Spatial Modelling and Prediction of Tropical Forest Conversion in the Isiboro Sécure National Park and Indigenous Territory (TIPNIS), Bolivia

By
Nelson Jery Sanabria Siles
February, 2009

Spatial Modelling and Prediction of Tropical Forest Conversion in the Isiboro Sécure National Park and Indigenous Territory (TIPNIS), Bolivia

by

Nelson Jery Sanabria Siles

Thesis submitted to the International Institute for Geo-information Science and Earth Observation in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation, Specialisation: (fill in the name of the specialisation)

Thesis Assessment Board

Chairman: Dr. Ir. C.A.J.M. de Bie (Department of Natural Resources, ITC)
External Examiner: Dr. J. Duivenvoorden (University of Amsterdam, UvA)
Supervisor: Dr. H.A.M.J. van Gils (Department of Natural Resources, ITC)

Supervisor:
Ir. B.G.C.M. Krol

INTERNATIONAL INSTITUTE FOR GEO-INFORMATION SCIENCE AND EARTH OBSERVATION
ENSCHEDE, THE NETHERLANDS

Disclaimer

A ti, que siempre estuviste cerca sin importar la distancia

Abstract

Forest conversion is occurring in the Isiboro Sécure National Park and Indigenous Territory (TIPNIS), Bolivia. Activities such as agricultural encroachment and forest extraction are leading to a rapid loss of primary forest and also have disturbed the traditional life of the indigenous communities. The present study has integrated the statistical approach of logistic regression and also that of artificial neural networks with GIS in an attempt to analyze and predict forest conversion in the TIPNIS.

Based on information obtained from land cover maps and satellite images, forest loss for the years 1976, 1986, 1991, 2001, 2004 and 2006 were calculated. According to the results of the study, during the period 1976 – 2006, 23% of primary forest has been lost in the southern part of the TIPNIS. The deforestation rates presented variations, they rose and fell and then rose again. The rates of deforestation were 0.005%.a $^{-1}$ until 1986, 1.3%.a $^{-1}$ until 1991, 0.5%.a^{-1} until 2001, and 2.3%.a^{-1} until 2004 and 3.5%.a^{-1} until 2006. This study revealed that the variations of deforestation rates in the TIPNIS coincide with the degree of control of coca (Erythroxylum coca) cultivation that the Bolivian government has permitted in the Chapare Province. When government controls of coca growing were more lax, the deforestation rates increased.

To model forest conversion this study considered the change that has occurred in the forest areas as a categorical dependent variable. The univariate tests of association Cramer's V was used to test five potential explanatory variables for forest conversion (*"Distance from Forest Edge", "Distance from Roads", "Distance from Settlements", "Landscape Position"* and *"Type of Settlement"). "Type of Settlement"* was excluded from the modelling because the data input (map) was too coarse.

Logistic regression analysis was used (i) to assess the relative significance of explanatory variables on forest change during the period 2001-2004; and (ii) to predict probability of forest change for the period 2004-2006. *"Landscape Position"* was the most significant explanatory variable, followed by the explanatory variables *"Distance from Forest Edge", "Distance from Roads", and "Distance from Settlements"*. Logistic regression prediction resulted in an Area Under a ROC Curve (AUC) of 85%.

Finally, the study made use of the artificial neural network Multi-Layer Perceptron (MLP) to improve the prediction of probability of forest change for the period 2004 to 2006. The prediction performed used the same data set used by the logistic regression prediction. The AUC obtained by MLP was 92%.

The predictive performance of both models proved successful. While MLP produces better prediction results in general, logistic regression analysis is still needed to understand the relative significance of the explanatory variables on the forest change.

Acknowledgements

I would like to express my gratitude to the International Institute for Geo-Information Science and Earth Observation (ITC), for the support of the fellowship, and also to thank all the staff for always extending a helping hand, a special thanks to Michael Weir.

My sincere thanks go to the Centro de Levantamientos Aeroespaciales y Aplicaciones SIG para el desarrollo sostenible de los recursos naturales (CLAS) for giving me the opportunity to be part of the Msc course. Thanks to its entire staff, and a special thanks to Enrique, Stephan, Sergio, Guy, Carlos and Nereyda.

And here at ITC I wish to acknowledge with deep gratitude:

- Hein van Gils for always giving me good advice, encouragement and enthusiastic support which made this work possible.

- Bart Krol for his excellent guidance and perceptive advice as well as for his friendship.

This document owes a debt of gratitude to the managers and guards at the TIPNIS, as well. Thanks to Vladimir Orzolini for providing the data to make this research possible and a special thanks to Raul Urquieta for his support during the data collection and fieldwork phases of this study. The guards of the park generously shared their knowledge about the TIPNIS and their company during the field work. I am indebted to them for keeping me clear of the cocaleros, as well.

To Patricia, Analia, Paul and Ronald: for their helping hand during the collection of data.

Special thanks to Louis for clarifying my ideas. I am deeply indebted to him for that.

To my parents and Francisco my most personal thanks for always understanding and supporting my decisions. To my siblings, for always being there for me.

To all my friends here in the ITC, especially Ruben, Airthon, Lucia, Carlitos, Bass, Luis, Ricardo, Marcela, Pilar and Fernando.

Table of contents

List of figures

List of tables

List of abbreviations

AIC	Akaike Information Criterion
ASCII	American Standard Code for Information Interchange
ASTER	Advanced Spaceborne Thermal Emission and Reflection Radiometer
AUC	Area Under the ROC Curve
CETEFOR	Centro Técnico Forestal (Forestry Technical Center)
CLAS	Centro de Levantamientos Aeroespaciales y Aplicaciones SIG para el desarrollo sostenible de los recursos naturales, UMSS. Cochabamba, Bolivia. (Aerospacial surveys and GIS applications Centre for sustainable management of natural resources)
DEM	Digital Elevation Model
FAO	The United Nations Food and Agriculture Organization
FRA	Global Forest Resources Assessment
GCPs	Ground Control Points
Landsat TM	Landsat Thematic Mapper
LRM	Logistic Regression Model
MLP	Multi-Layer Perceptron
PSAD 56	Provisional South American 1956
ROC	Relative Operating Characteristic
SRTM	Shuttle Radar Topographic Mission
SWIR	Shortwave Infrared
TIPNIS	Isiboro Sécure National Park and Indigenous Territory
TPI	Topographic Position Index
UTM	Universal Transverse Mercator
VNIR	Visible and Near Infrared

1. Introduction

1.1 Motivation and Problem Statement

According to the Global Forest Resources Assessment (FAO - FRA) (2005) [23], of the 20 countries clearing the most forest between 1990 and 2005, Bolivia occupied the twelfth position with a rate of 135,200 hectares of forest loss per year. In spite of Bolivia's environmental laws[1], deforestation[2][24] is occurring even in protected areas[44].

Isiboro Sécure National Park and Indigenous Territory (TIPNIS) is located in the Sub-Andean strip of the tropical Andes in Bolivia. Because there is a great diversity of species and ecosystems in the TIPNIS, the park is considered one of the most important places in the country to preserve fauna and flora [73]. The TIPNIS National Park is also called "*indigenous territory*" because three native ethnic groups traditionally live inside of it. In recent years the presence of additional and illegal settlers have been reported [72]. According to the managers of the park [59, 72], the TIPNIS faces two main threats: internal threats stemming from the non-rational and unplanned use of natural resources by the resident communities (e.g., agriculture, hunting, fishing, and forest extraction), and external threats resulting from political decisions at the departmental and national levels (e.g., building of new roads, oil concessions, etc.). Among the internal threats, activities such as agriculture and forest extraction are leading to a rapid loss of primary forest and also have disturbed the traditional life of the indigenous communities [60]. Due to the park's limited funding, there is no current information about how exactly the deforestation process has been and is still happening inside the TIPNIS [59].

Modelling the forest change occurring inside the park can be an important additional tool to help to understand the deforestation process in the TIPNIS. But, deforestation modelling is not simple because deforestation is simultaneously a dynamic, spatial, and socio-economic process [29]. Deforestation has multiple causes. The particular mix of causes varies from place to place [70] cited by[27]. The usual causes of deforestation are spatial pattern drivers (proximate causes such as agriculture expansion, population growth, roads, etc.) and drivers or forcing functions that explain root causes and pressure on the forest (underlying causes such as economic factors, institutions, national policies, etc.) [9, 27].

1 Environmental Law 1333 and General Regulation of Protected Areas DS 24781.

2 The United Nations Food and Agriculture Organization (FAO)-Global Forest Resources Assessment (FRA) define deforestation as "the conversion of forest to another land use or the long-term reduction of the tree canopy cover below the minimum 10 percent threshold".

Lambin (1994) [38] and Mas *et al*., (2004) [47] mention that deforestation models are motivated by the following potential benefits: (1) to provide a better understanding of how driving factors govern deforestation, (2) to generate future scenarios of deforestation rates, (3) to predict the location of forest clearing and, (4) to support the design of policy responses to deforestation.

According to Kaimowitz and Angels (1998) [37], one way to model deforestation is to make use of empirical models. Empirical models quantify the relationships between variables using empirical data and statistical methods. Several studies have analyzed land use change under these approaches [51, 63, 64, 68, 71]. In the particular case of deforestation, the spatial forest change is a categorical dependent variable which results from the interaction of several explanatory variables. Authors like [16, 20, 41, 42, 48, 69, 79] have worked successfully making use of logistic regression and Geographic Information Systems (GIS) as tools to analyze deforestation and its causes.

Logistic regression analysis has the advantage of taking into account several independent explanatory variables for prediction of a categorical variable [77]. In this case, the dependent variable is the change and no change that has occurred in the forest areas. Logistic regression analysis fits the data to a logistic curve instead of the line obtained by ordinary linear regression. In addition to the prediction, logistic regression is also a useful statistical technique that helps to understand the relation between the dependent variable (change) and independent variables (causes) [47].

Proximate causes of forest conversion such as roads or settlements can be identified making use of GIS tools [29]. Underlying causes are more difficult to identify than are proximate causes, sometimes because the information is not available and sometimes because the information is general in nature and so eludes easy measurement (insufficient socio-economic data, land tenure, influence of national policies, etc.). Thus, the ability to link underlying causes of forest conversion to spatial patterns obtained by GIS is another important tool to be considered [6, 56]. An interesting example of this is given by Van Gils and Loza (2006) [79] who identified the importance of an underlying driving force, land tenure in their case, based on the size, shape and spatial patterns of the parcels in the Carrasco province of Bolivia.

In recent years, improved performance in land use change modelling has been achieved successfully by combining statistical methods with other approaches such as decision trees [49], Bayesian methods [1] or with stochastic spatial models such as cellular automata [54] and artificial neural networks [46].

Other studies also related to land use change have reported that the artificial neural network “Multi-Layer Perceptron” has shown good potential for predicting future scenarios [43, 47, 55, 61]. According to Mas *et al*., (2004) [47], artificial neural networks are powerful tools for models because they have the ability to handle non-linear functions, to perform model-free function estimation, to learn from data relationships that are not otherwise known and, to generalize to unseen situations. In addition, the same authors also mention that artificial neural networks are able to directly take into account any non-linear complex relationship between the explanatory variables and deforestation.

Based on the availability of data, the present study has integrated the statistical approach of logistic regression and also that of artificial neural networks with GIS in an attempt to analyze and predict forest conversion in the TIPNIS.

1.2 Research Objectives

The main objective of this study is to analyze and predict processes of forest conversion in the Isiboro Sécure National Park and Indigenous Territory (TIPNIS) in Bolivia.

In order to reach the goal, the following specific objectives are considered:

- To determine and quantify forest changes that occurred in the TIPNIS from 1976 to 2006.
- To identify and analyze the most significant explanatory variables that lead to forest conversion in the TIPNIS.
- To establish a predictive model based on the validation and comparison of two approaches: Logistic Regression and Multi-Layer Perceptron.

1.3 Research Questions

1. What are the causative factors associated with forest conversion in the TIPNIS?

    - Agricultural Encroachment
    - Distance from Roads
    - Distance from Settlements
    - Distance from Forest Edge
    - Type of Settlement
    - Landscape Position
    - Underlying Forces

2. Can an artificial neural network improve the forest conversion prediction performed by the Logistic Regression Model?

1.4 Hypotheses

1. No hypothesis; depends on the research results.
2. H_1: The prediction of TIPNIS deforestation by an artificial neural network is significantly better than the prediction of deforestation by the logistic regression model
   H_0: The prediction of TIPNIS deforestation by an artificial neural network is not significantly better than the prediction of deforestation by the Logistic Regression Model.

1.5 Research Approach

To model and predict forest conversion in the TIPNIS, it was necessary to:

1. Analyze the forest conversion that occurred from 1976 to 2006.
2. Analyze the relationship between significant factors and forest change.
3. Select statistically the best predictor set of explanatory variables.
4. Predict forest conversion.

To carry out the study, three approaches were combined.

GIS was used to measure the forest loss for the years 1976, 1986, 1991, 2001, 2004 and 2006 and to provide dependent and explanatory variables as spatial data for the modelling of forest conversion that occurred during the period 2001 to 2004. Also, GIS provided spatial data for the validation of forest change prediction for the period 2004 - 2006.

Logistic regression analysis was used to: i) help to understand the relationship between the dependent variable and the explanatory variables based on the behaviour of forest change during the period 2001 - 2004; ii) select the best combination of explanatory variables for forest conversion prediction and iii) predict forest change for the period 2004-2006.

Finally, the study made use of the artificial neural network Multi-Layer Perceptron (MLP) to predict forest conversion for the period 2004 to 2006. Figure 1 illustrates the general research approach.

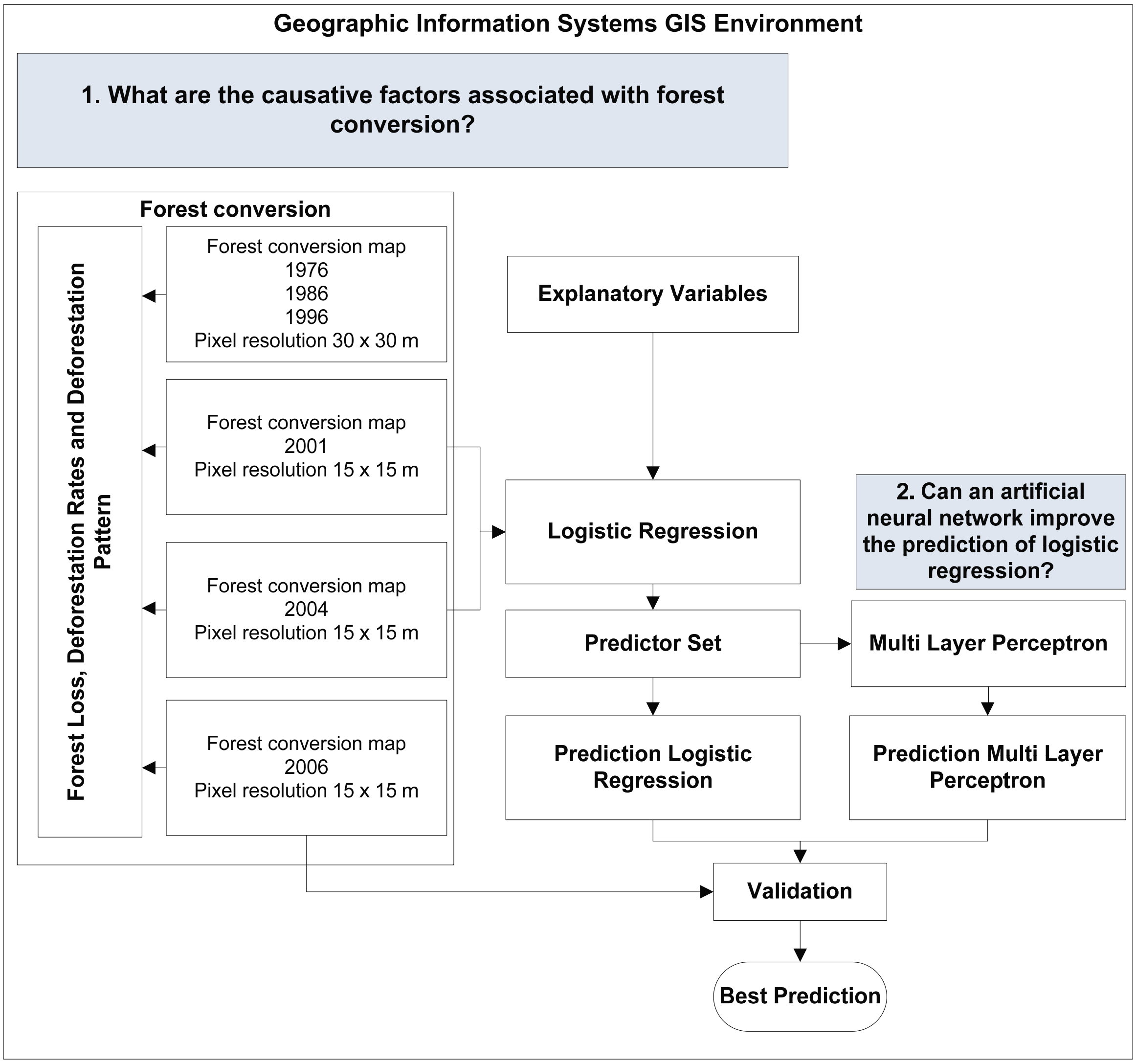


Figure 1 Modelling and Predicting Forest Conversion in the TIPNIS, General Approach

1.6 Thesis Structure

This study is divided into five chapters and includes references and appendices.

Chapter 1, Introduction, is a general introduction to the research and states the main objectives and research questions.

Chapter 2, Materials and Methods, describes the study area, and presents the materials and methods involved during the execution of the research.

Chapter 3, Results, presents the main results obtained by the research.

Chapter 4, Discussions, discusses the main findings of this study, analyzing the results in the context of other studies.

Chapter 5 Conclusions is dedicated to the presentation of the main findings of this research and includes some recommendations for future work.

The *References* Section details the sources used as references in this document.

The *Appendices* contain detailed information supporting the methods used in this document.

2. Materials and Methods

This chapter is divided into three sections:

- 2.1 *Study Area*; describes the characteristics of the study was carried out.
- 2.2 *Materials*, describes the materials used during the study.
- 2.3 *Methods*; describes the methods used to reach the objectives.

