Analysis of Landslide Susceptibility Zone using Frequency Ratio and Logistic Regression Method in Hambalang, Citeureup District, Bogor Regency, West Java Province

S Hidayat¹, H Pachri¹, and I Alimuddin¹

¹ Department of Geological Engineering, Hasanuddin University, Makassar, Indonesia
sandrihidayat@gmail.com, hendrapachri@hotmail.com, iailimuddin@hotmail.com

Abstract. Hambalang area is one of the regions susceptible to landslide events. This is due to unstable geological conditions and high rainfall. Administratively, research area included in Citeureup District, Bogor Regency, West Java Province. Astronomically, research area is located at the coordinates 106°51'30" - 106°53'30" East Longitude and 06°32' - 06°34' South Latitude. This study aims to determine the geological conditions of the study area and conduct susceptibility zoning in the Hambalang area using the method of Frequency Ratio and Logistic Regression. The geological conditions of the study area consist of geomorphology in the form of corrugated and flatland morphological units, stratigraphy consisting of claystone units, andesite units and alluvial units, and the structure of the Sumurbatu anticline. The parameters used to analyze the causes of landslide are slope, geology, rainfall, soil type, land use and distance from rivers. There were 19 landslide events identified in the study area. The relationship between occurrences of landslide and the parameters mentioned earlier can be quantified using the value of Frequency Ratio and Logistic Regression. Based on the results of the validation with the Frequency Ratio method, the AUC value of 0.6854 which shows the model of landslide susceptibility based on the selection of parameters and the adequacy of the landslide event data is good. The value of the frequency ratio of the model is divided into three zones of susceptibility, namely high (42.30%), moderate (20.34%) and low (37.36%). Based on the results of the validation with the Logistic Regression method, the AUC value of 0.762 also shows the landslide susceptibility model based on the selection of parameters and the adequacy of the landslide event data so that the Logistic Regression value of the model is divided into three susceptibility zones, namely high (56.89%), moderate (19.53%) and low (23.68%).

Keywords: geological conditions, landslide susceptibility, frequency ratio, logistic regression

1. Introduction
Landslides susceptibility are movements of rock or soil mass on a slope due to the influence of gravitational forces. Movement of rock or soil mass occurs because of a disruption to the balance of shear strength and shear stress that works on a slope. The force imbalance is caused by the force from outside the slope which causes the magnitude of the launch force on a slope to be greater than the retaining force [6].

Based on data from the Center for Volcanology and Geological Disaster Mitigation (PVMBG) in February 2018 about potential areas of landslide in West Java Province, all 27 cities / regencies in...
West Java each have several sub-districts that potentially for landslide to occur with potential movements including in the middle-high level [2].

One of the districts in West Java that has the potential for landslide is Bogor Regency. Based on data from the Bogor City Disaster Management Agency throughout January - October 2018, there were 129 landslides [7].

Hambalang Area, Citeureup District is one of the Bogor Regency areas that experienced landslide events as evidenced by the destruction of 1 athlete building unit in 2012 [1]. In addition, based on the 2018 PVMBG report, the occurrence of landslide in the Hambalang area threatened 70 heads of the family. Based on these events, efforts should be made to provide information about the description of the susceptibility level of landslide in the Hambalang area of Citeureup District, Bogor Regency [2].

Information on susceptibility level of landslide can be presented using a Geographic Information System combined with statistical calculations. The method used in this study is Frequency Ratio and Logistic Regression method. This method is a method that is useful to determine the influence level of each parameter on landslide events. Based on the above, it is necessary to conduct research on Landslide Susceptibility Zones.

2. Material and Methods

2.1 Study Area and Data
The research area is included in the Citeureup District, Bogor Regency, West Java Province. Located at coordinates 106°51’30” - 106°53’30” East Longitude and 06°32’ - 06°34’ South Latitude. Geomorphology of the research area in the form of corrugated hills and plain morphological units, stratigraphy consists of claystone units, andesite units and alluvial units, as well as the structure of the Sumurbatu anticline.

