Identifying Dynamic Changes in Megathrust Segmentation via Poisson Mixture Model

J Rizal¹, A Y Gunawan¹, S W Indratno² and I Meilano³

¹ Industrial and Financial Mathematics Research Group, Faculty of Mathematics and Natural Sciences, Institut Teknologi Bandung, Indonesia.
² Statistics Research Group, Faculty of Mathematics and Natural Sciences, Institut Teknologi Bandung, Indonesia
³ Geodesy Research Group, Faculty of Earth Science and Technology, Institut Teknologi Bandung, Indonesia

jose_rizal201@s.itb.ac.id

Abstract. Analysis of dynamic changes is one of the important information in earthquake prediction models. The earthquake frequency occurrence is often taken as data to identify the dynamic changes in large earthquake sources. The commonly used probability distribution for the counting process is the Poisson distribution. However, overdispersion relative to the Poisson distribution is often found when its average value is less than variance. The purpose of this research is to obtain an overview of the dynamic change mechanisms at large earthquake source in megathrust segment, due to possible overdispersion case. One method that can be used to solve this problem is a Poisson mixture model. This method is designed to identify unobserved heterogeneity or dynamics in the data. Here, we focus on the dynamic changes in the large earthquake sources in the Sunda megathrust segmentation on the island of Sumatra. Our results show that the Poisson mixture model is able to provide a description of the dynamic changes in that area. Especially for that area, the level of dynamic change can be classified into seven categories. For five large earthquake sources in that area, various categories of dynamic changes are found to be different from two until six categories.

1. Introduction

Indonesia as one of the world's most frequent large earthquakes has a source map and earthquake hazards. However, there is no probabilistic model of seismic hazard in large earthquake sources in megathrust segmentation [1]. Some approaches that can be used to model earthquake hazards are ground motion, rock stretching, and dynamic changes in tectonic plates. The analyzed ground motion using Seismic Hazard Analysis (SHA) method to measure earthquake hazard level can be seen in [2,3,4]. However, for dynamic change, only a few researchers are studying it, since the dynamic changes occurring beneath the surface of the earth is very difficult to observe directly. So far there is no technology that can provide an overview of the dynamic changes that occur in tectonic plates, especially in the seismic megathrust segments. The dynamic changes can be justified from the stress level or the strength of the electromagnetic signal [5].

One approach to get a description of tectonic plate dynamics is to estimate the probability function from the number of earthquake events to a finite dimension function. So, in this study, a linear combination of probability functions from the counting process will be determined. Poisson
distribution as a probability function of a discrete random variable counting process is used in approximation. A discrete random variable \( Y \) is said to have a Poisson distribution with parameter \( \lambda > 0 \), if for \( i = 0,1,2,... \) the density function of \( Y \) is given by the following equation and the property that the variance equals the mean.

\[
P(Y = i) = p(y_i; \lambda) = e^{-\lambda} \frac{\lambda^i}{i!}
\]

(1)

However, we sometimes find a condition of overdispersion relative to the Poisson distribution, that the value of the sample variance \( (s_y^2) \) is much greater than the average value of sample \( (\bar{y}) \). It is possible due to the result of the counting process comes from several different group characteristics.

2. Earthquake Data

Our research area is the Sunda megathrust segmentation on the island of Sumatra, as can be seen in figure 1. The selection of this area was based on the results of a study of the potential for a large 8.8 Mw earthquake around the Mentawai island in western Sumatra [6,7]. Similarly, the results of paleoseismic and geodetic analyzed indicated that large earthquakes (8.8 Mw) and tsunamis are highly likely to occur in the future to the northern of Mentawai island [8,9].

