not be analysed. Therefore, the occurrence of the recurring earthquakes on the same day as the previous one in this study would be given a small value. In the data set, there are 45 days when the recurring earthquakes happened on the same day with a varying number of recurring earthquake events. On average, there are 2 recurring earthquake events per day. Accordingly, by assuming the recurring earthquakes happen in a uniform time, the recurring earthquake duration on the same day then will be given as 0.5 day (12 hours) instead of 0.

From Table 3, it is also known that the mean of the duration of the recurring earthquake tends to be higher as the magnitude is greater. The mean time between an earthquake and a recurring earthquake with \( M \geq 5.0 \) is 6.82 days and has a median of 3 days, with a maximum of 70 days since the first earthquake. The minimum duration of a recurring earthquake is 0.5 days, which means it happens on the same day as the first earthquake. Furthermore, the mean and median time between the earthquake and a recurring earthquake with \( M \geq 6.3 \) increases to 44.81 and 33 days respectively, with the minimum and maximum time as 0.5 days and 123 days after the first earthquake with \( M \geq 6.3 \). Another
The maximum likelihood estimator (MLE) is used to estimate the probability distribution for the duration data of the recurring earthquake events with a specific magnitude. The GAMLSS package provides the MLE with various probability distributions that can be tested. In this study, there are eight probability distributions to be evaluated, namely: Exponential, Gamma, Gumbel, Logistic, Lognormal, Normal, Normal Family and Weibull distributions. Three of these distributions (the Gamma, Lognormal, and Weibull) are frequently used for earthquakes, because they share three properties commonly observed for earthquake inter-event times [26]. The probability distribution which has the highest suitability level for the distribution of the data set is the one that has the smallest Akaike Information Criterion (AIC) value among all probability distributions tested. Table 4 provides AIC values for all probability distributions evaluated of the recurring earthquake events with M \geq 5.0, M \geq 5.1, … and M \geq 6.3.

3.2 Estimating probability distribution of duration data

The maximum likelihood estimator (MLE) is used to estimate the probability distribution for the duration data of the recurring earthquakes with a maximum magnitude evaluated (6.3) might occur on the same day as a previous earthquake with a similar magnitude. The earthquake’s magnitude varies from one earthquake to the next and the earthquake does not necessarily repeat at regular intervals [8]. Slightly different, the mean and median of distances between the recurring earthquakes seem not to have constant increasing or decreasing patterns. However, the maximum distance between recurring earthquakes tends to be smaller in accordance with the increase of its magnitude.

The interesting fact is that the recurring earthquakes with a maximum magnitude evaluated (6.3) might occur on the same day as a previous earthquake with a similar magnitude. The earthquake’s magnitude varies from one earthquake to the next and the earthquake does not necessarily repeat at regular intervals [8]. Slightly different, the mean and median of distances between the recurring earthquakes seem not to have constant increasing or decreasing patterns. However, the maximum distance between recurring earthquakes tends to be smaller in accordance with the increase of its magnitude.

3.3 Hazard rate of the recurring earthquake events

According to the results of the estimation probability distribution with the smallest AIC criteria, the recurring earthquake data set for each magnitude evaluated in this study approximately follows one of the following probability distributions: Lognormal, Weibull, Gumbel, and Exponential distribution. However, the Lognormal distribution is more appropriate than the Gaussian (Normal) or Weibull distributions for data of actual repeat time/difference between repeat times [27]. Another study showed that the Lognormal model provided only a slightly better fit than the Gaussian (Normal) or Weibull models by choosing a set of models without any physical or theoretical basis [28]. The parameters of those distributions used in this study are estimated through MLE and shown in Table 5.

