

# Java Language Program Application on Predicting the Next Earthquake in Indonesia

Ario Muhammad and Ratih Nur Esti Anggraini

**Abstract**—This paper applied the use of Java programming language to create a program named Java Earthquake Program (JEP) used to forecast the next earthquake in Indonesia. The 133 earthquake events from United States Geological Survey (USGS) occurred in the period of 1954 to 2012 were used to build the program. Three probabilistic models were used to model the earthquake data in Indonesia which are normal, gamma, and beta distributions. The goodness fit tests result from Java earthquake program (JEP) confirmed that gamma distribution is reliable to model the earthquake data in Indonesia. The calculation result of maximum conditional probability from JEP indicated that the next earthquake will be occurred on the date range of August 27, 2012 to February 19, 2013 with  $M_w \geq 5.0$ . This exactly confirmed that JEP is reliable to predict the earthquake recurrence time in Indonesia.

**Index Terms**—Java earthquake program (JEP), gamma, earthquake recurrence time.

## I. INTRODUCTION

Geographically, Indonesia is located in the seismically active zone and in the region of active and potentially active volcanoes. This makes Indonesia becomes one of regions in the world which is hit regularly by the significant earthquake. At the same time, more than 200 million people lives in Indonesia. With the higher density population in most of the earthquake zone areas, some significant earthquake events previously caused in the great loss of life, social and economic.

To reduce the great loss of life, social, and economic due to the earthquake, the scientific community has developed the methods to predict the recurrence time of the next earthquake in seismically active zone areas. One of the methods to predict the time of next earthquake is probabilistic analysis. This method was firstly developed by Utsu [1], [2], Rikitake [3], and Hagiwara [4]. Nowadays, the researcher has developed the probabilistic analysis to predict the time of next earthquake in some specific areas such as Mexico, Mexico, Turkey, Teheran, Italy, Los Angeles and Taiwan [5]-[9].

The use of probabilistic analysis on predicting the recurrence time of earthquake needs to conduct the mathematical computations. Most of the researchers use the commercial programs to perform the probability model in their works such as MATLAB [10] and Ms. Excel [11].

However, those commercial programs have the limitation on advancing the statistical analysis and at the same time, it is not free. For the developing country like Indonesia, the development of free software from some basic language programs is extremely important. In this paper, the free Java programming language was used to build the preliminary program conducted the earthquake prediction time using probabilistic analysis. This program was then named as Java Earthquake Program (JEP). Three theoretical probabilistic distributions are used in JEP to model the earthquake data in Indonesia, which are normal, gamma, and beta distributions. The Indonesian earthquake data with  $M_w \geq 5.0$  was used in this work whereas the probability models were evaluated by goodness-of-fit test to find the appropriate model represented the earthquake data. Conditional probability of the chosen model is then employed to estimate the prediction time of the next earthquake.

## II. REVIEW OF PROBABILISTIC ANALYSIS

The Indonesia earthquake data with  $M_w \geq 5.0$  from 1854 to 2012 was used to develop the JEP. The data were taken from United States Geological Survey (USGS). Three theoretical probability distributions were employed to model the Indonesia earthquake data. Those three probability models and their properties are shown in Table I. Those three models are tested by two different goodness-fit-tests, Kolmogorov-Smirnov (KS) and Chi-Square tests. Once the JEP found the appropriate model representing the data distribution of Indonesia earthquake then the earthquake would be forecasted. The forecasting process was done by calculating the maximum conditional probability.