2.1. Study Area

Isiboro Sécure National Park and Indigenous Territory (TIPNIS) is located in Bolivia, covering parts of both the Departments of Beni and Cochabamba, between the geographic coordinates 65º08" - 66º35" West and 15º 37" - 16º 40" South (See Figure 2). In the Department of Cochabamba, it is located in the Chapare Province. TIPNIS was created in 1965 and has an approximate extension of 950,661 ha [72]. The climate is humid with annual rainfall ranging from 2000 to 3000 mm. The temperatures are highest (25 to 32 degrees Celsius) from December through January and coldest from May through June (15 to 25 degrees Celsius). Occasionally, the temperature can fall below 5°C during southern wind episodes. The area's location is on the eastern slopes of the Andes range. The topography of the TIPNIS is characterized by a mountain range to the west (up to 3000 m.a.s.l.) and large flood plains to the east (down to 300 m.a.s.l.). The most important rivers are the Isiboro, the Ichoa and the Sécure, all tributaries of the Mamoré River, which, in turn, forms part of the Amazon Basin [72] [60].

The altitudinal range of the TIPNIS fosters a very high species and ecosystem diversity, forming very distinct ecological systems, such as montane cloud forests, sub-Andean Amazonian forests, mid- to lowland evergreen rain forests, and flooded savannas. Each of these ecosystems harbours a unique flora and fauna [72] [60].

According to Parkswatch (2004) [60], the protected area is also legally recognized as *Indigenous Territory*, property of the natives of the region: Chimán, Yuracaré, and Moxeño (Trinitarios) ethnic groups. Most human settlements are located along the area's two most important rivers or near the park's boundaries. The south-eastern sector has been settled by colonists. This fourth group is composed of immigrants of Aymara and Quechua cultures who came from the country's highlands (altiplano), especially from the Cochabamba, Oruro, Potosí, and La Paz Departments in the second half of the 20th century.

The managers of the national park [60, 72] provided the following information about these four human groups:

> The Yuracaré ethnic group settled in the TIPNIS and its surrounding areas several centuries ago. At present, they have taken over almost all of the TIPNIS area, aside from the central zone, but have mainly settled in the southern zone and along the lower Isiboro and Sécure rivers.

The Moxeño people (Trinitarios) historically occupied a much smaller area than the Yuracaré near the confluence of the area's most important rivers (Isiboro and Sécure), where their settlement patterns emulated those of the Jesuit missions. Today, they occupy a much larger territory, having spread out into the Park's central region and to an important portion of the southern zone.

The Chimán are not native to the area, but are found throughout the south of the Beni Department and in the foothills and mountains of the upper Sécure River. In recent years, some Chimán families have settled in Santo Domingo, which was formerly an exclusive Moxeño area.

The colonists began to settle in the area in the 1960s. The new access roads built in the 1970s brought even more migrants. In the 1980s the closure of the Government-owned mines, the opening of the Cochabamba-Santa Cruz highway (See Figure 3) and the perspective brought about by the production of leaves of coca (Erytroxylum coca) for the drug trade brought large numbers of new colonists to the area.

The people who crop coca are called *´cocaleros´*. They have invaded the southern part of the park [72]. They are responsible for most of the forest conversion in the Chapare Province [44].

Because most of the encroachment process has occurred in the southern part of the park, the study area for this research has focussed in this southern zone. The actual area of the study is located to the north of the Isiboro River, between the rivers Corijota, Sasasama and Lipurcy (See Figure 2). The zone under analysis has an area of 143,700 ha.

The main access into the study area is from the south, at a distance of about 55 km from the primary road Cochabamba – Santa Cruz (See Figure 3).

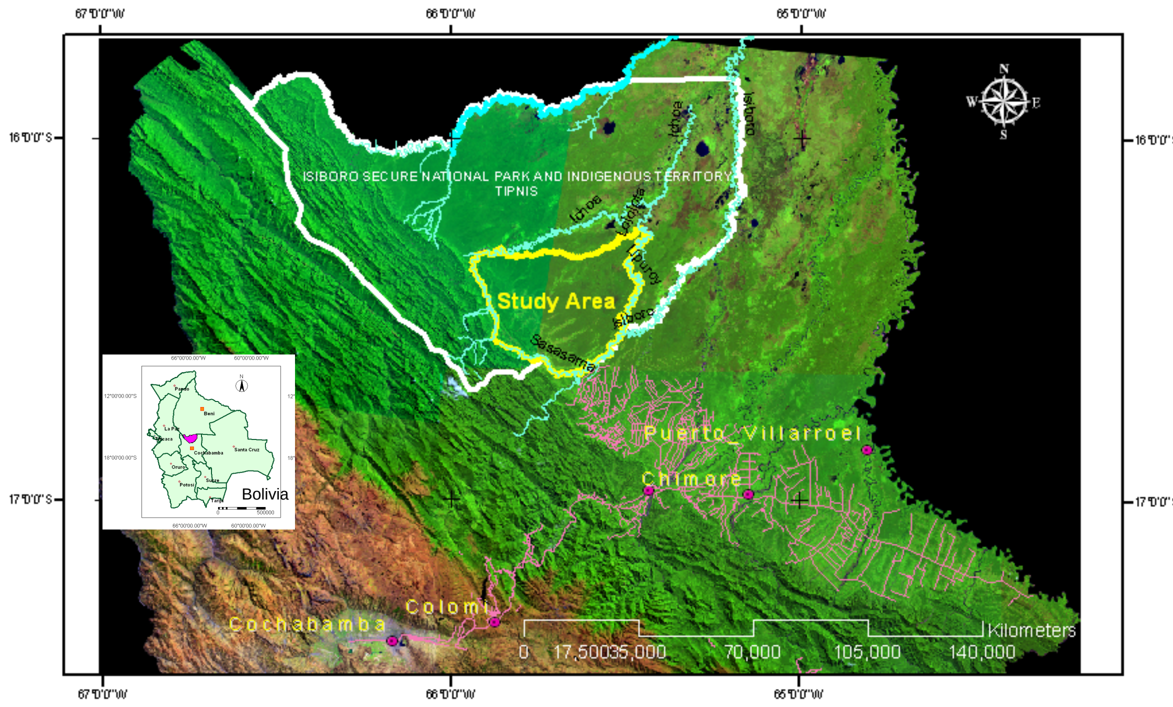


Figure 2 Study Area: Isiboro Sécure National Park and Indigenous Territory (TIPNIS) Cochabamba, Bolivia

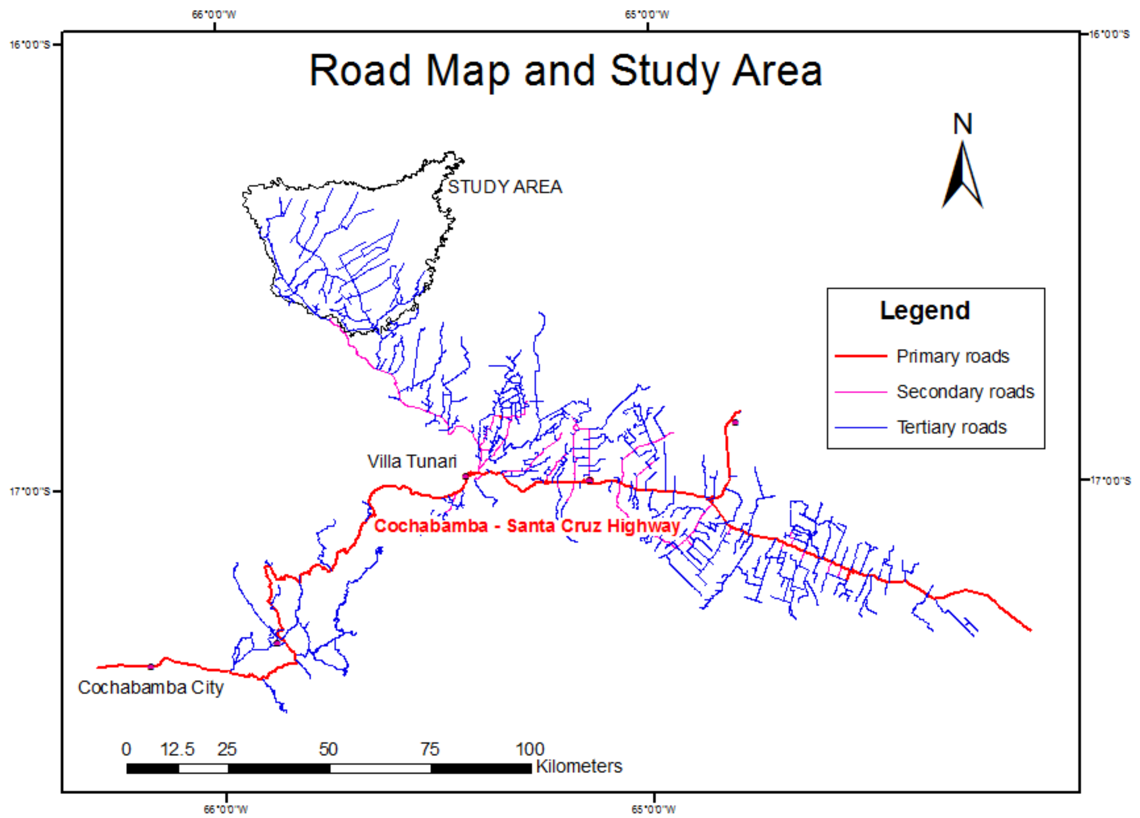


Figure 3 Roads Accessibility to the Study Area

2.2. Materials

To fulfil the objectives of this study, the following data was available.

2.2.1. Satellite Images

The Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) satellite has two types of Level-1 data; Level-1A and Level-1B data. Level-1A data are formally defined as reconstructed, unprocessed instrument data at full resolution. According to this definition the ASTER Level-1A data consist of the image data, the radiometric coefficients, the geometric coefficients and other auxiliary data without applying the coefficients to the image data to maintain the original data values. The Level-1B data are generated by applying these coefficients for radiometric calibration and geometric correction [18].

ASTER Level-1B images were obtained from the United States Geological Survey (USGS) Global Visualization Viewer (GLOVIS) [76]. The dates of the images are March 18th 2001, November 4th 2004 and October 17th 2006. The images are projected by default as UTM images and the datum default is World Geodetic System 1984 (WGS 84) [58]. The characteristics of the ASTER images are summarize in Table 1.

Landsat (Thematic Mapper) TM 5 Images for the years 1986 and 1991 were provided by the Aerospacial Surveys and GIS Applications Centre for Sustainable Development of Natural Resources (CLAS) [12]. They were already properly geo-referenced. The projection was Universal Transverse Mercator (UTM) and the Datum was Provisional South American 1956 (PSAD 56). The images were in ILWIS raster format. Table 1 also summarizes the characteristics of these images.

Table 1 Satellite Images Features

Table based on information provided by USGS 2008 [76] Mandgroup 2000 [53] and Yale 2007 [58]

Year	Satellite	Sensor	Spectral Range	Band #s	Scene Size	Pixel Res	Source
1986	Landsat 5	TM multi-spectral	0.45-0.52 µm 0.52-0.60 µm 0.63-0.69 µm 0.76-0.90 µm	1, 2, 3, 4	185 X 185 km	30 meter	CLAS [12]
1991	Landsat 5	TM multi-spectral	0.45-0.52 µm 0.52-0.60 µm 0.63-0.69 µm 0.76-0.90 µm	1, 2, 3, 4	185 X 185 km	30 meter	CLAS [12]
2001	ASTER	VNIR	0.52 - 0.60 µm 0.63 - 0.69 µm 0.78 - 0.86 µm	1, 2, 3N	60 X 60 km	15 meter	USGS GLOVIS [75]
		SWIR	1.600 - 1.700 µm	4	60 X 60 km	30 meter *	
2004	ASTER	VNIR	0.52 - 0.60 µm 0.63 - 0.69 µm 0.78 - 0.86 µm	1, 2, 3N	60 X 60 km	15 meter	USGS GLOVIS [75]
		SWIR	1.600 - 1.700 µm	4	60 X 60 km	30 meter *	
2006	ASTER	VNIR	0.52 - 0.60 µm 0.63 - 0.69 µm 0.78 - 0.86 µm	1, 2, 3N	60 X 60 km	15 meter	USGS GLOVIS [75]
		SWIR	1.600 - 1.700 µm	4	60 X 60 km	30 meter *	

* Image re-sampled to 15 meter

2.2.2. SRTM DEM

A Digital Elevation Model (DEM) is a digital representation of ground surface, topography or terrain. A DEM consists of a raster grid of regular elevation values (pixels) with a resolution of 3-arc seconds (90 m approx). The NASA Shuttle Radar Topographic Mission (SRTM) has provided DEMs for over 80% of the globe. These DEMs are available in both ArcInfo ASCII and GeoTiff format to facilitate their ease of use in a variety of image processing and GIS applications [11]. Data can be downloaded using a browser or accessed directly from the internet for free. The SRTM - DEM used in this study was provided by CLAS [12]. The SRTM - DEM was properly geo-referenced (UTM zone 20/ PSAD56).

2.2.3. Thematic Maps

The following thematic maps were provided by the managers of the TIPNIS [72] and CLAS [12] in vector and raster format.

- Deforestation maps for the years 1976, 1986, 1991, 2001 and 2004. These maps were obtained from 30 X 30m pixel size raster maps (Landsat5 TM).
- Maps of roads of the years 2004 and 2006; Esc: 1: 100,000
- A map of villages 2004, Esc: 1: 100,000
- A map of Type of Settlement 2004 Esc: 1: 250,000

These maps were also already geo-referenced in UTM projection and WGS84 datum.

2.3. Methods

This section describes the methods and criteria used to reach the objectives of this study. The section is divided into 4 subsections:

2.3.1 *General Description*

2.3.2 *Data Preparation*

2.3.3 *Forest Loss, Deforestation Rates and Deforestation Pattern*

2.3.4 *Modelling Forest Conversion*, describes:

- Creating dependent and independent variables for forest conversion
- Logistic Regression Model
- Multi-Layer Perceptron Model

2.3.1. General Description

In order to obtain information about forest conversion in the TIPNIS, this study was interested in identifying forest areas that have suffered changes between 1976 and 2006. To identify changes in the forest, the first step was to obtain maps of forest conversion using two main general categories: “Forest” and “Disturbed Forest”. In this study, the category **“Forest”** is assigned to forest areas relatively undisturbed by human activity, where human impacts have normally been limited to low levels of hunting, fishing and harvesting of forest products. The category “**Disturbed Forest”** consists of those areas where *forest disturbance or change*s have occurred resulting from anthropogenic activities.

In the *Data Preparation* Section, forest conversion maps were extracted from the land cover maps and satellite images. For land cover maps, a reclassification of land use categories into the new categories "Forest" and "Disturbed Forest" was performed. To extract forest conversion maps from satellite images, the following sequence was performed: Geo- referencing, supervised classification (Maximum Likelihood Algorithm), validation of the classification (Confusing Matrix) and, a final reclassification was carried out to obtain the "Forest" and "Disturbed Forest" categories.

Based on measuring forest areas in different but correlative years (1976, 1986, 1991, 2001, 2004 and 2006), in the *Forest Loss, Deforestation Rates and Deforestation Pattern* Section forest loss was obtained by comparing map areas. Deforestation rates were calculated using the formula proposed by Puyravaud, (2003) [65] . The deforestation pattern was determined based on the size of patches of disturbed forest, making use of the land metric "Area".

The *Modelling Forest Conversion* Section, i) describes how the dependent and independent variables were created in order to be analyzed in a GIS environment; ii) describes the statistical procedures to analyze the relationship between variables; iii) describes logistic regression, analysis and prediction, and, finally, iv) describes prediction of the artificial neural network. The *Modelling Forest Conversion Section* summarizes the procedures to be performed as follows:

> After the creation of variables, potential explanatory variables for forest conversion were tested with Cramer's V (Univariate tests of association).
>
> Forest conversion was analyzed using binomial logistic regression and GIS data using IDRISI 15 software. The purpose of modelling was (i) to assess the relative significance of explanatory variables on forest change during the period 2001-2004; and (ii) to predict probability of forest change for the period 2004-2006.
>
> An alternative model for forest conversion prediction was also performed using the IDRISI 15 software. The artificial neural network Multi-Layer Perceptron was executed to predict forest change for the period 2004-2006. The prediction performed used the same data set used by the logistic regression prediction. Finally, the comparison between the models was made using the Area Under the ROC Curve (AUC).

For purposes of description, the methodological approach has been divided into three main blocks.

1. Data Preparation (Subsection 2.3.2)
2. Forest Loss, Deforestation Rates and Deforestation Pattern (Subsection 2.3.3)
3. Modelling Forest Conversion (Subsection 2.3.4)

Table 2 summarizes the steps for each block and Figure 4 contains a methodological flowchart of the steps.

Table 2 Methodological Approach Building Blocks

Block	Step	Input	Methodological approach	Output
1. Data Preparation	**a. Satellite images: importing and geo-referencing**	Satellite images without geo references	-ITC exercise book guide	Geo-referenced images
	b. Land cover classification	Geo-referenced images	-Maximum Likelihood Algorithm (MLA)	Land cover map
	c. Validation of the classification	Land Cover map and Ground Control Points	-Cross tabulation assessment	Validated land cover map
	d. Land cover reclassification	Validated land cover maps	Reclassification	Binary maps: Forest and Disturbed Forest; 1976; 1986;1991;2001;2004 and 2006
2. Forest loss, Deforestation Rates and Deforestation Pattern	**a. Forest Loss and deforestation rates**	Binary maps (Output 1.d)	-Area calculation -Deforestation rates -calculation	Forest loss table, graphics, deforestation rates
	b. Deforestation pattern	Binary maps (Output 1.d)	-FRAGSTATS: Land metric area	Land metric indexes
3. Modelling Forest Conversion	**a. Creating dependent and independent variables for forest conversion**	-Binary maps (Output 1.d) -Roads maps (2004, 2006) -Villages map -DEM Digital -- Elevation Model -“Type of Settlement” map	-Creating the dependent variable -Creating explanatory variables -Tests of independence: Chi square test and Cramer’s V.	-Dependent variable -Explanatory variables
	b. Logistic Regression Model	-Dependent variable (Output 3.a) -Explanatory variables (Output 3.a)	-Calibration -Prediction -Validation	**Probability for forest conversion for the year 2006 LRM**
	c. Multi-Layer Perceptron Model	-Dependent variable (Output 3.a) -Explanatory variables (Output 3.a)	-Calibration -Prediction -Validation	**Probability for forest conversion for the year 2006 MLP**
	d. Comparison between models	-Probability for forest conversion for the year 2006 LRM (Output 3.b) -Probability for forest conversion for the year 2006 MLP (Output 3.c)	-AUC/ROC	**AUC/ROC results**

Figure 4 shows the sequence of the complete process for analysis and prediction of forest conversion in the TIPNIS.