Suppose \( \{X_t, t \in T\} \) random variable denotes the number of an earthquake with magnitudes \( M \geq 5 \) Mw occurring in the Sunda megathrust, Sumatra within a year. The observation period was selected from 1973 to 2017. The data source was taken from the catalogs of United States Geological Survey. Table 1. Description of large earthquake sources; coordinate, maximum magnitude, and mean-variance number of the earthquake (1973-2017)

| Large Earthquake Sources | Coordinate | Maximum Magnitude (Mw) | Number of Earthquakes |
|--------------------------|------------|------------------------|-----------------------|
| Latitude | Longitude | Mean | Variance |
| I – Aceh Andaman | [3.119,13.454] | [90.264, 96.452] | 9.2 | 19.711 | 329.560 |
| II – Nias-Simelue | [0.066, 3.119] | [94.856, 98.416] | 8.9 | 9.600 | 239.973 |
| III – Mentawai-Siberut | [-2.197, 0.666] | [96.504,100.261] | 8.7 | 3.956 | 18.362 |
| IV – Mentawai-Pagai | [-4.430,-2.197] | [98.503,101.646] | 8.9 | 5.422 | 74.659 |
| V – Enggano | [-7.101,-4.430] | [100.393,105.139] | 8.8 | 14.978 | 39.800 |
| Sunda megathrust, Sumatra | [-7.101,13.454] | [90.264, 105.139] | 9.2 | 53.667 | 1546.864 |

Based on the history of major earthquake events, earthquake researchers incorporated in PusGeN have mapped five major earthquake sources in the Sunda megathrust, Sumatra [1]. Some of the recorded earthquakes ever happened in the Sunda megathrust are the Great Sumatra-Andaman earthquake with magnitude 9.2 mw that has occurred in the segmentation I, the earthquake Nias-Simeulue March 28, 2005 with magnitude 8.7 Mw, the earthquake of 1797 with magnitude 8.7 Mw and 1833 with magnitude 8.9 Mw which occurred in Mentawai-Siberut, the Mentawai-Pagai earthquake occurred in 2007 and 2010 with magnitude of 8.5 Mw and 7.8 Mw respectively , and the earthquake that occurred in Bengkulu on September 12, 2007 (8.4 Mw) [10,11].

Table 1 shows that there is variation in maximum magnitude value, area, and earthquake occurrence in each major earthquake source (fact 1). Figure 2 shows that each major earthquake source has a histogram number of an earthquake with different skewness and deviation (fact 2). Thus, although from one subduction zone, each major earthquake source is indicated to have a different dynamic change. The ratio between variance and average for each large earthquake sources is more than 1, that is, there is a condition of overdispersion relative to Poisson (fact 3). From these facts, the Poisson mixture model can be applied to identify the dynamic changes analyzed from the number of earthquakes in the segment.
3. Poisson Mixture Model

The Poisson distribution is a probability distribution of discrete random variables that defines the number of events that occur within a given period of time. The easily recognizable parameters of the Poisson distribution are the sample mean and the sample variance has the same value. However, overdispersion relative to the Poisson distribution is often found, that is, the sample variance is larger than the sample mean. Suppose that probability density function of a random variable $X$ for some probabilities $0 \leq \delta_j \leq 1, j = 1, 2, ..., m,$ can be expressed in the form

$$P(X = i) = \sum_{j=1}^{m} \delta_j p_i(x_j, \lambda_j), i = 0, 1, 2, ...$$

(2)

with $\sum_{j=1}^{m} \delta_j = 1$, then $X$ is said to have a $m$-finite mixture density [12], where $\lambda_j$ is the accident rate in the $j$ th component of the mixture and $\delta_j$ is the mixing parameters. To specify the component, one needs a discrete random variable $C$ which performs the mixing:

$$C = \begin{cases} 1 & \text{with probability } \delta_1 \\ \vdots & \vdots \\ m & \text{with probability } \delta_m \end{cases}$$

(3)

Using equation (3), we can write equation (2) in the multiplication of conditional probabilities and marginal probabilities, as follows: [12]

$$P(X = i) = \sum_{j=1}^{m} P(X = i | C = j) P(C = j), i = 0, 1, 2, ...$$

(4)

The expectation of the mixture can be given in terms of the expectations of the component distributions. Using equation (1) we have,

$$E(X) = \sum_{j=1}^{m} P(C = j) E(X | C = j) = \sum_{j=1}^{m} \delta_j E(Y_j)$$

(5)

More generally, for mixture the $k$ th moment about the origin is simply a linear combination of the $k$ th moments of its components $Y_j$ :

$$E(X^k) = \sum_{j=1}^{m} \delta_j E(Y_j^k) \quad k = 1, 2, 3, ...$$

(6)
Using the properties of $Var(X) = E(X^2) - (E(X))^2$, equations (5) and equation (6) for $k = 2$, we get the variance of the mixture as follows:

\[ Var(X) = \sum_{j=1}^{m} \delta_j \left( Var(Y_j) + (E(Y_j))^2 \right) - \left( \sum_{j=1}^{m} \delta_j E(Y_j) \right)^2 \]

