Driving-factors identification of land-cover change in west java using binary logistic regression based on geospatial data

Udjianna Sekteria Pasaribu 1, Riantini Virtriana 2, Albertus Deliar 2, Irawan Sumarto 2

1 Faculty of Mathematics and Natural Sciences, Mathematics Study Program, Institute of Technology Bandung (ITB), 10 Ganesha Street, 40132 Bandung, Indonesia
2 Faculty of Earth Sciences and Technology, Geodesy and Geomatics Engineering, Institute of Technology Bandung (ITB), 10 Ganesha Street, 40132 Bandung, Indonesia

e-mail: udjianna@math.itb.ac.id

Abstract. The Land is a fundamental factor in production activity. Accordingly, it is closely related to economic growth—which supports the living needs of human beings. In many cases, human activities related to land use are often uncontrollable, impacting many negative effects on the environment, both locally and globally. More broadly, these activities will lead to some changes in land cover and some other physical features such as climate. In order to understand the phenomenon of land cover changes, we approach them through modelling. To detect any changes in land cover in a region, it is necessary to identify the driving factors causing land-cover change. The relation between driving factors and response variables can be evaluated by using regression analysis techniques. In this case, land cover change is a dichotomous phenomenon, i.e., binary. Binary Logistic Regression (BLR) model is one of the regression analyses which can be used to describe the nature of dichotomy. From the results of this study, the driving factors causing land-cover change in West Java were found, those are: the distances to the central business districts in some certain areas such as Bandung City, Bekasi Regency, Bekasi City, Bogor Regency, Karawang Regency, and Sukabumi Regency; the distance to the capital of the province; the distance to the main roads; the population numbers; and some physical features of the land such as slope, curvature, and height. This predictive model had an accuracy level of 49.79%, which equals to 1.827,217,44 ha area.

1. Introduction
Over the past few decades, land cover changes have been a phenomenal issue. Incessant, often uncontrollable even, land cover changes result in significant impact both locally and globally. Land cover changes will bring a negative impact on environmental conditions when vegetation changes into settlements. These changes are massive in scale. One of the negative impacts is local or global climate change (1, 2). Deforestation in the context of developing agricultural land or developed land use is one example of land cover changes. The changes may result in the decline of biodiversity, soil degradation, and other resources needed by humans (3).

Land cover data of 2000-2010 shows that West Java Province is one of the provinces with high land cover dynamics. In 2010, land cover data shows that the increase in settlements came from the vegetation area. Forest and field areas in West Java decreased by 288,203 ha in period 2000-2010. In period 2005-2010, forests were converted into fields, rice fields, and plantations which is mainly driven by efforts to meet the need for agricultural products and commodities (4).

Fundamentally, there are two main groups of factors driving that cause land cover changes, namely biophysical and socioeconomic factors (5-7). Biophysical factors are characteristics and processes that
occur in a natural environment such as due to climate change, changes in topography, geomorphological processes, and the impact of volcanic eruptions. Socioeconomic factors pertain to human activities such as demography, social, economy, culture, politics, industry, and technology (8). Both factors above are related to one another.

Based on previous theory and research, it is assumed that there are 13 (thirteen) variables affecting land cover change in West Java, representing physical and socioeconomic factors. Variables related to physical factors include slope, slope shape, land elevation, and distance to the main road. The selection of these factors is based on the heterogeneous condition of the topography of West Java Province. The southern part consists of mountains, sloping hillsides in the central part, and expansive land in the northern part.

2. Data and Methods

Land cover and land-use changes (LULCC) are a phenomenon that does not adhere to normal assumption (9). LULCC in a location and in a certain period of time considered as a dichotomous and binary phenomenon or event. In other words, the phenomenon only consists of two categories, changing and not changing. Due to its binary nature, land-use change may use Binary Logistic Regression (BLR) to presume the relation between land cover change and other factors influencing it.

BLR is a probability method of land cover change in one spatial unit (pixel) defined by the interaction between land cover with factors promoting the change. The relation between land cover probability and driving factors is based on statistical analysis on land cover data and involved factors. Thus, a particular formula is generated which would then be used in the model. The advantage of the BLR method is that it can predict the land cover change in an area by considering the driving factors of land cover change (10).