TABLE I: EMPLOYED STATISTICAL CUMULATIVE DISTRIBUTION FUNCTION AND THEIR PROPERTIES

Type	CDF*	Mean	Variance
Normal	$\frac{1}{2} \left[ 1 + \operatorname{erf} \left( \frac{x - \mu}{\sigma\sqrt{2}} \right) \right], x \in R$	$\mu$	$\sigma^2$
Gamma	$\frac{\gamma \left( k, \frac{x}{\theta} \right)}{\Gamma(k)}$	$\frac{\alpha}{\gamma}$	$\frac{\alpha}{\gamma^2}$
Beta	$I_x(\alpha, \beta)$	$\frac{\alpha}{\alpha + \beta}$	$\frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$

\*Cumulative distribution function

The Kolmogorov-Smirnov (K-S) test is based on the empirical distribution function (ECDF). Once we have the data with  $n$  ordered data points  $x_1, x_2, \dots, x_n$  then the ECDF is expressed as the following:

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I_{x_i \leq x} \quad (1)$$

Manuscript received January 5, 2013; revised April 5, 2013.

Ario Muhammad is with the Department of Civil Engineering, the Narotama University, Surabaya, Indonesia (e-mail: ariomust@gmail.com).

Ratih Nur Esti Anggraini is with Department of Informatics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia (e-mail: ratih.nea@gmail.com).

where  $I_{X_i \leq x}$  is the indicator function, equal to 1 if  $X_i \leq x$  and equal to 0 otherwise [12]. The probability model will be acceptable if the maximum difference,  $D_{n(max)}$ , of cumulative distribution function of theoretical probability model,  $F(x)$ , and ECDF is less than the critical value,  $D_n^\alpha$ , associated with the size of observed data and level of significance described by the following equation:

$$D_{n(max)} = \max |F_x - F_{n(x)}| < D_n^\alpha \quad (2)$$

The level of significance used to find the critical value for K-S test is 0.1. The critical value at the level of significance of 0.1 used in this research is found to be 0.1634. On the other hand, the framework of Chi-Square test compares the frequencies of theoretical probability model ( $t_i$ ) and the frequencies of observed model ( $o_i$ ). The summation of the difference ( $S$ ), between the observed and probability model is defined as the chi square. The probability model used to represent the Indonesia earthquake data will be acceptable if the summation value is less than chi square distribution value,  $\lambda^2$ , as presented in Eq. (3). This chi-square distribution value is determined based on the observed data size and level of significance. The level of significance of 0.1 is employed in this work as same as used for K-S test and it is confirmed that the chi square distribution value ( $\lambda^2$ ) is found to be 52.95.

$$S = \sum_{i=1}^n \frac{(o_i - t_i)^2}{t_i} < \lambda^2 \quad (3)$$

Based on the previous investigations conducted by Utsu [1], Hagiwara [4], Ferraes [5], Jafari [7], Tripathi [13], and Orfanogiannaki and Papadopoulos [14] the mode of conditional probability can be used to predict the recurrence time of the earthquake. The maximum value of conditional probability in this paper is defined as the prediction time ( $P_t$ ) of the next earthquake. The prediction time equation of each model is presented in Table II.

TABLE II: PREDICTION TIME OF EACH PROBABILITY MODEL

Prediction time ( $P_t$ )	
Type	
Normal	$P_t = \mu$
Gamma	$P_t = \frac{\alpha - 1}{\gamma}$
Beta	$P_t = \frac{\alpha - 1}{\alpha + \beta - 2}$

The square error calculation proposed by Benjamin and Cornell [15] is used in this work to evaluate the result of recurrence time prediction. Let the variance of the recurrence time from the earthquake catalog is defined as  $\sigma^2$  and the mean of the recurrence time of the earthquake is expressed as  $\mu$ , then the square error can be defined as:

$$\varepsilon^2 = \sigma^2 - (\mu - P_t)^2 \quad (4)$$

### III. JAVA EARTHQUAKE PROGRAM DESCRIPTION

#### A. Flowchart of the Built Program

Fig. 1 describes the flowchart of JEP. The data taken from USGS was firstly compiled to the txt before it would be used for building the JEP. The probabilistic parameters used to perform goodness fit tests and earthquake prediction time are cumulative distribution function (CDF) and frequency of

each probability model. The CDF was calculated previously before frequency of the earthquake in each magnitude was defined. The development of JEP for calculating the CDF and frequency was based on their equation as shown in Table I. Additionally, the K-S and Chi-Square test were included in the JEP using their equation as written in Eq. (2) and (3).

