J. Indones. Math. Soc.
Vol. 32, No. 01 (2026), pp. 1–15.

Analysis of Earthquake Potential along the Coastal Region of South Java using Semi-Markov Models as a Tsunami Mitigation

Athaya Rahma Puteri[1], Halimatus Sa'diyah[1], Alfia Nur Fauziah[1], Christina Agustin Raphonhita Simbolon[1], Ramadhani Latief Firmansyah[1], and Dwi Ertiningsih[1*]

[1]Department of Mathematics, Universitas Gadjah Mada, Indonesia

Abstract. This study applies a semi-Markov model to assess earthquake occurrence in the South Java coastal region. The main objective is to forecast earthquakes in this area, considering three key factors: geographic location, timing, and seismic magnitude. The South Java coastal region is chosen for this study due to its proximity to the island of Java, the economic hub of Indonesia. The study divides the South Java coastal region into five distinct zones and categorizes earthquakes into three magnitude groups. The results predict that earthquakes will occur in the South Coast regions of East Java, Central Java, or West Java between December 26, 2022, and November 20, 2023. Additionally, projections suggest that earthquakes are likely to occur in East Java, West Java, or Banten between November 21, 2023, and December 31, 2030. The estimated magnitudes range from 5 to 6 Mw. The findings also indicate that no tsunamis are expected along the South Java coast until 2030. Model validation using the Mean Absolute Percentage Error (MAPE) results in a value of 4.224%. This confirms the high accuracy of the predictions. Although no tsunamis are forecasted, the public must remain alert and prepared for the anticipated earthquakes. These findings provide important insights for disaster mitigation and emphasize the need for ongoing monitoring, early warning systems, and community preparedness to minimize potential risks.

Key words and Phrases: semi-Markov model, earthquake forecasting, tsunami, south coast of Java, Mean Absolute Percentage Error (MAPE).

*Corresponding author : dwi_ertiningsih@ugm.ac.id
2020 Mathematics Subject Classification: Primary: 60K25 Secondary: 60J27, 60G55, 90B22
Received: 25-10-2023, accepted: 17-02-2025.

1. INTRODUCTION

Reporting from the National Agency for Disaster Countermeasure (BNPB) page, Indonesia is an archipelagic country that is geographically located at the meeting point of four tectonic plates, namely the Asian continental plate, the Australian continental plate, the Indian Ocean plate, and the Pacific Ocean plate. This situation is one of the reasons why earthquakes often hit Indonesia. One of the causes of earthquakes is the movement of tectonic plates (tectonic movement). Tectonic movement is the movement of the world's tectonic plates which will cause two tectonic plates to collide and shift each other. This earthquake is called a tectonic earthquake.

The United States Geological Survey (USGS) recorded that from January 1 to September 30, 2023, Indonesia experienced 1,618 earthquakes. These earthquakes can cause losses to the community because it can cause building damage, fires, landslides, economic losses, and even loss of life. An earthquake can also cause a tsunami if it occurs in the deep sea which shifts the tectonic plates on the seabed and the earthquake has a magnitude of more than 7.6 on the Richter scale (CNN Indonesia, 2021).

Hiller and Lieberman [1] emphasize that operations research provides a powerful framework for making optimal decisions in complex systems. By integrating mathematical modeling, quantitative analysis, and computational tools, OR helps organizations allocate resources efficiently, evaluate alternatives rigorously, and improve overall performance. In case of the occurrence of earthquake, forecasting the occurrence of earthquakes is crucial to minimizing these potential losses. Several studies have been conducted on earthquake forecasting. Lewis [2] emphasizes that effective forecasting requires selecting methods suited to the data, the decision context, and the desired time horizon. No single technique is universally superior; rather, combining statistical approaches with managerial judgment yields the most reliable results, whereas Jafari [3] attempted to forecast an earthquake in Tehran, Iran, using several statistical models, but the resulting error rate was quite large.

Many classical studies such as Vere-Jones [4] and Harte [5] also emphasize that statistical forecasting remains challenging due to complex seismic patterns. Gkarlaouni et.al. [6] presents a stochastic comparison of earthquake activity in two Greek fault systems: the Mygdonia graben and the Gulf of Corinth. Using local seismic catalogues, the authors analyse magnitude, inter-event time and distance distributions, along with spatial clustering. The results show clear differences in seismic behaviour between the two regions, underscoring the value of statistical methods for characterizing fault-specific seismicity. Sadeghian [7] investigates how different zoning methods influence the accuracy of earthquake forecasting using semi-Markov models. The study shows that the choice of spatial zoning significantly affects predicted earthquake occurrences, highlighting the importance of appropriate zoning for reliable seismic hazard assessment.

