Using Geographically Weighted – Binary Logistic Regression to Analyze Land Cover Change Phenomenon (Case Study: Northern West Java Development Region)

Q R Muzdalifah¹, A Deliar¹, R Virtriana¹, A Naufal¹ and I S Ajie¹

¹Remote Sensing and Geographic Information Science Research Group, Bandung Institute of Technology (ITB), No. 10 Jalan Ganesha, 40132, Indonesia

qonitarahmamz@gmail.com

Abstract. Land is one of the important resources that can be used to supply the needs of human life. Uncontrolled land utilization will cause land cover change phenomenon. Land cover change phenomenon can be analyzed by using a model. To get an accurate result, the selection of models in the analysis of land cover change must be based on the characteristics of land cover change phenomenon itself. Land cover change is a binary phenomenon and strongly related to the local characteristics of a region. A model that can be used in the analysis of binary phenomena is Binary Logistic Regression (BLR) model. However, the application of BLR model has a disadvantage. BLR model is one of the global models which assumes that the analyzed phenomenon has homogeneous characteristics for the entire study area. This does not correspond to the characteristic of land cover change phenomenon. Therefore, we need another local model that is able to show local characteristic variations of land cover change. Geographically Weighted Regression (GWR) model is one of the local spatial regression techniques that can be used to analyze phenomena that have spatially heterogeneous characteristics. The application of GWR model for binary phenomena (dependent variable) such as land cover change is called Geographically Weighted – Binary Logistic Regression (GW–BLR) model. This research aims to analyze land cover change phenomenon in the Northern West Java development region using GW-BLR and compares the result to BLR model. The results of this research indicate that the analysis of land cover change in the Northern West Java development region using GW-BLR model has a higher level of accuracy compared to BLR model. The modeling results of land cover change using GW-BLR model has an overall accuracy value of 91.10% and using BLR model has an overall accuracy value of 84.09%. Therefore, it can be concluded that land cover change phenomenon in Northern West Java development region can be analyzed more accurately by considering its local spatial characteristics through using the GW-BLR model.

1. Introduction

The influence of human interaction in utilizing nature to fulfill their needs can be seen in various types of land cover [23]. Land cover is a biophysical state of the earth's surface, such as vegetation, waters, buildings, and so on [5]. If the process of interaction between humans and nature is not well controlled, it will lead to uncontrolled land cover change. Uncontrolled land cover change can cause various negative impacts on the environment [5], [17], [18], [20]. The negative impacts of land cover change are biodiversity loss, increment of greenhouse effect, land degradation, habitat destruction, water
pollution, and erosion. Land cover change also plays a role in the global carbon cycle. According to the Intergovernmental Panel on Climate Change (IPCC), in quantity, there are approximately 1.6 billion gigatons of carbon emitted annually due to land cover change [9].

The negative impact of land cover change can be anticipated by conducting a study or analysis [21]. Land cover change phenomenon can be analyzed by using a model. To get an accurate result, the selection of models in the analysis of land cover change must be based on the characteristics of land cover change phenomenon itself. A model that is often used in land cover change analysis is the Binary Logistic Regression (BLR) model [1], [2], [3], [11], [19], [23]. The BLR model is a regression method used to find the relationship between the binary dependent variable (y) with independent (x) variable. The BLR model is suitable for phenomena with binary properties such as land cover change [16]. In the BLR model, land cover change is analyzed as dependent variable which consists of two categories (change or not change).

However, the application of BLR model in land cover change analysis has a disadvantage. BLR is one of the global analytical models that is estimated by using all existing data with the same parameter weight to produce only one model equation for the entire study area [10], [24]. One model equation indicates that each independent variable of the model has the same influence on dependent variable in the entire study area. In the land cover change analysis, it means that each land cover change driving factor has the same influence for the entire study area so that the land cover change has homogeneous characteristics. This characteristic of BLR model does not correspond to the characteristic of land cover change phenomenon. Land cover change for different regions will have different dynamics as well. This depends on the local characteristics of each region [14], [22]. Therefore, land cover change and its driving factors have spatially heterogeneous characteristics.