The estimation of parameters of a mixture distribution is often performed by maximum likelihood. In general, the likelihood of a mixture model with $m$ components with $n$ observations is given [12],

\[ L(\lambda_1, \lambda_2, \ldots, \lambda_m, \delta_1, \delta_2, \ldots, \delta_m | x_1, x_2, \ldots, x_m) = \prod_{i=1}^{n} \sum_{j=1}^{m} \delta_j p_i(x_j, \lambda_j) \]

The maximum likelihood estimate associated with a sample of observations is a choice of parameters which maximizes the probability density function of the sample, that is, the likelihood function. Maximizing the likelihood function of equation (8) is difficult to resolve analytically because it involves the $2m - 1$ parameter to be estimated. Therefore, the estimation of parameters is easier to do with a numerical maximization of the likelihood (or its logarithm). In this research, we use flemix package in R which is useful for estimation on mixture models [13].

4. Results and Discussion

The last row of table 1, shows that the average ($\bar{x} = 53.667$) < variance ($\bar{\sigma}_x = 1546.864$), so there is an overdispersion condition relative to Poisson. Poisson probability density function with rate 53.667 has one peak and symmetry at that point. There is a discrepancy between bar plot of the observed counts (empirical) and Poisson probability density function (figure 3) and also the cumulative plotting of the density function empirical and fitted Poisson cumulative density function (figure 4). Thus, although \{\(X_t, t \in T\)\} is a random variable of the process of enumeration of an occurrence, the Poisson distribution cannot be used in determining the probability function of \{\(X_t, t \in T\)\}. Figure 3 shows that there is more than one peak of the plot of observed counts (empirical) this indicates the heterogeneity of realization data \{\(X_t, t \in T\)\}. The heterogeneity of the data in each \(t \in T\) can be considered as a dynamic change from \{\(X_t, t \in T\)\} [5,14].

**Figure 3.** Number of earthquake 1973-2017 bar plot of observed counts and fitted Poisson probability density function (theoretical)

**Figure 4.** Number of earthquake 1973-2017 cumulative density function empirical and fitted Poisson cumulative density function

The Poisson mixture model fits the observations much better than does a single Poisson distribution. This is seen from the means and variances of the seven models and the maximum log-likelihood value as presented in table 2. In computing the means and variances of the models we use equations (5) and equation (7). Table 2 shows that the mean for each mixture model is the same as
the average observation. However, the variance value of the mixture model approaching the sample variance is the Poisson mixture model with the number of components $m = 7$, so we can write the model is a 7-Poisson mixture model.

**Table 2.** Poisson mixture models fitted to the number of earthquake series

| Model $m$ | $i$ | Parameter $\delta_i$ | $\lambda_i$ | $-\log$-likelihood | Mean | Variance |
|-----------|-----|-----------------------|-------------|---------------------|------|----------|
| 1         | 1   | 1.000                 | 53.667      | 610.5399            | 53.667 | 53.667   |
| 2         | 2   | 0.238                 | 109.167     |                     |      |          |
| 3         | 1   | 0.688                 | 33.758      | 223.544             | 53.667 | 1356.886 |
| 2         | 2   | 0.245                 | 78.310      |                     |      |          |
| 3         | 3   | 0.067                 | 168.531     |                     |      |          |
| 4         | 1   | 0.645                 | 32.580      | 211.040             | 53.667 | 1419.563 |
| 2         | 2   | 0.212                 | 64.472      |                     |      |          |
| 3         | 3   | 0.077                 | 101.204     |                     |      |          |
| 4         | 4   | 0.066                 | 168.741     |                     |      |          |
| 5         | 1   | 0.282                 | 27.668      | 204.231             | 53.667 | 1480.752 |
| 2         | 2   | 0.391                 | 37.433      |                     |      |          |
| 3         | 3   | 0.195                 | 67.566      |                     |      |          |
| 4         | 4   | 0.110                 | 119.854     |                     |      |          |
| 5         | 5   | 0.022                 | 220.000     |                     |      |          |
| 6         | 6   | 0.092                 | 220.000     |                     |      |          |
| 7         | 7   | 0.092                 | 220.000     |                     |      |          |

Visually the matching curves of the mixture model fit the histogram of the number of the earthquake as shown in figure 5. Convergence along the Sumatra subduction zone forms a beveled pattern where tectonic strains are partitioned into dip-slip components in the megathrust segmentation and horizontal components in the fault zone of Sumatra [15]. With such seismotectonic arrangements, it is quite natural that the dynamic changes occurring in the subduction zone of Sumatra are categorized as highly volatile up to six levels.