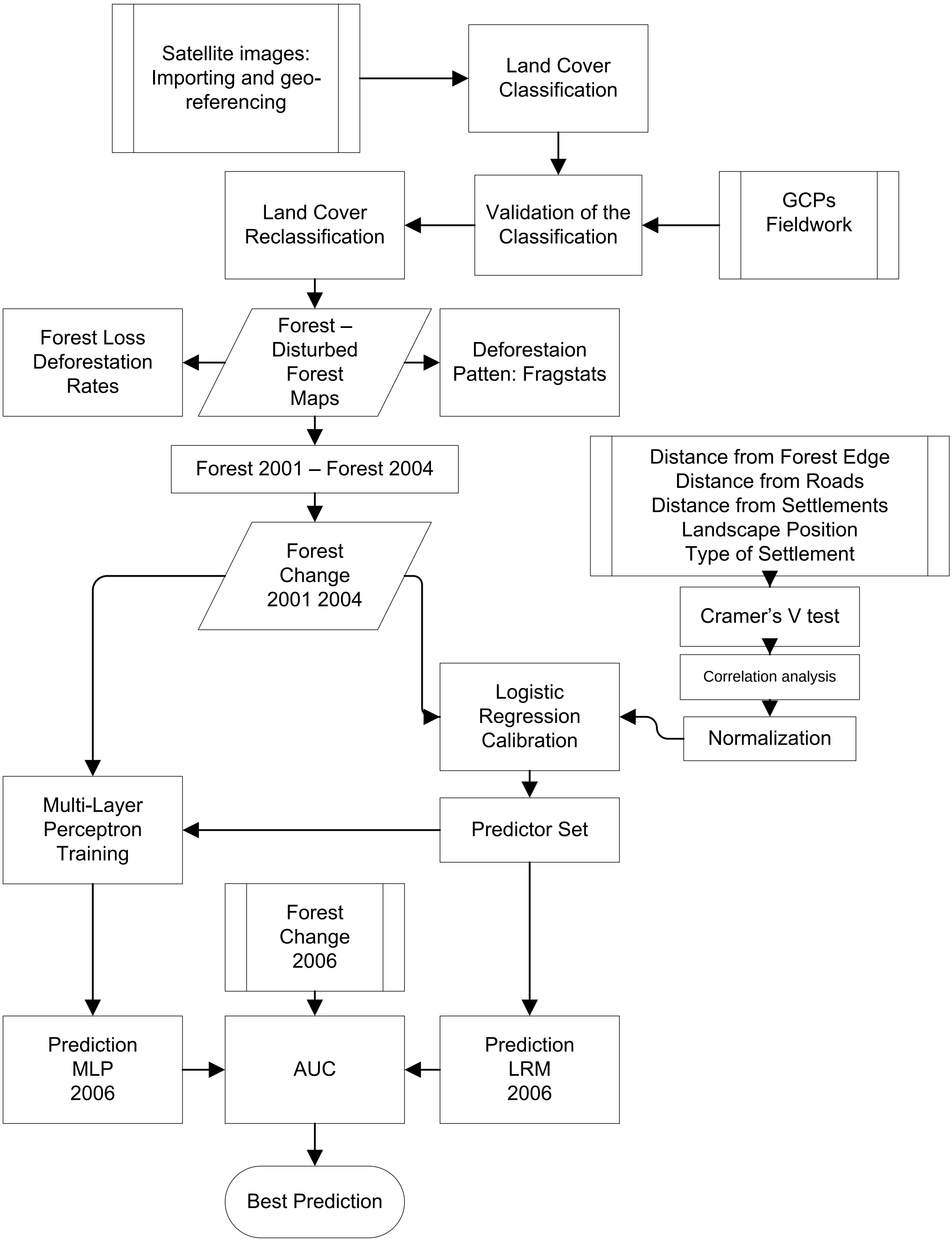


Figure 4 Methodological Approach Flowchart

2.3.2. Data Preparation

In this subsection, the areas of "Forest" and "Disturbed Forest" were determined making use of GIS software. GIS tools were used to standardize geographic reference systems, to classify satellite images, to validate the classification and finally to prepare data to be used as input in the modelling and prediction of forest conversion.

Subsection 2.3.2 procedures are described next in the following subsections:

2.3.2.1 Satellite Images: Importing and Geo-referencing
2.3.2.2 Land Cover Classification
2.3.2.3 Validation of the Classification
2.3.2.4 Land Cover Reclassification

2.3.2.1. Satellite Images Importing and Geo-referencing

Three ASTER Level-1B images for the years 2001, 2004 and 2006 were needed to extract forest conversion maps. All images were imported and geo-referenced into the same coordinate system (UTM/PSAD56) using ERDAS 9.2.

In order to be manageable in GIS software, ASTER Level-1B images had to be converted from Hierarchical Data Format (*.hdf' or *.hdf.met') to Tagged Image File Format (TIFF). The importing of ASTER Level-1B images (TIFF format) was performed using the Erdas 9.2 Import/Export option and the images were converted to Imagine format (IMG). Visible and Near Infrared (VNIR) and Shortwave Infrared (SWIR) bands were imported for the supervised classification. The sequence of the importation and geometric correction was preformed according to the ITC Exercises book guide. Natural Resources Management, Modules 3&4 RS and RS/GIS 2007 [35].

All the images were geometrically referenced to a common spatial reference system, as follows:

Projected Coordinate System: PSAD_1956_UTM_Zone_20S
Projection: Transverse Mercator
False Easting: 500000.00000000 meters
False Northing: 10000000.00000000 meters
Central Meridian: -63.00000000 degrees
Scale Factor: 0.99960000
Latitude of Origin: 0.00000000 degrees
Linear Unit: Meter

2.3.2.2. Land Cover Classification

Once the images were properly geo-referenced, ILWIS 3.3 software was used to perform a supervised classification procedure. The objective of the classification was to obtain two single categories “Forest” and “Disturbed Forest”. To classify with only two categories was too coarse. More categories were needed to differentiate the two forest categories.

To reach this goal, the *Maximum Likelihood Algorithm (MLA)* classifier *(Subsection 2.3.2.2.1)* was used to preliminarily classify six categories. The categories are described in *Subsection 2.3.2.2.1.1*. In order to assess the classification accuracy, the classified image was compared with ground control points (*Subsection 2.3.2.3).*

After the accuracy assessment, there were still water areas that needed to be separated. The ArcGIS 9.2 polygon editor was used to correct polygons of burnt areas misclassified as water. Finally, a reclassification of categories was performed to obtain only two categories, "Forest” and "Disturbed Forest" (*Subsection 2.3.2.4*)

As mentioned in *Section 2.3.1*, the final category ¨Forest¨ consisted of those areas with *no disturbed forest*: Primary Forest, Grassland, River Shores and Water. The final category ¨Disturbed Forest¨ consisted of those areas where *forest disturbance or change* had occurred.

The classification procedures are described below.

2.3.2.2.1 Supervised Classification of Land Cover Categories

The classifier selected for the classification was Maximum Likelihood Algorithm. The classification was performed using ERDAS 9.2 software.

Maximum Likelihood Algorithm (MLA) is a statistical decision rule applied to raster images that examines the probability function of a pixel for each of the classes and assigns the pixel to the class with the highest probability. MLA has been widely used in land cover classification because it usually provides high classification accuracies [13, 31]

The supervised classification resulted in six categories of land cover: ‘*Primary Forest’*, ‘*Secondary Forest’*, ‘*Crops and Cattle Pastures*’, ‘*Grassland’*, ‘*River Shores’*, and ‘*Water*’. These categories are described below (*Subsection 2.3.2.2.2*).

During the classification, the category ‘*Water*’ presented similar values of reflectance to forest burnt areas. The ArcGIS polygon editor was used to correct this issue. The water areas depicted in images earlier than 2006 were compared with the water areas in the 2006 image. If they remained the same, they were classified as water. If not, they were classified as burnt areas.

After these procedures were performed, a final map of land cover classification was obtained. Figure 11 (Subsection 3.1.1) depicts the categories of land cover in percentages.

The cover classification for the year 2006 was used as reference to help determine the classification of images of earlier years (2001 and 2004) based on the assumption that a Disturbed Forest cannot return to Primary Forest within 5 to 20 years. According to Etter et al (2006) [19], time intervals of less than 10 years are too short to regenerate effectively a forest cover from a cleared area. Brearley F. (2007) [8], also mentions that it takes 55 years for a secondary forest cleared for agriculture to achieve approximately 80% of the biomass of tropical primary forest.

The categories of classification were supported taking into account four basic references:

- Vegetation map of the TIPNIS 2004 [72]
- FAO CETEFOR Report 2002 [22]
- Ritter N (2006) [67]
- Observations during the field work.

To collect information about Land Cover in the TIPNIS, this study registered names of common (usual) plants in the same place in which the Ground Control Points were registered. Four TIPNIS park guards were the source of local knowledge to identify common species of plants. The land cover for forest areas was estimated using a method of transect sampling: the line intercept method also called the line-intercept ground sampling. Finally, the presence of cattle was registered by simple observation or by evidence such as cattle prints, cattle dung, fences, etc. Appendices Ia, Ib, Ic contain the form used to collect this data and the data collected.

2.3.2.2.1.1 Categories of Classification

Primary Forest

‘Primary Forest’ is defined as a forest ecosystem, relatively undisturbed by human activity, with the principal characteristics and key elements of native ecosystems such as complexity, structure, and diversity and an abundance of mature trees. Human impacts in such forest areas have normally been limited to low levels of hunting, fishing and harvesting of forest products. Such ecosystems are also referred to as "mature," "old-growth," or "virgin" forests. FAO 2002 [26].

The most common species of trees in the Primary Forest of the TIPNIS are: “Almendrillo”, (*Dypterix odorata*); “mapajo” (*Ceiba pentandra*); “bibosi” (*Ficus sp*.); “mara o caoba” (*Swietenia macrophylla K.*); “palo maría” (*Calophyllum brasiliensi*); “yesquero” (*Cariniana sp.*); “cedro” (*Cederla sp*.); “ochoó “ (*Hura crepitans L.*); “coquino” (*Chrysophyllum sereceun*); “coloradillo” (*Byrsonina sp.)*; “sangre de toro” (*Virola sp*.); “gabetillo” (*Aspidosperma sp.*); “ocoró” (*Reedia acuminata*); “blanquillo” (*Buchenavia oxicarpa*); “peloto” (*sapiun marmieri*); “verdolago” (*Terminalia amazónica*); “piraquina” (*Xylopia amazónica*); “chonta” (*Astrocarium chonta*), and “pachiuba” (*Socratea sp.).*SERNAP-TIPNIS 2004 [72] and field work.

Secondary Forest

'Secondary Forest' is rainforest that has been disturbed, naturally or unnaturally. Secondary Forest can be created in a number of ways from degraded forest recovering from selective logging to areas cleared by slash-and-burn agriculture that have been reclaimed by forest. Due to the lack of a full canopy, more light will reach the floor supporting vigorous ground vegetation. Butler, Rhett A. 2006 [10].

'Secondary Forest' in the TIPNIS, is characteristically dominated by *Cecropia peltata*, *Ochroma pyramidal,*, *Inga spp*., *Copaifera sp*., and *Sapium spp.* as mentioned by SERNAP-TIPNIS 2004 [72], Ritter 2006 [67] and field work.

Grassland

In this study, *Grassland* is defined as a land where grass or grass-like vegetation grows and is the dominant form of plant life. Grasslands occur generally in areas near rivers and lowlands. The most common plant species are: "grama Negra" (*Paspalum notatum*), "pata de gallo" (*Sporolobus poireti*), "pasto amargo" (*Paspalum conjugatun*),"goma negra" (*Eleusini indica*), "sujo" (*imperata brasiliensi*), etc. SERNAP-TIPNIS 2004 [72] and fieldwork.

Crops and Cattle Pastures

'Crops' are areas where agricultural activity is present. The most common crops are "coca" (*Erytroxylum coca*), "plátanos" and "bananos" (*Musa spp*.), rice (*Oryza sativa*), "yuca" (*Manihot utillisima*), "maiz" (*Zea maíz*), some fruits like "cacao" (*Theobroma cacao L.),* "cítricos" (*Citrus spp.*) and other species in lesser quantities.

'Cattle Pastures' are forest areas that have been burned and converted into pastures for cattle. These pastures contain grass, leguminous plants and weeds with a few trees. In general, cattle pastures in the TIPNIS frequently contain *Brachiara decumbens, Brachiaria brizantha, Axonopus scoparius, Desmodium ovalifolium, and Pueraria phaseoloides* SERNAP-TIPNIS 2004 [72] and fieldwork.

For purposes of this study, crops and cattle were combined into a single category 'Crops and Cattle Pastures' because the final objective of the Land Cover classification was to obtain maps that discriminate Primary Forest from Disturbed Forest caused by anthropogenic activities.

River Shores and Water

"River Shores" describes sandbanks near to the shores of rivers, which are common in the borders of the study area. The category '*Water*' is used to represent rivers and lagoons.

The result of the classification is depicted in Figure 11, Subsection 3.1.1.

2.3.2.3. Validation of the Classification

The assessment of land cover classes was performed using an ILWIS 3.3 Confusion Matrix, taking into account both Ground Control Points obtained through field work and the classified image obtained for the year 2006.

2.3.2.3.1 Ground Control Points (GCPs)

The stratified *clustered representative sampling scheme* [34], was used to determine the GCPs. The sampling is done by interpretation unit. In this method, an equal number of sample points is allocated to each legend unit. The size of the unit and the number of polygons that belong to the unit do not influence the number of sample points. The total area covered by one legend unit is not taken into account either.

According to the ITC, 2005 [34], ‘Cluster sampling’ means that sample points are concentrated in more accessible areas where several legend units occur within a short distance. The areas where observation sites are concentrated are called ‘Sample areas’. These sample areas should contain representative examples of all legend units delineated during the image interpretation or image classification. For each unit a relative equal number of samples is extracted in order to be validated during the field work with GPS marks. At least 80 GCPs were anticipated for this study. Due to social conflicts between “cocaleros” and TIPNIS managers, only 39 points could be taken during the field work (see Appendix Ib).

2.3.2.3.2 Confusion Matrix

The Confusion Matrix was calculated in ILWIS 3.3 and displays graphically the results of the comparison between the GCPs taken during the field work and the classified image. According to Gong 1997 [28] the Confusion Matrix is an N x N matrix of "observed" and "classified" cells corresponding to N land cover classes. The matrix depicts the land cover classification category versus the field-observed land cover type. The diagonal cells indicate correct observations, meaning that the observations were classified correctly according to the field observations. Any observation off the diagonal indicates a misclassified accuracy control point. The result of the Confusion Matrix showed an overall accuracy of 85.58%. Table 4 (Subsection 3.1.2) shows the matrix between classified and observed cells.

2.3.2.4. Land Cover Reclassification

The objective of the classification was to discriminate forest without disturbance from areas with forest conversion. Areas with forest needed to be separated from areas that have suffered change due to anthropogenic influence. Then, the classified land cover maps were reclassified into two categories as follows:

‘*Primary Forest*’, ‘*Grassland*’, ‘*River Shores*’, and ‘*Water*’= “Primary Covers”

‘*Secondary Forest*’, ‘*Crops and Cattle Pastures*’ = “Secondary Covers”

According to Figure 11 (Subsection 3.1.1), the categories ‘*Grassland*’, ‘*River Shores*’, and ‘*Water*’ occupy, respectively, 8.5%, 1.2 %, and 0.5 % of the total area of study. As these categories have not shown evidence of change due to human activities, they have been included as part of the category 'Forest' in order to generalize the information. The same logic has been applied to the categories '*Secondary Forest'* and '*Crops and Cattle Pastures*', which have been reclassified as part of the category 'Disturbed Forest'.

This reclassification resulted in three maps containing only the two area categories "Forest" and "Disturbed Forest" for each of the years 2001, 2004 and 2006 (see Subsection 3.1.3). Deforestation maps for the years 1976, 1986, 1991 were already properly geo-referenced and ready to be used. All these maps provided the information used to calculate deforestation rates and also to obtain binary maps of change and no change.

2.3.3. Forest Loss, Deforestation Rates and Deforestation Pattern

General Description

In this block, forest areas were measured for each year of study. Forest loss values for each year were calculated and then used for calculating deforestation rates. Finally, relationships between the patch area size and pattern deforestation were found.

The block procedures are described in the following three subsections:

- Forest Loss (Subsection 2.3.3.1)
- Deforestation Rates (Subsection 2.3.3.2)
- Deforestation Pattern (Subsection 2.3.3.3)

2.3.3.1. Forest Loss

Forest loss was quantified for five sequential periods: 1976-1986; 1986-1991; 1991-2001; 2001; 2004; 2004-2006. ArcGIS 9.2 was used to calculate forest area for each year. Forest loss for each period was obtained by simple subtraction of maps. The results are depicted in Table 5 in Subsection 3.2.1

2.3.3.2. Deforestation Rates

As pointed out by Armenteras et al 2006[2], this research also assumes that the deforestation rate does not remain constant and needs to be calculated for each year. During the literature review, three popular formulas were found: two of them provided by FAO (FAO 1996 [25] and FAO 1995 cited by [65]) and the third recently proposed by Puyravaud 2003 [65] (The formulas are listed below).

All three formulas were tested yielding similar results; see Table 5 (Subsection 3.2.1). In the end, the formula provided by Puyravaud 2003 [65] was selected because, according to its author, this formula is derived from the Compound Interest Law. It is also derived from the mean annual rate of change and, for this reason, is more intuitive than the formula used by FAO. Table 5 (Subsection 3.2.1) shows deforestation rates for each year and Figure 13 (Subsection 3.2.2) shows graphically deforestation rates for the periods 1976-1986, 1986-1991, 1991-2001, 2001-2004, 2004-2006.

FAO 1995 formula

$$q = ((A_2 / A_1)^{1/(t_2 - t_1)}) - 1$$

q = deforestation rate (% lost area/year)
A_1 = initial forest area
A_2 = final forest area
t_2-t_1 = interval in years during which change in land cover is being assessed

FAO 1996 formula

$$DR = 1 - (1 - (\frac{A_1 - A_2}{A_1}))^{\frac{1}{t}} \times 100$$

DR = deforestation rate (% lost area/year)
A_1 = initial forest area
A_2 = final forest area
t = interval in years during which change in land cover is being assessed

Puyravaud formula

$$r = \frac{1}{t_1 - t_2} \ln \frac{A_2}{A_1}$$

r = deforestation rate (% lost area/year)
A_1 = initial forest area
A_2 = final forest area
t_2-t_1 = interval in years during which change in land cover is being assessed

2.3.3.3. Deforestation Pattern

Forest patterns have been successfully identified using GIS [2, 6, 78]. Mertens and Lambin (2000) [51] contend that land-cover changes often exhibit high degrees of spatial and temporal complexity. To help clarify the spatial pattern of forest conversion in the study area, FRAGSTATS software was used to obtain forest fragmentation indexes.

FRAGSTATS is a spatial pattern analysis program for categorical maps. The landscape subject to analyze is user-defined and can represent any spatial phenomenon. FRAGSTATS simply quantifies the areal extent and spatial configuration of patches within a landscape [50]. While this software computes several statistics for each raster map, in this thesis only “class (patch type) in the landscape” is used to calculate a single class because information is being collected only for the patch type “Disturbed Forest”. Figure 14 Subsection 3.2.3 shows the frequency of patches of “Disturbed Forest” for each year of study.