In this research, several methods were used, including Exponential, Gamma, Lognormal, Pareto, Rayleigh, and Weibull models. Each method has its limitations, such as exponential and Weibull using only one parameter, the gamma only uses two parameters, lognormal has a large error value, the Pareto is not suitable for long-term distributions, and the Rayleigh requires data to come from a two-dimensional Gaussian distribution. Furthermore, Alarifi, et.al. [8] conducted research to forecast the magnitude of earthquakes in the Northern Red Sea region using the Artificial Neural Network (ANN) method and obtained forecasting results that were 32% better than other methods. This research produces predictions of the magnitude of earthquakes, while predicting the time of earthquake events is still not lacking because this method cannot capture time relationships with high accuracy. In the same year, Sadeghian [9] forecasted earthquakes in Tehran, Iran, using semi-Markov which was able to forecast the location, time, and magnitude of an earthquake simultaneously with a high level of accuracy. This method produces complete predictions compared to other methods.

Furthermore, in [8], a research has been conducted to forecast the magnitude of earthquakes in the Northern Red Sea region using the Artificial Neural Network (ANN) method and obtained forecasting results that were 32% better than other methods. This research produces predictions of the magnitude of earthquakes, while predicting the time of earthquake events is still not lacking because this method cannot capture time relationships with high accuracy. In the same year, Ramin Sadeghian forecasted earthquakes in Tehran, Iran, using semi-Markov which was able to forecast the location, time, and magnitude of an earthquake simultaneously with a high level of accuracy. This method produces complete predictions compared to other methods.

Considering the limitations of previous models and the success of the semi-Markov model in earthquake forecasting, this research adopts the semi-Markov method. This approach was chosen because this approach is more effective than other models. This research focuses on the south coast of the Java region because it is near Java Island, the center of the Indonesian economy, and if an earthquake or tsunami occurs in this region, it will cause massive loss to Indonesia. Thus, applying the semi-Markov model can estimate the possibility of an earthquake occurring when, where, and how strong the earthquake will be, and can find out which areas have the potential for a tsunami due to the earthquake so that it can minimize the impact of the earthquake, that is, the impact of the losses incurred.

2. METHODS

2.1. **Data Collection.**

This research was conducted using data from earthquake disasters site (the United States Geological Survey, USGS) recorded on the period January 1, 1910 to December 31, 2022 on the south coast of Java. This data consists of location, depth of earthquake ($\leq$ 60 km), time of occurrence of the earthquake, and magnitude ($\geq$ 5 Mw). The criteria for the depth and strength of the earthquake data are based on the characteristics of earthquakes that have the potential to cause tsunamis, which were obtained from earthquake events that caused tsunamis in the past. The south coast of Java is an area of investigation, bounded by longitudes 5°52′35"S, 11°S and latitudes 105°1′11"E, 114°40′22.8"E. The data were collected from the United States Geological Survey (USGS) site from January 1, 1910 to December 31, 2022. Regions are chosen as states. We divided the coastal area of the south coast of Java into five regions based on the division of provincial areas on the island of Java, i.e., the south coast of East Java (R_1), Central Java (R_2), Yogyakarta (R_3), West Java (R_4), and Banten (R_5). Furthermore, magnitudes were chosen as states. We divided the magnitudes into three categories based on the earthquake strength data pattern, i.e.,

$$M_1 : 5 \leq \text{Mw} < 6$$
$$M_2 : 6 \leq \text{Mw} < 7.6$$
$$M_3 : \text{Mw} \geq 7.6$$

where Mw is a unit of magnitude of earthquakes; it is called moment magnitude.

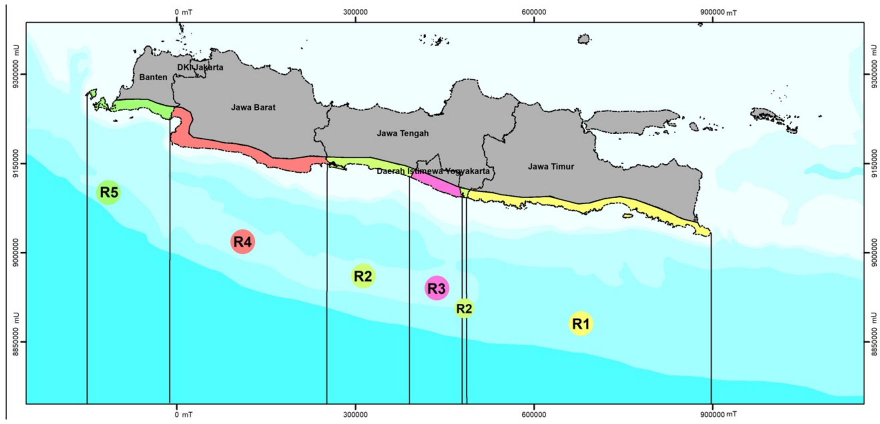

FIGURE 1. Map of South Coast of Java

2.2. **Analysis Method.**

To find forecasting for earthquakes, the Semi-Markov process is used, which is a development of the Markov stochastic process. Further explanation is given as follows.