Geographically Weighted Regression (GWR) model is one of the local spatial regression techniques that can be used to model spatially heterogeneous phenomena [6]. Unlike the global model, the GWR method will produce a different model for each data point in a study area. If there are 100 data points in a study area, then these 100 points will produce different GWR models. Different model equations indicate that each independent variable of the model has different heterogeneous influence for each data point in a study area. Therefore, the GWR model can describe in the local characteristics variations of the phenomenon [10]. The application of GWR model for binary phenomena (dependent variable) such as land cover change is called Geographically Weighted – Binary Logistic Regression (GW-BLR) model. This research aims to analyze land cover change phenomenon in the Northern West Java development region using GW-BLR and compares the result to BLR model.

2. Materials and Methods

2.1. Study area

This research was conducted in the Northern West Java development region which consist of 14 administrative districts, those are Bogor Regency, Bogor City, Depok City, Bekasi City, Bekasi Regency, Karawang Regency, Purwakarta Regency, Subang Regency, Sumedang Regency, Indramayu Regency, Majalengka Regency, Cirebon Regency, Cirebon City, and Kuningan Regency. The determination of administrative districts in the Northern West Java development region is regulated in the Regional Regulation No. 7 of 2012 on the Northern West Java Development. Development region was formed to manage the area based on the same local potential, increase productivity, develop economy and infrastructure. Figure 1 illustrates the study area of this research.
2.2. Data
This research uses 5 data, those are West Java Province land cover maps (2005 and 2010) sourced from the Indonesia Geospatial Information Agency (BIG), SRTM DEM (Shuttle Radar Topographic Mission) with 90 meters resolution, West Java Province administrative map (2005) sourced from BIG, roads network map (2010) sourced from BIG, and the human population data sourced from West Java Province Central Bureau of Statistics (BPS). This data will be used to define the dependent and independent variables in the analysis of land cover change using the BLR and GW-BLR method. The dependent variable in this study is land cover change, while there are 5 variables stand as independent variables, those are slope, height, distance to the main road, population, and distance to the district capital.

Land cover change data were obtained from the overlay of land cover data in 2005 and 2010. Land cover change data in the Northern West Java development region 2005 - 2010 are illustrated in Figure 2. The independent variables data were obtained from further analysis of some data. For example, spatial analysis of the administrative map to get distance to the district capital variable data. SRTM DEM spatial analysis to get height and slope data. The roads network map spatial analysis to get the distance to the main road data. Spatial analysis of human population data, land cover map, and administrative map to obtain population data.

Figure 1. Research study area.
2. Land cover change data in the Northern West Java development region 2005 - 2010.

Land cover change and its driving factors data in this study were processed using the BLR and GW-BLR method in the form of sample data. Sample data is taken with a grid pattern and the distance between sample points is 1800 m. In the Northern West Java development region, there is a 5291 sample data point. The sample data is used in this study because the existing population data is too large in size so processing time using population data will require a long duration. Therefore, sample data is used to reduce the duration of the study.

2.3. Binary logistic regression

Binary Logistic Regression (BLR) is a data analysis method that is used to find the relationship between the binary dependent variable (y) and the independent variable (x) [12]. The dependent variable in BLR is data that represents a phenomenon with binary properties. Binary is a data type, where data consists of only two category values such as success or failure, life or death, and change or not change. By using the BLR method, we will get a coefficient of each independent variable. This variable will be used to calculate the occurrence probability of a phenomenon (in this case, land cover change). This probability calculation will produce a range of values between 0 and 1. The binary logistic regression equation has the form (1) [15]:

\[
\logit(\pi_i) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n \\
\ln \frac{\pi_i}{1-\pi_i} = \alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n \\
\pi_i = \frac{\exp(\alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n)}{1 + \exp(\alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n)}
\]
2.4. Geographically weighted-binary logistic regression

Consider a global linear regression model written as (2) [6]:

\[ y_i = \beta_0 + \sum_{k=1}^{n} \beta_k x_{ik} \]  

(2)

The Geographically Weighted Regression (GWR) method is a spatial regression technique and the extension of linear regression framework by allowing local rather than global parameters to be estimated for each data point so that the GWR equation can be rewritten as (3) [6]:

\[ y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{n} \beta_k(u_i, v_i)x_{ik} \]  

(3)

Where \( \beta_0(u_i, v_i) \) denotes constant of the regression equation of the ith point in space, \( \beta_k(u_i, v_i) \) denotes independent variable coefficient of the ith point in space, \( i \) denotes regression point, and \( u_i, v_i \) denotes the coordinates of the ith point in space.

The GWR method is a local model that can be used to analyze spatially heterogeneous phenomena [6]. GWR model estimates parameters for each data point in the study area using a set of the nearest neighbor data [10]. The neighbor data point that has the closest distance to the regression point will have a greater influence on the estimation process than the further data point. The magnitude of this influence is manifested in the weight value given to each data point in the GWR model parameter estimation process. Estimation of parameters in the GWR model can be done through (4) [6]:

\[ \hat{\beta}_i = (X^T W(u_i, v_i) X)^{-1} (X^T W(u_i, v_i) Y) \]  

(4)

where the bold type denotes a matrix and \( W(u_i, v_i) \) is an n by n matrix whose off-diagonal elements are zero and whose diagonal elements denote the geographical weighting of each of the n observed data for regression point \( i \). The weighting scheme in the GWR method is based on the proximity to regression point. The weight value will be obtained through a weighting function commonly called a kernel [7]. In general, the weight value assigned to the GWR method will be affected by the kernel type and the weighting function used in the kernel. In this research, parameters estimation will be done by using the adaptive kernel type and Gaussian weighting function that can be rewritten as (5) [6]:

\[ w_{ij} = \exp \left[ -\frac{1}{2} \left( \frac{d_{ij}}{b} \right)^2 \right] \]  

(5)

Where \( d_{ij} \) denotes distance data point to regression point and \( b \) denotes bandwidth. According to [10], the type of kernel and the weighting function will affect the estimation results of GWR model even though it is practically not significant or even very small. However, the size of the kernel or bandwidth used in a kernel function will significantly influence the results of estimated parameters. Bandwidth is the size of the range which acts as a determinant of the nearest neighbor which has a significant influence on the regression point so that the point will be included in the process of estimating the GWR model parameters. Therefore, determining the optimal bandwidth is an important factor that needs to be considered in using the GWR model. There are several number of criteria that can be used for optimal bandwidth selection. One of them is by minimizing the value of Akaike Information Criterion (AICc). [13] define the AICc for GWR as (6):

\[ AICc = 2n \log_e \hat{\sigma} + n \log_e (2\pi) + n \frac{n + tr(S)}{n - 2 - tr(S)} \]  

(6)
Where \( n \) is the sample size, \( \hat{\sigma} \) is the estimated standard deviation, and \( \text{tr}(S) \) denotes the trace of the hat matrix which is a function of the bandwidth. The application of GWR model for binary dependent variable such as land cover change is called Geographically Weighted – Binary Logistic Regression (GW-BLR) model. The equation of the GW-BLR method can be written as (7) [4]:

\[
\text{Logit} (p_i) = \beta_0(u_i, v_i) + \sum_{k=1}^{n} \beta_k(u_i, v_i)x_{ik}
\]

(7)

3. Result and Discussion

3.1. Optimal bandwidth of land cover change in the Northern West Java Development Region

In this research, the Northern West Java development region will be represented by 5291 sample points. This sample point has a grid-shaped distribution with the distance between sample points is 1800 m. Based on the results of data processing that has been done using the GW-BLR method, it will be obtained the optimal bandwidth value of land cover change in the Northern West Java development region. The optimal bandwidth value of land cover change in the Northern West Java development region is 822 points. It means that land cover change phenomenon of the sample point in the Northern West Java development region will be influenced by the 822 nearest neighbor points around it. Figure 3 illustrates the optimal bandwidth value of land cover change in the Northern West Java development region.

![Figure 3. The optimal bandwidth value of land cover change in the Northern West Java development region.](image)

All data in figure 3 is land cover change data in the Northern West Java development region. If we take the red point as the regression point/ GW-BLR points, then the yellow point is the 822 nearest neighbor point around the red point that will be included to estimate the regression parameters from the red point. The neighbor data point (yellow point) that has closer distance to regression point (red point) will have
a greater influence on the estimation process than further neighbor datapoint (yellow point). It denotes the local characteristic of GW-BLR model. GW-BLR model estimates regression parameters for each data point in the study area using only a set of the nearest neighbor data with a certain bandwidth (in this research, the 822 nearest neighbors). Meanwhile, the global model (BLR) estimates regression parameters by using all existing data to produce only one model equation for the entire study area.