![Figure 1. Map of research location](image-url)
Table 1 Data Sources and Landslide Parameters

| DATA TYPE        | YEAR | DESCRIPTION                        | SOURCE | SCALE          | OUTPUT MAP                              |
|------------------|------|------------------------------------|--------|----------------|-----------------------------------------|
| DEM TerraSAR-X   | 2017 | • Elevation data                   | PVMBG  | Resolution 9 m | 1. Slope Map                            |
| Bogor Indonesia  |      | • Raster format                    |        |                | 2. Distanc from river Map                |
| Topographic      |      |                                    |        |                |                                         |
|                  |      | • Contour                          | PVMBG  | 1:25.000       | Observation Map                         |
|                  |      | • Way                              |        |                |                                         |
|                  |      | • Settlement                        |        |                |                                         |
|                  |      | • River                            |        |                |                                         |
|                  |      | • Elevation                        |        |                |                                         |
|                  |      | Shapefile vector format            |        |                |                                         |
|                  |      |                                    |        |                |                                         |
| Lithology        | 2017 | Shapefile vector format            | PVMBG  | 1:100.000      | Lithology Map                           |
| Land Use         | 2017 | Shapefile vector format            | PVMBG  | 1:100.000      | Landuse Map                             |
| Soil type        | 2017 | Shapefile vector format            | PVMBG  | 1:100.000      | Soil type Map                           |
| Rainfall intensity| 2016-2017 | • Bogor Regency monthly rainfall data | BPS    |                | Rainfall Intensity Map                  |
|                  |      | • Coordinate of rainfall stations  |        |                |                                         |

The factors that control landslides consist of 6 factors, namely slope, distance from rivers, lithology, land use and rainfall. These factors have been transformed into raster data using GIS. The slope factor is extracted from the DEM TerraSAR-X image with a resolution of 9 m. slope in the study area varied which was then divided into 5 zones based on zones 0˚ - 2˚, zones 2˚ - 4˚, zones 4˚ - 8˚, zones 8˚ - 16˚ Zones 16˚ - 35 °. Landuse of research land consists of settlements, plantations, rice fields, lakes, shrubs and rivers. The research area is divided into 3 units lithology, namely claystone units, andesite units and alluvial units. The distance from the river in the study area is divided into 8 zones, namely zones 0-100 m, zones 100-200 m, zones 200-300 m, zones 300-400 m to zones 700 - 800 m. The type of soil consists of two, namely clay and sandy clay soil. Rainfall is divided into 10 classes 3454 mm - 4519 m. All of these factors can be seen in figure 2 – figure 7.
2.2 Method
2.2.1 Frequency Ratio Analysis
The landslide parameters are transformed into raster data with a resolution of 9 m as shown in Table 2. The landslide parameter is overlapped with a landslide event point. Each zone of the factor has two attributes, namely the number of pixels in the zone and the number of landslides in the zone. These two attributes are then used to calculate the frequency ratio value of each zone according to formula (1) [8].

\[ FR = \frac{\sum_{i=1}^{n} Di/Ai}{\sum_{j=1}^{m} \sum_{i=1}^{n} Ai} \]  \[ \text{.........(1)} \]

FR = Frequency Ratio  
Di = number of landslide points in a class in certain factors  
Ai = number of areas in a class in certain factors
After the FR value of each landslide parameter is calculated, then each map is overlapped to produce a LHI (Landslide Hazard Index) value such as formula (2) [5].

\[ LHI = FR_1 + FR_2 + FR_3 + \ldots + FR_n \ldots \] (2)

Then the LHI value is classified into 3 zones, namely the low hazard zone, the medium hazard level zone and the high hazard level zone which is presented in the form of a Landslide Susceptibility map. The landslide susceptibility map is validated by the point of occurrence of landslide. This validation shows how well the model predicts landslides. The results of this validation will produce predictive accuracy values based on AUC (Area Under Curve).

