The application of R and ArcGIS software as a learning media to estimate the return period of the destructive earthquake in Aceh areas using maximum likelihood method

H Sofyan, Jumila and L Rahayu
Syiah Kuala University, Jl. Teuku Nyak Arief Darussalam, Banda Aceh 23111, Indonesia
E-mail: hizir@unsyiah.ac.id

Abstract. Aceh is one of the most vulnerable areas for earthquakes. The earthquakes occurred has caused much loss including lives. Therefore, earthquake occurrence time estimation is needed to mitigate preparedness to the earthquake disaster. The purpose of this research was to explain the measurement result of destructive earthquake repetitive period in Aceh and other surrounding areas using Likelihood Maximum method toward the utilization of R and ArcGIS software instructional media. The data was obtained on trial utilization of R and ArcGIS software on Spatial Statistics course. The participants are 18 students taking the spatial statistics class. The research instrument was the observation sheet used to evaluate the implementation of student worksheets in the class. The data sources the students used for the measurement were obtained from USGS and BMKG. The results found that the students were able to predict the earthquake correctly using student worksheet assisted by the R and ArcGIS software. That is the earthquake would occur in the next 5.2 years at the marine territory of the southern region of Aceh Singkil. Based on this result, it can be concluded that the student worksheet could be used in the spatial statistics class assisted by the R and ArcGIS software.

1. Introduction

The province of Aceh is a region that coincides with the meeting of the Eurasian plate with the Indo-Australian plate along the Indian Ocean. In this zone, there have been a number of major earthquakes that have caused severe damage in some areas around the Indian Ocean; one of which occurred in December 26, 2004 that had triggered the tsunami with a magnitude of 9.1 and the earthquake that occurred in Pidie Jaya on December 7, 2016 with a magnitude of 6.5.

Research on seismotectonic potential as a precursor of earthquake level in Aceh has analyzed the level of seismicity and the period of earthquake re-occurring in Aceh region using the relationship between the frequencies with earthquake magnitude using the Maximum Likelihood method with ZMAP software. The study concludes that between two to five years there would be an earthquake for the 6 Mw scale in Simeulue Island, Northwest Aceh and North Aceh (Andaman) [1]. The results of this research can be used as the context of the problem in Spatial Statistics course. Students are assigned to look for relevant data sourced from USGS and BMKG, then, analyze the period of the destructive earthquake in Aceh and surrounding areas using the Maximum Likelihood method with a learning media, the R and ArcGIS software.

Maximum Likelihood aims to estimate statistics of random variables by maximizing probabilities with existing samples. This method is often used in different applications and studies in various
situations [2]. In this study this method is used to analyze the relationship of frequency (N) and magnitude (M) earthquake which is known by Gutenberg Richter [3].

\[ \log N(M) = a - bM \]  

Where:
- \( N \) = the number of earthquakes
- \( a \) = a determination that depends on the period, total area and area of observation activity
- \( b \) = seismotectonic parameter of an area in which in an earthquake occur and depending on the local rock.
- \( M \) = magnitude

The value of the tectonic parameters in the research area can be seen by looking at \( b \)-value. The \( b \)-value can show the stress level of the rock. If \( b \)-value is low, the stress condition in the area is high [4]. \( b \)-value is said to be low if \( b \)-value has a smaller value, the average value of \( b \)-value in the entire study area [5]. The \( b \)-value approximation is correct if the standard deviation of \( b \)-value is less than one [6].

The equation for \( b \)-value is presented as follows [7].

\[ b = \frac{\log e}{M - M_0} \]  

Where:
- \( M \) = Average magnitude
- \( M_0 \) = Minimum magnitude
- \( \log e = \log 2.71828 = 0.434 \)

The error rate or standard deviation of the \( b \)-value calculation with the magnitude can be estimated using the equation in (3).

\[ \sigma_b = 2.30b^2 \sqrt{\frac{\sum_{i=1}^{N} (M_i - \mu)^2}{N(N-1)}} \]  

In addition, there is an \( a \)-value that shows the level of seismicity in a region. The smaller the \( a \)-value, the smaller the seismic activity [8]. The equation for \( a \)-value is shown in [1].

\[ a = \log N(M \geq M_0) + \log(b \ln 10) + M_0b \]  

Thus, to obtain the average value of earthquake return periods in a region can be calculated using the following equation.

\[ P(\theta) = \frac{1}{N_i(M \geq M_0)} \]  

Where \( \theta \) is the period of repeating earthquake for magnitude \( M \geq M_0 \) [1].

