IOP Conference Series: Earth and Environmental Science

PAPER • **OPEN ACCESS**

Multiple linear regression analysis of remote sensing data for determining vulnerability factors of landslide in PURWOREJO

To cite this article: Sudaryatno *et al* 2020 *IOP Conf. Ser.: Earth Environ. Sci.* **500** 012046

IOP Conf. Series: Earth and Environmental Science **500** (2020) 012046 doi:10.1088/1755-1315/500/1/012046

Multiple linear regression analysis of remote sensing data for determining vulnerability factors of landslide in PURWOREJO

Sudaryatno[1], Prima Widayani[1], Totok Wahyu Wibowo[1], Bayu Aji Sidiq Pramono[1], Zulfa Nur'aini 'Afifah[1], Awit Dini Meikasari[1], Muhammad Rizki Firdaus[1]

[1] Dept. of Geographic Information Science, Universitas Gadjah Mada, Kaliurang Street, Sleman District, 55281 Daerah Istimewa Yogyakarta Regancy, Indonesia

e-mail: sudaryatno@ugm.ac.id

Abstract. Purworejo is one of the potential area that could be experiencing landslides, because the geomorphological conditions which are included in Menoreh Hills are geographically sloping to very steep. Based on the Indonesian Disaster Information Data (DIBI) and the National Disaster Management Agency (BNPB) in the last five years from 2014 to April 2019 there have been 64 landslides in Purworejo. The research on landslide vulnerability mapping has been done with various spatial modeling methods, one of them is using Information Value Model (IVM). There are four landslide factors arranging the model, such as elevation, slope, slope direction and vegetation index (NDVI). The purpose of this research is to determine the most influence factors towards landslide vulnerability levels thorugh remote sensing data. Multiple regression analysis is used to determine the most influential factors. In this research, dependent variable represented by eight landslide factors, and the independent variable is vurnerability level of landslide in Purworejo. The results of this study explain that the predictor variables that most influence the occurrence of landslides in Purworejo are elevations with regression values that are quite dominant among other variables.

1. Introduction

Purworejo is one of the districts in Indonesia that is prone to landslides. Landslides can be found in areas that have steep slope physiographic conditions. This is in accordance with the topography in the western and northern regions of Purworejo. Therefore, modeling of landslide vulnerability is important to do. Landslides are often referred to as slope instability caused by geomorphological factors, especially the slope angle [1].

Research on landslide vulnerability mapping has been carried out by several previous researchers with various spatial modeling methods [2, 3]. The use of the Information Value Model method rooted in statistics for mapping landslide vulnerabilities [4]. The result is quite satisfactory because it is able to produce landslide hazard maps with accuracy of up to 81.8% [5]. IVM is a model developed from information theory to conduct statistical analysis, which considered the factor of landslide parameters. Each parameter used, has the same level of equality. Based on the results of the IVM modeling, further analysis was carried out to find the most influential parameters for landslide events. The statistical analysis used was multiple linear regression. Regression analysis is a statistical technique that can determine the mathematical relationship between the dependent variable (x axis) and predictor variable (y axis). The parameters used are elevation, slope, aspect and NDVI. Those parameters are the value result from IVM modeling, not the original value of image extraction. The purpose of this study was to determine the most influential parameters for landslide events.

2. Study Area

The study area is Purworejo as shown at figure 1 below. Purworejo as Regency in the part of Central Java Province was due to the fact that in the last five years from 2014 to April 2019 have some 65 landslide with the most enormous years of the event is 2019 or have 15 landslide occurrence. Administratively the Purworejo district consists of 16 sub-districts, which are divided into 494 villages. Geographically, Purworejo Regency is located in the south-central part of Java Island, more precisely located at 8º 30 '- 7º 20' South Latitude, and 109º 47' 28”- 111º8' 20” East Longitude. Administratively, the Purworejo borders on several regions at the north side is Wonosobo Regency, Central Java Province, at the west side is Kebumen Regency, Central Java Province, at the east side is Special Region of Yogyakarta Province at the south side is Indian Ocean.