2.2.1. *Stochastic Process.*

A stochastic process is a collection of random variables $\{X(t)\}$ indexed by the index t running through the set T, where T is a set of non-negative integers. The random variable $X(t)$ represents the state of a system at time t. Stochastic processes are divided into 2 types, namely, if $T = 0, 1, 2, \ldots$ then it is called a stochastic process with discrete parameters and is denoted by $\{X(n)\}$, whereas if $T = \{t \mid t \geq 0\}$ then it is called a stochastic process with continuous parameters and is denoted by $\{X(t) \mid t \geq 0\}$. This research was conducted to investigate the systematic nature of earthquake events related to space and place, time distribution, and regional earthquake magnitude. This approach is suitable for application in urgent needs because it can utilize all available information and develop hypotheses about the systematics governing seismicity at all scales, such as groupings expressed as consecutive earthquake events. Using a stochastic approach to seismicity will reveal additional implications about the seismic behavior of fault systems that should be applicable in any seismotectonic setting.

2.2.2. *Semi-Markov Process.*

The semi-Markov process is an extension of the stochastic Markov process. In a semi-Markov process, the Markov property is no longer fulfilled, meaning that predicting the future state is not only based on the current state but also on the length of time in the current state before moving to the future state. Consider a stochastic process with a state space $\{1, 2, \ldots, n\}$ and the system is in the initial state $X(0)$ at time $T(0)$. The process remains in that state for a certain amount of time m_0 and then moves to state $X(1)$ at time $T(1)$. The process remains in that state for a certain amount of time m_1. Then move to state $X(2)$ at time $T(2)$, and so on until the n-th state. This process is called a semi-Markov process. One of the applications of the semi-Markov process is to predict disasters.

2.2.3. *Transition Matrix.*

The transition matrix is denoted by G which is obtained by calculating the probability of a semi-Markov process is in state i and then makes a transition to state j. The transition matrix is calculated based on two classifications, namely based on the location of the earthquake and the earthquake strength scale. Thus, there are two transition matrices, namely the transition matrix for the location of the earthquake and the earthquake strength scale. The transition matrix must satisfy the following two conditions.

$$\begin{aligned} G_{ij} &\geq 0; \\ \sum_{j=1}^{N} G_{ij} &= 1; \quad i, j = 1, 2, \ldots, N, \end{aligned} \tag{1}$$

where G_{ij} is the ij-th element of the transition matrix and N is the number of states in the system.

2.2.4. *Holding Time Matrix.*

The time during which the semi-Markov process remains in state i within time t_{ij} before transitioning to state j is called holding time. In this case, the unit of time used is 30 days. Continue with calculating the mass holding time function for transitions between earthquake locations and the earthquake strength scale. The holding time mass function is the probability mass function of T_{ij} in t_{ij}, as given below:

$$\Pr\{t_{ij} = m\} = T_{ij}(m); \quad m = 1, 2, \ldots, n, \quad i, j = 1, 2, \ldots, N, \tag{2}$$

where t_{ij} is the time during which the semi-Markov process remains in state i before transitioning to state j, T_{ij} is the mass holding time function, and n is the number of time intervals.

2.2.5. *Core Matrix.*

The core matrix is denoted by $C(m)$ which is the probability of two joint events in which a system in state i at time 0 transitions to state j after m holding time as follow:

$$C_{ij}(m) = G_{ij}T_{ij}(m); \quad m = 1, 2, \ldots, n, \quad i, j = 1, 2, \ldots, N, \tag{3}$$

where $C_{ij}(m)$ is the ij-th element of the core matrix at time m, G_{ij} is the ij-th element of the transition matrix, and $T_{ij}(m)$ is the ij-th element of the holding time matrix at time m.

2.2.6. *Waiting Time Mass Function.*

After calculating the core matrix, continue by adding up each row of the core matrix for each time difference (holding time) to get the mass waiting time function. That is

$$\sum_{j=1}^{N} C_{ij}(m) = \sum_{j=1}^{N} G_{ij}T_{ij}(m) = w_i(m), \tag{4}$$

where $w_i(m)$, namely the probability that the waiting time for the i-th state is equal to m. The cumulative probability distribution of waiting time is given as follow:

$$LEw_i(n) = \sum_{m=1}^{n} w_i(m), \tag{5}$$

where $LEw_i(n)$ is the probability that the waiting time for the i-th state is less than or equal to n. Meanwhile, the complement of $LEw_i(n)$ is

$$Gw_i(n) = \sum_{m=n+1}^{\infty} w_i(m), \tag{6}$$

where $Gw_i(n)$ is the probability that the waiting time for the i-th state is greater than n.