Three software programs were used to format data for input to FRAGSTATS. First, ArcGIS 9.2 was used to resample categorical raster maps of deforestation from 15 m into 50 m pixel. Then, ERDAS 9.2 was used to transform the maps into *.GIS format. Finally, LANDISVIEW beta 1.0 9 [5] was used to import the formats into ASCII. Use of ASCII format facilitates importing data into FRAGSTATS.

The land metric *Area* is described as follows:
Expression: aij = area (m2) of patch ij.
Units: Hectares
Range: The range in *area* is limited by the grain and extent of the image. In this case, the minimum area was 0.25 ha because the pixel size was resampled to 50m since FRAGSTATS was not able to work with the original pixel size of 15m and the large number of pixels in the raster of the study area. Loza (2004) reported the same issue with pixels of 30 by 30m. The method of resampling was ‘nearest

neighbour', which consists of assigning coordinate map values and pixel size to the nearest pixel. When using nearest neighbour resampling, the value of the input pixel closest to a new output pixel is used as the output value [33].

2.3.4. Modelling Forest Conversion

2.3.4.1. General Description

Before the modelling procedures could be performed, the dependent variable was defined as the presence or absence of forest change between two observation years (Subsection 2.3.4.2.1) Then, the explanatory variables were defined and also their association with the dependent variable was tested using Cramer's V. Finally, linear regression analysis was performed to verify independence between explanatory variables (Subsection 2.3.4.2.2). The explanatory variables were: "Distance from Forest Edge", "Distance from Roads", "Distance from Settlements", "Landscape Position" and "Type of Settlement".

The Logistic Regression Model (LRM) (Subsection 2.3.4.3) was used (i) to assess the relative significance of explanatory variables on forest change during the period 2001-2004; and (ii) to predict probability of forest change for the period 2004-2006 based on the forest conversion that occurred during the period 2001-2004.

The artificial neural network Multi-Layer Perceptron (MLP) was used in an attempt to improve the prediction of probability of forest change for the period 2004 to 2006. The prediction performed made use of the same data set used by the logistic regression prediction (Subsection 2.3.4.3).

The probability of change was assessed comparing the results of both models to the real change that occurred between the two years 2004 and 2006. The Area Under the ROC Curve (AUC) was used as the tool of assessment. The whole modelling process was performed in an IDRISI 15, Andes Edition environment.

The Modelling procedures are described in the following subsections:

- Creating Dependent and Independent Variables for Forest Conversion for Forest conversion *(Subsection 2.3.4.2)*
    - o Dependent Variable *(Subsection 2.3.4.2.1).*
    - o Explanatory Variables *(Subsection 2.3.4.2.2)*
- Logistic Regression *(Subsection 2.3.4.3)*
- Multi-Layer Perceptron *(Subsection 2.3.4.4)*

2.3.4.2. Creating Dependent and Independent Variables for Forest Conversion

As mentioned before, the analysis and prediction of forest conversion was performed using two models: LRM and MLP. Both models require that the input data must be in a continuum or ranked scale. The dependent variable and independent variables were created to respond to this requirement.

In order to fulfil this requirement, as first step, all data were imported to IDRISI 15 software, using the Import option. This step is very important and must be done very carefully. IDRISI 15 requires that the user specify the exact number of rows and columns for the raster data and geo-reference in order to have the proper pixel size for all the raster data. The following parameters were used:

Columns	:	3661
Rows	:	3061
Reference System	:	UTM Zone 20 S
Reference Units	:	Meters
Distance Units	:	1.00
Minimum X	:	185665.00
Maximum X	:	240580.00
Minimum Y	:	8155985.00
Maximum Y	:	8201900.00
Y Resolution	:	15
X Resolution	:	15

2.3.4.2.1 Dependent Variable

The dependent variable is a binary presence or absence event, where 1=change and 0=no change. According to Rossiter & Loza (2008) [69], this is a logical response variable, which takes only two values: True or False. Thus, a Boolean map with categories of Change and No change was needed. This map was obtained for the period 2001 – 2004.

Change and No change (2001 – 2004)

> IDRISI 15 was used to subtract the spatial distribution of forest for the year 2004 from the areas of forest for the year 2001.
> Map 2001= Forest area
> Map 2004= Forest area
> Map 2001 minus Map 2004 = Change 2001 2004

Figure 16 in Subsection 3.3.1 contains the map of the dependent variable can be visualized.

2.3.4.2.2 Explanatory Variables

"Distance from Forest Edge", "Distance from Roads", "Distance from Settlements", "Landscape Position" and "Type of Settlement" were considered as potential explanatory variables of forest conversion. This section explains why these variables have been considered as explanatory variables for forest conversion.

Distance from Forest Edge

Forest borders have a high probability to be deforested [42, 80] and experience has shown that deforestation tends to start from the edge of existing forest [15]. Figure 5 illustrates that deforestation near the border is pronounced but that after 1,500 meters the amount of deforestation drops off to virtually nothing, which means that the relationship between forest change and forest edge is non-linear. After testing with Cramer's V was completed (Subsection 2.3.4.2.3), the independent variable "Distance from Forest Edge" was used as input for the models.

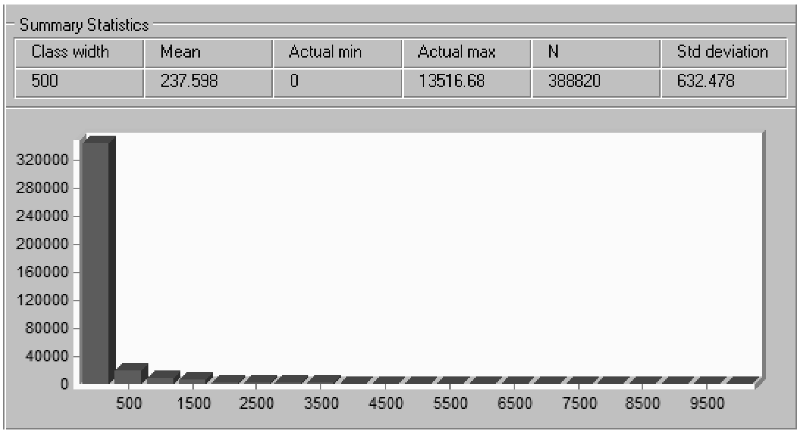


Figure 5 Frequency of Occurrence of Forest Change from Forest Edge

Distance from Roads and Settlements

Deforestation is also highly related to proximity to roads and urban areas [42, 80]. The frequency of forest change near roads is pronounced but also drops off to virtually nothing for roads after 6.5 km and after around 1 km for settlements (Figure 6 and Figure 7). Thus, these variables were tested with Cramer's V (Subsection 2.3.4.2.3), and then they were incorporated in the model as well.

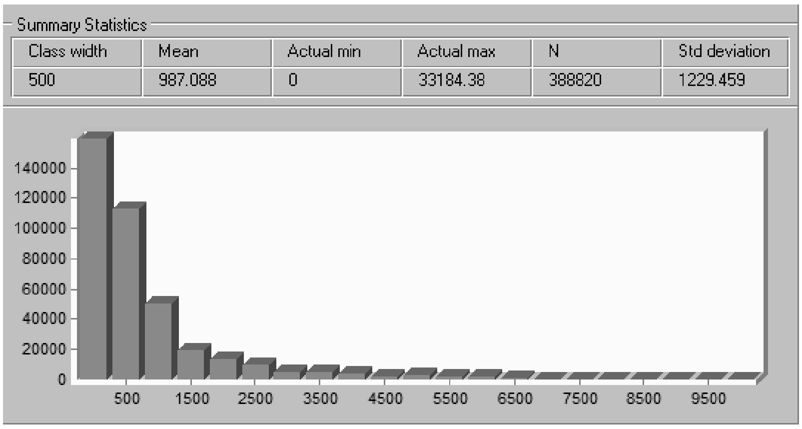


Figure 6 Frequency of Occurrence of Forest Change from Roads

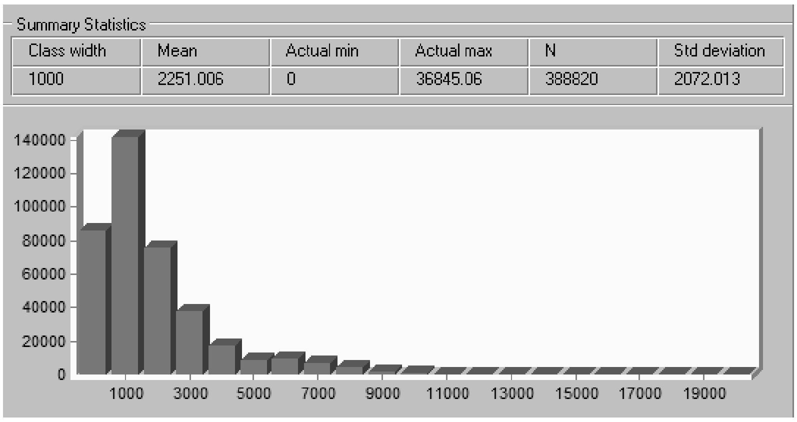


Figure 7 Frequency of Occurrence of Forest Change from Settlements

Landscape Position

According to Figure 8, the frequency of deforestation for each category of "Landscape Position" is unequal. The deforestation tends to occur markedly in areas of gentle slope. Topography seems to have some influence on deforestation of the TIPNIS. Thus, "Landscape Position" was also tested to see the influence on the forest clearing (Subsection 2.3.4.2.3), and then it was incorporated in the modelling as an explanatory variable.

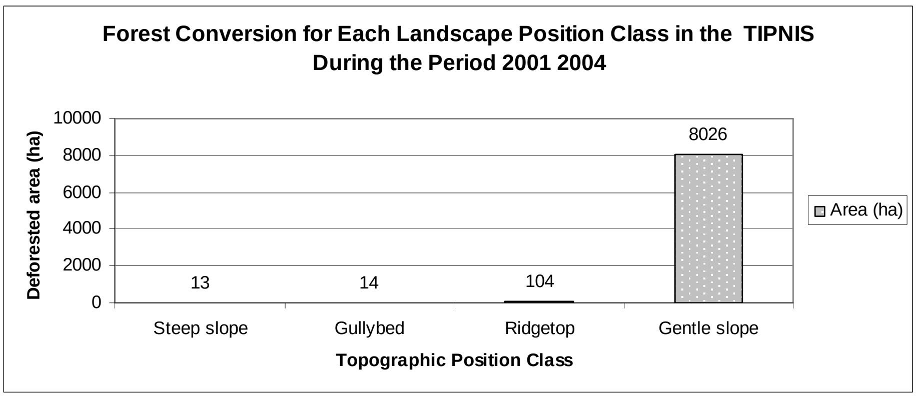


Figure 8 Frequency of Occurrence of Forest Change for Each "Landscape Position"

Type of Settlement

As mentioned earlier, TIPNIS is occupied by four different ethnic groups. But in the *Study Area* (southern part) only three ethnic groups were identified. The areas that they occupy were defined as a categorical variable called "Type of Settlement¨. The frequency of deforestation for each group is depicted in Figure 9.

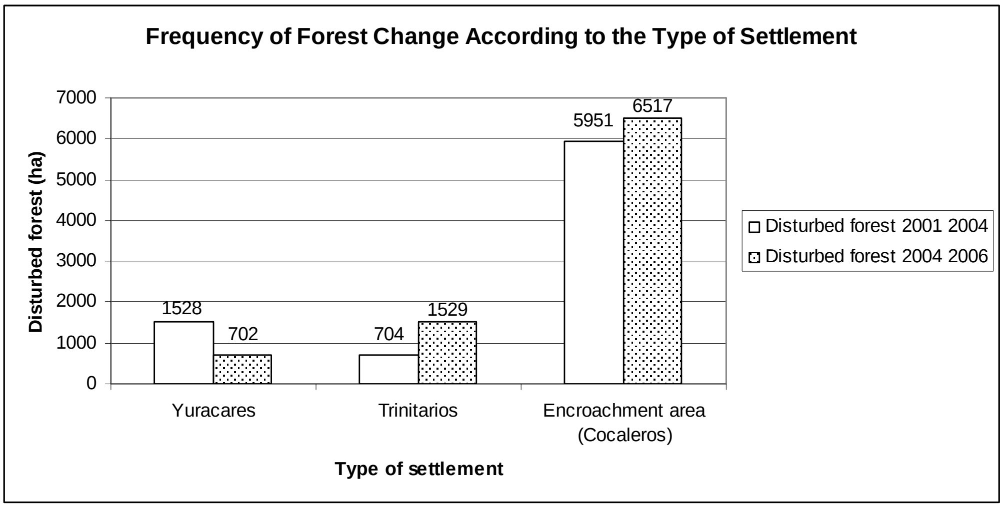


Figure 9 Frequency of Occurrence of Forest Change According to the "Type of Settlement"

The relationship between the "Type of Settlement" variable and change areas was also evaluated (Subsection 2.3.4.2.3), but the categorical variable has not been included as an input for the model (The reasons are discuss in Chapter 4).

2.3.4.2.2.1 Creating Continuous Variables

Distance from Forest Edge

The following procedure was implemented for "Distance from Forest Edge":

- The vector files (shp) Forest2001 and Forest2004 were imported.
- Raster files were created from each of the vector files.
- To obtain forest edge:
  A PATTERN command was applied to the Boolean raster map "Forest2001" making use of the option "Center versus neighbours" and a window size of 3*3. The result was a raster map with coarse forest borders called "Forest pattern". To make the forest borders thinner, "Forest pattern" and "Forest2001" were multiplied making use of the OVERLAY command. As a result, a raster map of forest borders "Forest edge" was obtained. Finally, the *Distance from Forest Edge* was obtained with the DISTANCE command. The map of *Distance from Forest Edge* is depicted in Figure 17, Subsection 3.3.4.

Distance from Settlements and Distance from Roads

The map of settlements and the map of roads were provided by the managers of the park (TIPNIS).

The following procedure was used to obtain the variable distances:

- The settlements and roads vector files (shp) were imported.
- Raster files were created from each of the vector files.
- The Operator DISTANCE was applied.

All maps obtained can be seen in Figure 17, Subsection 3.3.4.

2.3.4.2.2.2 Creating Categorical Variables

Landscape Position

The Topographic Position Index (TPI) was used to characterize the landscape position. TPI is an algorithm developed by Andrew Weiss 2001 [36] that has been incorporated into ArcGIS 9.2 to separate a region into landscapes classes.

The degree to which a pixel is higher or lower plus the slope of the pixel can be used to classify the pixel into slope position. If it is significantly higher than the surrounding neighbourhood, then it is likely to be at or near the top of a hill or ridge. Significantly low values suggest the cell is at or near the bottom of a valley. TPI values near zero could mean either a flat area or a mid-slope area, so the cell slope can be used to distinguish the two grids from elevation grids. TPI values provide a simple and repeatable method to classify the landscape into slope position and landform category [36].

Making use of the slopes obtained from a SRTM DEM (91x91 m), the Topographic Position Index (TPI) extension of ArcGIS 9.2 was performed to classify the study area. Four categories were obtained (See Table 3). The threshold parameters were assigned based on the default ranges of the ArcGIS software. The map of *Landscape Position* is depicted in Figure 17, Subsection 3.3.4.

Table 3 Landscape Categories Obtained from the TPI

TPI Threshold Parameters	Landscape Category
- 10 m	Gully bed
+ 12 m	Ridge top
- 6 m	Gentle slope
+ 6 m	Steep slope

Type of Settlement

Based on the “Type of Settlement” map provided by the managers of the park (TIPNIS 2007), three ethnic groups were identified in the study area: “Trinitarios”, “Yuracarés” and “Cocaleros”. It is important to mention that the fourth ethnic group present in the TIPNIS, Chimanes, live in the northern part of the park. Because this study was carried out in the southern part of the park only, the Chimanes were not included as part of the study. The “Type of Settlement” map is depicted in Figure 17, Subsection 3.3.4.

2.3.4.2.3 Measuring the Strength of the Relationship between Explanatory Variables and Forest Change (Univariate tests of association Cramer’s V)

To assess preliminarily the relationship between the categorical variables and occurred change, Chi square test of independence was performed for categorical variables in SSPS 17. For the two categorical variables “Landscape Position” and “Type of Settlement”, the significance level was < 0.05 (see Subsection 3.3.2.1 and Appendix II).

Chi-square says that there is a relationship between variables, but it does not say just how significant this relationship is. Cramer’s V is a post-chi square test that was used to provide this additional information.

Cramer's V is a statistic that transforms chi-square (for a contingency table larger than two rows by two columns) to a range of zero to one, where unity indicates complete agreement between two nominal variables [39] and [81]. V is calculated by first calculating chi-square, then using the following calculation:

$$V = \sqrt{\frac{X^2}{(N)(k-1)}}$$

Where:
χ^2 = square root of chi square
N = number of cases in the table
k = the lesser of the number of rows and columns.

In 2006, IDRISI 15 was released including a new tool called the "explanatory variable test procedure". According to IDRISI Help 2006 [21], the explanatory variable test procedure is based on Cramer´s V contingency table analysis. This procedure can test the strength of the association between the dependent variable and both quantitative and qualitative variables.

Quantitative variables (‘*Distance’ variables in our case*) are binned to 256 categories in order to conduct this test. “Distance from Forest Edge”, “Distance from Roads” and “Distance from Settlements” were tested using IDRISI’s explanatory variables test. The test results obtained are shown in Table 7 Subsection 3.3.2.2.

For qualitative variables, the procedure uses the native categories of the variable to test association with the distribution of land covers in the later land cover map.

Before performing the explanatory test procedure, the qualitative variables “Landscape Position” and “Type of Settlement” needed to be transformed from nominal to numeric values. The Evidence of Likelihood tool was used to perform this transformation. This procedure looks at the relative frequency of pixels belonging to the different categories of that variable within areas of change. In effect, it asks the question of each category of the variable, "How likely is it that you would have a value like this if you were an area that would experience change?" [12, 21]

After the classification, the ‘Landscape Position’ map was imported to IDRISI 15 and was transformed with the “Evidence of Likelihood” tool in order to get quantitative data input. Then, the map was separated into four individual maps, one per category. Each Landscape category was tested with the explanatory variable test based on Cramer’s V, using as dependent variable the Boolean map “Change 2001-2004”. The results are shown in Table 7 Subsection 3.3.2.2.

The map “Type of Settlement” was imported to IDRISI and also transformed with the “Evidence of Likelihood” tool. The types of settlement were separated into three individual categories which were tested with the explanatory variable test based on Cramer’s V, using as dependent variable the Boolean map “Change 2001-2004”. The results are shown in Table 7 Subsection 3.3.2.2.

