Spatial model of deforestation using Geographic Information System (GIS) and logistic regression in Besitang forest

N Sulistiyono1,2*, P Patana1 and A Susilowati3
1Department of Forest Conservation, Faculty of Forestry, Universitas Sumatera Utara, Medan, Indonesia.
2Center of Excellence for Mangrove, Universitas Sumatera Utara, Medan, Indonesia.
3Department of Forest Management, Faculty of Forestry, Universitas Sumatera Utara, Medan, Indonesia.

E-mail: *nurulinsulistyono@usu.ac.id

Abstract. The existence of Besitang forests is under threat of deforestation so that its area continues to decline. Besitang forest is part of the Gunung Leuser National Park. Efforts to prevent deforestation need to be carried out with a proper deforestation policy. As a first step, it is crucial to identify the drivers of deforestation. This study aims to obtain information on the drivers of deforestation in Besitang forests from 2000 to 2016. The method used is to do spatial modeling using GIS and binary logistic regression. The results of the research show that elevation factors, distance from settlements, and distance from oil palm plantations are factors that trigger deforestation that occurs in Besitang forests.

1. Introduction
The application of GIS has been widely used for environmental modeling, one of which is for modeling deforestation. Geographical information, in its simplest form, is information related to a particular location [1]. A geographic information system is a tool in processing spatial data into information [2]. On the other hand, binary logistic regression is a statistical method that is part of a general linear model (GLM). GLM can be used when the dependent variable does not follow a normal distribution. The application of GIS and binary logistic regression to modeling deforestation has been widely used by other researchers [3-6]

A model is an abstraction of reality. A simple model can be interpreted as any form of presenting reality or a simplified form of reality [7]. Real-world models can make it easier for humans to understand the study of the selected application area by reducing the number of complexities. The definition and understanding of a simple model can be understood that a model is essentially a simplified form of reality or real-world phenomena. The deforestation model is part of the land-use change model. The phenomenon of land-use change can be modeled in many ways so that various models can be obtained. In this study, the deforestation model was modeled spatially by using a binary logistic regression statistical approach.

Besitang Forest is a habitat for Sumatran tigers, Sumatran elephant, and Sumatran pongo, whose existence continues to experience pressure. This can be seen from a study conducted by [8], which reports the deforestation rate of 224.14 ha per year. On the other hand [9] also says that forest degradation on a massive scale of 527.85 ha in 2008 - 2016.
It is important to know the knowledge of the spatial factors that trigger deforestation through the creation of a spatial model of deforestation. The spatial model of deforestation is useful in forest management activities, especially for mitigating future deforestation events. This study aims to obtain a spatial model of deforestation in Besitang forests using a binary logistic regression approach.

2. Materials and methods

2.1. Research location
This research was conducted in the Management Section of the National Park (MSNP) region VI Besitang. The research location is part of Gunung Leuser National Park (GLNP), Langkat Regency, North Sumatra Province (Figure 1).

![Figure 1. Map of the study area in Besitang](image)

2.2. Spatial model of deforestation
Land cover information is obtained from Landsat 5 satellite imagery recording date of 5 July 2000 and Landsat 8 image recording date of 1 July 2016, path row 129/057, and path row 129/058. Elevation and slope data are derived from Shuttle Radar Topography Mission (SRTM). Data of distance from the road, distance from the settlement, distance from the river, and distance from oil palm plantation were obtained using the euclidian distance [8,10]. Whereas the total population and total of farm households data for 2017 were obtained through the Central Statistics Agency.

Binary logistic regression was used to obtain a deforestation model. Randomly selected sample pixels are used to build a model of deforestation. The dependent variable in the logistic regression is binary (presence or absence of events). Value of 1 means that there has been deforestation and a value of 0, which means non-deforestation in 2000-2016. In general, the formulated of logistic regression equation can be followed [11].

\[
\pi(x) = \frac{\exp(\beta_0 + \beta_1x_1 + \cdots + \beta_px_p)}{1+\exp(\beta_0 + \beta_1x_1 + \cdots + \beta_px_p)}
\]  

Information:
\(\beta_0 = \text{intercept}\)
\[ \beta_1 = \text{coefficient of variable } X \]
\[ X_{1-p} = \text{variable } X \]
\[ \pi = \text{chance of change in forest land (0-1)} \]
\[ X_1 = \text{distance from river (m)} \]
\[ X_2 = \text{distance from the road (m)} \]
\[ X_3 = \text{elevation (m)} \]
\[ X_4 = \text{slope (\%)} \]
\[ X_5 = \text{total of the population (person)} \]
\[ X_6 = \text{distance from settlement (m)} \]
\[ X_7 = \text{distance from the oil palm plantation (m)} \]
\[ X_8 = \text{total of farm household (family)} \]

The process of selecting the drivers of deforestation is carried out using the stepwise method by reducing or removing unaffected trigger factors from the model. The model validation test was carried out using the overall accuracy (OA) by comparing the predicted deforestation with the actual deforestation. Processing of the spatial distribution model of deforestation in raster data format was carried out using Arc GIS 10.7 software.

3. Results and discussion

3.1. The driving factors of deforestation
Drivers of deforestation are factors that can trigger deforestation. Determination of the driving factors for deforestation refers to previous studies [3,5]. The spatial data of the factors that triggered deforestation in 2000 - 2016 in the early years can be seen in Figure 2 below.

![Image of driving factors of deforestation](image)

**Figure 2.** The driving factors of deforestation: a) distance from rivers, b) distance from roads, c) elevation, d) distance from oil palm plantation, e) slope, f) total of population, g) distance from the settlement, h) total of farm households

3.2. Deforestation model
Based on the multicollinearity test, four variables have a high correlation with other independent variables so that they are excluded from the model, the variables are the distance from the road, distance from the river, slope, and a total of population. The multicollinearity test is intended to see the relationship between the drivers of deforestation as a predictor variable. A good logistic regression model should avoid multicollinearity problems.
The total number of pixels used to build a spatial model of deforestation in the Besitang forest landscape is 392 pixels consisting of 194 deforestation event pixels (1) and 198 non-deforestation pixels (0) randomly selected. The test results using the stepwise model produced three variables that triggered deforestation in Besitang forests during 2000 - 2016. The binary logistic regression equation formed is presented as follows:

$$
\pi = \frac{\exp(39.764 - 0.004 x6 + 0.003 x7 - 0.096 x3)}{1 + \exp(39.764 - 0.004 x6 + 0.003 x7 - 0.096 x3)}
$$

The results of the spatial model of deforestation in Besitang forests during 2000 - 2016 showed that there were three trigger factors, namely elevation, distance from settlements, and distance from oil palm plantations. Meanwhile, the factor of the number of farm households was excluded from the model based on the stepwise method.

Deforestation in Besitang forest is suspected to be caused by forest opening areas for oil palm plantations and settlements. The odds ratio value of the distance from the settlement is 0.996, meaning that pixels near to the settlement have a probability of 1.004 greater than pixels that are farther residential. The elevation odds ratio elevation value is 0.908, meaning that the lower elevation pixels have a deforestation probability of 1.101 higher than the pixels at higher elevations. The odds ratio value for the distance from the plantation is 1.003, which means that pixels far from oil palm plantations have a deforestation probability of 1.003 higher compared to pixels near to oil palm plantations. This indicates that the deforestation that occurs in Besitang forest is mostly in low elevation areas, near to existing settlements, and far from existing oil palm plantations.

The results of this study are different from the results of the study [3], which mentions the factors of population density, elevation, distance from the road, the number of farm households, which say the factors that trigger deforestation on the island of Java. The research result [12] states that the driving factors for deforestation in the Romanian Carpathian Mountains are the distance from the forest edge, forest fragmentation, slope, elevation, distance from roads, and distance from housing. The results of this study further confirm the opinion of [5] that the driving factors for deforestation for each region are specific so that policies to address deforestation problems cannot be generalized for all regions.

![Figure 3](image-url)

**Figure 3.** a) value of deforestation probability based on logistic regression models, b) deforestation estimate based on logistic regression models, c) actual deforestation

Testing the feasibility of the model using the Hosmer and significance test Lemeshow in the deforestation model is 0.998 (> 0.05) so that it can be said that the deforestation model
is fit (accept $H_0$), which shows the model can explain the data or is appropriate so that the deforestation model formed is acceptable or feasible to use. Deforestation that occurs in Besitang forest landscapes is influenced by factors such as the amount of distance from the settlement, distance from the oil palm plantation, and elevation. The coefficient of determination ($R^2$) is 0.985, which means 98.5% of deforestation can be explained by variables of distance from settlements, distance from oil palm plantation, and elevation.

In spatial modeling using binary logistic regression, deforestation is identified as pixels with an opportunity value of more than 0.5. In contrast, non-deforestation is identified as pixels with a value of less than 0.5. The distribution of the opportunities value for deforestation in this study can be seen in figure 3a. In contrast, the comparison of deforestation from the model and the actual value can be seen in figures 3b and 3c. The result of the model validation test in the form of overall accuracy shows a value of 98.7%, which means that the pixel classification error rate based on the model is only 1.3%.

4. Conclusions
A spatial model using GIS and binary logistic regression can be used to model deforestation in Besitang forests from 2000 - 2016. The elevation factors, distance from settlements, and distance from oil palm plantations trigger deforestation in Besitang forests.

References
[1] Martin and David 1996 Geographic Information System (Routledge, London)
[2] DeMers and Michael N 1997 Fundamentals for Geographic Information System (USA: John Wiley & Sons Inc.)
[3] Prasetyo LB, Kartodihardjo H, Okarda B, Adiwibowo S and Setiawan Y 2009 Spatial model approach on deforestation of Java Island, Indonesia JIFS 6 pp 37 - 44.
[4] Arsanjani JJ, Helbich M, Kainz W and Darvishi B A 2013 Integration of logistic regression, Markov chain and cellular automata models to simulate urban expansion International Journal of Applied Earth Observation and Geoinformation. 21 pp 265-75
[5] Sulistiyono N, Jaya INS, Prasetyo L and Tiryana T 2015 Spatial Model of Deforestation in Sumatra Islands Using Typological Approach Journal of Tropical Forest Management 21(3) pp 99-109.
[6] Bavaghar MP 2015 Deforestation modelling using logistic regression and GIS Journal of Forest Science 61 (5) pp 193-9
[7] Kuby M, Tierney S, Roberts T and Upchurch C 2005 A comparison of geographic information systems, complex networks, and other models for analyzing transportation network topologies (Hampton, Virginia, US : NASA Langley Research Center)
[8] Sulistiyono N, Ginting BSP, Patana P and Sisulowati A 2019 Land Cover Change and Deforestation Characteristics in The Management Section of National Park (MSN) VI Besitang, Gunung Leuser National Park Journal of Sylva Indonesiana (JSI) 02(02) pp 91 – 100
[9] Sulistiyono N, Nifrody T, Patana P and Sisulowati A 2019 Estimation of forest degradation distribution using landsat satelite imagery in Besitang forest landscape The 8th International Symposium for Sustainable Humanosphere IOP Conf. Series: Earth and Environmental Science (2019)
[6] Sulistiyono N, Rambe BA, Patana P and Purwoko A 2020 Spatial model of the Sumatran tigers (Panthera tigris sumatrae) prey habitat suitability index in Besitang International Conference on Agriculture, Environment and Food Security (AEFS) 2019
[11] Hosmer DW and Lemeshow S 2000 Applied Logistic Regression. (USA: Wiley)
[12] Kucsicca G and Dumitriță C 2019 Spatial modelling of deforestation in Romanian Carpathian Mountains using GIS and Logistic Regression Journal of Mountain Science 16(5)

Acknowledgements
This study was fully supported by DRPM Grant 2017 from the Ministry of Research, Technology and Higher Education, Republic of Indonesia, and GLNP on the support of secondary data.