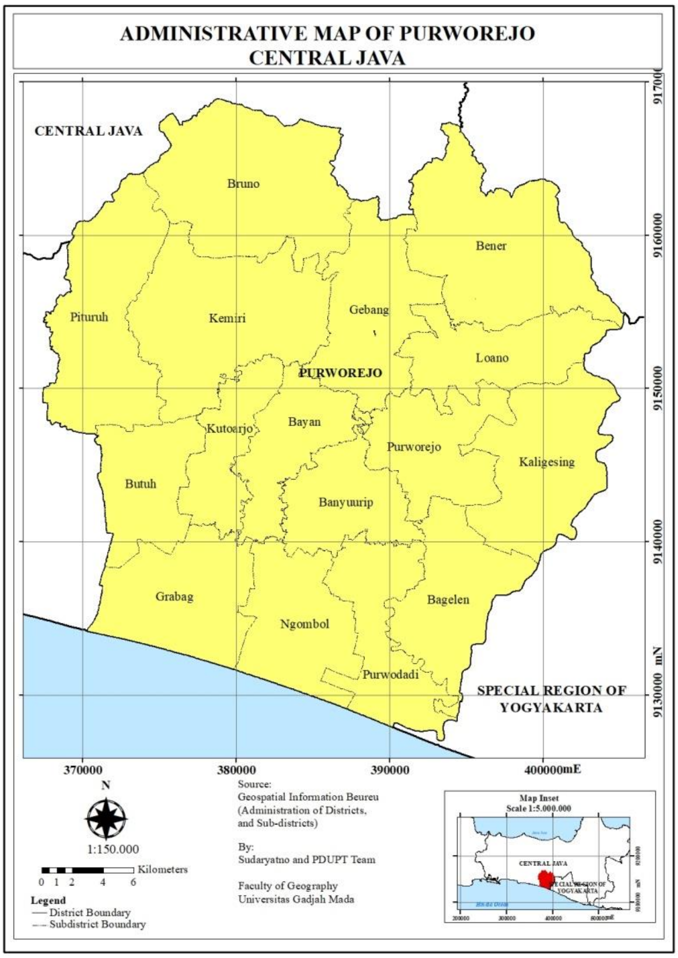

Figure 1. Administrative map of study area.

IOP Conf. Series: Earth and Environmental Science **500** (2020) 012046 doi:10.1088/1755-1315/500/1/012046

3. Methodology

IVM modeling for landslide vulnerability involves 9 parameters, they are; land use, elevation, aspect, slope, distance from fault, distance from road, distance from hydrographic network and vegetation density. Based on these 9 parameters, 4 subjectively selected parameters are considered to be the most influential on landslide events, they are; aspect, slope, elevation and vegetation density. The four parameters are extracted from remote sensing data, where aspect, slope, and elevation are extracted from digital elevation modeling (DEM), while vegetation density is obtained using the NDVI method.

IVM was developed from information theory to carry out statistical analysis. Based on this model, the information value of the landslide parameter factor will be used for landslide modeling. The following formula shows the information value I (xi, H) of each avalanche parameter factor xi (I = 1,2,3, ..., n).

$$I(x_i, H) = ln \frac{N_i/N}{S_i/S} \quad (1)$$

In this case H indicates the possibility of a landslide, S is the total mapping unit, N is the area of the landslide in the study area. Si is the number of mapping units in the presence of xi predisposing factors. Whereas, Ni is the area of landslides in the presence of xi predisposing factors. The total information I of each mapping unit can be calculated by summing the information values of all predisposing factors. The result could be a positive or negative value, with the following class. If I <0, the probability of a landslide is lower than the average; If I = 0, the probability of a landslide is equal to the average; And if I> 0, the probability of landslides is higher than the average [5]. In other words, the higher the value of information, the higher the probability of landslide.