The results of the explanatory test procedure for each variable were Cramer’s V values. According to Eastman, 2006 [15], variables that have a Cramer’s V of about 0.15 or higher are useful while those with values of 0.4 or higher are good. The *p* value expresses the probability that the Cramer’s V is not significantly different from 0. Thus, a low value of *p* is not a good indicator of a variable’s worth, but a high value is a sure sign that it can be rejected.

Although Cramer’s V assessed the relationship between an individual explanatory variable and forest change (Subsection 3.3.2.2), a deeper analysis was needed in order to see the significance of each variable when there are other variables involved in the forest change process. Logistic regression was used to give a better insight (Subsection 2.3.4.3).

2.3.4.2.4 Relationship between Explanatory Variables

To avoid ***multicollinearity***, correlation analysis between the independent variables was performed. None of the variables showed high correlation. The results are described in (Subsection 3.3.3).

2.3.4.3. Logistic Regression Model (LRM)

Forest conversion was modelled and analyzed using Logistic Regression Model in IDRISI 15, Andes edition. The purpose of modelling was (i) to assess the relative significance of four explanatory variables on forest change during the period 2001-2004; and (ii) to predict probability of change for the year 2006.

Logistic regression is a variation of ordinary regression which is used when the dependent (response) variable is a dichotomous variable.

In this study, as mentioned before, the dependent variable is a binary presence or absence event, where 1= forest change and 0= no change, for the period 2001–2004. The logistic function gives the probability of forest change as a function of the explanatory variables. In other words, the probability of forest change for each pixel is a function of the values that the other variables have for the same pixel.

According to Schneider and Pontius 2001 [62] the function is a monotonic curvilinear response bounded between 0 and 1, given by a logistic function of the form:

$$(1)\ \ p = E(Y)\frac{\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + ...)}{1 + \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + ...)}$$

Where: p is the probability of forest loss in the cell, $E(Y)$ the expected value of the binary dependent variable Y, β_0 a constant to be estimated, β_i a coefficient to be estimated for each independent variable X_i. The logistic function can be transformed into a linear response with the transformation:

$$(2)\ \ p' = \log_e\left(\frac{p}{1-p}\right)$$

hence

$$(3)\quad p' = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots..$$

The transformation (Eq. (2)) from the curvilinear response (Eq. (1)) to a linear function (Eq. (3)) is called a logit or logistic transformation. The transformed function allows linear regression to estimate each $\beta_{i.}$ Since each of the observations is a pixel, the final result is a probability score (p) for each pixel.

In logistic regression the significance of the coefficients β_i is tested with the Wald test, which is obtained by comparing the maximum likelihood estimate of every β_i with its estimated standard error (Hosmer and Lemeshow, 1989 cited by [32]). A coefficient is significant if the tested null hypothesis that the estimated coefficient is 0 can be rejected at a 0.01 (or 0.05) significance level [77].

According to Ayalew and Yamagishi 2005 [3], in order to appropriately interpret the meaning of Eq. (1), one has to use the coefficients as a power to the natural log(*e*). The result represents the odds ratio or the probability that an event will occur divided by the probability that it fails to do so. If a coefficient is positive, its transformed log value will be greater than one, meaning that the event is more likely to occur. If a coefficient is negative, the transformed log value will be less than one and the odds of the event occurring decreases. A coefficient of 0 has a transformed log value of 1, and it does not change the odds one way or the other. For a positive coefficient, the probability plotted against the values of an independent variable follows an S-shaped curve. A mirror image will be obtained for a negative coefficient (Menard, 1995 cited by [4]).

2.3.4.3.1 Calibration of the Model

The Logistic Regression Model requires that the variables be linearly related to the forest change. Thus, before introducing the variables in the model, all the variables were normalised between 0.1 and 0.9. The natural log transformation was performed for continuous variables (Distances). The Evidence of Likelihood transformation was applied for the categorical explanatory variable "Landscape Position".

To calibrate the Logistic Regression Model, the explanatory variables measured for the year 2001 were incorporated in the IDRISI's Logistic Regression Module as independent variables. The forest change for the period 2001-2004 was incorporated as the dependent variable. In order to decrease processing time and to reduce the negative spatial interdependence, the stratified sampling was selected with a 10% sampling proportion.

The stepwise method was used to select the best set of predictor variables since the study considered 6 different predictor sets (Table 9, Subsection 3.3.5). Finally, and following the methodology used by van Gils and Loza 2006 [79], the selection of the best-fitted model with the minimum amount of predictors was done by means of the Akaike Information Criterion (AIC) index. The smaller the AIC is, the better the fit of the model (Table 10, Subsection 3.3.5).

The results were the regression equation of the best-fitted predictors set (Subsection 3.3.5) and a map of probability of deforestation for the year 2004 "Calibration 2004" (see Figure 18 Subsection 3.3.5).

2.3.4.3.2 Prediction

The prediction for forest change between the year 2004 and the year 2006 was performed using the obtained probability of deforestation for the year 2004. For the new prediction, the dynamic variables "Distance from Forest Edge" and "Distance from Roads" were changed as long as they were in the year 2004. The variables "Distance from Settlements" and "Landscape Position" remained the same. The result was a new map of probability of forest change for the year 2006 (see Figure 19 Subsection 3.3.5).

2.3.4.3.3 Validation

The real "forest change" map for the year 2006 (Figure 22 Subsection 3.3.7) was used to assess the prediction of probability of forest change performed by the model.

The validation was performed with the Relative Operating Characteristic (ROC) curve, which is an effective and widely used method for evaluating the discriminating power of a statistical model [19, 30, 62]. Eastman 2006 [15], also mentions that ROC can be used to determine how well a continuous surface predicts the locations given the distribution of a Boolean variable. (In this study, Forest Change is the Boolean variable).

A ROC curve is a graph of the True Positive Fraction (sensitivity) vs. False Positive Fraction (1-specificity). The Area Under an ROC Curve (AUC) is a measure of overall performance. The maximum area is 1.0: The test is useless if the diagonal line is from 0.0 to 1.0 and the area under ROC=0.5, so a more meaningful measure is the area in excess of 0.5. As test performance improves, the curve moves towards the upper left corner and the area under ROC increases. The obtained AUC is shown in (see Figure 23 Subsection 3.3.7).

2.3.4.4. Multi-Layer Perceptron Model (MLP)

To make a parallel prediction of forest conversion from the year 2004 to the year 2006, the IDRISI's Multi-Layer Perceptron operator was used. IDRISI's MLP was created to undertake the classification of remotely sensed imagery through an MLP neural network classifier using the feed back propagation algorithm. In the year 2006, MLP was incorporated into the Land Change Modeler for Ecological Sustainability in IDRISI 15.0 (the Andes Edition), offering an alternative tool for cover change modelling.

Lippitt 2008 [40] mentions that MLP has three primary components: an input layer, an output layer, and one or more hidden layers. Each layer is composed of a user-defined number of neurons (mathematical functions conceived as an abstraction of biological neurons). Output neurons represent the classes specified by the calibration data. Input variables and hidden layer neurons are randomly weighted and assigned membership to an output neuron.

According to Nefeslioglu *et al* 2008 [57], the feed-forward back-propagation learning algorithm is a well recognized procedure for training neural networks (Multi-Layer Perceptron—MLPs topology). It is based on searching a performance surface (error as a function of neural network weights) using gradient descent for point(s) with minimum error.

In other words, the input data (explanatory variables set) is repeatedly presented to the neural network. With each presentation the output of the neural network is compared to the desired output (forest change 2001-2004) and an error is computed. This error is then fed back (back-propagated) to the neural network and used to adjust the weights such that the error decreases with each iteration and the neural model gets closer and closer to producing the desired output [47]. This process is known as "training".

The purpose of training the network is to get the proper weights both for the connection between the input and hidden layer, and between the hidden and the output layer for the classification of the unknown pixels [15]. The training result is a parameter file (*.bpn), where weights and other information are recorded. This information is visualized in this study in a map of propensity of deforestation for the year 2004. The parameter file (*.bpn) is then used to make the prediction for the year 2006.

2.3.4.4.1 Calibration

Similar to the LRM, the explanatory variables measured for the year 2001 were incorporated in the IDRISI's Multi-Layer Perceptron module as independent variables. The forest change for the period 2001-2004 was incorporated as the dependent variable.

A sample set was needed to create both the training subsets and test subsets. According to Nefeslioglu *et al* 2008 [57], it is expected that the training data include all the data belonging to the problem domain. Certainly, the training subset is used in the training stage of the model development to update the weights of the network. On the other hand, the test data should be different from those used in the training stage. The main purpose of the test subset is to check the network performance using untrained data and to confirm its accuracy. No exact mathematical rule to determine the required minimum size of any of these subsets exists.

This was a critical point during the process as the number of training samples for all subsets affects the accuracy of the training results. According to Eastman 2006 [15], too few samples may not represent the population for each category, while too many samples may cause samples to overlap, leading to a possible over training of the network. Additionally, too many iterations can also cause over training.

Finally, after several attempts the sample set was determined to be in the range 180,000 to 200,000, and the training stage was performed. The training process entails running patterns through the network until the network has "learned" how to apply the data in the future. The parameters used for the training stage are shown in Figure 10 (Nex page).

After 5,000 iterations, an accuracy rate of 84.24% was obtained. According to Eastman 2006 [15], one should achieve an accuracy rate in the vicinity of 80%. If the range is lower than 75%, the neural network must be trained again. Finally, the classify option was performed and the result was a single output: the map of probability of deforestation for the year 2004 (Figure 20 Subsection 3.3.6).

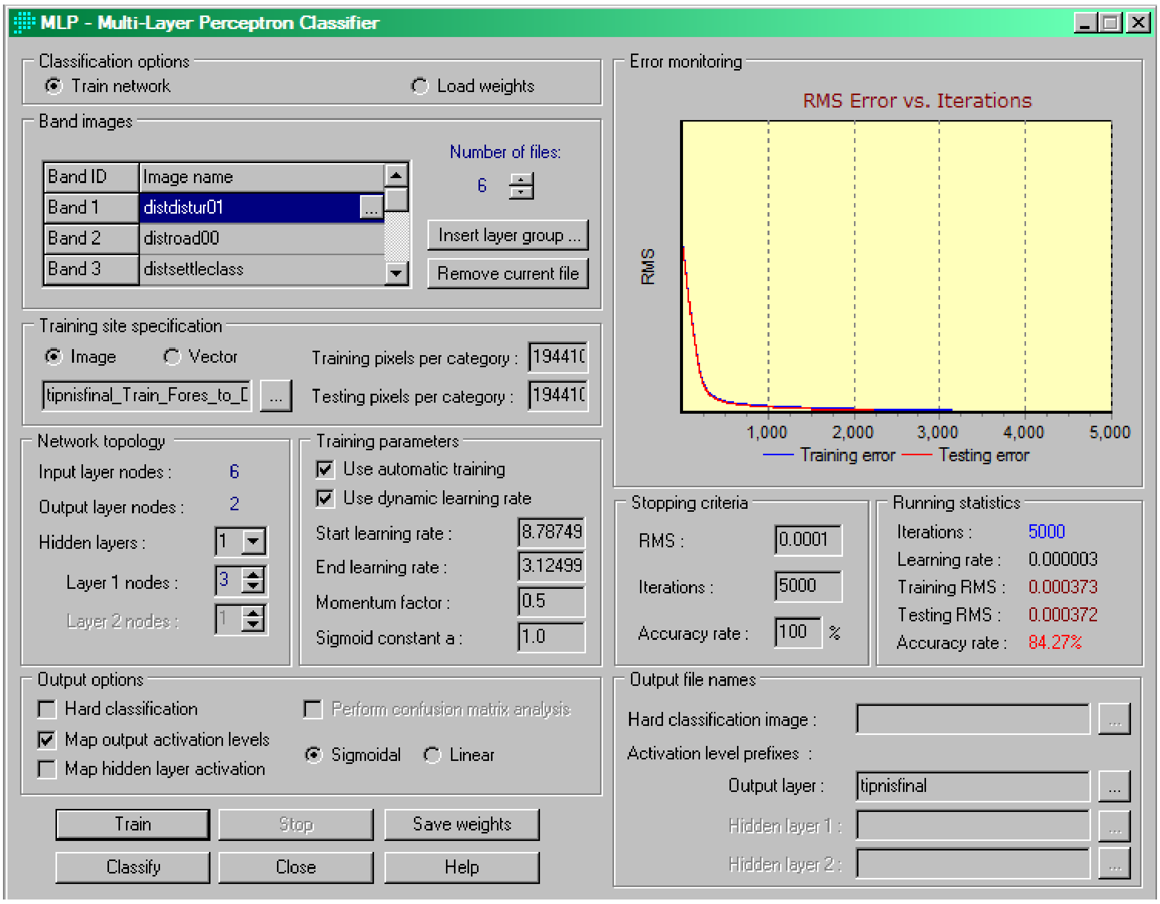


Figure 10 Parameters Introduced to the MLP before the Training Stage

2.3.4.4.2 Prediction

Once the network has been trained, new data can be run through it. The network classifies new data based on its previous training experience. If an exact match cannot be found, the network will select the closest match found in memory (parameter file *.bpn). The prediction for the year 2006 was also based on the replacement of the variables "Distance from Forest Edge 2004" and "Distance from Roads 2004" with the new variables "Distance from Forest Edge 2006" and "Distance from Roads 2006". The output of the MLP is an activation value which expresses, for each pixel, the propensity to deforestation. The data layer consists of cells with continuous scores varying from zero to one: the higher the pixel score, the higher the probability of change for that pixel [43]. The result is then a "fuzzy" deforestation map that portrays gradations of the probability of being deforested for the year 2006 (Figure 21 Subsection 3.3.6)

2.3.4.4.3 Validation

The result of the prediction for the year 2006 was compared to the real "forest change" map for the year 2006. Validation was performed using the Area Under the ROC Curve (AUC). The obtained AUC is shown in Figure 24 Subsection 3.3.7.

2.3.4.5. Comparison between Models

The AUC curves obtained for each model during the validation of the prediction show which model has performed the best prediction (see Figure 25 Subsection 3.3.8)

3. Results

3.1. Results for Land Cover Classification

3.1.1. Classification Results

The classification results (Figure 11) show that 66.7 % of the forest's cover is still Primary Forest and the category covering the smallest area in 2006 is River Shores (0.5 %).

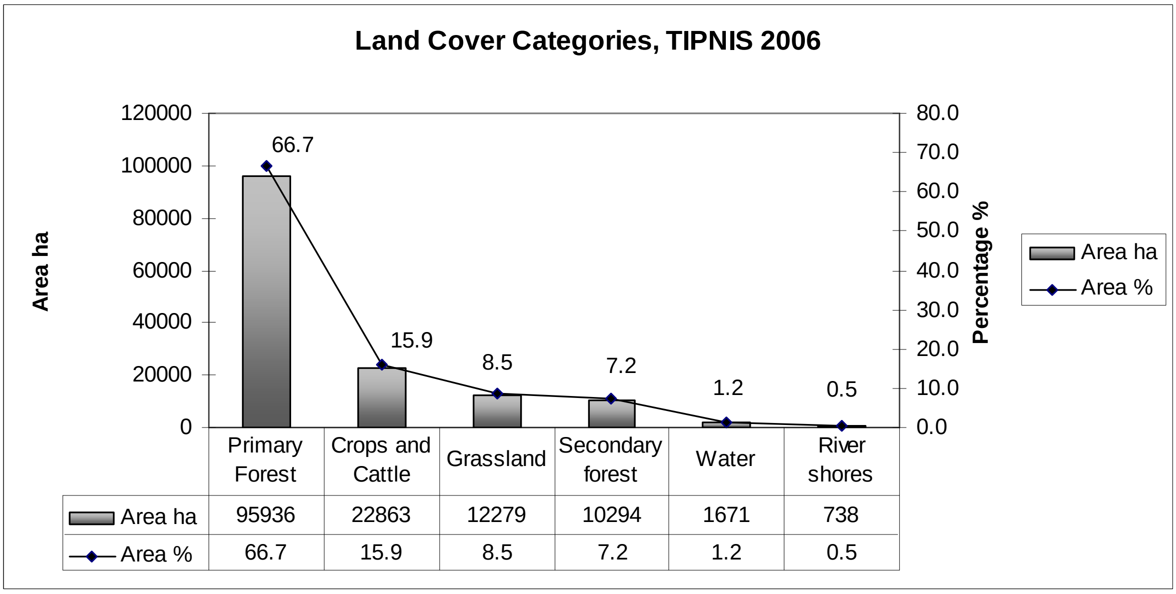


Figure 11 Land Cover Categories, TIPNIS 2006

The methodological approach is described in Subsection 2.3.2.2.

3.1.2. Validation Results

Table 4 shows the Confusion Matrix comparing the classified categories (rows) to the Ground Control Points (columns) taken during the field work. 81.58% was the overall accuracy obtained by the Land cover classification. The methodological approach is described in Subsection 2.3.2.3.

Table 4 Confusion Matrix Output ILWIS 3.3 of Classified and Observed Cells

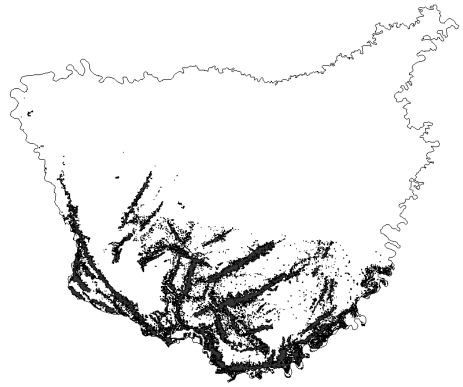
Confusion Matrix	Forest	Crops and Cattle Pastures	Grassland	Secondary Forest	ACCURACY
Forest	9	0	1	0	0.9
Crops and Cattle Pastures	0	9	0	1	0.9
Grassland	0	0	5	0	1
Secondary Forest	1	3	1	8	0.62
RELIABILITY	0.9	0.75	0.71	0.89	
Overall Accuracy = 81.58 %					

3.1.3. Reclasification Results

Figure 12 Final Maps with Two Categories: Forest and Disturbed Forest. Years: 1986, 1991, 2001, 2004 and 2006.