2.2.2 Logistic Regression Analysis
Logistic regression analysis is done by combining all the causes of landslides and landslide events. The combined data results are then converted into excel tables which are then carried out by the logistic regression analysis using SPSS software. The logistic regression calculation uses the following formula:

\[ Y = b_0 + b_1x_1 + b_2x_2 + \ldots + b_nx_n \]

where Y is the dependent variable representing the presence or absence of landslides, b0 intercept of the model, b1 ... bn the regression coefficient of each landslide parameter, x1 ... xn is the landslide parameter.

The results of the regression analysis are in the form of a constant value and the coefficient value of each landslide parameter. Then to test the accuracy of the regression model, validation is done. Validation is done through the making of the ROC (Relative Operating Characteristic).

Logistic regression, which is a multivariate analysis model, is useful for predicting the presence or absence of characteristics or results based on the value of a set of predictor variables. The advantage of logistic regression is that, through adding appropriate link functions to ordinary linear regression models, variables can be continuous or discrete, or a combination of the two types, and they do not always have a normal distribution [4].

3. Analysis and Discussion
3.1 Frequency Ratio
The frequency ratio method is based on the relationship between the spread of the landslide point and each of the landslide control factors. This relationship explains the contribution of each factor to the occurrence of landslides. The value of the frequency ratio of each landslide control factor zone is shown in Table 2.

| Factor | Zone | Class | Factor | Landslide | Tot_LS | Class_Dens | Map_Dens | FR |
|--------|------|-------|--------|-----------|--------|------------|----------|----|
| Slope  | 0 - 2| 14115 | 0.08   | 2         | 2.08   | 0.00015    | 0.00012  | 1.24|
|        | 2 - 4| 20485 | 0.12   | 2         | 2.12   | 0.0001    | 0.00012  | 0.87|
|        | 4 - 8| 52937 | 0.32   | 10        | 10.32  | 0.00019    | 0.00012  | 1.63|
|        | 8 - 16| 69193 | 0.41   | 4         | 4.41   | 0.00006    | 0.00012  | 0.54|
|        | 16 - 35| 10969 | 0.07   | 1         | 1.07   | 0.0001    | 0.00012  | 0.81|
| Rainfall| 3454-3608| 7637 | 0.05   | 0         | 0.05   | 0.0001    | 0.00012  | 0.05|
|        | 3608-3708| 15329 | 0.09   | 2         | 2.09   | 0.00014    | 0.00012  | 1.15|
|        | 3708-3795| 22461 | 0.13   | 2         | 2.13   | 0.00009    | 0.00012  | 0.80|
|        | 3795-3883| 24320 | 0.14   | 2         | 2.14   | 0.00009    | 0.00012  | 0.74|
The value of the frequency ratio of each zone in the study area varies from 0.050 - 2.03. In slope factors, slopes with more steep slopes have a higher probability of landslides. The slope of 4-8 has a frequency ratio value of 1.63 which indicates the probability of a higher landslide occurrence. The higher the slope, increasing the shear stress of the slope constituents so that slope stability is disturbed and the potential for landslides [5].

Determination of the landslide susceptibility zone is based on the parameters that influence the landslide event in the study area. Determination of these parameters is based on the AUC (Area Under Curve) value. The higher the AUC value, the higher the level of influence of parameters on landslide events. The AUC threshold value is 0.5 so the parameters that have an AUC value of less than 0.5 will be ignored for determining the landslide susceptibility zone. The AUC value of each parameter can be seen in table 3.
Table 3. AUC values for each parameter validation test results

| Parameter          | AUC Value  |
|--------------------|------------|
| Slope              | 0.634466   |
| Rainfall           | 0.649031   |
| Lithology          | 0.611689   |
| Type of Soil       | 0.524925   |
| Land Use           | 0.605526   |
| Distance from River| 0.588419   |

Based on the AUC value of the parameters above, it can be seen that there are four parameters that can be validated. The purpose of this validation is to ensure the landslide susceptibility model of the selected parameters can be accounted for. Validation is done by first calculating the value of LHI. To calculate the LHI value, each frequency ratio value for each zone is calculated by formula (2). The LHI values found in the research area vary. The LHI value is then classified into 3 zones, namely low zone (37.36%), moderate zone (20.34%) and high zone (42.30%) shown in Figure 8.