2.2.7. *Interval Transition Probability Matrix.*

The interval transition probability matrix $F(n)$ is obtained through a recursive procedure, with $F(0)$ being the identity matrix. So we will obtain an interval transition probability matrix for transitions between locations and earthquake strength scales. The interval transition probability matrix formula is

$$F(n) = GW(n) + \sum_{m=0}^{n} (G \cdot T(m))F(n-m) = GW(n) + \sum_{m=0}^{n} C(m)F(n-m), \tag{7}$$

where m and n are natural numbers representing the time interval, G is the transition matrix, $T(n)$ is the holding time matrix, $C(m)$ is the core matrix, and $GW(n)$

is a diagonal matrix where the i-th element in the matrix is the same as the element in the matrix $Gw_i(n)$ in Eqn. (6). By using Eqn. (7), it can be obtained that $F_R(n)$ and $F_M(n)$ respectively express the interval transition probability matrix for transitions between earthquake locations and the interval transition probability matrix for transitions in the earthquake strength scale.

2.2.8. *Earthquake Forecasting Matrix.*

Based on Bruin [10], the earthquake forecasting matrix, $\hat{F}_{RM}$, is obtained by multiplying the interval transition probability matrix $F_R(n)$ with the interval transition probability matrix $F_M(n)$ using the following equation:

$$\hat{F}_{RM}(d) = F_{R_{r_0,r_1}}(d) \cdot F_{M_{m_0,m_1}}(d), \tag{8}$$

where $F_{R_{r_0,r_1}}$ represents the interval transition probability matrix for transition from the region $r_0$to the region r_1, and $F_{M_{m_0,m_1}}$ represents the interval transition probability matrix for transitions from an earthquake magnitude of m_0 to m_1. For example, the entries of the $F_{RM}(d)$ matrix, denoted as $F_{RM_{r_i m_j}}(d)$, express the probability of an earthquake occurring in region r_i with a magnitude m_j within in the time interval n. By using Algorithm 1, we can deterministically forecast earthquake occurrences over the next several time periods.

Definition 2.1. (***i*-th order maximum**). *The i-th element in a sorted as well as decreased list, in which none of the elements are equal to each other, is named i-th order maximum.*

Algorithm 1 Deterministic Forecasting Algorithm

Step 0: Begin
Step 1: Use the past n data to determine $\hat{F}_{RM}$ (n is the number of total data)
Step 2: Determine $\hat{F}_{RM}$
Step 3: $\hat{F}_{RM}1(i) = \hat{F}_{RM}(i) \forall i = 1, \ldots, k$ (is the number of future time intervals that can be forecasted)
Step 4: $M_j = \{t\text{-th order maximum in } \hat{F}_{RM}1(j), \forall j = 1, \ldots, k\}$
Step 5: $\hat{F}_{RM}2(j) = \left[\dfrac{\hat{F}_{RM}1(j)}{M_j} \right], \forall j = 1, \ldots, k$
(The highest integer number that is less than the real number x is obtained by using $[x]$.)
Step 6: For all element $x \in \hat{F}_{RM}2(j)$ with $x \geq 1$ will be replaced by 1 and name the resulted matrices $FRMD(j), \forall j = 1, \ldots, k$
Step 7: An earthquake with magnitude m in region r will occur in the j-th time period if $FRMD_{rm}(j) = 1$; else, no earthquake with magnitude m in region r will occur in the j-th time period.
Step 8: End

2.2.9. *The Model Validation.*

To determine the level of accuracy of earthquake forecasting results, a validation process must be carried out. Many indicators can be used to measure the level of accuracy of forecasting, including Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Deviation (MAD), and Mean Absolute Percentage Error (MAPE). The forecasting results are always attempted to be close to actual events. In other words, the error value produced by the forecasting results should be small. Mean Absolute Percentage Error (MAPE) can be calculated by using the absolute error for each period divided by the actual observed value for that period. The advantage of the MAPE approach is that it can express the percentage error in forecasting results regarding actual demand during a certain period, providing information on whether the percentage error is too high or too low, thereby enhancing accurate. However, a significant deficiency of MAPE is that the calculation will yield an infinite or undefined value if the actual value is zero or near zero (see [11]). We can determine the level of accuracy of earthquake forecasting results using Algorithm 2. If the actual value is zero, Algorithm 2 bypasses this calculation and reduces the count r and m by one.

Algorithm 2 Validation Algorithm

Step 0: Begin
Step 1: $d = 0$
Step 2: $n_2(0) = 0$
Step 3: $d = 1$
Step 4: Use the $(n_1 + n_2(d))$ first data to determine $\hat{F}_{RM}$ and F_{RM} where n_1 is the number of the first data that used to forecast the next n_2 data and $n_2(d)$ is the number of the actual data that occurred during d periods time after the first n_1 data.
Step 5: Determine $\hat{F}_{RM}$ and F_{RM}
Step 6: If $F_{RM_{ij}} \neq 0$, then

$$MAPE(d) = \frac{\sum_{i=1}^{r} \sum_{j=1}^{m} \left(\frac{|F_{RM_{ij}} - \hat{F}_{RM_{ij}}|}{F_{RM_{ij}}} \right)}{r \cdot m} \times 100\% \tag{9}$$

Step 7: $d = d + 1$
Step 8:

$$MAPE = \frac{\sum_{d=1}^{p} MAPE(d)}{p} \tag{10}$$

Step 9: End

The results of the MAPE calculation are usually in percentage form. The smaller the percentage, the better the accuracy level. The MAPE calculation criteria to measure the accuracy of the forecasting results are given in Table 1.