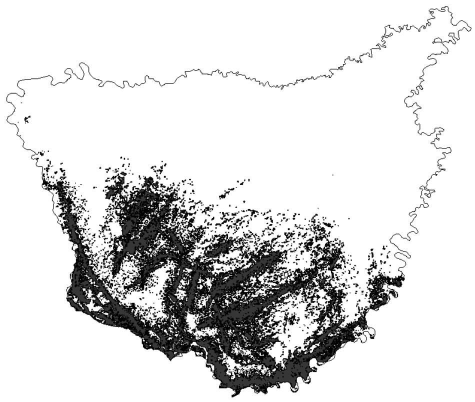
1986

1991

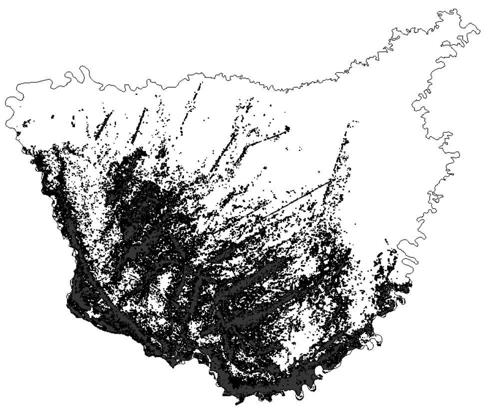


2001

2004



2006



Areas in white represent Forest.
Areas in black represent the extent of Disturbed Forest.

Figure 12 shows the forest loss for each year of study. The methodological approach is described in Subsection 2.3.2.4.

3.2. Results for Forest Loss and Deforestation Rates

3.2.1. Forest Loss Results

Table 5 summarizes the forest loss for 6 years during the period 1976-2006. The year with the least deforestation was 1986 (134 ha/year). The most deforestation (4066 ha/year) was registered for the year 2006. Both Table 5 and Figure 13 (Next page) illustrate that the forest area has been decreasing constantly during the period 1976 to 2006. In 1976, almost 100% of the area was forest. In 2006, only 77% of the area remained forest.

Table 5 Forest Area and Deforestation Rates During the Period 1976 – 2006

Year	Forest Area Km2	Forest ha	% Forest Area	% Disturbed Forest	ha/year Forest Loss	Forest Loss Rate Puyravaud	Deforestation Rates		
							r Puyravaud	q FAO 2000	DR FAO 1996
1976	1424.7	142469.6	99.09	0.91					
1986	1424.0	142402.1	99.04	0.96	134	0.005	-0.005	-0.005	-0.005
1991	1336.4	133644.9	92.95	7.05	1751	-1.27	-1.27	-1.26	-1.20
2001	1270.6	127062.8	88.37	11.63	658	-0.51	-0.51	-0.50	-0.48
2004	1187.6	118758.1	82.60	17.40	2768	-2.25	-2.25	-2.23	-2.13
2006	1106.3	110625.3	76.94	23.06	4066	-3.55	-3.55	-3.48	-3.37

Table 5 also lists Deforestation Rates calculated using three different formulas. There is a slight difference between them. The difference tends to be bigger as the size of the calculated areas increases. The biggest difference (0.18) was observed for the year 2006 between “r Puyravaud” and “DR FAO 1996”.

Table 6 depicts the variation in deforestation rates for different periods. First, the rate for the entire period of study 1986 – 2006 is presented, which is -1.9. Then, the table lists the rate for 1986 and 2001 (-0.9), and finally, Table 6 specifies the average deforestation rate of -2.9 for the last period 2001 to 2006.

Table 6 Average Deforestation Rates for Different Periods

Period	Deforestation Rates Average
1986 - 2006	-1.9
1986 - 2001	- 0.9
2001 - 2006	- 2.9

3.2.2. Deforestation Rates Results

According to Figure 13, the forest area in the TIPNIS has been decreasing constantly from 1976 to 2006. The deforestation rates rise and fall and then rise again. The rate of deforestation until 1991 rose from -0.005%.a $^{-1}$ (period 1976 – 1986) to -1.3%.a $^{-1}$ for the year 1991. Then, the deforestation rate fell from -1.3 - 0.5%.a $^{-1}$. For the year 2001, it rose again from -0.5 -2.3 %.a $^{-1}$. For the period 2004 -

2006, the rate continued increasing from -2.3 to -3.5. %.a $^{-1}$. The deforestation rate average for the period 1986 – 2006 is -1.9 %.a $^{-1}$ and for 2001-2006 is - 2.9%.a $^{-1}$.

The methodological approach is given in Subsection 2.3.3.

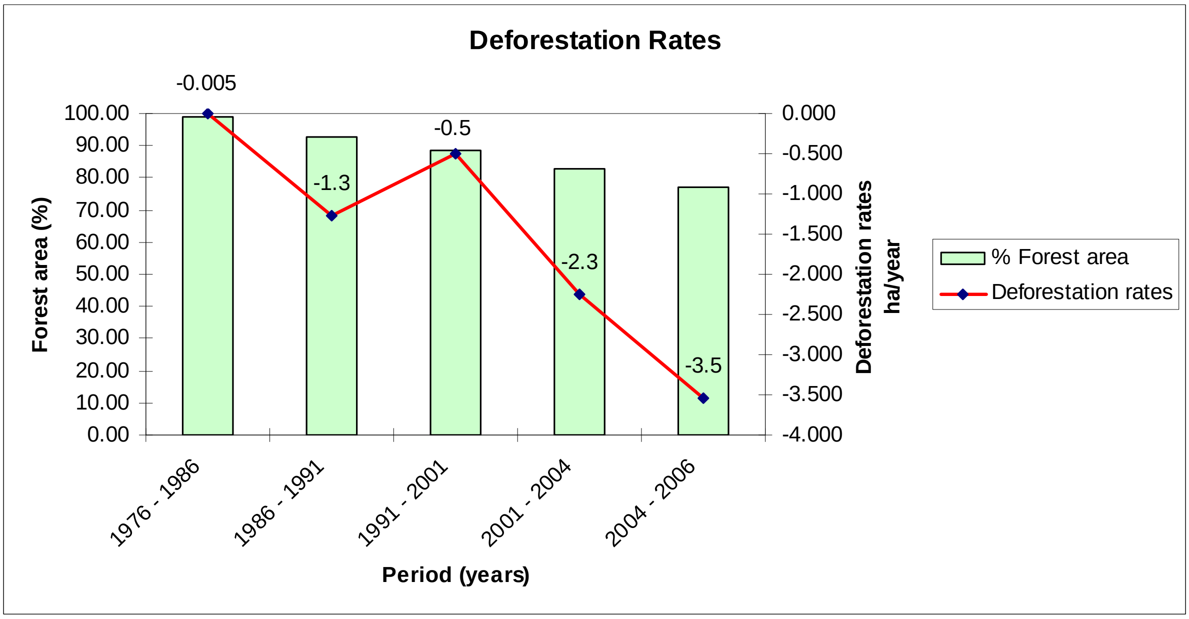


Figure 13 Decreasing Forest Area for the Period 1976 – 2006 and Deforestation Rates

3.2.3. Deforestation Pattern Results

Figure 14 illustrates the frequency of patches of "Disturbed Forest" for each year of study. A pattern of a large and growing number of small patches of deforestation can be observed each year. Note the difference between small patches (0.5 up to 5 ha) of deforestation and larger deforested areas.

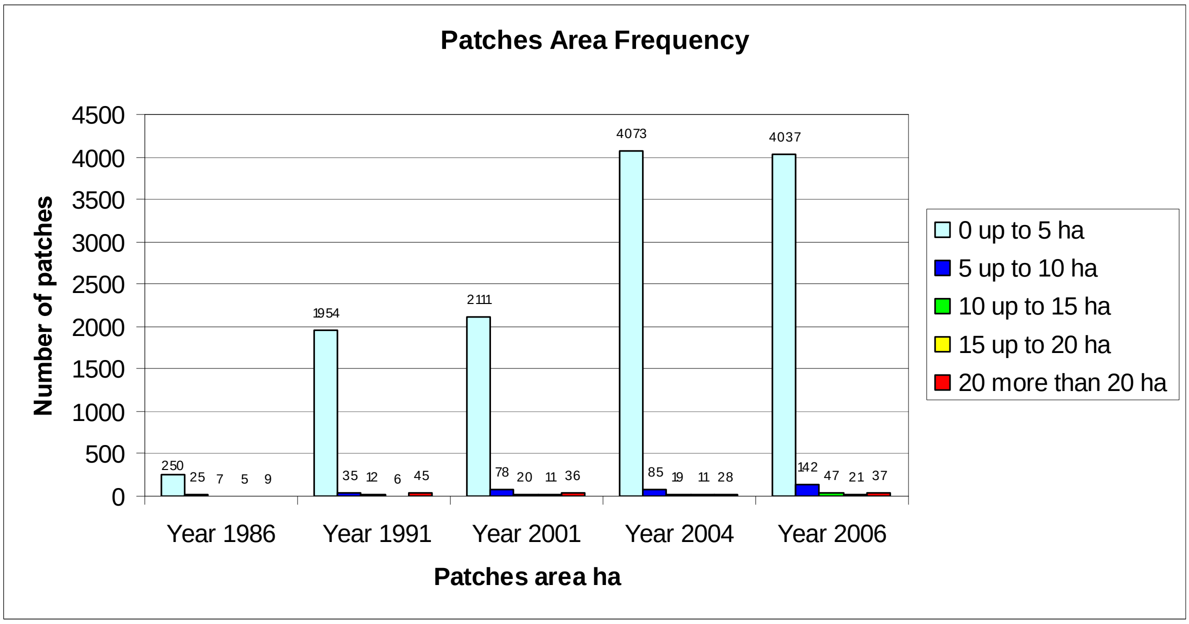


Figure 14 Patches Area Frequency Obtained from FRAGSTATS Results

Figure 15 depicts the number of patches that were present in five distance ranges from the nearest road during the year 2004. The major number of patches is located between 250 and 750 m from the nearest road.

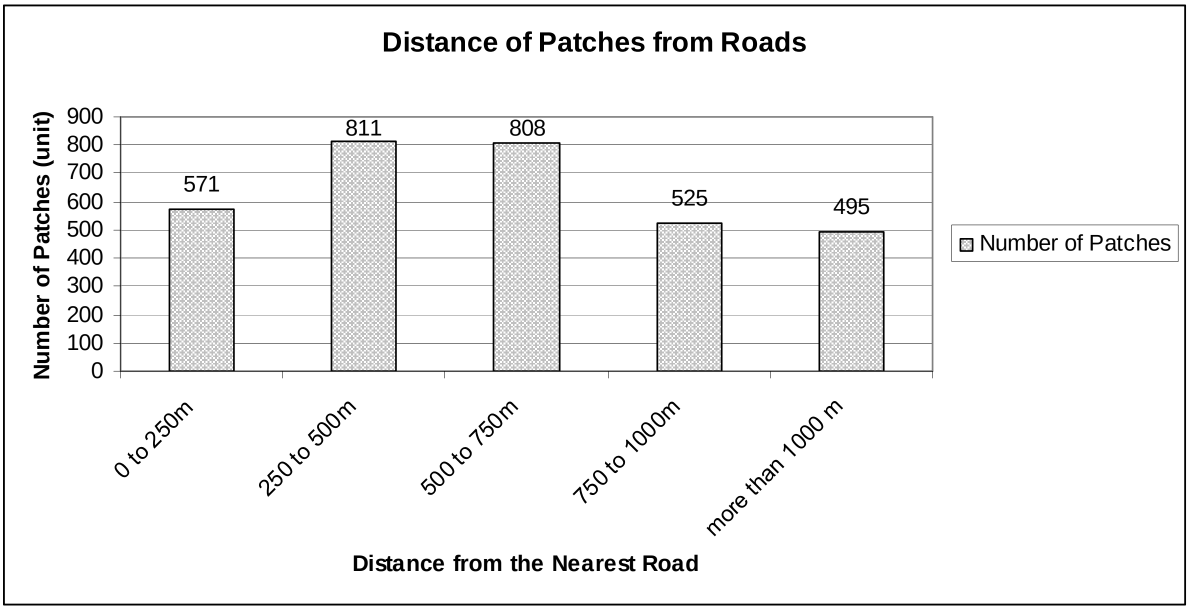


Figure 15 Distance of Patches from the Nearest Road

3.3. Modelling Conversion Results

3.3.1. Dependent Variable Results

Figure 16 shows the forest change that occurred during the periods 2001 to 2004 and 2004 to 2006. Areas of change are represented in black. Forest change for the period 2001 to 2004 was the dependent variable used by the logistic regression analysis.

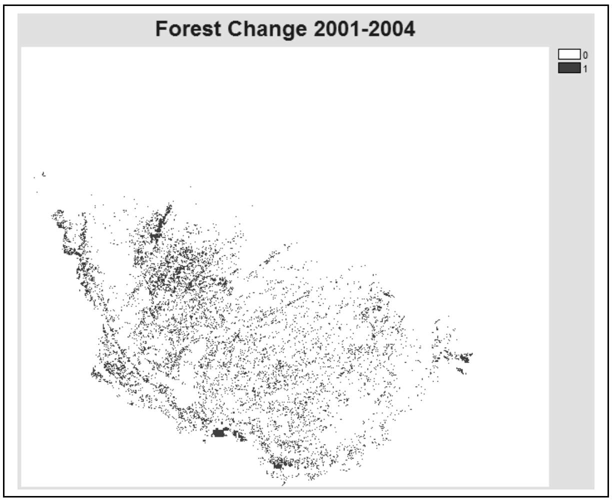


Figure 16 Forest Change That Occurred During the Periods 2001-2004 (1 = Change; 0 = No change)

3.3.2. Univariate Test of Association Results

3.3.2.1. Chi Square Test of Independence, Categorical Variables

For both variables "Landscape Position" and "Type of Settlement", the Pearson Chi-square calculated was a *p* value of 0.000, which is less than the significance level of 0.05. (See the Tables in Appendix II.)

3.3.2.2. Explanatory Variable Test Cramer's V

Table 7 shows the strength of association that each explanatory variable has with Forest Change. The measurement of association used is Cramer's V. For continuous variables, V values are between 0.5 and 0.6 with a *p* value of 0.0, which means a good association exists.

Table 7 Relationship between Explanatory Variables and Forest Change Using Cramer's V

Explanatory Variable	Cramer's V	*p* value
Distance from Forest Edge	0.6003	0.0000
Distance from Roads	0.6147	0.0000
Distance from Settlements	0.5179	0.0000
Landscape Position		
Gully bed	0.9571	0.0000
Gentle slope	0.9597	0.0000
Ridge top	0.0299	0.0000
Steep slope	0.0000	1.0000
Type of Settlement		
Cocaleros	0.9892	0.0000
Trinitarios	0.7293	0.0000
Yuracares	0.0000	1.0000

Cramer's V values for "Landscape Position" show a strong association with forest change for the classes "Gully bed" and "Gentle slope" (0.96). On the other hand, class "Ridge Top" has a very low value (0.023) and "Steep Slope" presents a V value of 0 and a *p* value of 1, meaning there is a very poor association.

For "Type of Settlement", Yuracares class has a V value of 0.00 and *p* value equal to 1, which means there is a very poor association. Trinitarios have a V value of 0.73, a good association, while the Cocaleros class shows a strong association with forest change with a V value of 0.98 and p=0.

The methodological approach is described in Subsection 2.3.4.2.3.

3.3.3. Relationship between Explanatory Variables Results

According to Table 8, the variables related to “distance from roads” and “distance from settlements” presented correlation coefficients between 0.3 and 0.5. This is understandable since settlements have some dependence on roads and vice versa. Similar results occurred when Distance from Forest Edge and Distance from Roads were correlated, reflecting some relationship between roads and ease of entry into the forest. Note that the lowest coefficients (no more than 0.147) were found when the variable Landscape Position was crossed with the distance variables confirming that no relationship exists between these variables. See graphics in Appendix III.

Table 8 Correlation Coefficients between Independent Variables

Variable 1	Variable 2	Correlation Coefficient
Distance from Forest Edge	Distance from Roads	0.554
Distance from Forest Edge	Distance from Settlements	0.417
Distance from Forest Edge	Landscape position	0.047
Distance from Roads	Distance from Settlements	0.367
Distance from Roads	Landscape Position	0.135
Distance from Settlements	Landscape Position	0.147

3.3.4. Creating Variables Results

Figure 17 depicts the explanatory variables of this study.

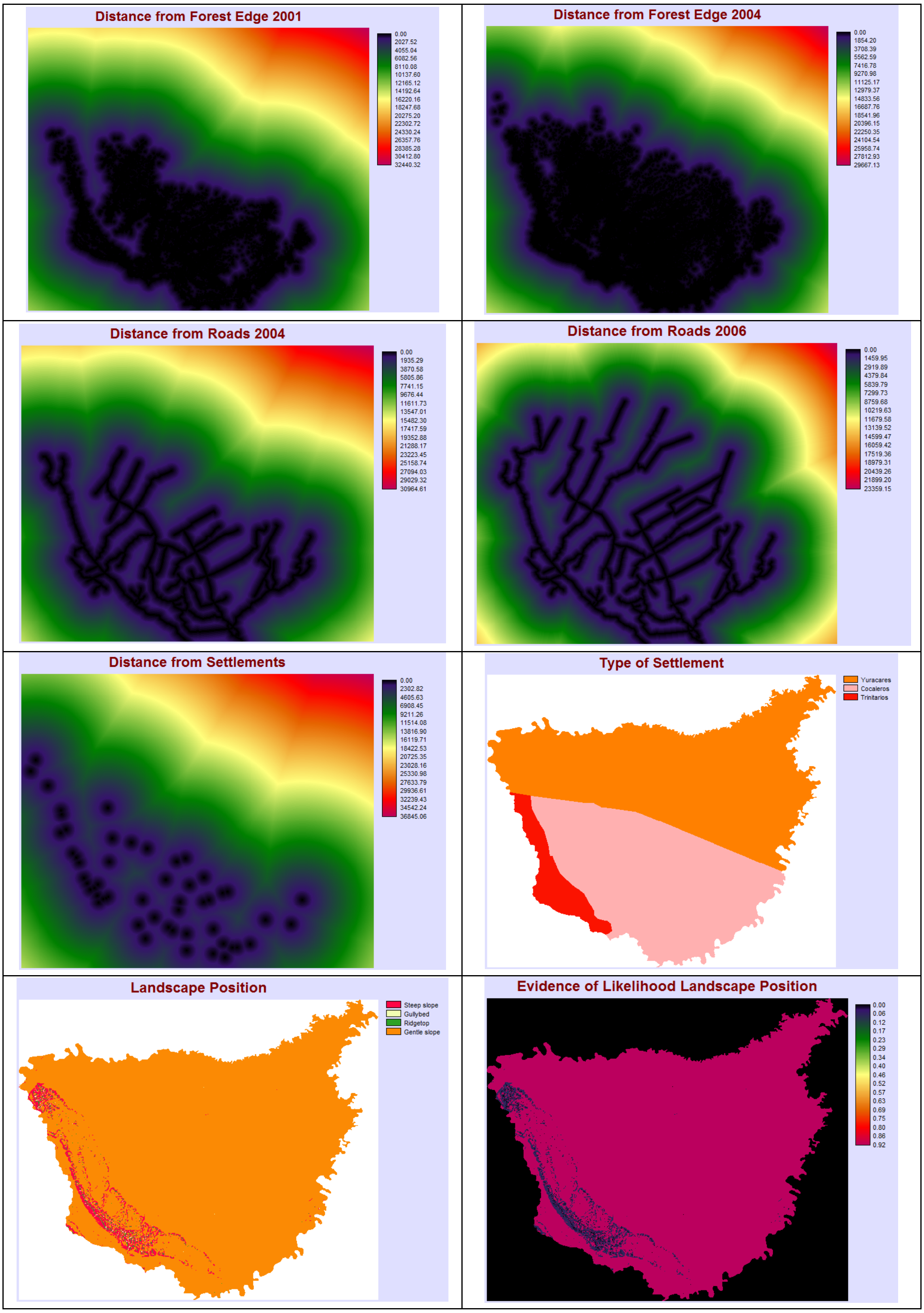


Figure 17 Explanatory Variables Used as Input in the Modelling

The methodological approach is described in Section 2.3.4.2.2

3.3.5. Logistic Regression Model Results

Table 9 and Table 10 summarize the results obtained by the Logistic Regression Model for six sets of predictor variables.