### Table 4. AIC values of the examined probability distributions for duration data between the recurring earthquake events at all evaluated magnitudes

| Magnitude (M) | EXP   | GA    | GU    | LG    | LOGNO | NO    | NOF   | WEI   |
|---------------|-------|-------|-------|-------|-------|-------|-------|-------|
| M \geq 5.0    | 2671.12 | 2654.56 | 3656.16 | 3148.21 | 2616.73 | 3241.44 | 3243.44 | 2646.21 |
| M \geq 5.1    | 2127.11 | 2101.63 | 2816.82 | 2481.11 | 2087.65 | 2546.45 | 2548.45 | 2096.05 |
| M \geq 5.2    | 1696.86 | 1666.60 | 2226.88 | 1978.01 | 1654.84 | 2038.78 | 2040.78 | 1660.81 |
| M \geq 5.3    | 1332.14 | 1300.85 | 1769.43 | 1542.19 | 1295.80 | 1603.65 | 1605.65 | 1296.23 |
| M \geq 5.4    | 975.87  | 943.09  | 1266.46 | 1127.68 | 940.90  | 1167.29 | 1169.29 | 939.99  |
| M \geq 5.5    | 715.37  | 695.70  | 910.56  | 815.68  | 699.13  | 842.26  | 844.26  | 694.75  |
| M \geq 5.6    | 616.59  | 599.51  | 755.84  | 695.56  | 606.78  | 708.56  | 710.56  | 600.15  |
| M \geq 5.7    | 387.68  | 384.88  | 454.84  | 426.75  | 393.61  | 431.16  | 433.16  | 385.71  |
| M \geq 5.8    | 267.29  | 268.16  | 299.58  | 288.96  | 276.31  | 289.10  | 291.10  | 268.64  |
| M \geq 5.9    | 234.85  | 235.72  | 255.90  | 250.33  | 245.05  | 249.25  | 251.25  | 236.41  |
| M \geq 6.0    | 186.22  | 186.62  | 200.32  | 196.73  | 195.14  | 195.72  | 197.72  | 187.61  |
| M \geq 6.1    | 142.58  | 142.76  | 158.93  | 155.49  | 146.97  | 154.80  | 156.80  | 143.34  |
| M \geq 6.2    | 103.53  | 102.65  | 121.88  | 115.99  | 104.92  | 116.94  | 118.94  | 103.01  |
| M \geq 6.3    | 78.84   | 79.66   | 88.04   | 85.76   | 82.91   | 85.58   | 87.58   | 80.18   |

\*indicates the smallest AIC value

According to the selection criteria, it is known that there are four probability distributions that have the smallest AIC value among all magnitudes evaluated. Those probability distributions are Lognormal, Weibull, Gumbel, and Exponential. Therefore, the duration of recurring earthquake events among all magnitudes evaluated could be estimated to follow one of those four probability distributions. Based on information provided at Table 4, it is known that the recurring earthquake events for M \geq 5.0, M \geq 5.1, M \geq 5.2 and M \geq 5.3 are approximated to follow Lognormal distribution, the recurring earthquake events for M \geq 5.4 and M \geq 5.5 are approximated to follow the Weibull distribution, the recurring earthquake events for M \geq 5.6, M \geq 5.7 and M \geq 6.2 are approximated to follow the Gumbel distribution, while the rest follow the Exponential distribution.
Table 5. Parameter estimations of the selected probability distributions at all evaluated magnitudes

| Magnitude (M) | Probability Distribution Estimated | Parameter | μ  | σ  |
|---------------|-----------------------------------|-----------|----|----|
| M ≥ 5.0       | Lognormal, LOGNO (μ, σ)           |           | 1.166 | 0.2734 |
| M ≥ 5.1       | Lognormal, LOGNO (μ, σ)           |           | 1.380 | 0.3582 |
| M ≥ 5.2       | Lognormal, LOGNO (μ, σ)           |           | 1.541 | 0.4091 |
| M ≥ 5.3       | Lognormal, LOGNO (μ, σ)           |           | 1.761 | 0.4690 |
| M ≥ 5.4       | Weibull, WEI (μ, σ)               |           | 2.745 | -0.3967 |
| M ≥ 5.5       | Weibull, WEI (μ, σ)               |           | 2.950 | -0.3722 |
| M ≥ 5.6       | Gamma, GA (μ, σ)                  |           | 3.390 | 0.2840 |
| M ≥ 5.7       | Gamma, GA (μ, σ)                  |           | 3.591 | 0.1924 |
| M ≥ 5.8       | Exponential, EX (μ)               |           | 3.913 | -    |
| M ≥ 5.9       | Exponential, EX (μ)               |           | 4.062 | -    |
| M ≥ 6.0       | Exponential, EX (μ)               |           | 4.117 | -    |
| M ≥ 6.1       | Exponential, EX (μ)               |           | 4.021 | -    |
| M ≥ 6.2       | Gamma, GA (μ, σ)                  |           | 4.077 | 0.2909 |
| M ≥ 6.3       | Exponential, EX (μ)               |           | 3.802 | -    |