Table 1. MAPE Criteria for Model Evaluation

MAPE	Forecasting power
$< 10\%$	Highly accurate forecasting
$10\% - 20\%$	Good forecasting
$20\% - 50\%$	Reasonable forecasting
$> 50\%$	Weak and inaccurate forecasting

3. MAIN RESULTS

3.1. **Transition matrix.**

The transition probability matrix of region-to-region (G_R) transitions and magnitude-to-magnitude (G_M) were determined in Eqn. (1).

$$G_R = \begin{array}{c} \\ R_1 \\ R_2 \\ R_3 \\ R_4 \\ R_5 \end{array} \begin{array}{c} \begin{array}{ccccc} R_1 & R_2 & R_3 & R_4 & R_5 \end{array} \\ \begin{bmatrix} 0.6565 & 0.0404 & 0.0455 & 0.1667 & 0.0909 \\ 0.2759 & 0.1724 & 0 & 0.4483 & 0.1034 \\ 0.2 & 0.04 & 0.24 & 0.32 & 0.2 \\ 0.1759 & 0.0553 & 0.0301 & 0.6231 & 0.1156 \\ 0.2836 & 0.0746 & 0.0597 & 0.3134 & 0.2687 \end{bmatrix} \end{array}$$

The elements in the G_R matrix represent the probability of an earthquake occurring in region j, given that the previous earthquake occurred in region i. For example, the probability that an earthquake will occur in region R_4 after the previous earthquake occurred in region R_2, is 0.4483.

$$G_M = \begin{array}{c} \\ M_1 \\ M_2 \\ M_3 \end{array} \begin{array}{c} \begin{array}{ccc} M_1 & M_2 & M_3 \end{array} \\ \begin{bmatrix} 0.9206 & 0.0751 & 0.0043 \\ 0.72 & 0.28 & 0 \\ 1 & 0 & 0 \end{bmatrix} \end{array}$$

The elements in the G_M matrix represent the probability of an earthquake occurring with magnitude j, given that the previous earthquake occurred with magnitude i. For example, the probability that an earthquake will occur with magnitude M_1 after the previous earthquake occurred with magnitude M_2, is 0.72. Furthermore, it should be noted that if an earthquake occurs, whether with a magnitude of M_1, M_2, or M_3, the aftershocks will tend to have a magnitude of M_1. In other words, the intensity of the subsequent earthquake will tend to decrease compared to the initial earthquake.

3.2. **Holding Time.**

The largest time interval between the times of earthquake occurrences is 2978 days, so by selecting 30 days for time unit, forecasting the next $\left\lceil \frac{2978}{30} \right\rceil = 100$ time units in each forecasting will be possible. Therefore, holding time mass functions

$T_R(m)$ and $T_M(m)$ are obtained using Eqn. (2).

$$T_R(1) = \begin{array}{c} \\ R_1 \\ R_2 \\ R_3 \\ R_4 \\ R_5 \end{array} \begin{array}{c} \begin{array}{ccccc} R_1 & R_2 & R_3 & R_4 & R_5 \end{array} \\ \begin{bmatrix} 0.8769 & 0.375 & 0.5556 & 0.2424 & 0.3333 \\ 0.625 & 0.4 & 0 & 0.5384 & 0.6667 \\ 0.2 & 0 & 0.6667 & 0.125 & 0.2 \\ 0.3142 & 0.7272 & 0.6667 & 0.8145 & 0.2174 \\ 0.4736 & 0.6 & 0.25 & 0.3333 & 0.4444 \end{bmatrix} \end{array}$$

$$T_R(2) = \begin{array}{c} \\ R_1 \\ R_2 \\ R_3 \\ R_4 \\ R_5 \end{array} \begin{array}{c} \begin{array}{ccccc} R_1 & R_2 & R_3 & R_4 & R_5 \end{array} \\ \begin{bmatrix} 0.0538 & 0.25 & 0.2222 & 0.4545 & 0.2778 \\ 0.125 & 0.4 & 0 & 0.1538 & 0.3333 \\ 0 & 1 & 0.1667 & 0.125 & 0.2 \\ 0.2857 & 0.0909 & 0 & 0.0806 & 0.1304 \\ 0.1579 & 0 & 0.25 & 0.2381 & 0.1111 \end{bmatrix} \end{array}$$

and so on until $T_R(100)$.

$$T_M(1) = \begin{array}{c} \\ M_1 \\ M_2 \\ M_3 \end{array} \begin{array}{c} \begin{array}{ccc} M_1 & M_2 & M_3 \end{array} \\ \begin{bmatrix} 0.6550 & 0.4286 & 0 \\ 0.4167 & 0.2143 & 0 \\ 1 & 0 & 0 \end{bmatrix} \end{array}$$

$$T_M(2) = \begin{array}{c} \\ M_1 \\ M_2 \\ M_3 \end{array} \begin{array}{c} \begin{array}{ccc} M_1 & M_2 & M_3 \end{array} \\ \begin{bmatrix} 0.1375 & 0.2286 & 1 \\ 0.1944 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \end{array}$$

and so on until $T_M(100)$.