The JEP will start to determine the appropriate model representing the Indonesia earthquake data after evaluating the result from those two goodness fit tests. The mode of conditional probability of the chosen model will then automatically perform in JEP to calculate the earthquake prediction time. The built program of maximized conditional probability was conducted using the equation in Table II. The range period of earthquake recurrence time is run in JEP by calculating the error of the probability model using Eq. (4). From this step, then JEP can directly announce the range time of the next earthquake in Indonesia.

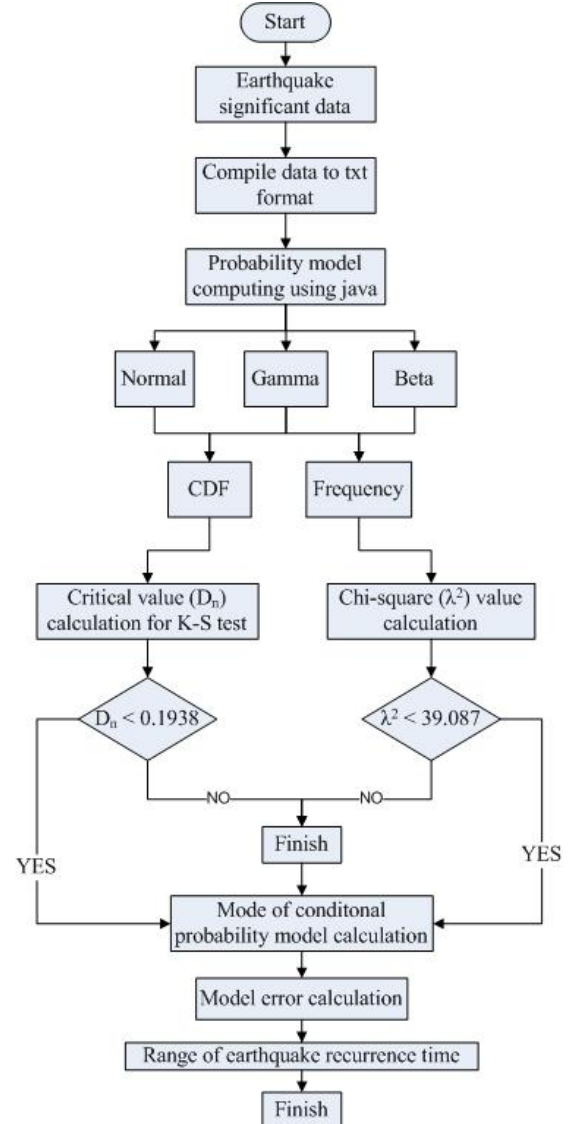


Fig. 1. Procedure of earthquake recurrence time prediction.

#### B. Input, Running Program, and Output

Fig. 2 shows the screen view of Java-based program. The figure performs the calculation of PDF and ECDF of observed data and the determination of CDF and frequency for each probability model. The observed data from USGS only needs the total number of data ( $N$ ) in order to calculate probability density function (PDF) and CDF.

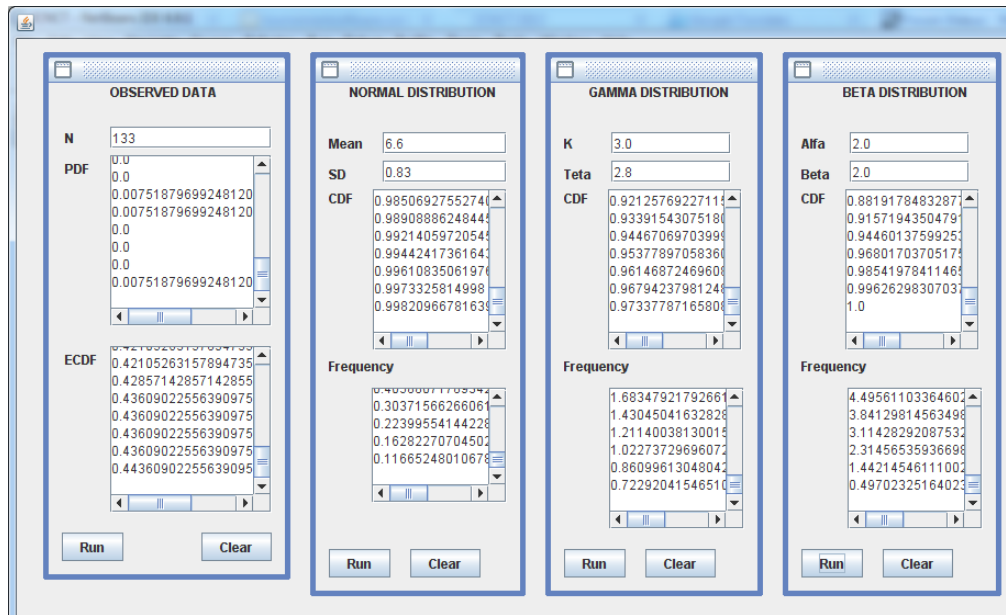


Fig. 2. Screen view of Java-based program to perform the probabilistic analysis

The input parameters to calculate CDF and frequency of normal distribution are N, mean, and standard deviation (SD). The N, mean, and SD of the data are 59, 6.60, and 0.83 respectively. The value of k and teta are the input of gamma distribution. For predicting the recurrence time of Indonesia earthquake, the value of k and teta used in this work are 3.0 and 2.8. in addition, beta distribution needs the value of alfa and beta as the input parameters to calculate CDF and frequency. In this paper, 2.0 is the value for both alfa and beta. To run the program of JEP, the “Run” button is just simply clicked whereas to clear the input parameters the “Clear” button is used.

#### IV. RESULT AND DISCUSSIONS

##### A. Result of Goodness Fit Tests

TABLE III: SUMMARY OF TWO GOODNESS FIT TESTS

Probability Model	Normal	Gamma	Beta
At manitude	7.0	5.8	7.0
$D_{n(max)}$	0.24	0.05	0.23
The suitability	No	Yes	No
$\lambda^2$	85.33	51.70	67.79
The suitability	No	Yes	No

The comparison value of cumulative probability models and empirical data using to simulate the K-S test is presented in Fig. 3. The maximum critical value ( $D_{n(max)}$ ) of each model and its comparison to the limit of critical value at the level of significance of 0.1 is shown in Table III. The table precisely shows that the suitable probability model to represent the distribution of Indonesia earthquake data is gamma distribution model. The maximum critical value ( $D_{n(max)}$ ) of gamma distribution is 0.05 which is less than 0.1634 as the critical value ( $D_n^\alpha$ ) at the level of significance 0.1. Contrary with the maximum critical value ( $D_{n(max)}$ ) of gamma distribution, the maximum critical value ( $D_{n(max)}$ ) of normal

and beta distributions are found to be 0.24 and 0.23 respectively. These values are greater than the critical value ( $D_n^\alpha$ ) at the level of significance 0.1 indicating that those two models are inappropriate to simulate the Indonesia earthquake data.

The comparison of observed and theoretical probability model frequencies which will be representing the chi square goodness fit test is presented in Fig. 4 while the value of chi square distribution of each model is described in Table 3. The chi-square goodness fit test result also shows the same tendency with the result of K-S test. The chi square distribution value ( $\lambda^2$ ) of gamma distribution is found to be 44.77 (Table III) which is less than the chi square distribution value ( $\lambda^2$ ) at the level of significance of 0.1. The chi square distribution values of other 2 models are 83.30 and 68.15 represented the model of normal and beta respectively. Noted that for both beta model, the chi square distribution value ( $\lambda^2$ ) calculation does not include the zero value since it results in the infinity number hence those values above only use for non zero frequency. Based on those two test results, it is exactly confirmed that the gamma distribution is the appropriate model to represent the Indonesia earthquake data.

##### B. Earthquake Recurrence Time Prediction

Since the result of two goodness tests shows that gamma distribution model is an appropriate model to represent the Indonesia earthquake data, this model is then be used to predict the earthquake recurrence time. The data must be regroup from the first earthquake occurrence on November 2, 1954 to the last earthquake event on July 23, 2012. The recurrence time is then calculated between two earthquake events start from the first to the last. The mean ( $\mu$ ) of the earthquake data is 5.25 days and standard deviation ( $\sigma$ ) is found to be 2.57 resulting on the variance ( $\sigma^2$ ) of 6.60. Table 2 expressed the prediction time ( $P_t$ ) of gamma distribution which is determined by the following equation:

$$P_t = \frac{\alpha-1}{\gamma} \quad (5)$$

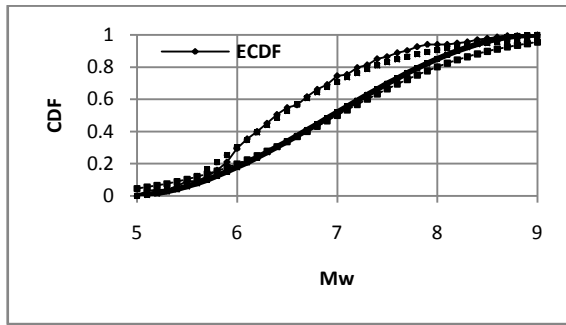


Fig. 3. Comparison of observed and five theoretical cumulative probabilities

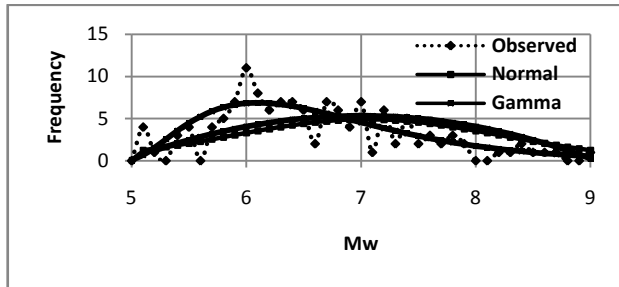


Fig. 4. Comparison of observed and five theoretical frequencies

Eq. (5) precisely indicates that the value of  $\alpha$  and  $\gamma$  must be firstly calculated before the earthquake prediction time is estimated. Those values can be determined using the equation of mean ( $\mu$ ) and variance ( $\sigma^2$ ) of gamma distribution taken from Table 1. Using the equation of both mean and variance of gamma distribution, the  $\alpha$  and  $\gamma$  are calculated which are 4.17 and 0.79 respectively. Those two values are then employed to Eq. (5) in order to calculate the earthquake prediction time. It is found that the prediction time ( $P_t$ ) is 4 months with error of 2.86 months (86 days). The result of mode conditional probability of gamma distribution showed that the recurrence time of next earthquake was predicted occurred on the date range of August 27, 2012 to February 19, 2013 with  $M_w \geq 5.0$ .

## V. CONCLUSIONS

The application of java programming language to predict the recurrence time of Indonesia earthquake has successfully conducted in this paper. The goodness fit tests result from Java earthquake program (JEP) confirmed that gamma distribution is reliable to model the earthquake data in Indonesia. The calculation result of maximum conditional probability from JEP indicated that the next earthquake will be occurred on the date range of August 27, 2012 to February 19, 2013 with  $M_w \geq 5.0$ . This exactly confirmed that JEP is reliable to predict the earthquake recurrence time in Indonesia.

## REFERENCES

- [1] T. Utsu, "Large earthquakes near Hokkaido and the expectancy of the occurrence of a large earthquake of Nemuro," *Report of the Coordinating Committee for Earthquake Prediction*, 1972a, vol. 7, pp. 7-13.
- [2] T. Utsu, "Aftershocks and earthquake statistics (IV)," *Journal of the Faculty of Science, Hokkaido University Series VII Geophysics*, 1972b, vol. 4, pp. 1-42.
- [3] T. Rikitake, "Probability of an earthquake occurrence as estimated from crustal strain," *Tectonophysics*, 1974, vol. 23, pp. 299-312.
- [4] Y. Hagiwara, "Probability of earthquake occurrence as obtained from a Weibull distribution analysis of crustal strain," *Tectonophysics*, 1974, Vol. 23, pp. 323-318.
- [5] S. G. Ferraes, "A probabilistic prediction of the next strong earthquake in the Acapulco-San Marcos segment, Mexico," *Geofisica Internacional*, 2005, vol. 44, pp. 347-353.
- [6] V. Yilmaz, M. Erisoglu, and H. E. Celik, "Probabilistic prediction of the next earthquake in the NAFZ (North Anatolian Fault Zone), Turkey," *Dogus Universitesi Dergisi*, 2004, vol. 5, pp. 243-250.
- [7] M. A. Jafari, "Statistical prediction of the next great earthquake around Tehran, Iran," *Journal of Geodynamics*, 2010, vol. 49, pp. 14-18.
- [8] S. Laura and H. Lawrence, "A test of physical-based strong ground motion prediction methodology with the 26 September 1997,  $M_w = 6.0$  Colfiorita (Umbria-Marche sequence), Italy earthquake," *Tectonophysics*, 2009, vol. 476, pp. 145-158.
- [9] J. P. Wang, Y. M. Wu, T. L. Lin, and L. Brant, "The uncertainties of a Pd3-PGV onsite earthquake early warning system," *Soil Dynamic and Earthquake Engineering*, 2012, vol. 36, pp. 32-37.
- [10] G. Preethi and B. Santhi, "Study on techniques of earthquake prediction," *International Journal of Computer Application*, 2011, vol. 29, no. 4, pp. 55-58.
- [11] J. P. Wang and D. Huang, "RosenPoint: A Microsoft Excel-based program for the Rosenbluth point estimate method and an application in slope stability analysis," *Computer & Geosciences*, 2012.
- [12] I. M. Chakravarti, R. G. Laha, and J. Roy, *Handbook of Methods of Applied Statistics*, John Wiley and Sons, 1967, ch. 1, pp. 392-394.
- [13] J. N. Tripathi, "Probabilistic assessment of earthquake recurrence in the January 26, 2001 earthquake region of Gujarat," *Indian Journal of Seismology*, vol. 10, pp. 119-130, 2006.
- [14] K. Orfanogiannaki and G. A. Papadopoulos, "Conditional probability approach of the assessment of Tsunami potential: application in three Tsunamigenic regions of the Pacific Ocean," *Pure Appl. Geophysics*, 200, vol. 164, pp. 593-603.
- [15] J. R. Cornell and C. A. Benjamin, *Probability, Statistic, and Decision for Civil Engineers*, McGraw-Hill, 1970.



Ario Muhammad was born in North Moluccas, Indonesia on September 14, 1987. He received his bachelor in civil engineering from Universitas Muhammadiyah Yogyakarta (UMY) in 2009, Indonesia and received National Taiwan University of Science and Technology (NTUST) Scholarships for pursuing his master degree in Construction Engineering. He graduated from NTUST at 2011. Now he is working as a junior lecturer at Department of Civil Engineering, Narotama University in Surabaya – Indonesia. His research interest is in probabilistic analysis of earthquake engineering.



Ratih Nur Esti Anggraini was born in East Java, Indonesia on December 10 1984. She received her bachelor in informatics at the Department of Informatics, Institut Teknologi Sepuluh Nopember, Surabaya (ITS) Indonesia in 2008. In 2010, she was then awarded the scholarship to pursue her master degree in Computer Science at the National Taiwan University of Science and Technology (NTUST). She graduated from NTUST in 2012. Now she is working as a junior lecturer at Department of Informatics, Institut Teknologi Sepuluh Nopember, Surabaya (ITS) – Indonesia. Her research interest is in software engineering.