The probability of earthquakes occurring is the possibility of destructive earthquakes in one area over a period of time. Assuming the time interval follows the exponential form \( e^{-NT} \), the probability of earthquake occurrence is provided as follows.

\[ P(M, T) = 1 - e^{-N_i(M \geq M_0)T} \]  

Where \( P(M, T) \) is the probability of an earthquake with magnitude \( M \) and period \( T \).
The condition of Aceh that is often stricken by an earthquake disaster can be used as the context of the problems to be solved by students in spatial statistics class that is one of the courses in Statistics Department of Mathematics and Science Faculty at Syiah Kuala University. To analyze the calculation of earthquake return period required a learning media that can facilitate the students. Learning media is a tool in teaching process that is a physical means to deliver learning content/material [9]. Learning media is also a means of communication in the form of print and hearing-sight, including hardware technology [10].

Over time, there are many learning media in the form of software that students can use to solve various problems that occur in the universe, such as QGIS that has been used as Development of Web GIS for information of renewable energy in Aceh province after the rehabilitation and reconstruction process [11]. Besides, Xplore, GRASS, and ERDAS have also been used to classify satellite imagery data using the fuzzy clustering method [12].

Learning media used in this research were R and ArcGIS software. R is one of the languages in statistical programming based on the S programming language developed by Bell Laboraties. R was first introduced in 1996 by two statistics professors, Ross Ihaka and Robert Gentlemen, from the University of Auckland in New Zealand. They released the code as a free software package. ArcGIS is the most widely used software by users of Geographic Information System (GIS). One of the packages in ArcGIS is ArcMap, which is the main application for most GIS process and computer mapping. ArcMap has major capabilities for visualization, developing new spatial databases, selecting (querying), editing, creating map designs, analyzing and creating the final look in activity reports. In this study, ArcMap is used to visualize the results obtained from calculations with the R software into map form.

The spatial statistic learning with the utilization of R and ArcGIS software could maximize by using student worksheet that guides students in predicting the return period of an earthquake. Thus, this study aims to investigate the implementation of the worksheet assisted by R and ArcGIS software in spatial statistic class.

2. Method
This research generally aims to develop learning tools in spatial statistics class on Statistic Study Program of Mathematics and Natural Sciences Faculty that train students to solve problems. The learning tools were developed by following the research phases of controlling the Plomp Model (1999). This article focuses on instructional media used to help solve the given problem of "estimate return period earthquake." The learning media used is R and ArcGIS software. The explanation given is limited to the results obtained by the students in analyzing the repeated periods of earthquakes in Aceh and surrounding areas by the Maximum Likelihood method based.

The data used were earthquake occurrences from January 1907 to December 2017. The data obtained from the Agency of Meteorology, Climatology, and Geophysics (BMKG) were the data in 2009 while the data in 1907 to 2008 were obtained from the catalog of earthquakes in the United States Geological Survey (USGS). The selected data were earthquake data located on longitude 92° - 98.5° East Longitude and latitude 0° - 8° North latitude which coincide in Aceh province and at the maximum depth of 80 km. Research area is divided into 1° x 1° grid to obtain 48 grids with earthquake incidences as much as 9202.

The result of students’ calculation when solving the worksheet based on the learning materials used was collected. Trials were conducted to measure the practicality of the learning materials. The implementation of the worksheet with R and ArcGIS software in this trial was examined based on students’ calculation. Students are expected to predict the repeated period of earthquake correctly, which is 5.2 years. If the students get the expected result, it can be said that the worksheet can be implemented in the spatial statistic class.
3. Result and discussion

At the beginning of the lesson, students were instructed to learn information about earthquakes through BMKG and USGS website. The detailed instruction is written on the worksheet. The data obtained were earthquake data from The USGS and BMKG earthquake catalog, and the largest magnitude earthquake ever recorded in that period was in 2004 and 2012 in Simeulue. Earthquakes continue to occur from 1907 to 2017 with the cumulative incidence annually seen in Figure 1.

![Figure 1. Cumulative frequency of earthquake events.](image1)

Figure 1 is a graph visualized by the students using R software. The graph shows that the earthquake increased significantly after 2000, where there was a major earthquake that caused the tsunami in 2004, and there had been many subsequent earthquakes.