3.3. **Core Matrix.**

The core matrix is obtained using Eqn. (3).

$$C_R(1) = \begin{array}{c} \\ R_1 \\ R_2 \\ R_3 \\ R_4 \\ R_5 \end{array} \begin{array}{c} \begin{array}{ccccc} R_1 & R_2 & R_3 & R_4 & R_5 \end{array} \\ \begin{bmatrix} 0.5757 & 0.0152 & 0.0253 & 0.0404 & 0.0303 \\ 0.1724 & 0.0689 & 0 & 0.2414 & 0.0689 \\ 0.04 & 0 & 0.16 & 0.04 & 0.04 \\ 0.0553 & 0.0402 & 0.0201 & 0.5075 & 0.0251 \\ 0.1343 & 0.0447 & 0.0149 & 0.1044 & 0.1194 \end{bmatrix} \end{array}$$

$$C_R(2) = \begin{array}{c} \\ R_1 \\ R_2 \\ R_3 \\ R_4 \\ R_5 \end{array} \begin{array}{c} \begin{array}{ccccc} R_1 & R_2 & R_3 & R_4 & R_5 \end{array} \\ \begin{bmatrix} 0.0353 & 0.0101 & 0.0101 & 0.0757 & 0.0253 \\ 0.0344 & 0.0689 & 0 & 0.0689 & 0.0344 \\ 0 & 0.04 & 0.04 & 0.04 & 0.04 \\ 0.0502 & 0.0050 & 0 & 0.0502 & 0.0151 \\ 0.0447 & 0 & 0.0149 & 0.0746 & 0.0298 \end{bmatrix} \end{array}$$

and so on until $C_R(100)$.

Elements in the matrix C_R represent the probability of a joint event. The first event is an earthquake occurring in region R_j, where previously an earthquake occurred in region R_i, while the second event is the transition carried out after a holding time m. For example, the value 0.1724 in the matrix $C_R(1)$ represent the probability that an earthquake will occur in region R_1, where an earthquake previously occurred in region R_2, and the earthquake took place within a time interval of one period or 30 days.

$$C_M(1) = \begin{array}{c} \\ M_1 \\ M_2 \\ M_3 \end{array} \begin{array}{c} \begin{array}{ccc} M_1 & M_2 & M_3 \end{array} \\ \begin{bmatrix} 0.6030 & 0.0322 & 0 \\ 0.3 & 0.06 & 0 \\ 1 & 0 & 0 \end{bmatrix} \end{array}$$

$$C_M(2) = \begin{array}{c} \\ M_1 \\ M_2 \\ M_3 \end{array} \begin{array}{c} \begin{array}{ccc} M_1 & M_2 & M_3 \end{array} \\ \begin{bmatrix} 0.1267 & 0.0171 & 0.0043 \\ 0.14 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \end{array}$$

and so on until $C_M(100)$.

Elements in the matrix C_M represent the probability of a joint event. The first event is an earthquake occurring at magnitude M_j, given that a previous earthquake occurred at magnitude M_i. The second event is the transition that occurs after a holding time of m. For example, the value 0.3 in the matrix $C_M(1)$ indicates the probability that an earthquake will occur at magnitude M_1, given that a previous earthquake occurred at magnitude M_2, and this occurred within a time interval of one period or 30 days.

3.4. **Waiting Time Mass Function.**

The waiting time mass function is obtained using Eqn. (5).

$$w_{i_R}(1) = \begin{array}{c} R_1 \\ R_2 \\ R_3 \\ R_4 \\ R_5 \end{array} \begin{bmatrix} 0.6868 \\ 0.5517 \\ 0.28 \\ 0.6482 \\ 0.4179 \end{bmatrix}$$

$$w_{i_R}(2) = \begin{array}{c} R_1 \\ R_2 \\ R_3 \\ R_4 \\ R_5 \end{array} \begin{bmatrix} 0.1565 \\ 0.2069 \\ 0.16 \\ 0.1206 \\ 0.1642 \end{bmatrix}$$

and so on until $W_{i_R}(100)$.

$$w_{i_M}(1) = \begin{array}{c} M_1 \\ M_2 \\ M_3 \end{array} \begin{bmatrix} 0.6352 \\ 0.36 \\ 1 \end{bmatrix}$$

$$w_{i_M}(2) = \begin{matrix} M_1 \\ M_2 \\ M_3 \end{matrix} \begin{bmatrix} 0.1480 \\ 0.14 \\ 0 \end{bmatrix}$$

and so on until $W_{i_M}(100)$.