The second analysis, we also identified dynamic changes in large earthquake sources on Sunda megathrust segmentation. Figure 2 and table 1 show that, although from one subduction zone, the value of the number of earthquake events and the maximum magnitude varies between earthquake sources. There is a strong indication that not all large earthquake sources will have the same dynamic change characteristics. The location of major earthquake sources as shown in table 1, represents the occurrences of major earthquakes that have occurred before.
In general, each large earthquake source produces Poisson mixture models with a different number of components, as shown in Table 3. The Aceh-Andaman's major earthquake source has a higher dynamic change level than other large earthquake sources, that is, up to six levels while Mentawai-Siberut's major earthquake source has only two levels of dynamic change.

![Figure 5](image)

**Figure 5.** The number of earthquake data; the histogram of counts compare to Poisson mixture model of $m = 1$ until $m = 7$

Figure provides a description of the dynamic changes, generally in the Sunda megathrust segmentation (figure 6) and partially in large earthquake sources (figure 7-11). Visually it appears that, from 2004 to 2005, large earthquake sources were in dynamic change at the highest level. While from 2007 to the following years, the dynamic changes from each source of the earthquake tend to decrease towards the lowest level.

**Table 3.** Poisson mixture models fitted to the number of earthquake series in large earthquake source

| Large Earthquake Source | Number of Components ($m^*$) | Poisson Mixture Model – (Log-likelihood) | Mean | Variance | Observations | Mean | Variance |
|-------------------------|------------------------------|----------------------------------------|------|----------|--------------|------|----------|
| I                       | 6                            | $m = 1$                                | 370.907 | 164.202 | 19.711 | 330.504 | 19.711 | 329.560 |
| II                      | 5                            | $m = m^*$                              | 405.548 | 137.528 | 9.600  | 239.125 | 9.600  | 239.973 |
| III                     | 2                            | $m = m^*$                              | 141.514 | 103.883 | 3.956  | 18.515  | 3.956  | 18.362  |
| IV                      | 4                            | $m = m^*$                              | 252.890 | 115.682 | 5.422  | 74.550  | 5.422  | 74.659  |
| V                       | 3                            | $m = m^*$                              | 155.571 | 140.440 | 14.978 | 40.624  | 14.978 | 39.800  |
Based on the earthquake catalog of the USGS, Sumatra Island, especially the Aceh-Andaman region hit by a massive earthquake above 8 Mw in 2004. The 2002 Sumatra earthquake that occurred in the large earthquake source of Aceh-Andaman segmentation is believed to be the preliminary earthquake prior to the 2004 Sumatra earthquake [15]. These are similar to our results, as shown in figure 7, the dynamics occurring in Aceh-Andaman's major earthquake source began in 2001 increased to the highest level in 2005. After the 2004 and 2005 Sumatra earthquakes, there were several major earthquake events off the coast of the Andaman Islands, Nias-Simeuleu Islands with a relatively smaller magnitude and frequency. For example, an earthquake with magnitude 8.7 Mw occurred around the Nias-Simeulue region in 2005. This earthquake was not considered aftershocks, although it was close to the epicenter of the 2004 earthquake. However, the quake was most likely triggered by a stress change associated with the events of 2004 [10]. These result can also be explained by the resulting dynamic changes, as can be seen in figure 7 and figure 8. The dynamic changes in Nias-Simelue's major earthquake source have jumped three levels from the previous one in 2005.

**Figure 6.** Visualization of dynamic changes based on the number of earthquakes; the Sunda megathrust segmentation

**Figure 7.** Visualization of dynamic changes based on the number of earthquakes; Aceh–Andaman’s major earthquake source

**Figure 8.** Visualization of dynamic changes based on the number of earthquakes; Nias–Simelue’s major earthquake source

**Figure 9.** Visualization of dynamic changes based on the number of earthquakes; Mentawai–Siberut’s major earthquake source
The dynamic changes occurring at the large earthquake sources of Mentawai Island have different characteristics, as shown in figure 9 and figure 10. The difference is the number of dynamic levels and patterns of change. These indicate that on Mentawai Island there are differences in movement characteristics of tectonic plates. Although the dynamic changes in Mentawai Island are not as extreme as Aceh-Andaman's major earthquake source, however, there are other factors that result in the still great potential earthquake of Mentawai Island [6,7]. The Enggano's major earthquake source has three levels of dynamic change, as shown in figure 11. The most extreme dynamic changes occurred in 2000, this was the impact of the earthquake of Bengkulu that occurred on June 4, 2000. From large earthquake sources on the island of Sumatra, the dynamic changes occurring in this segment are not as volatile as others. Seismic activity and large earthquake potential in Enggano and surrounding is lower than the large earthquake sources on Sumatra Island [10].

5. Conclusion
In this paper, we have applied a Poisson mixture model to the number of earthquakes to analyze dynamic changes in the Sunda megathrust segmentation on the island of Sumatra and large earthquake sources. The performance of Poisson mixture model in describing dynamic changes can be considered in the process of determining the number of components. The results showed that Poisson mixture model is able to provide a description of the dynamic changes in the Sunda megathrust segmentation. Especially for the subduction zone of Sumatra, the level of dynamic change can be classified into seven categories. When we analyzed five large earthquake source segmentation in that area, various categories of dynamic changes are found to be different from two until six categories. The Aceh-Andaman segmentation had the highest dynamic changes among the major earthquake sources on the Sunda megathrust. While the Mentawai-Siberut segmentation had a low dynamic change only two categories

References
[1] PusGeN 2017 Peta Sumber dan Bahaya Gempa Indonesia (Badan Peneliti dan Pengembangan Kementerian Pekerjaan Umum dan Perumahan Rakyat, Indonesia)
[2] Boore D M and Atkinson G M 2008 Earthquake Spectra 24(1) 99-138
[3] Bommer J J, Douglas J, Scherbaum F, Cotton F, Bungum H, and Fah D 2010 Seismol. Res. Lett. 81(5) 783-793
[4] Bozorgnia Y, Hachem M M, and Campbell K W 2010 Earthquake Spectra 26(1) 1-23
[5] Yip C F, Ng W L, and Yau C Y 2018 Stochastic Enviromental Research and Risk Assessment 32(5) 1415-1434
[6] Newman A V, Hayes G, Wei Y, and Converse, J 2011 Geophys. Res. Lett. 38(5)
[7] McCloskey J, Antoniolo A, Piatanesi A, Sieh K, Steachy S, Nalbant S, Cocco M, Giunchi C, Huang J, and Dunlop P 2008 Earth Planet. Sci. Lett. 265(1-2) 61-81
[8] Sieh K, Natawidjaja D H, Meltzner A J, Shen C C, Cheng H, Li K S, Suwargadi B W, Galetzka J, Philibosian B, and Edwards R L 2008 Science 322(5908) 1674-1678
[9] Chlieh M, Avouac J P, Sieh K, Natawidjaja D H, and Galetzka J 2008 Journal of Geophysical Research: Solid Earth 113(B5)
[10] McCaffrey R 2009 Annu. Rev. Earth Planet. Sci. 37 345-366
[11] Ambikapathy A, Catherine J K, Gahalaut V K, Narsaiah M, Bansal A, and Mahesh P 2010 J. Earth Syst. Sci. 119(4) 553-560
[12] Zucchini W, MacDonald I L, and Langrock R 2017 Hidden Markov Models for Time Series: an Introduction using R (CRC Press: Boca Raton)
[13] BenagliaT, Chauveau D, Hunter D, and Young D 2009 Journal of Statistical Software 32(6) 1-29
[14] Orfanogiannaki K, Karlis D, and Papadopoulos G A 2014 Research in Geophysics 4(1) 1-6
[15] Vallee M 2007 Bull. Seismol. Soc. Am. 97(1A) S103-S114.