Table 9 Summary of Regression Equation Coefficients of 6 Sets of Predictor Variables

	Regression Coefficients					
Explanatory Variable	**M1**	**M2**	**M3**	**M4**	**M5**	**M6**
Intercept	0.24	1.38	2.70	1.55	0.09	-0.10
Distance from Forest Edge	-0.58	-0.46	-0.42	-0.34	-0.40	
Distance from Roads		-0.27	-0.23	-0.21	-0.26	-0.61
Distance from Settlements			-0.23	-0.27		
Landscape Position				1.27	1.20	1.66

Table 10 Statistics of the 6 Predictor Sets Obtained by Logistic Regression

Statistics	M1	M2	M3	M4	M5	M6
Total Number of Pixels	11202660	11202660	11202660	11202660	11202660	11202660
$-2\ln L$ (L=likelihood)	251767	249364	246749	242639	246651	257327
$-2\ln L_0$	322696	323750	322424	321145	324237	320673
Model Chi-square	70928	74386	75675	78506	77586	63346
Goodness of Fit	600369	570442	566964	506280	519634	571994
Pseudo R^2	0.22	0.23	0.23	0.24	0.24	0.20
AUC/ROC	0.88	0.88	0.88	0.89	0.89	0.87
Odds Ratio	3.97	3.56	4.25	4.23	3.82	5.04
AIC	251769.31	249367.97	246754.64	242646.83	246656.99	257330.73

This study selected the set predictor M4 as the best combination to be used in the prediction. The selection procedure was performed as follows:

According to Ayalew and Yamagishi 2004 [3], a key starting point could be the model chi-square, whose value provides the usual significance test for logistic regression. It is a difference between −2ln L (L=likelihood) for the best-fitting model (Predictor set) and −2ln L0 for the null hypothesis in which all the coefficients are set to 0. The value measures the improvement in fit that the independent variables brought into the regression.

In this study, the high value chi-square (for the predictor set M4) indicates that the occurrence of forest change is far less likely under the null hypothesis (without forest conversion influencing parameters) than the full regression model (where the parameters are included).

The goodness of fit is an alternative to model chi-square for assessing the significance of Logistic Regression Models. It is calculated based on the difference between the observed and the predicted values of the dependent variable. The smaller this statistic is, the better fit it indicates. Model M4 has a value of 506,280 which is the smallest “Goodness of fit” statistic among the model sets.

The pseudo R2 value, which can be calculated from 1−(ln L/ln L0), indicates how the logit model fits the dataset (Menard, 1995 cited by [3]). Thus, pseudo R2 equal to 1 indicates a perfect fit, whereas 0 shows no relationship. When a pseudo R2 is greater than 0.2, it shows a relatively good fit (Clark and Hosking, 1986 cited by [4]). The pseudo R2 of the M4 predictor set is 0.24.

Under ROC, the M4 predictor set obtained an accuracy of 89% and provided the smallest AIC index making it the best-fitted predictor set.

Regression Equation best-fitted M4 predictor set[3]:

Linear probability (logit(change0104)) = 1.55
- 0.34*distance from disturbance log
- 0.21*distance from roads log
- 0.27*distance from settlements log
+ 1.27*Landscape Position

The relative significance of the explanatory variables can be assessed using the corresponding coefficients in the Logistic Regression Model. According to Eastman 2006 [15], the intercept can be thought of as the value for the dependent variable when each independent variable takes on a value of zero. The coefficients indicate the effects of each of the explanatory variables on the dependent variable.

Figure 18 and Figure 19 show the results of the calibration and the prediction of the LRM. The colour in the figures indicates the degree of probability of deforestation. Areas in red show high probability for forest conversion, while, areas in other colours have decreasing probability for deforestation.

[3] The statistics of the Predictor ser M4 can be visualized in Appendix IV.

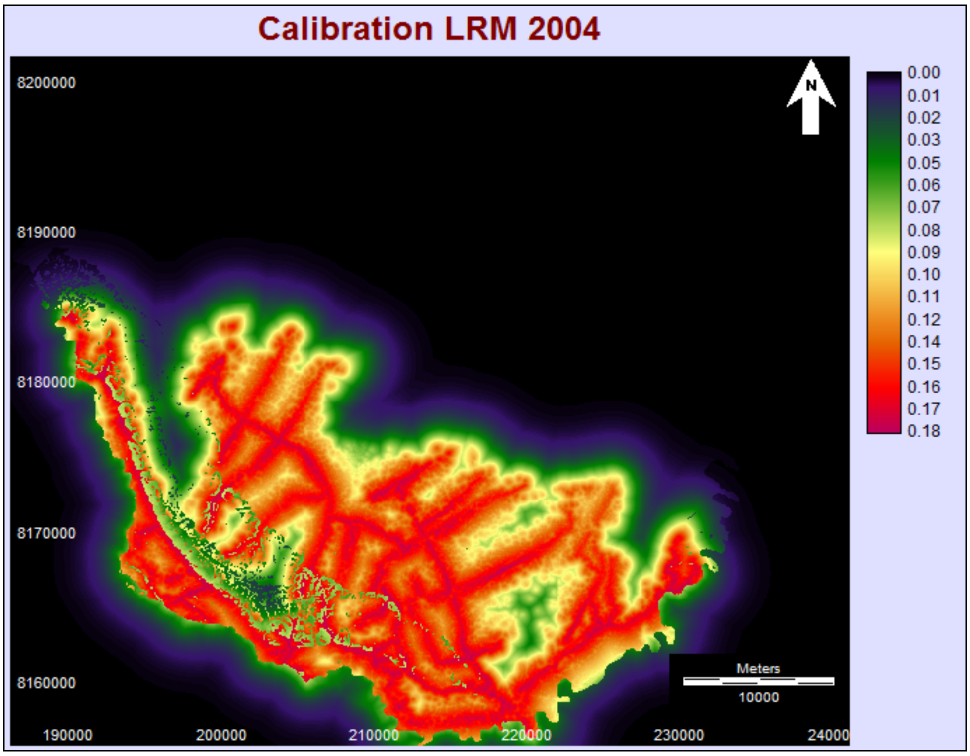


Figure 18 Map of Probabilities of Deforestation Obtained by LRM (Calibration 2004)

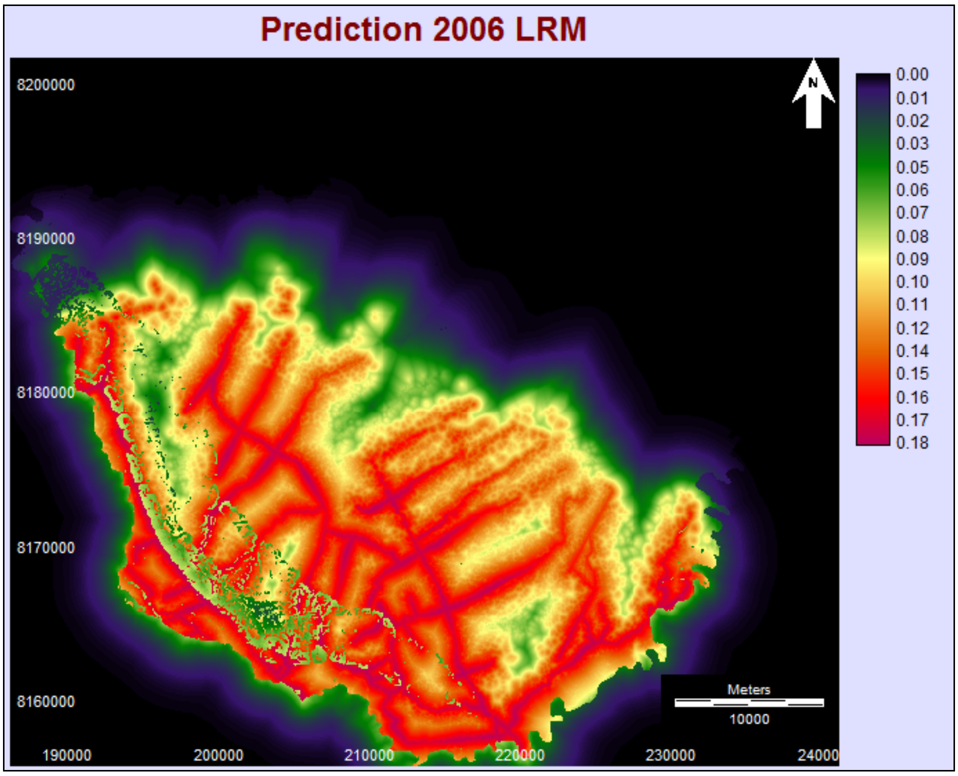


Figure 19 Map of Probabilities of Deforestation Obtained by LRM (Prediction 2006)

The methodological approach is described in Subsection 2.3.4.3

3.3.6. Multi-Layer Perceptron Model Results

Figure 20 and Figure 21 show the results of the calibration and the prediction of the MLP Model. The colour in the figures indicates the degree of probability of deforestation. Areas in red show high probability for forest conversion, while, areas in other colours have decreasing probability for deforestation. Black dots in the maps represent areas that are already deforested.

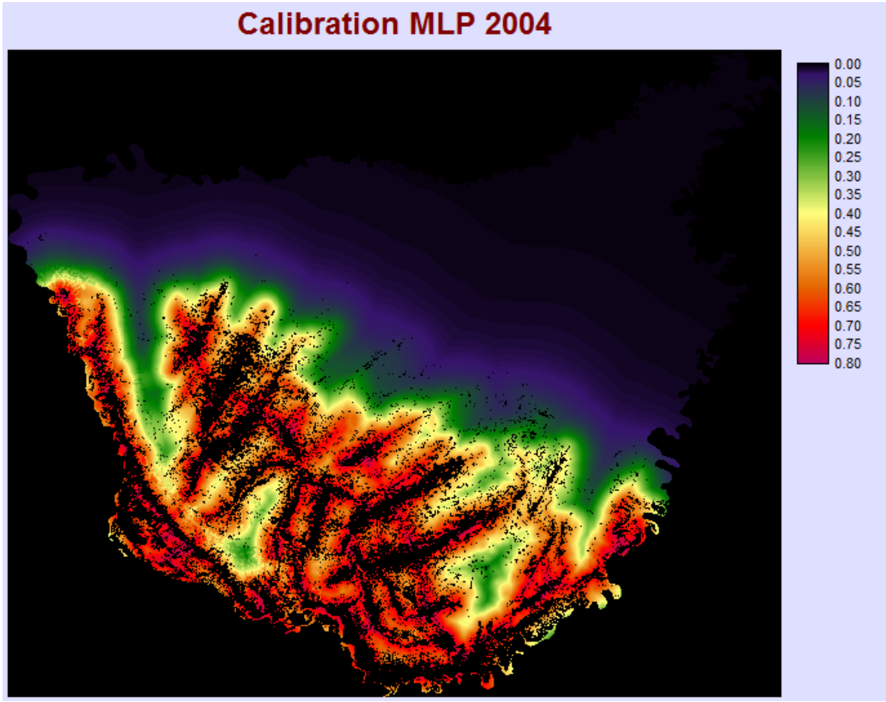


Figure 20 Map of Probabilities of Deforestation Obtained by MLP (Calibration 2004)

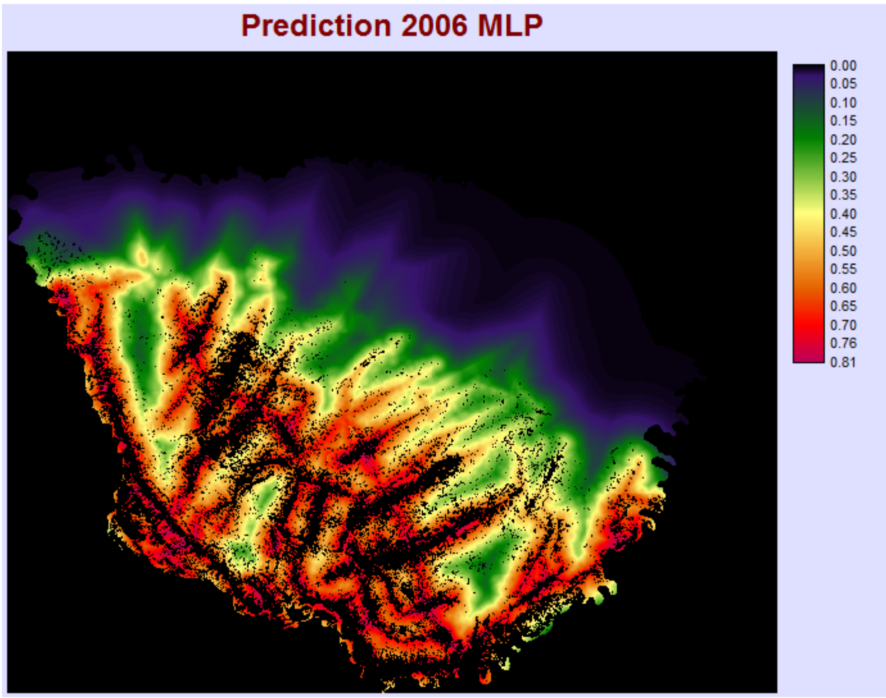


Figure 21 Map of Probabilities of Deforestation Obtained by MLP (Prediction 2006)

3.3.7. Validation for the Prediction for the Year 2006, Logistic Regression Model and Multi Layer Perceptron

Figure 22 illustrates the real change occurred for the period 2004 to 2006, areas in black are areas of change.

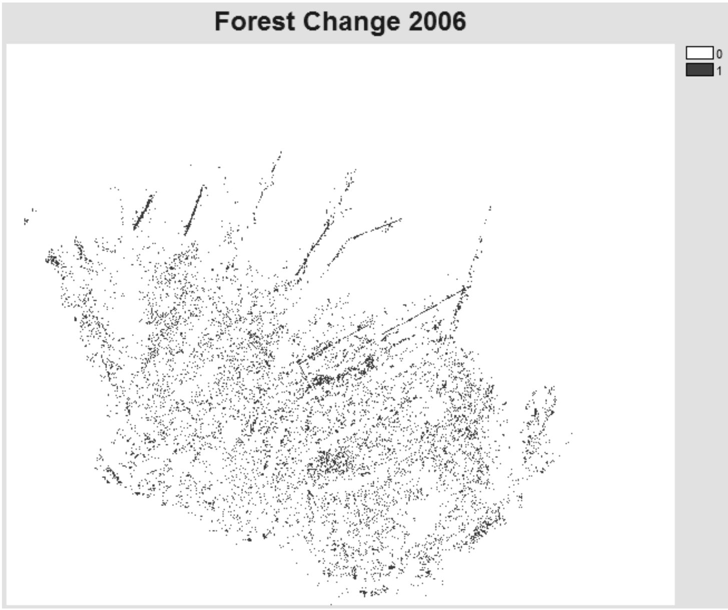


Figure 22 Forest Change Year 2006 (1 = Change; 0 = No change)

Figure 23 illustrates the AUC/ROC curve for the LRM. The Area Under the ROC Curve is 0.852 (Table in Appendix V), which gives an accuracy of 85%.

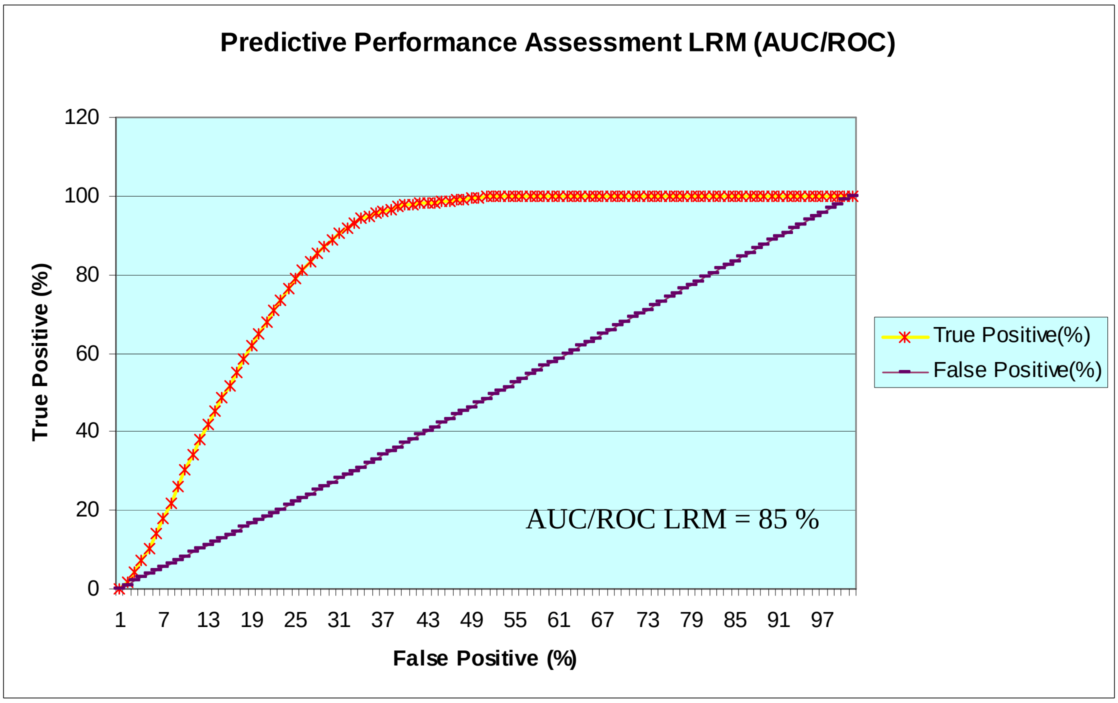


Figure 23 Predictive Performance Assessment LRM (AUC/ROC)

Figure 24 illustrates the AUC/ROC curve for the MLP. The Area Under the ROC Curve is 0.92 (Table in Appendix VI), which gives an accuracy of 92%.