3.5. **Interval Transition Probability Matrix.**

The interval transition probability matrix is obtained using Eqn. (7).

$$F_R(1) = \begin{matrix} & R_1 & R_2 & R_3 & R_4 & R_5 \\ R_1 & 0.8889 & 0.0151 & 0.0252 & 0.0404 & 0.0303 \\ R_2 & 0.1724 & 0.5172 & 0 & 0.2413 & 0.0689 \\ R_3 & 0.04 & 0 & 0.88 & 0.04 & 0.04 \\ R_4 & 0.0553 & 0.0402 & 0.0201 & 0.8592 & 0.0251 \\ R_5 & 0.1343 & 0.0447 & 0.0149 & 0.1044 & 0.6716 \end{matrix}$$

$$F_R(2) = \begin{matrix} & R_1 & R_2 & R_3 & R_4 & R_5 \\ R_1 & 0.7136 & 0.0296 & 0.0481 & 0.1415 & 0.0661 \\ R_2 & 0.2222 & 0.3614 & 0.0102 & 0.3072 & 0.0968 \\ R_3 & 0.0495 & 0.0440 & 0.7432 & 0.0865 & 0.0754 \\ R_4 & 0.1385 & 0.0482 & 0.0296 & 0.7328 & 0.0499 \\ R_5 & 0.1943 & 0.0347 & 0.0353 & 0.1937 & 0.5084 \end{matrix}$$

and so on until $F_R(100)$.

$$F_M(1) = \begin{matrix} & M_1 & M_2 & M_3 \\ M_1 & 0.9365 & 0.0591 & 0.0043 \\ M_2 & 0.4483 & 0.5516 & 0 \\ M_3 & 0.9678 & 0 & 0 \end{matrix}$$

$$F_M(2) = \begin{matrix} & M_1 & M_2 & M_3 \\ M_1 & 0.9257 & 0.0716 & 0.0026 \\ M_2 & 0.5033 & 0.4953 & 0.0013 \\ M_3 & 0.9366 & 0.0591 & 0.0043 \end{matrix}$$

and so on until $F_M(100)$.

3.6. **Forecasting.**

Widiyantoro et.al. [12] analyze seismic gaps south of Java to assess the potential for future megathrust earthquakes and tsunamis. The study identifies regions of accumulated strain, suggesting heightened seismic and tsunami risk, and underscores the need for improved monitoring and disaster preparedness in the area. Based on the data, the last earthquake occurred on December 25, 2022, in the Southern Coastal Region of Central Java (R_2) with a magnitude of 5.13 Mw (M_1). We have $\hat{F}_{RM}(m)$ as a probabilistic forecasting matrix for 1 to 100 time periods for the South Coast of Java using Eqn. (8).

$$\hat{F}_{RM}(1) = \begin{array}{c} \\ R_1 \\ R_2 \\ R_3 \\ R_4 \\ R_5 \end{array} \begin{array}{c} \begin{array}{ccc} M_1 & M_2 & M_3 \end{array} \\ \begin{bmatrix} 0.16686 & 0.00555 & 0 \\ 0.50059 & 0.01665 & 0 \\ 0 & 0 & 0 \\ 0.23361 & 0.00777 & 0 \\ 0.06674 & 0.00221 & 0 \end{bmatrix} \end{array}$$

$$\hat{F}_{RM}(2) = \begin{array}{c} \\ R_1 \\ R_2 \\ R_3 \\ R_4 \\ R_5 \end{array} \begin{array}{c} \begin{array}{ccc} M_1 & M_2 & M_3 \end{array} \\ \begin{bmatrix} 0.20814 & 0.01313 & 0.00095 \\ 0.33856 & 0.02136 & 0.00155 \\ 0.00958 & 0.00060 & 0.000043 \\ 0.28772 & 0.01816 & 0.00132 \\ 0.09071 & 0.00572 & 0.000415 \end{bmatrix} \end{array}$$

and so on until $\hat{F}_{RM}(100)$.
By using Algorithm 1, deterministic forecasting during the future 100 time periods, for the South Coast of Java zoning is determined as shown in Table 2.

TABLE 2. Deterministic forecasting matrix in South Coast of Java

Periods	South Coast of Java zoning method
1 to 11	R_1M_1, R_2M_1, R_4M_1
12 to 100	R_1M_1, R_4M_1, R_5M_1

TABLE 3. Real occurrences during next 10 time periods after last earthquake occurrences in South Coast of Java

Periods	South Coast of Java zoning method
1	R_1M_1
2	R_5M_1
3	R_1M_1, R_2M_1
4	-
5	-
6	R_3M_1
7	-
8	R_5M_1
9	-
10	R_4M_1

3.7. **The Model Validation.**

Using Algorithm 2, we obtain the value of MAPE is 4.224%. Based on Table 1, the forecasting power is highly accurate. Moreover, Tables 2 and 3 show that, in the zoning of the south coast of Java, 57% of the earthquakes have been accurately forecasted in both location and magnitude. In addition, 43% of the forecasted earthquakes occurred in nearby areas, but their magnitudes were correct.