The statistical analysis method used in this study is Multiple Linear Regression (MLR). MLR is a statistical analysis technique that can measure statistically the relative influence of several factors and explain objectively how values depend on predictor variables. The dependent variable used in MLR is the IVM value of landslide susceptibility, while the predictor variables are aspect, slope, elevation and vegetation density. The landslide vulnerability IVM value is placed on the x axis, while aspect, slope, elevation and vegetation density values are placed alternately on the y axis. So that it produces 4 regression charts. Based on the four graphs, it can be seen that the highest value is the parameter that affects the landslide.

4. Result and Discussion

Remote sensing data is closely related to statistics. The type of data in remote sensing is rasio. So that many statistical methods are used to anlyzed remote sensing data. Remote sensing data is commonly used for disaster studies. Research has been carried out regarding landslide vulnerability in Purworejo [6]. The method used in this study is the Information Value Model (IVM) method. In this study, eight factors were assumed to affect the vulnerability of landslides in Purworejo District. However, these factors do not all use remote sensing data. In this study selected parameters obtained from remote sensing data to determine the most influential parameters using the multi linear regression.

Table 1. Result of Regression Statistics

Regression Statistics	
Multiple R	0.803722
R Square	0.645969
Adjusted R Square	0.645902
Standard Error	1.351801
Observations	42436

IOP Conf. Series: Earth and Environmental Science **500** (2020) 012046 doi:10.1088/1755-1315/500/1/012046

The first result is summary output of multiple linear regression between predictor variables and dependent variables. Multiple R shows the size of the closeness of the relationship between the dependent variable and all the independent variables simultaneously. In the case of one predictor variable, r can be positive or negative (between -1 to 1), but for two or more variables, the R value is always positive (between 0 and 1). The greater the value of R (- or +), the greater the relationship between these variables. R Square (R^2), often referred as the coefficient of determination, is measuring goodness of fit from the regression equation; that is giving the proportion or percentage of total variation in the dependent variable explained by the independent variable. R^2 value is between 0 - 1, and the suitability of the model is said to be better if R^2 is getting closer to 1. (further description of R2 see the discussion below)

Adjusted R Square. An important characteristic of R^2 is that its value is a function that never decreases from the number of independent variables in the model. Therefore, to compare two R^2 of two models, one must take into account the number of independent variables in the model. This can be done using "adjusted R square". The term adjustment means that the value of R^2 has been adjusted by the number of variables (degrees of freedom) in the model. Indeed, this adjusted R^2 will also increase with increasing numbers of variables, but the increase is relatively small. It is often recommended, if there are more than two independent variables, you should use adjusted R square.

Standard Error. It is a standard error of the estimated dependent variable (in our case is a demand). This number is compared to the standard deviation of the demand. The smaller the standard number of this error compared to the standard deviation number of the request, the regression model is more precise in predicting demand.

Table 1. Shows multiple values of R that are quite large and close to 1, which is 0.803722. This means that the relationship between predictor variables and variables is closely tied. Each variable has a significant impact on the occurrence of landslides in Purworejo. If viewed from the adjusted R square value which is 0.645902 it also results in the conclusion that all predictor variables have a significant effect on the dependent variable. This is because the value is close to a value of 1. In accordance with the rules of the adjusted R square, the closer it is to 1, the higher the significance. Likewise, the value of R2 is 0.6459696, which further reinforces that each predictor variable has a significant influence on the dependent variable.

Table 2. ANOVA Result

ANOVA	*df*	*SS*	*MS*	*F*	*Significance F*
Regression	4	4052.695541	1013.174	200.0106152	3.059E-170
Residual	42431	214938.4976	5.065601		
Total	42435	218991.1931			

The Analysis of Variance (ANOVA) table tests model acceptability from a statistical perspective in the form of diversity source analysis. This ANOVA is often also translated as a variance analysis. From the ANOVA table, it was revealed that the diversity of actual data of the dependent variable (demand) originated from the regression model and from the residuals. In a simple sense for our case, the variation (rise and fall) of demand is caused by variations in price and income (regression models) and from other factors that influence demand that we do not include in the (residual) regression model.