Once the parameters of those distributions are known, the p.d.f. and c.d.f. of the distributions are able to be estimated through the equations in Table 1. Furthermore, the hazard rate is also derivable from p.d.f. and c.d.f. due to its relation as described in Equation (1). The earthquake behavior could be observed through the hazard rate of the recurring earthquake duration. Along with the hazard definition, the hazard rate in this study implies the instantaneous rate of recurring earthquake events with a specific magnitude evaluated after not having any event with this magnitude within \( t \) days. From another perspective, the hazard rate indicates the number of failure events over a specific time period. In this study, the hazard rate shows the number of recurring earthquake events per day at the evaluated time \( t \) days. Figure 2 shows the comparison of hazard rates for all magnitudes evaluated. The X-axis represents the duration (days), while Y-axis represents the hazard rate.

As shown in Fig. 2, the recurring earthquakes for each magnitude evaluated have a different hazard rate. There are four main trends of the hazard rate behavior, those which increase until a certain time \( t \) and then decrease for the rest of the time, those which decrease over time \( t \), those which increase over time \( t \) and those which remain constant.
which are constant with time \( t \). This result implies that, in general, the hazard rate is dependent on time \( t \). In other words, the number of possible earthquake events per day is different within all times \( t \) for a specific magnitude evaluated. Note that the length of the hazard rates over the time \( t \) are different for each magnitude evaluated due to the fact that those magnitudes have different maximum time to happen as shown in Table 3. The longer time period, the longer the length of the hazard rate.

Specifically, the hazard rates for recurring earthquakes with magnitudes bigger than 5.0 but smaller than 5.4 would likely increase significantly until around time ten days after the initial earthquake and then decrease for the following time \( t \) to the maximum time that it could possibly happen, which also can be seen in Figure 2. Among those magnitudes, lower magnitudes tend to have a bigger hazard rate (i.e. would more likely to happen). Also, recurring earthquakes for \( M \geq 5.4 \) and \( M \geq 5.5 \) tend to have decreasing hazard rate trends. However, the rates are relatively small compared to the previous magnitudes evaluated. The following two bigger magnitudes (\( M \geq 5.6 \) and \( M \geq 5.7 \)) have contrast behavior which increases with time \( t \) at the maximum hazard rate of 0.20 for \( M \geq 5.6 \) and 0.008 for \( M \geq 5.7 \).

These results confirm that the earthquake phenomenon in the Sumatra territory could happen anytime with different hazard rates for magnitude under 5.8. While for recurring earthquakes with bigger magnitudes, hazard rates are constant for any point in time \( t \) except for \( M \geq 6.2 \).

These results imply that the hazard rate for these specific magnitudes under 5.8 are dependent on time, while bigger magnitudes are independent of time. Some studies reported time-dependent seismic hazard assessment in several regions [29–31], while some other researchers investigated both time-dependent and time-independent seismic hazard assessments in some areas like California [32] and Greece [33]. Furthermore, an earthquake with bigger magnitude tends to occur at a lower rate compared to earthquakes with smaller magnitudes. However, for a massive magnitude such as Sumatra-Andaman earthquake in 2004, it will take at least 300 to 400 years before an earthquake with similar magnitude will occur again [7].