4. CONCLUDING REMARKS

This research was conducted by processing earthquake data from the United States Geological Survey website for the period from January 1, 1910 to December 31, 2022. According to the results, earthquakes occurred between the southern coastal areas of East Java, Central Java, and West Java with a magnitude of $M_1(5 \leq \text{Mw} < 6)$ from December 26, 2022 to November 20, 2023. Additionally, earthquakes occurred between the southern coastal areas of East Java, West Java, and Banten, with a magnitude of $M_1(5 \leq \text{Mw} < 6)$ from November 21, 2023, to December 31, 2030. In conclusion, it is forecasted that there will be no tsunami in the southern coastal area of Java until 2030. Our model validation using MAPE calculations indicates a highly accurate forecasting level.

Acknowledgement. This research was supported by BELMAWA DIKTI. We would like to express our gratitude to the Directorate of Student Affairs UGM, PKM Center UGM, and BMKG for their invaluable assistance and resources throughout this research.

REFERENCES

[1] C. W. Churchman, R. L. Ackoff, and E. L. Arnoff, "Introduction to operations research.," 1957. https://doi.org/10.2307/2227681.

[2] N. Meade, "Industrial and business forecasting methods, lewis, cd, borough green, sevenoaks, kent: Butterworth, 1982. price:£ 9.25. pages: 144," 1983. https://doi.org/10.1002/FOR.3980020210.

[3] M. A. Jafari, "Statistical prediction of the next great earthquake around Tehran, Iran," *Journal of Geodynamics*, vol. 49, no. 1, pp. 14–18, 2010. https://dx.doi.org/10.1016/j.jog.2009.07.002.

[4] D. Vere-Jones, "Forecasting earthquakes and earthquake risk," *International Journal of Forecasting*, vol. 11, no. 4, pp. 503–538, 1995. https://doi.org/10.1016/0169-2070(95)00621-4.

[5] D. S. Harte, "Log-likelihood of earthquake models: evaluation of models and forecasts," *Geophysical Journal International*, vol. 201, no. 2, pp. 711–723, 2015. https://doi.org/10.1093/gji/ggu442.

[6] C. Gkarlaouni, E. Papadimitriou, S. Lasocki, G. Lizurek, V. Karakostas, and A. Kilias, "Stochastic analysis of earthquake activity in two seismogenic fault systems in greece.," *Bulletin of the Geological Society of Greece*, vol. 47, no. 3, pp. 1099–1108, 2013. https://doi.org/10.12681/bgsg.10965.

[7] R. Sadeghian, "The effects of zoning methods on forecasting of the next earthquake occurrences, according to semi-Markov models," *Annals of Geophysics*, vol. 53, no. 5-6, pp. 83–97, 2010. https://pdfs.semanticscholar.org/bbe6/fbe66c0f51b9daa74235a397e6da0aef8f33.pdf.

[8] A. S. N. Alarifi, N. S. N. Alarifi, and S. Al-Humidan, "Earthquakes magnitude predication using artificial neural network in northern red sea area," *Journal of King Saud University-Science*, vol. 24, no. 4, pp. 301–313, 2012. https://doi.org/10.1016/j.jksus.2011.05.002.

[9] R. Sadeghian, "Forecasting time and place of earthquakes using a semi-Markov model (with case study in Tehran province)," *Journal of Industrial Engineering International*, vol. 8, no. 1, p. 20, 2012. https://link.springer.com/article/10.1186/2251-712X-8-20.

[10] P. C. M. de Bruin, "Semi-markov model applied to the estimation of earthquake occurrences," B.S. thesis, 2019. https://studenttheses.uu.nl/bitstream/handle/20.500.12932/32997/Thesis%20(Patricia%20de%20Bruin).pdf.

[11] S. Kim and H. Kim, "A new metric of absolute percentage error for intermittent demand forecasts," *International Journal of Forecasting*, vol. 32, no. 3, pp. 669–679, 2016. https://doi.org/10.1016/j.ijforecast.2015.12.003.

[12] S. Widiyantoro, E. Gunawan, A. Muhari, N. Rawlinson, J. Mori, N. R. Hanifa, S. Susilo, P. Supendi, H. A. Shiddiqi, A. D. Nugraha, *et al.*, "Implications for megathrust earthquakes and tsunamis from seismic gaps south of java indonesia," *Scientific reports*, vol. 10, no. 1, p. 15274, 2020. https://www.nature.com/articles/s41598-020-72142-z.