![Landslide Susceptibility Map using The Frequency Ratio Method](image)

Figure 8. Landslide Susceptibility Map using The Frequency Ratio Method

The results of the landslide susceptibility analysis are then validated using existing landslide data. The LHI value obtained is then divided into 256 classes from the highest to the lowest. Each LHI class is calculated as a percentage of the number of landslides and its cumulative percentage such as the rate
change curve made with the LHI value as the x axis and the cumulative percentage of landslide events as the y axis as shown in Figure 9. AUC (Area Under Curve) is calculated from the rate change curve indicating predictive accuracy from the frequency ratio model. The AUC value ranges from 0.6854, which means that the predicted accuracy value is 68.54%. The AUC value of the curve is > 0.5 which states that the frequency ratio model has a good predictive accuracy value and can be applied in the study area.

![Figure 9. Validation Test Model of Landslide Susceptibility](image)

### 3.2 Logistic Regression

Logistic regression analysis is used to determine the relationship of landslide events to factors that are parameters of landslides. This statistical analysis consists of two variables, namely the dependent variable and the independent variable. The dependent variable or dependent variable in this study is a landslide event that is categorical dichotomy with a value of 1 for landslide pixels and 0 for pixels without landslides. While the independent variables or independent variables used in this study are the factors that become the parameters in the analysis of landslide events. The independent variables in this study are slope, lithology, rainfall, soil type, land use and distance from the river. The dependent variable used in the logistical regression analysis is the location of landslide events in the form of shapefile data. This data is then converted into raster data that has a resolution of 9 meters with an attribute of value 1 for the location of the landslide and a value of 0 for locations that do not occur landslides.

First, logistic regression calculation is done by combining the dependent variable and the independent variable. The results of this process are in the form of raster data where each pixel contains information about the class or subparameter that causes landslide and the presence or absence of landslide events.

The next step is data conversion, where raster data is converted into ASCII data format which is then imported into the SPSS application for logistic regression calculations. The purpose of logistic regression analysis is to find out the parameters that are the main causes of landslide events in the study area. The results of the analysis in table 4 show that the parameters that have a significant effect on the occurrence of landslide in the study area are landuse and lithology.

#### Table 4

| Landslide Parameter         | β      | Sig. | Exp(β) |
|-----------------------------|--------|------|--------|
| Land Use                    | -0.254 | 0.024| 0.775  |
| Slope                       | 0.007  | 0.969| 1.007  |
| Distance from the river     | -0.263 | 0.215| 0.769  |
| Lithology                   | 1.038  | 0.033| 2.823  |
| Landslide Parameter   | β       | Sig.  | Exp(β) |
|----------------------|---------|-------|--------|
| Rainfall Intensity   | -0.090  | 0.384 | 0.914  |
| Constants            | -4.675  | 0.004 | 0.009  |

Based on the table above logistic regression analysis, the regression model making only involves two parameters, namely land use parameters and lithology parameters. So that the equation of the logistic regression model is included in the ground motion susceptibility analysis, namely:

\[ Z = -4.675 + (-0.254 \times \text{Land Use}) + (1.038 \times \text{Lithology}) \]

The logistic regression model that has been done becomes a reference for calculating the probability of landslide events in the study area. Calculations are carried out using the technique overlay of each parameter causing landslides. Furthermore, the landslide susceptibility zone (figure 10) was divided into 3 classes, namely the high zone (56.89%), the moderate zone (19.53%) and the low zone (23.58%).

![Landslide Susceptibility Map](image)

**Figure 10.** Landslide susceptibility map using the logistic regression method

The accuracy of the landslide susceptibility prediction model is done by validating the landslide zoning map. Validation is done by making the ROC curve by plotting the sensitivity value (True
positive rate) on the y axis and value 1 - specificity (positive negative rate) on the x axis. The results of plotting values of sensitivity and 1 - specificity are presented in table 5 and figure 11.