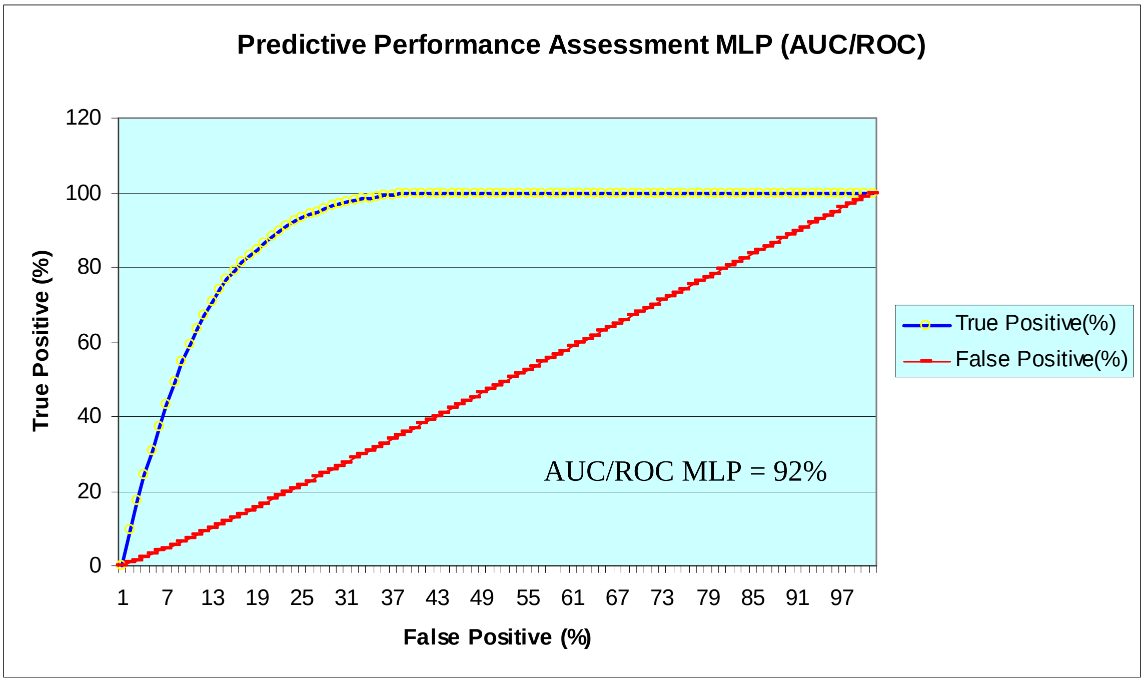


Figure 24 Predictive Performance Assessment MLP (AUC/ROC)

3.3.8. Comparison between LRM vs. MLP (AUC/ROC)

Figure 25 shows the difference between the ROC curves obtained by the two models: Logistic Regression and Multi-Layer Perceptron

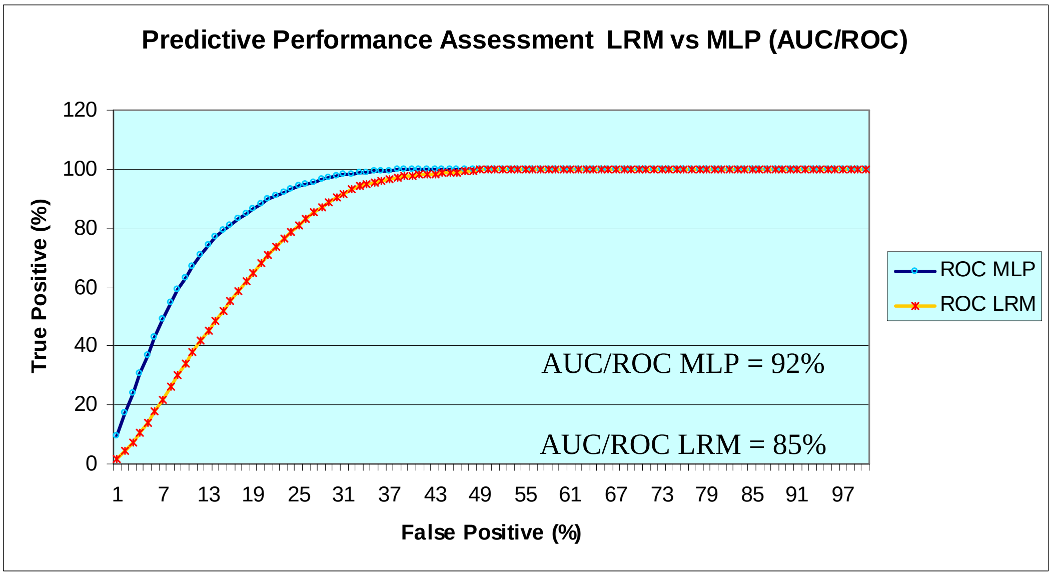


Figure 25 Predictive Performance Assessment LRM vs. MLP (AUC/ROC)

4. Discussions

Classification

Classification of the image for 2006 was acceptable with an overall accuracy of 81.83 % (see Table 4 Subsection 3.1.2). This accuracy was relatively good taking into account that only 39 points were sampled. In this study, the sampling designed was 'Cluster sampling' and it was expected that at least 80 points were sampled. That goal was impossible to reach because of the social conflicts caused by cocaleros in the TIPNIS. The acceptable accuracy of the classification even though only 39 points were used is perhaps due to the coarseness of the classified categories. The author of this study is aware that more samples were needed to have better confidence in the performed classification.

Forest Loss, Deforestation Rates and Deforestation Pattern

In the southern part of the TIPNIS, 23 %.a $^{-1}$ of Primary Forest has been lost from 1976 to 2006.
The average deforestation rate for the same period was 1.9 %.a $^{-1}$. However, the annual rate registered for the period 2001-2006 was 2.9%.a $^{-1}$ (Table 5 and Figure 13 Subsections 3.2.1 and 3.2.2). These rates are in the range obtained by other studies related to tropical deforestation in the country, which registered annual rates of 2.6 % for the north-eastern part [74] and 1.5%–3.1% for the central part [79].

The deforestation in the Chapare Province is linked to the cultivation of coca plant [7, 44]. In 1988, Law 1008 (the Regulation of Coca and Controlled Substances Law) became effective which made coca cultivation illegal in the Chapare Province as well as in the rest of Bolivia except for the Yungas region in La Paz [44, 52]. Introduction of the United States Agency for International Development (USAID) Program[4], and cooperation by the Bolivian Government reduced the coca cultivation until the middle of the 1990s [66]. In 2002, the Movement Towards Socialism (Movimiento al Socialismo MAS) political party led by the leader cocalero Evo Morales, obtained about 21% of representation in the National Congress. Their influence has introduced less control of coca cultivation and a result of this has been increased levels of forest clearing [44]. Thus, the variations in the deforestation rates observed in this study can be explained at least in part as a consequence of Government policies and the cultivation of coca.

A pattern of a large and growing number of small patches (0.5 up to 5 ha) of deforestation for each year of the study was observed (Figure 14 Subsection 3.2.3). This suggests that deforestation starts with small patches that grow in size over time. The particular pattern of small spots of cleared forest may be related to the cropping of coca (Erythroxylum coca), which, as explained above, is controlled by the Bolivian government. Cocaleros traditionally work around government controls. The crop has been and is still cultivated in a semi-hidden way a moderate distance from the roads, generally 250 to

[4] USAID's assistance program in Bolivia focuses on poverty reduction and supports the Government of Bolivia's National Development Plan (NDP). Efforts to increase business, agricultural, and trade opportunities for the poor, plus support for the sustainable use of natural resources, and assistance to farmers to produce alternatives to coca (USAID Bolivia, 2007 http://bolivia.usaid.gov/US/7Faq.htm).

750 m (Figure 15 Subsection 3.2.3). Figure 14 (Subsection 3.2.3) also shows that the small patches have increased since 2001, which coincides with the political strength of the 'cocaleros' representatives in the government.

Forest Conversion Modelling

The modelling of forest conversion considered five explanatory variables: "Distance from Forest Edge", "Distance from Roads", "Distance from Settlements", "Landscape Position" and "Type of Settlement".

The univariate test of association Cramer's V (Table7, Subsection 3.3.2.2), revealed that the three "Distance" variables had good association with forest change (V value between 0.5 and 0.6). "Landscape Position" variable showed a strong association with forest change only for the classes "Gully bed" and "Gentle slope" (0.96), the classes "Steep slope" and "Ridge Top" showed poor association (0.03 and 0.000).

On the other hand, in spite of the fact that the variable " Type of Settlement" has two classes that showed a strong association with forest change (V Cocaleros=0.98 and V Trinitarios=0.72), it has not been included as an explanatory variable in the model for two main reasons:

- Cocaleros clear forest ignoring the borders delimited by the managers of the park. As a result, deforestation is happening outside the area that the model would consider as Cocaleros territory, as well as inside.

- The class "Yuracares" showed a V value of 0.000 and p value of 1. According to Eastman 2006 [15], a high p value is a sign that the variable can be rejected. This is logical, since the model carried out by this study considers only potential variables that have shown that they have an effect on the deforestation. In fact, the "Yuracares" tribe live near to the rivers in the northern part of the study area, and because of their way of life, they do not cause deforestation.

If the results of Cramers' V are close to 1, the model assigns a higher probability of deforestation. If the results are close to 0, the model assigns a lower probability of deforestation. If the class "Yuracares" (with value 0) had been included in the model, it would have been assigned a very low probability of deforestation. As "Type of settlement' map (Figure 17 Subsection 3.3.4) illustrates, the "Yuracares" class occupies the northern part of the park, which at this point is still Primary Forest. This is where it is anticipated by the author that most future deforestation will occur. If values of 0 were assigned to this particular area, the model would not consider the area to have a potential for forest conversion. Such a conclusion would be incorrect and would diminish the model´s performance. To avoid this circumstance and because it would have been necessary to have the precise places where "Yuracares" live, information that was lacking, it was decided not to take into account this variable.

Cramer's V provided a quick view of the degree of association between the dependent variable and each explanatory variable. The test helped to evaluate whether the explanatory variable was worthy to be included in the model or not. This test is useful as a measure of strength of relationship [14]. And it has been applied successfully in other studies related to spatial predictions [45, 77]

In the Logistic Regression analysis (Subsection 3.3.5), six predictor sets were compared. The best-fitted predictor set was a combination of all the variables incorporated into the model. For this combination, the AUC was 89% and the AIC index was the lowest for the tested predictor sets.

“Landscape Position” was the best single predictor for forest change (2001 – 2004), with a β value of +1.27. This means that this variable has the strongest positive influence on forest change among the explanatory variables. The model assumes that the probability of deforestation is high in areas of “Gentle Slope” and “Gully Beds” and low in areas of “Steep Slope” and “Ridge Top”.

"Distance from Forest Edge” is the second best single predictor. It is negative (β= -0.34). This means that the probability of forest change decreases in direct proportion to the increase in distance from the borders. In other words, the model assigns higher values of probability of change to areas which are closer to the forest borders.

“Distance from Roads” and “Distance from Settlements” have the same negative value (β= -0.21 β= -0.27). The model assigns the similar significance to these two variables. The negative value means that the probability of forest change decreases in direct proportion to the increase in distance from roads and settlements. In other words, the model assigns higher values of probability of change to areas which are closer to roads and settlements.

“Landscape Position” is a static variable that remains constant because it does not change dynamically during short periods of time and it was considered as a constant for the prediction. The “Distance” variables are, on the contrary, very dynamic. Thus, these variables are the ones that will be used to calculate predictions for future years.

The findings obtained in this study are in contrast to the study carried out by Loza A. (2004) [41] in the Carrasco Province, which is next to the Chapare Province. Loza A. found that the topography was not a significant factor for forest conversion. In the TIPNIS, deforestation is less frequent in areas with steep slope and this study found that flat areas are strongly susceptible to forest conversion. The topography of Loza’s study area presents mostly hills (lower altitude) and flat areas. Mountains are located at the southern part, and, as the author mentions in page 58, elevation could play an important role. But the highest mountains are located in the southern part of the Carrasco Province and they are covered with clouds in the images used by Loza. The author also mentions that in these areas Closed Forest was well represented. Perhaps the study carried out by Loza would have had other results without the interference of the clouds.

The variables “Distance from Roads” and “Distance from Settlements” are significant factor for forest conversion in this study, as well as other researches [16, 20, 21, 27, 41, 80] have found, In the particular case of the deforestation in the TIPNIS, it is believed that first people (e.g., the cocaleros) settle land reached beyond existing roads and then they develop roads to reach the already taken lands. However, this is difficult to verify with the data and the analysis provided by this study.

The best-fitted predictor set obtained by the Logistic Regression analysis was used as input for the artificial neural network Multi-Layer Perceptron incorporated in IDRISI 15. MLP does not provide information about the relative significance of the explanatory variables on the forest change. MLP only performs the prediction based on the algorithm Back Forward Propagation. The user of the MLP in IDRISI can modify basic parameters such as network topology, training parameter, and stopping criteria. For the purpose of this study, it was advisable to keep the default parameters (Figure 10). The critical point was to assign the number of samples to train the neural network. This was determined by trial and error, and finally with an accuracy of training of 84.3%, the model was calibrated. The weights assigned to the calibrated model are expressed in a specific format (*.bpn) that could be only visualized in IDRISI as a matrix of byte values for each sampled pixel. This means that data were not easy to interpret. Other studies which had worked with neural networks, even in a more complex way than this study, report the same limitation [17, 47]

During the comparison of prediction performance (Subsection 3.3.7), the LRM prediction performed from the year 2004 to the year 2006, obtained an AUC of 85.2%. The AUC for the same period obtained by the MLP model was 92%.

5. Conclusions

During the period 1976 – 2006, 23% of Primary Forest has been lost in the southern part of the TIPNIS. The deforestation rates presented variations, 1.3% until 1991, 0.5% until 2001 and 2.9 until 2006. These variations coincide with the degree of control of coca cultivation that the Bolivian government has permitted for the same periods in the Chapare Province. When government controls of coca growing were more lax, the deforestation rates increased. This suggests that the main reason of forest conversion in the park has been the cropping of coca.

The clearing of the forest also follows a particular spatial pattern. For each year of observation, an increasing number of small patches appear in the middle of the forest near to the roads or near to already deforested areas. This also could be related to the coca cultivation since the cocaleros traditionally avoid government controls.

The modelling was successful in predicting forest change in the TIPNIS from 2004 to 2006.

The best predictor set obtained by the LRM was composed of the explanatory variables: “Distance from Forest Edge”, “Distance from Roads”, “Distance from Settlements” and “Landscape Position”. The variable “Type of Settlement” was not included in the modelling of forest change because doing so could have led to contradictory performance of the predictive models.

The predictive performance of both models proved successful. While MLP produces better prediction results in general (AUC=92 %). Logistic regression analysis is still needed to understand the relative significance of the explanatory variables on the forest change.

The predictor set selected in this study is currently able to perform reliable predictions of forest change by updating the dynamic variables “Distance from Forest Edge”, “Distance from Roads” and “Distance from Settlements”.

This study worked with explanatory variables that were able to be represented spatially (maps, etc.). While this study can use models to deal effectively with spatial factors, deeper analysis of the forest conversion in the TIPNIS requires factoring in more elements, like socioeconomic data or more detailed information about the explanatory variables.

Suggestions for Further Studies

TIPNIS has threats such as the construction of a road across the area and the granting of oil exploration concessions. The processes followed in this study could be used with other explanatory variables to create scenarios of areas of forest change where particular wild life and/or endemic species could be affected.

The aim of this thesis was to predict probabilities of forest conversion. However, areas of change (not only probabilities) can be predicted by incorporation of methods such as Markov Chains and Cellular automata.

While this study considered only two categories, Forest and Disturbed Forest, further studies could model additional categories of land cover.

This study did not consider the variable “Type of Settlement” because of the poor quality of the available input. Better results can be expected if the model is tested with better quality data for this variable.

6. References

1. Agarwal, D.K.; Silander, J.J.A.; Gelfand, A.E.; Dewar, R.E.; & Mickelson, J.J.G.. 2005. Tropical deforestation in Madagascar: analysis using hierarchical; spatially explicit; Bayesian regression models. Ecological Modelling 185:105-131
2. Armenteras D.; Rudas G.; Rodriguez N.; Sua S.; & Romero M. 2006. Patterns and causes of deforestation in the Colombian Amazon. Ecological Indicators 6:353-368
3. Ayalew L. & Yamagishi H. 2005. The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains; Central Japan. Geomorphology 65:15-31
4. Ayalew L.; Yamagisehi H.; Marui H.; & Kanno T. 2005. Landslides in Sado Island of Japan: Part II. GIS-based susceptibility mapping with comparisons of results from two methods and verifications. Engineering Geology 81:432-445
5. Birt A.G.; Wang L.; & Xi W. 2006. LANDISVIEW Beta v1.0 Software. Engineering Laboratory at Texas A&M University. http://kelab.tamu.edu/standard/restoration/restoration_tools.htm, accessed November 2008
6. Borreno R. 2008. Linking spatial patterns of land use to agents of deforestation in the Brazilian Amazon. Thèse présentée en vue de l'obtention du grade de Docteur en Sciences. Université Catholique de Louvain: pp. 22-23
7. Bradley A.V. & Millington A.C. 2008. Coca and colonists: quantifying and explaining forest clearance under coca and anti-narcotics policy regimes. Ecology and Society 13:1-31
8. Brearley F.Q. 2007. Tropical disturbance and recovery. Frontiers in Ecology and the Environment. The Ecological Society of America. Write back. p 2. http://www.csu.edu.au/faculty/science/ses/staff/watson_papers/Watson_Beyond_impact_factors.pdf, accessed February 2009.
9. Bürgi M.; Hersperger A.; & Schneeberger N. 2005. Driving forces of landscape change - current and new directions. Landscape Ecology 19:857-868
10. Butler R. 2007. Types of rainforests. http://rainforests.mongabay.com/0103.htm, accessed August 2008
11. CGIAR-CSI. 2008. SRTM 90m Digital Elevation Data. Consortium for Spatial Information (CGIAR-CSI) http://srtm.csi.cgiar.org/, accessed on January 2008.
12. CLAS. 2007. Landsat TM 5 images. Years 1986 and 1996. Centro de Levantamientos Aeroespaciales y Aplicaciones SIG para el desarrollo sostenible de los recursos naturales. Universidad Mayor de San Simón. Cochabamba; Bolivia.
13. Couturier S.; Gastellu-Etchegorry J.P.; Patiño P.; & Martin E. 2008. A model-based performance test for forest classifiers on remote-sensing imagery. Forest Ecology and Management 257:23-37
14. De Vaus D. 2002. Surveys in Social Research. Routledge. Series Editor: Martin Bulmer.
15. Eastman J.R. 2006. IDRISI 15 Andes. