

# Forest cover dynamics analysis and prediction modelling using logistic regression model (case study: forest cover at Indragiri Hulu Regency, Riau Province)

**Irmadi Nahib\* and Jaka Suryanta**

Geospatial Information Agency

Jl Raya Jakarta Bogor KM 46 Cibinong, Jawa Barat, 16911, Indonesia

E-mail:irmnahib@gmail.com, irmadi.nahib@big.go.id

**Abstract.** Forest destruction, climate change and global warming could reduce an indirect forest benefit because forest is the largest carbon sink and it plays a very important role in global carbon cycle. To support Reducing Emissions from Deforestation and Forest Degradation (REDD +) program, people pay attention of forest cover changes as the basis for calculating carbon stock changes. This study try to explore the forest cover dynamics as well as the prediction model of forest cover in Indragiri Hulu Regency, Riau Province Indonesia. The study aims to analyse some various explanatory variables associated with forest conversion processes and predict forest cover change using logistic regression model (LRM). The main data used in this study is Land use /cover map (1990 – 2011). Performance of developed model was assessed through a comparison of the predicted model of forest cover change and the actual forest cover in 2011. The analysis result showed that forest cover has decreased continuously between 1990 and 2011, up to the loss of 165,284.82 ha (35.19 %) of forest area. The LRM successfully predicted the forest cover for the period 2010 with reasonably high accuracy (ROC = 92.97 % and 70.26 %).

## 1. Introduction

Forest is a natural resource that is very important and useful for life and living either directly or indirectly. Direct benefits from the existence of the forest are timber, non-timber products and wildlife. While the indirect benefits are environmental services, such as watersheds, aesthetic function, an oxygen supplier and carbon sink [1].

Indonesia has the highest deforestation rates in the world, exceeding even Brazil while having only a quarter of Brazil's forest area [2]. The average annual deforestation for the period 2000-2012 was 671,420 hectares, accounts for 525,516 ha of deforestation in mineral land 600 and 145,904 ha of deforestation in peat land. During this period, more than 80 percent of deforestation occurred in Kalimantan and Sumatra, while Sulawesi and Papua follow with 9 percent and 6 percent, respectively [3]. Meanwhile, refers to [4] the rate of deforestation in the Riau Province for the period 2000-2009 was 200,290.8 hectares.

Forest destruction, climate change and global warming reduce the indirect benefit of forest because forest is the largest carbon sink and play a very important role in global carbon cycle and can hold carbon at least 10 times greater than other vegetations prairie grass, crops and tundra [5]. Assessing the conversion of a forested landscape may help us to understand the way of natural resources



extraction occurs, and consequently the human influences on the forest ecosystem services. For better understanding of the impact of forest cover change, factors affecting it must be fully studied [6].

Forest cover dynamics is rate, pattern, spatial distribution, and quantity of change from forest cover to other land coverages due to either natural condition or human induced causes. The constant interplay of various human induced disturbances along with topographic and climatic factors can gradually degrade a healthy forest cover or change it to other land use/land cover category. Understanding of the causes of land use change has moved from simplistic representations of two or three driving forces too much more profound understanding that involves situation-specific interactions among a large number of factors at different spatial and temporal scales. Studies related to forest cover change using satellite-derived information help in understanding the phenomena like carbon dynamics, climate change and threat to biodiversity [6].

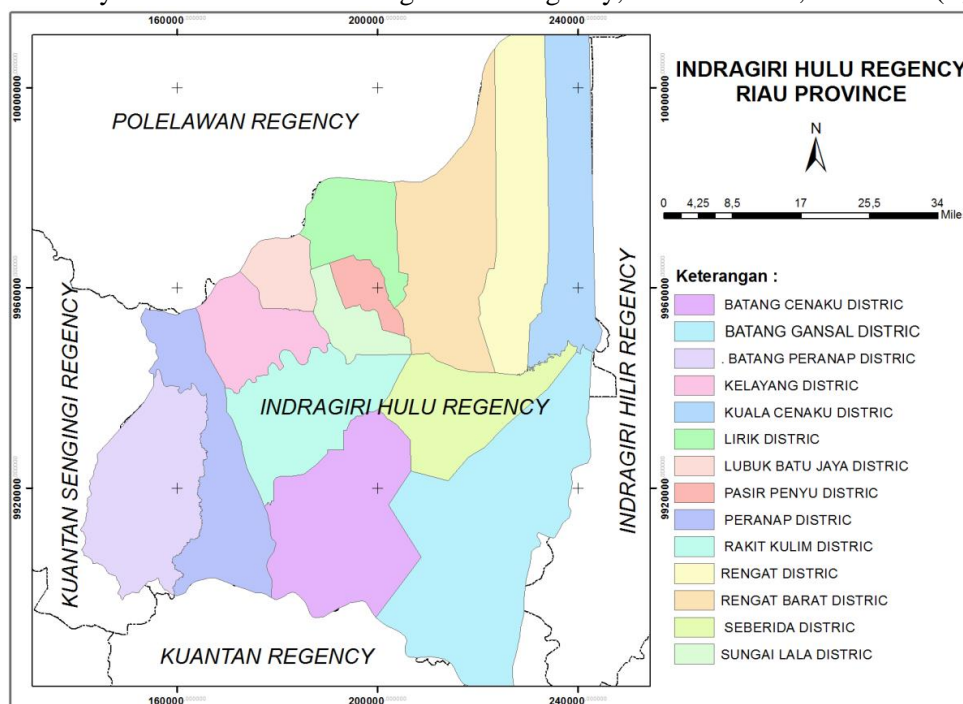
Recent advancement in remote sensing and GIS methods also enable researchers to model and predict land use/land cover more efficiently than ever-before. Several approaches have also been developed to model and predict the dynamics of land use / land cover [7].

The present study aims to analyze explanatory variables associated with forest conversion process and to model the forest cover change using logistic regression model.

## 2. Methods

### 2.1. Study area and data

The present study was carried out in Indragiri Hulu Regency, Riau Province, Indonesia (figure 1).



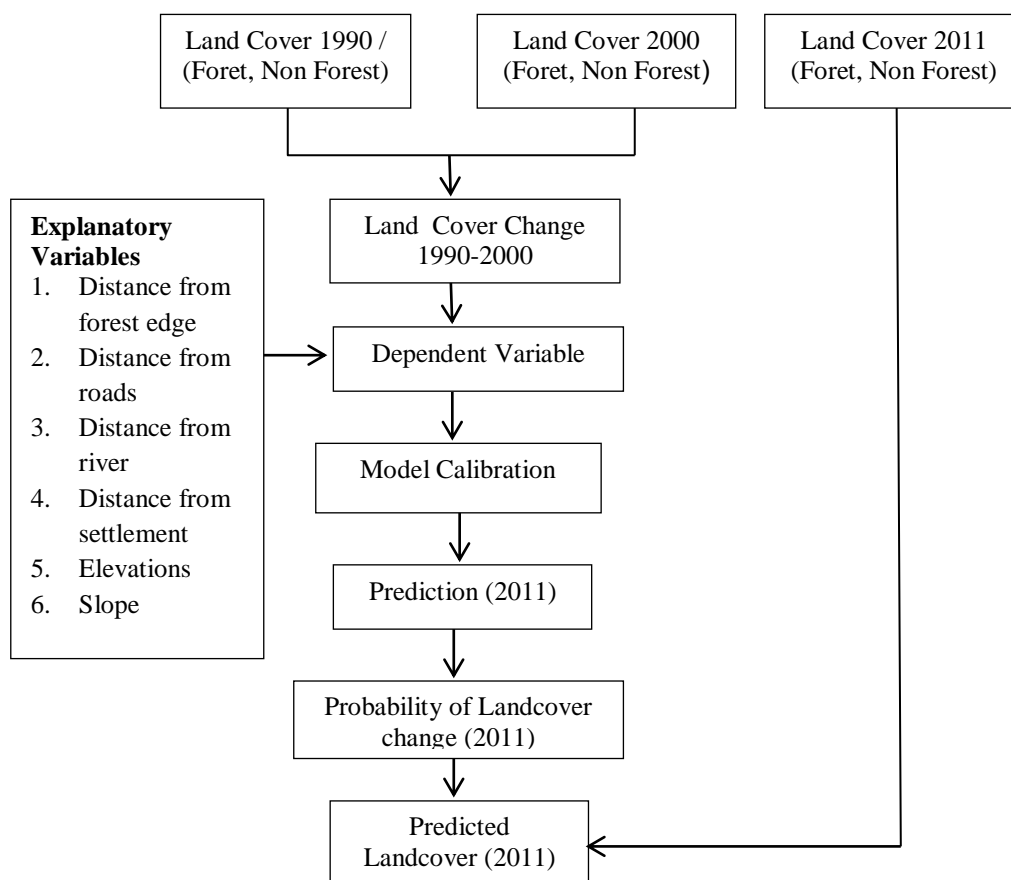
**Figure 1.** The map of study area

Indragiri Hulu is a regency of Riau, Indonesia. It has an area of 819,826 ha consists of Forest zone 189,368 ha and Forest for others (APL) 630,458 ha. The regency is divided into 14 districts, the seat of the regency is located at Rengat. The Forest zone of 189,368 Ha, consists of 84,753 National Park of Bukit Tigapuluh (NP BT), protected forest 34,758 and 69,856 wildlife reserve. Data used in the assessment:

- Indonesian Topographic Map, Scale 1: 50,000, Geospatial Information Agency (BIG)

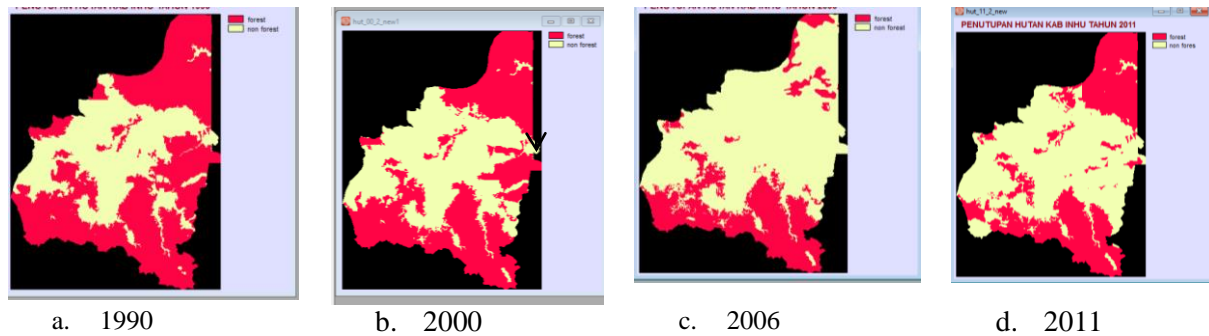
- Forest Area, Inland Water, Coastal and Marine Ecosystem Map of Indragiri Hulu Regency, Riau Province, Scale 1: 250,000 Directorate General of Forestry Planning Ministry of Forestry.
- Land Use/Cover Map of Indragiri Hulu Regency, Riau Province, Scale 1: 250,000 (1990, 2000, 2006 and 2011), Directorate General of Forestry Planning, Ministry of Environment and Forestry.
- Indragiri Hulu Regency Administration Map, Planning and Regional Development Agency (*Bappeda*) of Indragiri Hulu Regency, Riau Province.

Landcover change analysis was performed by a comparison method of landcover map series produced by Directorate General of Forestry Planning, Ministry of Environment and Forestry. This land cover map was produced through a visual interpretation of the 324 Landsat image. The determination of land cover area used the spatial analysis which is done by Land Change Modeler in Indrissi TerrSet process of the these map of Indragiri Hulu Regency in 1990, 2000, 2006 and 2011. Flowchart stage research activities are presented figure 2.



**Figure 2.** Flowchart stage research

The row data (Land Use/Cover Map) within 23 landcover catagories. For simplify of comparison, the catagories have been reduced to two categories of forest and non-forest. Forest is consist of the dryland forest, swamp forest, mangrove forest and timber forest plantation. The other catagories are included in the non-forest. Finally a boolean map of forest and non-forest was generated for all the four periods (figure 3).

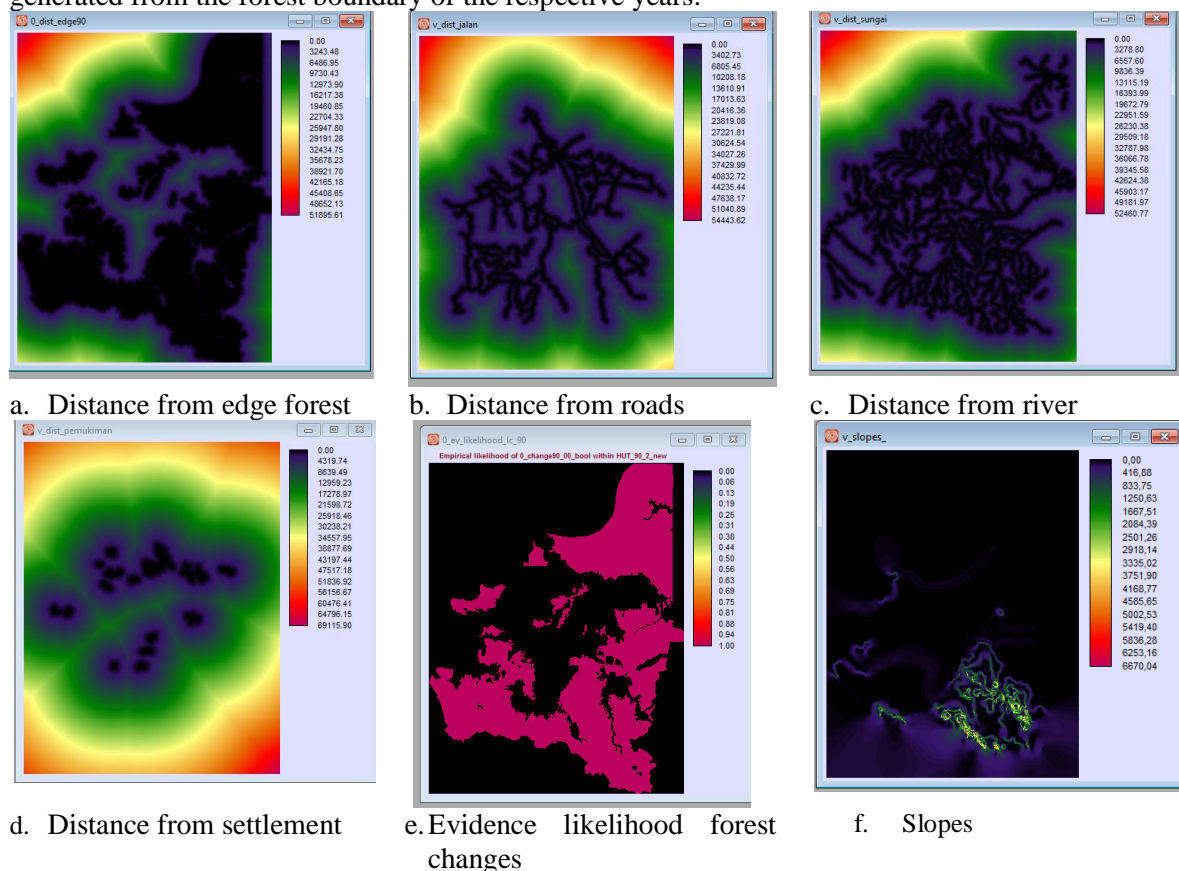


**Figure 3.** Forest cover during (a) 1990, (b) 2000, (c) 2006 and (c) 2011.

Based on Indonesian Topographic Map which collected from Geospatial Information Agency, we obtained information of road, river, settlement, and elevation layer. Moreover, we considered four distance parameter as the independent variables: distance from forest edge, roads, river, settlements, and slope; and forest cover change as the independent variable (figure 4).

## 2.2. Logistic regression model

A Boolean image with the categories ‘forest change’ (forest to non-forest) and ‘no change’ (forest remained unchanged) were generated for the period 1990–2000 and 2000–2011. Distance measures the euclidean distance between each cell and the nearest of a set of target features. Deforestation tends to start from the edge of existing forest [8]. Hence, Distance from forest edge was considered as one of the explanatory variables of forest cover change. The distance from forest edge for 1990 were generated from the forest boundary of the respective years.



**Figure 4.** Independent variable (environmental variable)

LRM was used to model and analyze the forest change in IDRISI TerrSet. The objective of the present study was to assess the importance of the explanatory variables on forest change from 1990 to 2000 and predicting the probability of change by 2011.

The explanatory variable test procedure is based on Cramer's V contingency table analysis which can test the strength of the association between the dependent variable. The binary presence or absence is the dependent variable (1 = forest change and 0 = no change) for the periods 1990–2000.

The probability of forest change is considered to be a function of the explanatory variables. It is a monotonic curvi linear response bounded between 0 and 1 [10] and defined by the logistic function:

$$p = E(Y) \frac{e^{\beta_0 + \beta_1\chi_1 + \beta_2\chi_2 + \beta_3\chi_3 + \beta_4\chi_4}}{1 + e^{\beta_0 + \beta_1\chi_1 + \beta_2\chi_2 + \beta_3\chi_3 + \beta_4\chi_4}} \quad (1)$$

where

$p$  = the probability of forest change,

$E(Y)$  = the expected value of the dependent variable  $Y$ ,

$\beta_0$  = a constant to be estimated,

$\beta_i$  = the coefficient to be estimated for each explanatory variable  $\chi_i$ .

This logistic function (equation 1) can be transformed (equation 2) into linear function (equation 3) which is called a logit or logistic transformation

$$\text{logit}(p) = \log_e \left( \frac{p}{1-p} \right) \quad (2)$$

$$\text{logit}(p) = \beta_0 + \beta_1\chi_1 + \beta_2\chi_2 + \beta_3\chi_3 + \beta_4\chi_4 \quad (3)$$

### 2.3. Model calibration and prediction

All the independent variables were normalized between 0.1 and 0.9 before introducing them in the model. The natural log transformation was done for the continuous variables (distances). For the categorical explanatory variable (slope position class), the evidence of likelihood transformation was applied. The regression equation of the best-fitted predictor set and the probability of forest change for 2000 were generated. The probability of forest change for 2000 was used to predict the change in forest between 2000 and 2010. The regression equation of the best-fitted predictor set and the probability of forest change for 2011 were generated. The threshold value was selected from the residual file of probability of forest change of 2010 and it was used to generate change - no change binary image (0 = forest to forest, non forest to non forest) and 1 = forest to non-forest). The resulted image is the predicted forest cover for the period 2011.

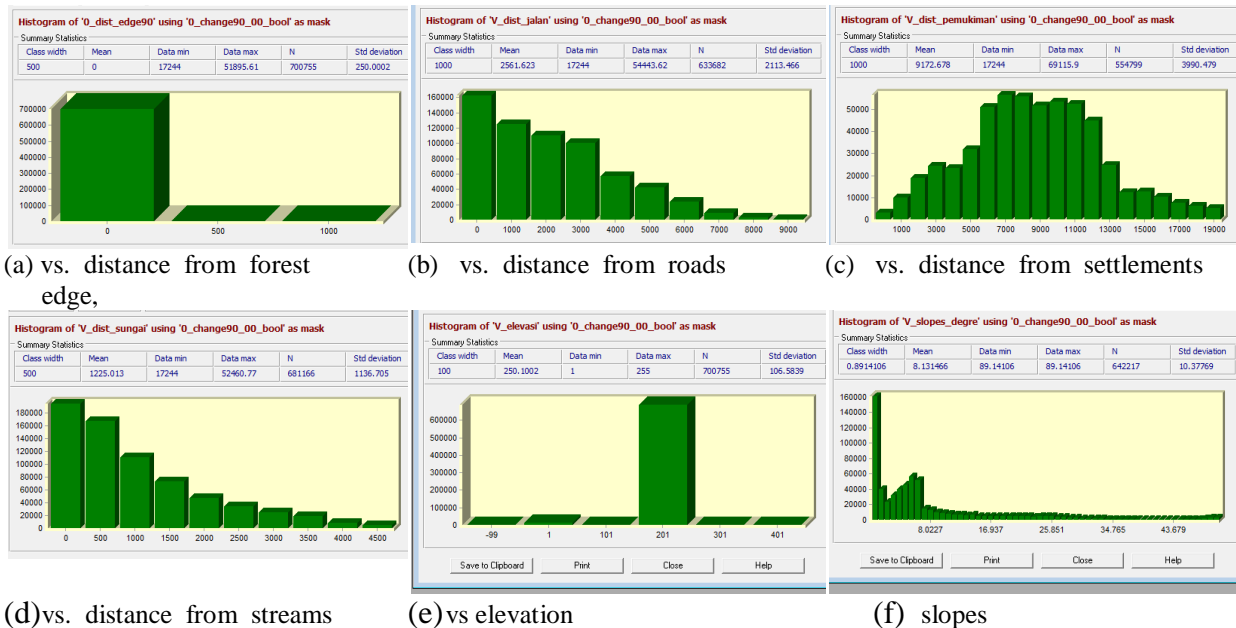
### 2.4. Validation of prediction

The predicted forest cover of 2010 was validated using Relative Operating Characteristic (ROC) /Area Under Curve (AUC) module of IDRISI TerrSet. The ROC module is comparing a suitability image depicting the likelihood of that class occurring (the input image) and a boolean image showing where that class actually exists (the reference image).

The ROC curve is the true positive fraction vs false positive fraction and the AUC is a measure of overall performance. The predicted forest cover map of 2011 was compared with actual forest cover map of 2011 using 100 sampling points. A ROC/AUC curve was generated between the false positive (%) and true positive (%).

## 3. Results

Figure 5 shows the relationship between dependent variable and the explanatory variables. This relationship was useful for selection of the explanatory variables for forest change in the study area.



**Figure 5.** Relationship between dependent and explanatory variables Forest cover change (a) vs. distance from forest edge, (b) vs. distance from roads, (c) vs. distance from settlements, (d) vs. slope position classes, (e) vs elevation, (f) slopes

Most of the forest change has taken place near the forest edges. Occurrences of forest change within 500 m of forest edge were noticed to be highly distinct but forest change dropped off to virtually nil beyond that, indicating a non-linear relationship between forest edge and forest change. The incidence of forest change near roads was found to be noticeable but dropped off to zero after 7 km from the roads. Influence the distribution of the settlement to the forest changes occurred in around 1 km up to 13 km from the settlements.

Therefore, these variables were tested with Cramer's V and then included in the model.

**Table 1.** Association between dependent variable (forest cover change) and explanatory variables using Cramer's V

Explanatory variable	Forest changes 1990-2000		Forest changes 2000-2006	
	Cramer's V	P value	Cramer's V	P value
Dist from forest edge	0.5549	0.0000	0.4035	0.0000
Dist from roads	0.3796	0.0000	0.2126	0.0000
Dist from settlements	0.4412	0.0000	0.3261	0.0000
Distance from streams	0.2046	0.0000	0.1193	0.0000
Slope position	0.11682	0.0000	0.3284	0.0000
Evidence likelihood forest changes	0.5981			

V values for variables : distance from roads, settlements and forest edge, were found to be between 0.1 and 0.5 with a p value of 0.0, indicating a poor association, there is no strong relationship with forest change. Logistic regression modeling refers to equation 3,

$$Y_1 = -0.1406 - 0.1661x_1 - 0.0001068x_2 - 0.00001650x_4 - 0.0005861x_5 - 0.0024x_6 \quad (4)$$

$$Y_2 = -1.0715 - 0.7204x_1 + 0.0001487x_2 - 0.0001362x_3 + 0.00008544x_4 + 0.0077x_5 \quad (5)$$

where : $x_1$  = distance from edge forest,  $x_2$  = distance from roads,  $x_3$  = distance from settlement

$x_4$  = distance from streams,  $x_5$  = elevation,  $x_6$  = slope

Based on table 2, The predictor based on forest changes 1990-2000, more hight (of model chi-square) than forest changes 1990-2000. A high value indicates that the forest change was less expected under the null hypothesis than the full regression model. Pseudo  $R^2$  value greater (0.5946 and 0.3750) than 0.2 indicates that the model is a relatively good fit for the data.

**Table 2.** Statistic of logistic regression

	Based on forest changes 1990-2000	Based on forest changes 2000-2006
Number of total observations	15,113,120	15,113,120
-2logL0	1,111,834.2390	54,2066.4826
-2log(likelihood)	450,748.1071	338,814.2830
Goodness of fit	395,226.7384	572,979.8677
Pseudo $R^2$	0.5946	0.3750
Chi-square ( $df=5$ )	661,086.1318	203,252.1996
ROC	0.9614	0.9287
AUC	0.9044	0.7026

The area under forest and nonforest during 1990, 2000, 2006 and 2010 (actual and predicted) is shown in table 3.

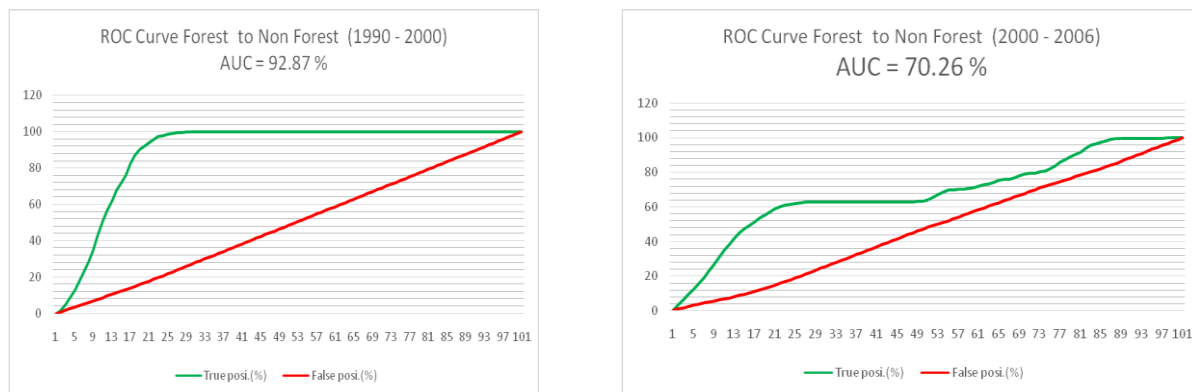
**Table 3.** Areas under forest cover during 1990, 2000, 2006, and 2010 (actual and predicted).

Area (ha)	Area (Percentage)				
	1990	2000	2006	2011	2011 (predicted)
Forest	60.93	52.75	30.09	39.48	44.92
Non Forest	39.07	47.25	69.91	60.52	55.08
Total	100.00	100.00	100.00	100.00	100.00

It was observed that the study area experienced a continuous loss of forest cover, leading to the loss of 165,284.82 ha (35.19 %) of forest during 1990–2011. Average the rate of deforestation in Indragiri Hulu Regency for the period 1990-201 was 7,870.71 ha hectares.

The accuracy estimate of the forest cover maps of 1990, 2000, 2006 and 2010 is shown in table 3. All the maps were having accuracy of more than 85%. ROC/AUC graph generated between model predicted forest cover and the actual forest cover of 2011 is shown in figure 6. The area under ROC curve are 92.97 % and 70.26 % (based on forest changed 1990-2000 and 2000 – 2006) which gives an accuracy of 92.97 % and 70.26 % for the predicted forest cover of 2011.





a. Based on forest changes 1990-2000

b. Based on forest changes 2000-2006

**Figure 6.** Validation of LRM prediction (AUC/ROC)

The driving factors of forest cover change may vary from place to place. In the present study, the selected explanatory variables encompass a substantial share of the factors driving forest cover changes. Specifically, the accessibility variables seem to be more important than the topographical ones. Many studies have indicated that most of these factors were also found to be important in other areas. Proximity to road, town and forest edge were found to be the important factors of forest change in southern Cameroon [10]

Refer to [11] modeled distance from roads, settlements and forest edge, and topography for tropical forest conversion and found that all these explanatory variables were highly significant predictor of forest conversion (ROC = 87%). Meanwhile [7] studied the effects of six factors : distance from roads and settlements, forest fragmentation index, elevation, slope and distance from the forest edge, on deforestation. An inverse relationship of forest change with distance from roads, settlements, forest edge and elevation was observed. The ROC of prediction of forest cover change was found to be 96%.

#### 4. Conclusions

The analysis of forest cover change revealed that the forest area has undergone continuous change leading to the loss of 165,284.82 ha (35.19 %) forest in the past 1990–2011. The explanatory variables, i.e., distance from forest edge, roads and settlements, slope position classes were significantly associated with the forest cover change in the study area.

The LRM efficiently modelled all the explanatory variables associated with forest change and helped in analyzing them for their relative significance on forest change process in the area. The model for predicting the forest cover change :

$$Y_1 = -0.1406 - 0.1661x_1 - 0.0001068x_2 - 0.00001650x_4 - 0.0005861x_5 - 0.0024x_6 \quad \text{and} \\ Y_2 = -1.0715 - 0.7204x_1 + 0.0001487x_2 - 0.0001362x_3 + 0.00008544x_4 + 0.0077x_5.$$

It predicted the forest–non forest cover of the area for 2011 with an accuracy of 92.87 % and 70.26 %.

#### Acknowledgments

The authors are thankful to the Centers for Research, Promotion and Cooperation, Geospatial Information Agency (BIG) for the data and financial support.

#### References

- [1] Nahib I and Widjojo S 2016 *The Impact of Landcover Changes on Carbon Stock :A Study Case In Central Kalimantan Forest (ACRRS)*
- [2] Margono B A 2014 Primary forest cover loss in Indonesia over 2000-2012 *Nature Climate Change* **4** 730-735



- [3] Government of Indonesia 2014 National Forest Reference Emission Level for Deforestation and Forest Degradation in the Context of the Activities Referred to Decision 1/CP.16. Paragraph 70 (REDD+) Under the UNFCCC. Submission by Indonesia
- [4] Sumargo W, Nanggara S G, Nainggolan F A and Apriani I 2011 *Potret Keadaan Hutan Indonesia 2000-2009 Edisi I* (Forest Watch Indonesia)
- [5] Adiriono T 2009 Method of Measuring Carbon (Carbon Stock) on Industrial Plantation Forest Crassiparva Acasia Type. *Thesis*. Faculty of Forestry Programs Graduate. University of Gadjah Mada. Yogyakarta. not published
- [6] Kumar S, Nandy S, Agarwal R and Kushwaha S P S 2014 Forest cover dynamics analysis and prediction modeling using logistic regression model *Ecological Indicators* **45** 444–455
- [7] Arekhi S 2011 Modeling spatial pattern of deforestation using GIS and logistic regression: a case study of northern Ilam forests, Ilam province, Iran *Afr. J. Biotechnol* **10** 16236–16249
- [8] Eastman J R 2006 *IDRISI 15 Andes. Guide to GIS and Image Processing* (Worcester, MA: Clark University)
- [9] Pontius Jr R G and Schneider L C 2001 Land-cover change model validation by an ROC method for the Ipswich watershed, Massachusetts, USA *Agric. Ecosyst. Environ.* **85** 239–248.
- [10] Mertens B and Lambin E F 2000 Land-cover-change trajectories in southern Cameroon *Annals of the association of American Geographers* **90** 467-494
- [11] Siles N J S 2009 Spatial Modelling and prediction of tropical forest conversion in the Isiboro Secure National Park and Indigenous Territory (TIPNIS), Bolivia. *Graduate Thesis*. International Institute for Geoinformation Science and Earth Observation, Enschede, The Netherlands.