Furthermore, the students created illustrations of earthquake distribution in Aceh and surrounding areas in 1907 to 2017, presented in Figure 2, visualized students with ArcGIS software using data obtained from BMKG and USGS.

![Figure 2. Map of spreading earthquakes.](image2)

![Figure 3. Map of b-value spatial variation.](image3)
The red dots in Figure 2 represent 4750 points of shallow earthquakes in Aceh and surrounding areas in 1907 to 2017 with a minimum magnitude of 5 amounting to 4750 events.

The next learning activity was b-value analysis to know the tectonic condition of the research area. The spatial b-value analysis conducted by the students in the research area with the number of earthquakes of 4750 events obtained the b-value value of 0.720 and the standard deviation obtained for 0.316 which is smaller than 1. This finding suggests that the calculation of b-value was close to the true value.

Furthermore, the students calculated spatial variation analysis of b-value. Spatial variation of b-value was determined by dividing the research area located at 92° – 98° East Longitude and 0° – 8° West Longitude into 832 parts with grid 1° × 1°. The result of the analysis shows that b-value ranges from 0.208 – 1.717 with average 0.705. 25 grids that had a low b-value is a region with b-value smaller than the average. In addition, there were 19 grids that had a high b-value, i.e., an area with b-value more than the average. Furthermore, 6 grids could not be analyzed because of the absence of earthquake data. However, for mapping, the area with an earthquake frequency of more than 10 was used, resulting in 34 areas with an average b-value of 0.7066. The variation of b-value can be seen in Figure 3.

Figure 3 is a contour map visualized by students with ArcGIS software, which shows that the blue region is a low b-value area. The region had high levels of stress and rare earthquakes, but in case of an earthquake, there will be earthquakes with a large magnitude. This is due to the buildup of energy in the region and when energy was released; it resulted in large-scale earthquakes. The red area is a region with high b-value which indicates that earthquakes were frequent in the area, with small magnitude due to the earthquake immediate energy released. The level of accuracy of the calculation can be seen based on the standard deviation value of b-value. Overall the standard deviation of b-value was less than 1, which means the calculation of b-value was close to the true value.

After the spatial analysis of b-value, the students conducted a temporal analysis of b-value. The temporal b-value analysis was performed by distributing earthquake occurrence time. The data used were the data from the BMKG earthquake catalog. Data grouping was obtained by combining the earthquake data on a quarterly basis each year. The first quarter starts from January to March, the second quarter from April to June, the third quarter from July to September, and the fourth quarter from October to December. The purpose of the temporal variation analysis of b-value was to look at random signs before a major earthquake. The values obtained for temporal variation b-value is visualized in Figure 4.

![Figure 4. Temporal variation of b-value.](image)

Figure 4 illustrates that temporarily in the event of a b-value earthquake decreased until a major earthquake occurred. This can be seen in 2012 and 2016 where there had been a major earthquake in
Simeulue and Pidie Jaya. In 2011, b-value increased after the earthquake in Simeulue in 2010. Also, b-value increased in 2017 after the earthquake in Pidie Jaya at the end of 2016.

After obtaining the value of b-value, the students calculated the return period of earthquakes using R software and mapped them using ArcGIS software. The return period of earthquakes in Aceh and surrounding areas with $M_w \geq 5$ is illustrated in the following contour maps.

![Earthquake Return Period](image)

**Figure 5.** Map of contours of return periods of earthquakes.

Figure 5 shows that red areas are areas with return periods of earthquakes in the shortest period of five to ten years when the occurrence of repeated periods of earthquakes continued to increase every change in red, yellow, green to blue. While the white color indicates the area that there is an earthquake frequency of less than 10; thus, no mapping was obtained. The period of the earthquake can be observed in Table 1.

| Grid | Period (years) | Year estimates of return period | Occurrence probability |
|------|----------------|---------------------------------|------------------------|
| 12   | 5.2            | 2023                            | 0.632                  |
| 17   | 6.3            | 2024                            | 0.633                  |
| 33   | 8.2            | 2026                            | 0.633                  |

Table 1 shows that the most rapidly estimated earthquake will occur in the 12th grid that coincides in the sea area of Southern area of Aceh Singkil with a repeat period of 5.2 years, in the 12th grid that coincides in Simeulue with a repeat period of 6.3 years, and in the 33rd grid that coincides in sea areas of Northern area of Aceh of with a return period of approximately 8.2 years. The probability of earthquake occurrences in the grid to 12, 15 and 33 is about 63%. The estimated year the earthquake repeated period was calculated after 2017 where the last earthquake event data were used, so earthquakes are expected to occur in the 12th grid in 2022, the 17th grid in 2023, and the 33rd grid by 2026.
As the validity of return period analysis, an analysis of the probability of earthquake occurrence in the near-estimated time, earthquakes will occur again in the same areas. The probability of earthquake occurrence can be seen in Table 2, which is the result of student calculation using the R software.