Table 5. Sensitivity values and 1-Specific landslide models

| Sensitivity | 1 - Specificity | Sensitivity | 1 - Specificity |
|-------------|----------------|-------------|----------------|
| 1,000       | 1,000          | 0,474       | 0,204          |
| 0,947       | 0,840          | 0,421       | 0,157          |
| 0,895       | 0,511          | 0,368       | 0,141          |
| 0,842       | 0,422          | 0,316       | 0,129          |
| 0,789       | 0,380          | 0,263       | 0,128          |
| 0,737       | 0,340          | 0,211       | 0,074          |
| 0,684       | 0,254          | 0,158       | 0,070          |
| 0,632       | 0,250          | 0,105       | 0,060          |
| 0,579       | 0,224          | 0,053       | 0,055          |
| 0,526       | 0,208          | 0,000       | 0,000          |

Figure 11. Ploting results of sensitity values and 1- specificity regression models on the ROC curve

Based on the results of plotting on the ROC curve the value of the area under the curve (Area Under Curve / AUC) was obtained. The calculation results obtained value 0.762. the accuracy of the prediction models produced is in sufficient categories, so that the model is feasible to be applied to the research area.

4. Conclusion
Based on frequency ratio analysis, the parameters that affect the occurrence of landslides in the study area consist of four parameters, namely slope, rainfall, land use and lithology. Meanwhile, based on logistic regression analysis, the parameters that affect the occurrence of landslides in the study area consist of two parameters, namely lithology and land use.
Based on frequency ratio analysis, there are three landslide susceptibility zones, namely a high susceptibility zone with an area of about 42.30%, moderate susceptibility zone with an area of about 20.34% and low susceptibility zone with an area of 37.36% of the research area. Meanwhile, based on logistic regression analysis, the high susceptibility zone has an area of about 56.89%, moderate susceptibility zone with an area of about 19.53% and low susceptibility zone with an area of around 23.58% of the total research area.

References

[1] Center for Volcanology and Geological Disaster Mitigation 2012. Brief report on landslides in Hambalang Village, Citeureup District, Bogor Regency, Bandung, Badan Geologi

[2] Center for Volcanology and Geological Disaster Mitigation 2018. Brief Report on Examination of Landslides in Citeureup District, Bogor Regency, West Java Province, Accessed on November 5, 2018, http://www.vsi.esdm.go.id/index.php/gerakan-tanah/kejadian-gerakan-tanah/1936-laporan-singkat-pemeriksaan-gerakantanah-di-kec-citeureup-kab-bogor-provinsi-jawa-barat.

[3] Lee S and Pradhan B 2007 Landslide Hazard Mapping at Selangor, Malaysia Using Frequency Ratio and Logistic Regression Models, Landslide Vol. 4, p. 33-41.

[4] Lee S dan Sambath T 2006. Landslide susceptibility mapping in the Damrei Romel area, Cambodia using frequency ratio and logistic regression models. Environmental Geology 50 (6), 847–856

[5] Pradhan B 2010 Landslide Susceptibility Mapping of a Catchment Area Using Frequency Ratio, Fuzzy Logic and Multivariate Logistic Regression Approaches, Journal Indian Society Remote Sensing Vol. 38, 301-320 p

[6] Prawiradisastra S 2013 Identification of Landslide Disaster Prone Areas in Lampung Province. Indonesian Science and Technology Journal Vol. 15 No. 1.

[7] Regional Disaster Management Agency 2018 Disaster Event Data. Accessed on November 5, 2018, https://bpbd.kotabogor.go.id/index.php/data-kejadian.

[8] Tazik E, Jahantab Z, Bakhtiari M, Rezaei A and Kazem Alavipanah S 2014 Landslide Susceptibility Mapping by Combining the Three Methods Fuzzy Logic, Frequency Ratio and Analytical Hierarchy Process in Dozain Basin. Tehran. The International Arithmetic of Photogrammetry, Remote Sensing and Spatial Information Sciences.

Acknowledgements

Acknowledgments were submitted to the Center for Volcanology and Geological Disaster Mitigation (PVMBG) for assistance and permission in collecting parameters of landslide events and landslide analysis in the study area.