3.2. Land cover change prediction model in the Northern West Java Development region

In land cover changes prediction analysis, there are several important parameters that need to be considered such as the Overall Accuracy (OA) and threshold. Overall accuracy is a parameter that denotes model accuracy. Model accuracy is a value that describes the level of suitability between the modeling result with actual conditions in the field [8]. Model accuracy testing can be done in various ways. One of them is by making comparisons through the error matrix or confusion matrix. OA is obtained by comparing the number of data on the main diagonal with the total number of data in the error matrix. OA shows the value of the ratio between the number of data that are modeled precisely according to the real conditions with the total number of data used in the testing process. Meanwhile, threshold is a cut-off parameter that indicates the lowest (smallest) value of land cover change that can still be detected [23]. Table 1 explains the OA value of BLR and GW-BLR models. Meanwhile, table 2 explains the threshold value of BLR and GW-BLR model.

Table 1. The overall accuracy value of BLR and GW-BLR model.

| Reference Data | BLR Model | GW-BLR Model | Total | Overall Accuracy |
|----------------|------------|--------------|-------|-------------------|
| Not Change (0) | 4383       | 4748         | 4782  | 84.09%            |
| Change (1)     | 399        | 34           | 4782  |                   |
| Jumlah         | 4426       | 465          | 4449  |                   |
| Berubah (1)    | 443        | 437          | 509   | 91.10%            |
| Jumlah         | 5185       | 106          | 4820  |                   |

Table 2. The threshold value of BLR and GW-BLR model.

| Model       | Threshold |
|-------------|-----------|
| BLR         | 0.1280    |
| GW-BLR      | 0.6017    |

Table 1 shows that the modeling of land cover change using local model GW-BLR method has the overall accuracy value of 91.10% and using BLR method has the overall accuracy value of 84.09%. The local model (GW-BLR) has a higher overall accuracy value than the global model (BLR). It proves that land cover change phenomenon in Northern West Java development region can be analyzed more
accurately by considering its local spatial characteristics through using the GW-BLR model. Meanwhile, the threshold value of BLR model in table 2 is 0.1280 and the threshold value of GW-BLR model is 0.6017. This means that the minimum value of land cover change detected in Northern West Java development region is 0.1280 by using BLR model and 0.6017 by using GW-BLR model. Besides generating those parameters, the BLR and GW-BLR methods also produce land cover change prediction map. The land cover change prediction map of BLR and GW-BLR model can be seen in Figure 4 and figure 5.

![Figure 4. The land cover change prediction map of BLR model.](image1)

![Figure 5. The land cover change prediction map of GW-BLR model](image2)
4. Conclusion
Based on the results and discussion, it can be concluded that the land cover change phenomenon in Northern West Java development region will be influenced by the 822 nearest neighbor points around it. Land cover change phenomenon in Northern West Java development region can be analyzed more accurately by considering its local spatial characteristics. This is proven through the overall accuracy value produced based on both models. The local model (GW-BLR) has a higher overall accuracy value than the global model (BLR). The modeling of land cover change using local model GW-BLR method has an overall accuracy value of 91.10% and using BLR method has an overall accuracy value of 84.09%.

Acknowledgement
The authors would like to thank the Institute for Research and Community Service (LPPM), Bandung Institute of Technology (ITB), which has provided research funding through ITB Research Program 2019.