Table 2. The probability of the occurrence of destructive earthquakes.

| Grid | Probability 5 years | Probability 10 years | Probability 30 years | Probability 50 years |
|------|----------------------|-----------------------|-----------------------|----------------------|
| 12   | 0.61809              | 0.85414               | 0.9969                | 0.99993              |
| 17   | 0.55073              | 0.79815               | 0.99178               | 0.99967              |
| 33   | 0.45656              | 0.70467               | 0.97424               | 0.99775              |

Table 2 shows that the probability of an earthquake re-occurrence increased by increasing time. In the 12th grid, it is estimated that the probability of an earthquake is 61.8% in the next five years and 85.4% in the next ten years. The results are in accordance with the estimated return period of the earthquake that occurred within 5.2 years with the probability of occurrence of 63.2% (Table 1).

In addition, Table 2 shows that the highest probability of earthquakes with an estimated 12th grid within 30 years will occur with a probability of 99.69% and within 100 years with a probability of 100%.

Based on the students’ answer, it can be concluded that the students successfully predicted the return period of earthquakes as expected. This finding implies that the worksheet with the utilization of R and ArcGIS software could help students achieve the learning objective. In another word, the worksheet utilized by R and ArcGIS software could be implemented in spatial statistics class.

4. Conclusion

Based on the estimation of return period of the destructive earthquake in Aceh area using the maximum likelihood method, the most rapid destructive earthquake is expected to occur in the sea area of Southern of Aceh Singkil in 5.2 years, which means that the earthquake is expected to occur in 2023. This estimation was obtained by students through the worksheet assisted by the R and ArcGIS software. It can be concluded that the R and ArcGIS software are ones of learning media for students to solve the case of spatial statistics. The use of the software needs to be assisted by the worksheet. The worksheet developed in this study meets the criteria of practicality in the learning on spatial statistics class.

References

[1] Tauladani R 2015 Kajian Seismisitas dan Periode Ulang Gempa Bumi di Aceh (Skripsi)
[2] Wang J P, Xu Y and Wu Y M 2017 Earthquake probability in Taipei based on non-local model with limited local observation: Maximum likelihood estimation Soil Dynamics and Earthquake Engineering 99 150
[3] Gutenberg B and Richter C F 1936 Magnitude and energy of earthquakes Science 83 183
[4] Nuanmin P Kuhlânek O and Persson L 2012 Variations of b-values preceding large earthquakes in the Andaman-Sumatra subduction zone J Asian Earth Sci 61 237
[5] Kuhlânek O Nuanmin P and Persson L 2018 Variation of b-values Proceeding Large Earthquake in The Shallow Subduction Zones of Cocos and Nazca Plates South Am Earth Sci 82 207
[6] Happrobo S Variasi b-value terhadap Ruang dan Waktu sebagai Prekursor Gempa Bumi Daerah Andaman Sumatera [Skripsi] 2016
[7] Han Q, Wang L, Xu J, Carpinteri A and Lacidogna G 2015 A robust method to estimate the b-value of the magnitude-frequency distribution of earthquakes Chaos, Solitons and Fractals 81 103
[8] Kanamori H 1981 The nature of seismic patterns before large earthquakes. In earthquake prediction: An international review Maurice Ewing Ser 1981 4 1
[9] John B B, Kevin F and Collis 1982 *Evaluating the Quality of Learning: the solo Taxonomy* (New York: Academic Press)
[10] National Education Association 1969 *Audiovisual Instruction Department New Media and College Teaching* (Washington D C: NEA)
[11] Nizamuddin, Hizir, Ardiaysyah, Pertiwi D and Handayani P 2017 *Development of Web GIS for information of renewable energy in Aceh Province after rehabilitation and reconstruction process* In: Earth and Environmental Science
[12] Sofyan H, Said MAM, Affan M and Bawahidi K 2014 *The Application of Fuzzy Clustering to Satellite Images Data* In: Remote Sensing. Venice