BLR yields an odds ratio which represents the odds that an outcome will occur compared to the odds that an outcome will not occur. Odds Ratio equation can be expressed as follows:

$$Odds \text{ Ratio} = \frac{\pi_l}{1-\pi_l} \tag{1}$$

where:

- $\pi_l$: the probability of an event $i$ occurring (land cover change)
- $1-\pi_l$: the probability of event $i$ not occurring

Based on the relation, the probability of change $\pi(x)$ or BLR equation can be expressed as follows:

$$logit \pi(x) = ln \frac{\pi_l}{1-\pi_l} = \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n \tag{2}$$

$$\pi(x) = \frac{e^{\alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n}}{1 + e^{\alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n}} \tag{3}$$

2.1. Data

In this study, thirteen (13) predictor variables presumed to drive the land cover change in West Java are used. All variables will then be tested using multivariate analysis.
Table 1. Predictor Variables.

| No | Variables                  | Data Source                      | Agency          |
|----|---------------------------|----------------------------------|-----------------|
| 1  | Slope                     | DEM SRTM                         | USGS            |
| 2  | Curvature                 | DEM SRTM                         | USGS            |
| 3  | Elevation                 | DEM SRTM                         | USGS            |
| 4  | Distance to Roads         | Route Map                        | BIG             |
| 5  | Distance to Capital City  | Administrative Border Map        | BIG             |
| 6  | Population density        | Population Data and Land Cover Map | BPS and BIG    |
| 7  | Distance to Settlements   | Land Cover Map                   | BIG             |
| 8  | Distance to Sukabumi Regency | Administrative Boundary Map | BIG             |
| 9  | Distance to Bekasi City   | Administrative Boundary Map      | BIG             |
| 10 | Distance to Bandung City  | Administrative Boundary Map      | BIG             |
| 11 | Distance to Karawang Regency | Administrative Boundary Map | BIG             |
| 12 | Distance to Bogor Regency | Administrative Boundary Map      | BIG             |
| 13 | Distance to Bekasi Regency | Administrative Boundary Map      | BIG             |

The thirteen factors above are hereafter called as predictor variables, which are derived from five main data namely elevation data (DEM SRTM) with a resolution of 30 meters, Land Cover Thematic Map in 2005, 2010 and 2015, Administrative Border Map in 2005, and Route Map in 2005 sourced from the Geospatial Information Agency (BIG), as well as Population Social Data in 2005 and 2010 sourced from the Central Statistics Agency (BPS) (figure 1).
Figure 1. Primary data used in this study: (a) Land cover data (b) Elevation data (c) Road network data (d) Administrative boundary data, and (e) Population density data.

As an effort to solve the research problem and achieve the intended goal, the research described into this methodology which can be best summarized by the following diagram (figure 2).
3. Results and Discussion

In table 2 correlation coefficient of each predictor variable on land cover classification is known. Predictor variables are classified according to physical and socioeconomic factors. Of the 8 (eight) land cover classes used, only 6 (six) land cover classes will be included in the correlation calculation. Waterbody and pond classes were not included because in the 2005-2010 period the classes did not significant changes. Thus, for the subsequent process, land cover classes included in the calculation are forest, field, plantation, settlements, rice fields, and shrub classes.

Table 2. The correlation between predictor variables (independent) with response variables for each land cover class. This table shows the relationship between predictor variables on each land cover class based on the correlation coefficient.

| Predictors                  | Forest | Field | Plantations | Settlement | Rice Fields | Shrubs |
|-----------------------------|--------|-------|-------------|------------|-------------|--------|
| Physical                    |        |       |             |            |             |        |
| Slope                       | -0.065 | -0.167| -0.025      | -0.051     | 0.014       | -0.084 |
| Curvature                   | -      | 0.009 | -0.002      | 0.016      | -0.006      | -      |
| Elevation                   | -0.072 | -0.079| -0.149      | -0.185     | -0.036      | 0.076  |
| Distance to Roads           | -0.066 | -0.029| 0.101       | -0.054     | -0.024      | -0.018 |
| Distance to Capital         | 0.003  | 0.025 | 0.114       | -0.12      | 0.037       | -0.167 |
| Population Density          | 0.256  | 0.172 | 0.036       | -0.656     | 0.319       | -0.167 |
| Distance to Settlements     | -0.02  | -0.085| -0.114      | 0.53       | -0.019      | -0.099 |
| Distance to Sukabumi Regency| -0.106 | -0.008| -          | -0.286     | 0.033       | -0.12  |
| Distance to Bekasi City     | -0.132 | -0.087| -0.05       | -0.343     | -          | -0.285 |
| Distance to Bandung City    | 0.046  | 0.055 | 0.253       | 0.237      | 0.039       | -0.145 |
| Distance to Karawang Regency| -0.107 | -0.093| 0.143       | -0.244     | 0.023       | -0.322 |
| Distance to Bogor Regency   | -0.117 | -0.069| 0.006       | 0.334      | -0.013      | -0.240 |
| Distance to Bekasi Regency  | -0.124 | -0.086| 0.07        | -0.328     | 0.006       | -0.304 |

There are other variables that have no correlation between one another. The absence of any correlation does not necessarily mean that the correlation value is zero, it means that the correlation value obtained has significance bigger than 0.1. This is due to the alpha quantity (α) used is 10%, thus the probability value shows H0 is accepted (significance value bigger than 0.1 indicates that the hypothesis stating that there is a correlation between two variables tested is rejected).

Population density has the biggest close relation with 65.6% on settlements change toward negative correlation. The correlation coefficient value means that if the population density is big, then the settlements change probability value is small. In this regard, stronger population pressure will not change the settlements into other classes. It just indicates that the needs in settlements will increase due to population pressure. Changes in settlements class on driving factors have a greater correlation coefficient value compared to other land cover classes. However, the small correlation value still
shows the relationship between the two variables tested because it is proven to be significant. Elevation factor representing physical factors has a relatively similar relation closeness on changes in each land cover class (7% - 18%). Other physical factors represented by curvature and distance to roads have smaller correlation coefficient value with 1% to 8%. Overall, the results show that the correlation between land cover change on factors related to socioeconomic has greater closeness value compared to factors related to topography condition of West Java.

However, the correlation coefficient value between predictor variables with land cover change does not show any causal relationship due to a fundamental difference between correlation and causality. If the two variables are said to be correlated, the one variable does not necessarily affect the other variable, or in other words, show causality. Causality relationships occur if the predictor variables affect the response variable. If the two variables are treated symmetrically (the measurement value remains the same if the roles of the variables are exchanged), then the two correlated variables cannot be said to have a causality relationship. Thus, the causality relationship may be absent even in two correlated variables. Correlation can only be used as one proof of the possibility of a causal relationship, but it cannot provide an indication of what the causal relationship is like if it does occur in the variables studied.

Correlation coefficient values which are proven to be insignificant indicate that there is no relationship between the variables tested. Thus, these variables will not be involved in the next process to be identified for the causal relationship. Therefore, defining the correlation value of the variables used will serve as a “filtering” phase for the next process, where only variables that have correlation and significance will be studied for their causal relationships.

After the correlation or relation between predictor variables and land cover class identified, the next process is identifying the causal relation or effect between predictor variables and land cover class through the regression equation. In BLR equation, all predictors function as independent variables, while land cover change data of 2005-2010 function as response variables. Table 3 contains regression coefficient value ($\beta$) and significance value (p-value) from every variable tested in BLR method. The principle of hypothesis decision is if the p-value > 0.10, H0 is accepted. And if the p-value < 0.10, H0 is rejected. This refers to data reliability of 90%, where the p-value to be met should be <0.10. If this condition is met, then the predictor variable will be used for further calculations.
### Table 3. Regression coefficient and significance value from the regression process. This table shows the effect of predictor variables on each land cover class based on the regression coefficient.

| Predictor Variables          | Forest                      | Field                       | Plantations                  | Settlement                    | Rice Fields                   | Shrubs                      |
|-----------------------------|-----------------------------|-----------------------------|------------------------------|------------------------------|------------------------------|------------------------------|
|                            | β                           | p-value                     | β                            | p-value                      | β                            | p-value                     |
| Slope                      | -3.27E-02                   | 2.829E-221                  | -6.02E-02                    | 0                            | -0.42E-02                    | 4.308E-08                   |
| Curvature                  | -18.51E-02                  | 6.422E-25                   | 2.27E-02                     | 0                            | 0.00052701                   | 367.89E-02                  |
| Elevation                  | -0.04E-02                   | 1.7014E-33                  | 0.01E-02                     | 3.719E-23                    | -0.15E-02                    | 0.0001342                  |
| Distance to Roads          | -0.01E-02                   | 9.387E-230                  | 0.004E-02                    | 0                            | 0.005E-02                    | 1.383E-11                  |
| Distance to Capital        | 0.003E-02                   | 7.0554E-52                  | 0.007E-02                    | 0                            | -0.002E-02                   | 5.9224E-11                 |
| Population                 | 228.7E-02                   | 0                           | 57.83E-02                    | 0                            | 0.55E-02                     | 17.84E-02                  |
| Distance to Settlements    | -0.002E-02                  | 0.00021984                  | -0.006E-02                   | 0                            | 0.009E-02                    | 7.0996E-19                 |
| Sukabumi Regency CBD       | -0.0009E-02                 | 6.549E-10                   | 0.005E-02                    | 0                            | 0.005E-02                    | 9.8147E-35                 |
| Bekasi City CBD            | -0.03E-02                   | 0                           | -0.007E-02                   | 0                            | 0.008E-02                    | 5.9791E-29                 |
| City          | CBD          | Latitude | Longitude | Adj Latitude | Adj Longitude | Adj Constant | Constant |
|--------------|--------------|----------|-----------|--------------|---------------|--------------|----------|
| Bandung City | -0.004E-02   | 1.7312E-50 | -0.006E-02 | 0            | 0.001E-02     | 1.736E-114   | -0.0034E-02 | 0        | 0.04E-02 | 5.8296E-15 |
| Karawang     | 0.01E-02     | 8.453E-111 | 0.004E-02 | 0            | 0.003E-02     | 1.260E-41    | 0.005E-02   | 0        |          |            |
| Bogor Regency CBD | 0.01E-02 | 0         | -0.009E-02 | 0            | -0.006E-02    | 1.258E-173   | 0.0005E-02  | 5.921E-08 | 0.14E-02 | 2.384E-24 |
| Bekasi Regency CBD | 0.009E-02 | 2.1333E-87 | 0.007E-02 | 0            | -0.004E-02    | 4.3326E-08   | 0.004E-02   | 3.485E-70 | 0.02E-02 | 0.01802  |
| Constant     | 730.4E-02    | 0        | -173.58E-02 | 0            | -523.38E-02   | 2.9615E-16   | -116.94E-02 | 9.131E-15 | 51.789E-02 | 4.394E-32 |

The Fifth International Conference of Indonesian Society for Remote Sensing
IOP Conf. Series: Earth and Environmental Science 500 (2020) 012003
doi:10.1088/1755-1315/500/1/012003
The effect of the driving factors of land cover change on the land cover class can be seen in Table 3 above. Each driving factor has various effects for each land cover class. This is characterized by positive and negative effects. All predictor variables that were proven to be uncorrelated with the land cover change in West Java were not included in the regression calculations. However, hypothesis testing for the regression coefficient test is still carried out with the same significance level, which is 90%. Even though the curvature variable correlated with rice field class, the resulting effect was not significant. Similarly, slope and distance to Karawang Regency have no significant effect on shrub class. Therefore, those variables were not proven to be the driving factors that cause land cover change.

Population density has a positive effect on land cover change for forest, fields, plantations and rice fields class. The greater number of populations in an area, the tendency of land cover classes to change to other classes is higher. Population density negatively impacts settlements, meaning that the population pressure is marked by an increase in the population in the region, will result in a potential increase in settlements around it. Once the effect of the driving factors on land cover change is identified through the regression coefficient, the next step is to predict future land cover. The initial stage in the prediction process is determining the probability of land cover change from 2010 to 2015 by using a regression coefficient from the partial BLR results and predictor variables in 2010 as input in the BLR equation.

Land cover change probability map shows the odds ratio related to the value of each predictor. Odds ratio mentioned is the percentage of the land cover change probability in an area by considering existing variables. Event odds is defined as the odds of an outcome occurring, divided with the odds of an outcome (land cover change) not occurring. Generally, the odds ratio is a group of odds divided by other odds. The odds ratio for predictors is defined as the relative number, outcome odds will increase (odds ratio > 1) or decrease (odds ratio < 1) when predictor variable value increase by 1 unit.

Table 4. The threshold value of each land cover class in West Java with BLR. Threshold value serves as a cut off in determining land covers that undergo or does not undergo changes.

| Class      | Threshold |
|------------|-----------|
| Forests    | 0.1594    |
| Fields     | 0.2384    |
| Plantations| 0.2697    |
| Settlements| 1.0000    |
| Rice fields| 0.0888    |
| Shrubs     | 0.6423    |

After the clarification process was done for each land cover change class, land cover change prediction map in West Java is obtained in this stage with change and not change categories. Figure 3 shows land cover change prediction map in West Java in 2010-2015 which is the result of modelling using BLR method.
Figure 3. The land cover change probability map using BLR. The land cover change probability map is classified into changing class (value 0) and unchanging class (1), with probability limit obtained from BLR calculation in each land cover class.

Land cover change prediction model verification was done by comparing it to actual land cover data of 2015. The comparison is in the form of cross-tabulation of the two maps to calculate the accuracy of the prediction model. In this stage, the degree of truth was determined with overall accuracy (OA) value. OA value represents the total comparison of pixels that are correctly classified in each class on reference pixel total (observation). The value indicates the truth accuracy or model accuracy resulted from land cover change classification done. Using data recapitulation, the resulting OA value is 49.79%. This value indicates that the prediction result (change or not change) reflects the actual condition accurately.

4. Conclusions
Considering the OA value obtained, BLR method should be used in identifying the driving factors of land cover change in West Java. The results of future land cover prediction by calculating the driving factors of change that occur in each land cover class are also proven to have a better level of accuracy compared to the prediction results in the study area as a whole. The identification of land cover change driving factors done in each land class in West Java provides a more detailed idea regarding influential factors (either driving or static). The results showed that the 13 driving factors tested in this study had non-uniform effects or "impetus" on each land cover class.

5. References
[1] Joshi K, Suthar C, editors. Changing urban land use and its impact on the environment (a case study of Jaipur City). Proceeding Asian Conference on Remote Sensing (ACRS); 2002.
[2] Koomen E, Rietveld P, de Nijs T. Modelling land-use change for spatial planning support. Springer; 2008.
[3] Lambin EF, Geist HJ, Lepers E. Dynamics of land-use and land-cover change in tropical regions. Annual review of environment and resources. 2003;28(1):205-41.
[4] Neufeldt H, Dawson IK, Luedeling E, Ajayi OC, Beedy T, Gebrekirstos A, et al. Climate change vulnerability of agroforestry. ICRAF Working Paper: World Agroforestry Centre; 2012.
[5] Briassoulis H. Analysis of land use change: theoretical and modeling approaches, the web book of regional Science. Regional research institute, West Virginia University, USA. 2000.
[6] Goldewijk KK, Ramankutty N. Land cover change over the last three centuries due to human activities: The availability of new global data sets. GeoJournal. 2004;61(4):335-44.
[7] Turner BL, Skole D, Sanderson S, Fischer G, Fresco L, Leemans R. Land-use and land-cover change: science/research plan. [No source information available]. 1995.
[8] Deliar A. Pemodelan Hibrid dalam Prediksi Dinamika Perubahan Tutupan Lahan (Studi Kasus: Wilayah Bandung). In: Teknik Geodesi dan Geomatika ITB, editor. Bandung2010.
[9] Xie Y, Mei Y, Guangjin T, Xuerong X. Socio-economic driving forces of arable land conversion: A case study of Wuxian City, China. Global Environmental Change. 2005;15(3):238-52.
[10] Hilbe JM. Logistic regression models: Chapman and hall/CRC; 2009.