References
[1] Ainiyah N 2016 Land Cover Change Driving Factor Analysis using Spatial Binary Logistic Regression Method in West Java Development Area (Master’s Thesis) (Bandung: Institut Teknologi Bandung)
[2] Ammarullah A 2015 Predictor Variable Analysis of Land Cover Change based on Binary Logistic Regression in the West Java Province (Master’s Thesis) (Bandung: Institut Teknologi Bandung)
[3] Arsanjani J J, Helbich M, Kainz W, and Boloorani A D 2013 Integration of Logistic Regression, Markov Chain and Cellular Automata Models to Simulate Urban Expansion. International J. of Appl. Earth Observ. and Geoinfo.
[4] Atkinson P M, German S E, Sear D A, Clark M J 2003 Exploring The Relations Between Riverbank Erosion and Geomorphological Controls Using Geographically Weighted Logistic Regression Geographic. Anal 35(1), 58–82
[5] Briassoulis H 2000 Analysis of Land Use Change: Theoretical and Modeling Approach
[6] Brunsdon C, Fotheringham A S, and Charlton M E 1996 Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity Geograph. Anal 28(4) 281–298
[7] Cleveland W S and Devlin S J 1988 Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting J. of the American Stat. Assoc. 83 596-610.
[8] Congalton, R G 1991 A Review of Assessing The Accuracy of Classifications of Remotely Sensed Data Remote Sensing of Envi. 37 35-46
[9] Denman K, Brasseur G, Chidthaisong A, Ciais P, Cox P M, Dickinson R E, Hauglustaine D, Heinze C, Holland E, Jacob D, Lohmann U, Ramachandran S, Dias P L D, Wofsy S C and Zhang X 2007 Couplings between Changes in The Climate System and Biogeochemistry: The Physical Science Basis (Contribution of Working Group I to The Fourth Assessment Report of The Intergovernmental Panel on Climate Change) (United Kingdom: Cambridge University Press)
[10] Fotheringham A S, Brunsdon C, and Charlton M 2002 Geographically Weighted Regression: The Analysis of Spatially Varying Relationships (Chichester: John Wiley & Sons)
[11] Fox J, Vogler J B, Sen O L, Giambelluca T W, and Ziegler A D 2012 Simulating Land-Cover Change in Montane Mainland Southeast Asia Environmental Manageme. 49(5) 968-979
[12] Hosmer and Lemeshow 2000 Applied Logistic Regression Second Ed. DOI: 10.1002/0471722146
[13] Hurvich C M, Simonoff J S, and Tsai C-L 1998 Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion J. of the Royal Statis. Socie. 60 271–93
[14] Lambin E 2004 Modelling Land-Use Change in Wainwright J and Mulligan M Environmental Modeling: Finding Simplicity in Complexity (John Wiley) 430
[15] Larsen R J 2006 Introduction to Mathematical Statistics and It’s Applications (Pearson/PrenticeHall)
[16] Liao Q, Li M, Chen Z, Shao Y, and Yang K 2010 Spatial Simulation of Regional Land Use Patterns based on GWR And CLUE-S Model 18th Int. Conf. on IEEE 2010 (China) 1–6
[17] Meyer W B and Turner B L 1996 Land Use/ Land Cover Change: Challenge for Geographers BioSci. 44 300304
[18] Moser S C 1996 A Partial Instructional Module on Global and Regional Land Use/ Cover Change: Assessing the Data and Searching for General Relationships GeoJurn. 39(3) 241-283
[19] Muzdalifah Q R 2018 Land Cover Change Driving Factor Analysis based on Spatial Planning Zonation (Case Study: West Java Province) (Bachelor’s Thesis) (Bandung: Institut Teknologi Bandung)
[20] Riebsame W E, Meyer W B, Turner B L 1994 Modelling Land Use and Cover as Part of Global Environmental Change Climate Cha. 28 45–64.
[21] Turner B L, Skole D, Sanderson S, Fischer G, Fresco L, dan Leemans R 1995 Land-Use and Land-Cover Change Science/Research Plan (International Human Dimension Programme on Global Environmental Change)
[22] Verburg P H, van Eck J R, de Nijs T, Schot P, and Dijst M 2004 Determinants of Land Use Change Pattern in the Netherlands Environ. and Plan. 31(1) 125-130
[23] Virtriana R 2016 Combination of Binary Logistic and Markov Chain Regression Models Based on Cellular Automata for Land Cover Dynamics Prediction in West Java (Doctoral Dissertation) (Bandung: Institut Teknologi Bandung)
[24] Wu L, Deng F, Xie Z, Hu S, Shen S, Shi J, and Liu D 2016 Spatial Analysis of Severe Fever with Thrombocytopenia Syndrome Virus in China Using a Geographically Weighted Logistic Regression Model Inveron. Res. and Pub. Health 13, 1125