Degree of Freedom (df) or free degree of total is n-1, where n is the number of observations. Because our observations have 10, the total free degree is 9. The free degree of the regression model is 2, because there are two independent variables in our model (price and income). The free degree for residuals is the rest, the total free degree - regression free degree = 9 - 2 = 7.

The SS (Sum of Square) column or the number of squares for regression is obtained from the sum of the squares of the predictions of the dependent variable (demand) minus the average value of the demand from the actual data. First of all, we manually search for the average request from our original data. Then each demand prediction (see the residual output table below) is reduced by the average then squared. Furthermore, all the results of the calculation are added together.

Table 3. Multiple Linear Regression Result

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-0.52486	0.013	-39.3835	0	-0.550985794	-0.49874	-0.55099	-0.49874
Elevation	-0.533509	0.019	-27.4545	1.6937E-164	-0.571597158	-0.49542	-0.5716	-0.49542
Slope	0.036868	0.008	4.57257	4.83137E-06	0.021064704	0.052672	0.021065	0.052672
NDVI	0.014255	0.047	0.300682	0.763658246	-0.078669735	0.107181	-0.07867	0.107181
Aspect	0.163745	0.043	3.749382	0.000177506	0.078146041	0.249344	0.078146	0.249344

Based on table 3. the variable that most significantly affects the occurrence of landslides in Purworejo is elevation. This is indicated by the smallest P-value and very far from 5% (consequence of using a 95% significance level). In addition, the coefficient value of the elevation is most away from zero. Compared to other variables, elevation shows values that are in accordance with the requirements of a variable that most influences the dependent variable.

5. Conclusion

Multiple linear regression analysis can be applied to analyze remote sensing data. Because remote sensing data is quantitative data with a data type ratio. Therefore, the analysis method is very suitable. The use of various statistical methods in data analysis is a strategy that must be done. Because the statistical method is able to give a brief explanation of any amount of data. So that it is easy to understand and if there is anomalous data, it can be anticipated immediately.

The use of multiple linear regression methods can also be applied specifically to determine the most influential variables on the occurrence of a landslide disaster. Predictor variables used are variable elevation, slope, NDVI, and aspect. This study concludes that the most influential variable on the occurrence of landslides in Purworejo is the elevation with an adjusted R2 value of 0.645902. In addition, it is supported by the P-value of the elevation variable which is 1.6937E-164 and the coefficient value is 0.522509. The P-value and coefficient of the elevation variable are the most satisfying values for deciding that a variable is a predictor variable that has the most influence on a dependent variable.

6. Acknowledgments

Authors wishing for acknowledge Department of Geographic Information Science, Universitas Gadjah Mada for giving us guide and financial support to continue our research in Purworejo Regency.

7. References

[1] Aman, S.N.A., Z.A. Latif, and B. Pradhan. Spatial probabilistic approach on landslide susceptibility assessment from high resolution sensors derived parameters. in IOP Conference Series: Earth and Environmental Science. 2014. IOP Publishing.

IOP Conf. Series: Earth and Environmental Science **500** (2020) 012046 doi:10.1088/1755-1315/500/1/012046

[2] Hadmoko, D.S., et al., Landslide hazard and risk assessment and their application in risk management and landuse planning in eastern flank of Menoreh Mountains, Yogyakarta Province, Indonesia. Natural Hazards, 2010. **54**(3): p. 623-642.
[3] Nugraha, H., et al., Geomorphometric characteristics of landslides in the Tinalah Watershed, Menoreh Mountains, Yogyakarta, Indonesia. Procedia Environmental Sciences, 2015. **28**: p. 578-586.
[4] Sharma, L., et al., Development and application of Shannon's entropy integrated information value model for landslide susceptibility assessment and zonation in Sikkim Himalayas in India. Natural hazards, 2015. **75**(2): p. 1555-1576.
[5] Ba, Q., et al., An improved information value model based on gray clustering for landslide susceptibility mapping. ISPRS International Journal of Geo-Information, 2017. **6**(1): p. 18.
[6] Sudaryatno, S., et al. Evidence Based Landslide Hazard Mapping in Purworejo using Information Value Model Approach. in Forum Geografi. 2019.