### 3.4 Comparison analysis of duration and distance between recurring earthquake events

In this section, the duration and distance between recurring earthquake events for all magnitudes evaluated are compared by using Analysis of Variance (Anova) or Kruskal-Wallis test when applicable. The first step in this analysis is to check whether or not the data set meets the normality assumption by using the Shapiro-Wilk test and homogeneity variances by Levene’s test. Shapiro-Wilk test is generally more powerful than the Kolmogorov-Smirnov test [34,35]. Shapiro-Wilk test can be used when the sample size is between 3 and 5,000 [36]. The null hypothesis in the test states that the data set follows normal distribution. In this study, the null hypothesis will likely reject when the p-value is smaller than 0.10. It is important to note that, from the previous section we know that the duration data set for all magnitudes evaluated are not normally distributed. Hence, the Shapiro-Wilk test is only employed for the distance data set. The results of the Shapiro-Wilk test of distance data set for all magnitudes evaluated are presented in Table 6.

Table 6. Shapiro-Wilk test for distance between the recurring earthquake events at all evaluated magnitudes

| Magnitude (M) | Distance (km) | Statistics | p-value | Decision |
|---------------|--------------|------------|---------|----------|
| \( M \geq 5.0 \) | 0.921 | 0.001* | Not Normal |
| \( M \geq 5.1 \) | 0.924 | 0.001* | Not Normal |
| \( M \geq 5.2 \) | 0.933 | 0.001* | Not Normal |
| \( M \geq 5.3 \) | 0.917 | 0.001* | Not Normal |
| \( M \geq 5.4 \) | 0.908 | <0.001* | Not Normal |
| \( M \geq 5.5 \) | 0.932 | <0.001* | Not Normal |
| \( M \geq 5.6 \) | 0.947 | 0.005* | Not Normal |
| \( M \geq 5.7 \) | 0.946 | 0.047* | Not Normal |
| \( M \geq 5.8 \) | 0.919 | 0.038* | Not Normal |
| \( M \geq 5.9 \) | 0.919 | 0.065* | Not Normal |
| \( M \geq 6.0 \) | 0.849 | 0.008* | Not Normal |
| \( M \geq 6.1 \) | 0.842 | 0.017* | Not Normal |
| \( M \geq 6.2 \) | 0.862 | 0.081* | Not Normal |
| \( M \geq 6.3 \) | 0.821 | 0.048* | Not Normal |

* indicates the test is significant.

Similar to duration data in Table 4, Table 6 shows that all distance data sets do not meet the normality assumption due to the p-value of the Shapiro-Wilk test that is less than 0.10. Moreover, Levene’s test is used to check whether the variances of \( k \) samples are equal. Levene’s test is powerful and robust to non-normality and has become a very popular tool for checking the homogeneity of variances [37]. The null hypothesis of Levene’s test states that the variances of \( k \) samples are equal (homogenous) and will likely reject when the p-value is smaller than 0.10. The Levene statistics for duration and distance data set are 1.778 and 52.649 respectively. The null hypothesis for both variables is rejected due to small p-values. The p-values for duration...
and distance are 0.041 and <0.001 respectively. Those values are smaller than 0.10 which means that the variances of duration and distance data on 14 magnitudes evaluated are not equal.

Accordingly, the nonparametric Kruskal-Wallis test is used as an alternative method. Kruskal-Wallis test is a preferred procedure in many situations to compare more than two independent samples [38]. The Kruskal-Wallis test can be used as an alternate method to Anova [39]. In this case, the null hypothesis of the test states that all magnitudes evaluated have a similar mean rank instead of a mean value. Rejecting the null hypothesis means that there are at least two magnitudes evaluated that have different mean rank in duration or distance data. In this study, the null hypothesis will likely be rejected when the p-value is smaller than 0.10. Table 7 presents the result of the Kruskal-Wallis test.

| Statistics | Duration time | Distance |
|------------|---------------|----------|
| $H$        | 193.244       | 7.763    |
| p-value    | <0.001*       | 0.933    |

* indicates the test is significant

According to Table 7, only duration data has a different mean rank between each magnitude evaluated due to a small p-value (<0.001). While for distance data, there is no statistical evidence that different magnitudes evaluated are not equal.

| Table 7: Kruskal-Wallis test for duration and distance between recurring earthquake events |

This matrix is symmetric due to the fact that comparison for example between M ≥ 5.0 and M ≥ 5.5 is similar to comparison between M ≥ 5.5 and M ≥ 5.0, and so forth. The diagonal does not have a p-value since the Mann-Whitney U test only compares two independent groups. Table 8 shows that the comparison between the magnitudes of up to M ≥ 5.9 each other will conclude a similar decision (they have different median), except for sequential magnitude comparison such as between M ≥ 5.9 and M ≥ 5.8 or between M ≥ 5.8 and M ≥ 5.7. Those pairs with small p-values have a different median of duration. Additionally, the comparison between any magnitudes lower than M ≥ 5.5 (M ≥ 5.0, M ≥ 5.1, ..., M ≥ 5.4) and the rest (M ≥ 5.5, M ≥ 5.6, ..., M ≥ 6.3) are also significant as marked with sign (*). In contrast, the comparison between recurring earthquakes with bigger magnitudes (M ≥ 6.0) are not significant due to larger p-values (>0.10). This finding proves that
recurring earthquakes with bigger magnitude (M ≥ 6.0) do not necessarily happen for a different length of time. As presented in Table 3, the median duration between recurring earthquakes with lower magnitudes have an increasing trend, which supports the Mann-Whitney U test. Therefore, it can be concluded that the recurring earthquakes for lower magnitude (under 6.0) will happen for a different duration. The lower the magnitude, the faster the recurring earthquake event will happen, while recurring earthquakes for bigger magnitudes (M ≥ 6.0) tend to happen for similar durations.

4 Conclusions

In the present study, the earthquake behavior in the Sumatra territory was assessed based on the duration and distance between an earthquake and the next (later) earthquake event (termed as recurring earthquake) which had a magnitude greater than or equal to the previous one. The survival analysis approach revealed that recurring earthquakes with specific magnitude evaluated have different hazard rates following their probability distributions. Earthquakes with magnitude under 5.8 could happen anytime with different rates, while the hazard rates are constant for larger magnitudes except for those events with M ≥ 6.2. Furthermore, the duration as well as the distance between the recurring earthquakes at any magnitude evaluated are compared with each other and it was found that the recurring earthquake events tend to happen at a similar average distance for any magnitude. However, the duration between recurring earthquake events is different particularly for lower magnitude under 6.0. This research has significant implications for earthquake researchers and the Indonesian government in understanding the phenomenon of earthquakes in Sumatra territory.

References

1. D. H. Natawidjaja, Gempahumi Dan Tsunami Di Sumatra Dan Upaya Untuk Mengembangkan Lingkungan Hidup Yang Aman Dari Bencana Alam (Jakarta, 2007)
2. R. Paris, F. Lavigne, P. Wassmer, and J. Sartoched, Mar. Geol. 238, 93 (2007)
3. J. Jatnika, A. D. Nugraha, and Wandono, J. Meteorol. Dan Geofis., 116, 113 (2016)
4. L. D. Rifai and I. P. Pudja, J. Meteorol. Dan Geofis. 11, 147 (2010)
5. S. Anwar, J. Lingkung. Dan Bencana Geol. 12, 33 (2021)
6. T. Lay, H. Kanamori, C. J. Ammon, M. Nettles, S. N. Ward, R. C. Aster, S. L. Beck, S. L. Bilek, M. R. Brudzinski, R. Butler, H. R. DeShon, G. Ekström, K. Satake, and S. Sipkin, Science (80-. ). 308, 1127 (2005)
7. F. Løvholt, H. Bungum, C. B. Harbitz, S. Glismland, C. D. Lindholm, and G. Pedersen, Nat. Hazards Earth Syst. Sci. 6, 979 (2006)
8. K. Satake and B. F. Atwater, Annu. Rev. Earth Planet. Sci. 35, 349 (2007)
9. S. Stein and E. A. Okal, Nature 434, 581 (2005)
10. C. Subarya, M. Chlieh, L. Prawirodirjio, J.-P. Avouac, Y. Bock, K. Sieh, A. J. Meltzner, D. H. Natawidjaja, and R. McCaffrey, Nature 440, 46 (2006)
11. R. Paris, M. Cachão, J. Fournier, and O. Voldoire, Géomorphologie Reli. Process. Environ. 16, 109 (2010)
12. NSF, (2005)
13. M. Walton, (2005)
14. S. Anwar, J. Lingkung. Dan Bencana Geol. 10, 159 (2019)
15. E. Lusiani, S. Anwar, and M. F. Nugraha, J. Meteorol. Dan Geofis. 19, 71 (2018)
16. J. J. Clague, C. Yorath, R. Franklin, and B. Turner, At Risk: Earthquakes and Tsunamis on the West Coast (Tricouni Press, Vancouver, 2006)
17. Á. González, J. B. Gómez, and A. F. Pacheco, J. Seismol. 10, 131 (2006)
18. D. Shanker and A. Harbindu, in 13 Th World Conf. Earthq. Eng. (Vancouver, B.C., Canada, 2004), p. Paper No. 3024
19. S. Zacks, Introduction to Reliability Analysis : Probability Models and Statistical Methods (Springer, New York, 1992)
20. R. Jiang, Introduction to Quality and Reliability Engineering (Springer, Berlin, Heidelberg, 2015)
21. H. Akaike, IEEE Trans. Autom. Control 19, 716 (1974)
22. R. Rigby, M. Stasinopoulos, G. Heller, and F. De Bastiani, Distributions for Modelling Location, Scale and Shape: Using GAMLSS in R (2017)
23. D. C. Montgomery and G. C. Runger, Applied Statistics and Probability for Engineers (John Wiley & Sons, Inc., Hoboken, New Jersey, 2003)
24. N. Nachar, Tutor. Quant. Methods Psychol. 4, 13 (2008)
25. J. Jatnika, A. D. Nugraha, and W. Wandono, J. Meteorol. Dan Geofis. 16, (2016)
26. A. J. Michael, Bull. Seismol. Soc. Am. 95, 1594 (2005)
27. B. C. Papazachos and C. A. Papaioannou, Pure Appl. Geophys. PAGEOPH 140, 593 (1993)
28. S. P. Nishenko and R. Buland, Bull. Seismol. Soc. Am. 77, 1382 (1987)
29. C.-H. Chan, Y.-M. Wu, C.-T. Cheng, P.-S. Lin, and Y.-C. Wu, Nat. Hazards Earth Syst. Sci. 13, 1143 (2013)
30. G. F. Panza, A. Peresan, A. Magrin, F. Vaccari, R. Sabadini, B. Crippa, A. M. Marotta, R. Spedondere, R. Barzaghi, A. Borghi, L. Cannizzaro, A. Amadio, and S. Zoffoli, Nat. Hazards 69, 1179 (2013)
31. G. P. Biassi and S. C. Thompson, Bull. Seismol. Soc. Am. 108, 39 (2018)
32. M. D. Petersen, T. Cao, K. W. Campbell, and A. D. Frankel, Seismol. Res. Lett. 78, 99 (2007)
33. C. A. Papaioannou and B. C. Papazachos, Bull. Seismol. Soc. Am. 90, 22 (2000)
34. Z. Hanusz, J. Tarasinska, and W. Zielinski, REVSTAT-Statistical J. 14, 89 (2016)
35. N. M. Razali and Y. B. Wah, J. Stat. Model. Anal. 2, 21 (2011)
36. P. Royston, J. R. Stat. Soc. Ser. C (Applied Stat. 44, 547 (1995)
37. J. L. Gastwirth, Y. R. Gel, and W. Miao, Stat. Sci. 24, 343 (2009)
38. A. Vargha and H. D. Delaney, J. Educ. Behav. Stat. 23, 170 (1998)
39. R. R. Pagano, Understanding Statistics in the Behavioral Sciences, Fourth Edi (WestPublishing Company, Eagan, Minnesota, 1994)
40. S. Ross, A First Course in Probability, Eighth Edi (Prentice Hall, Inc., New Jersey, 2010)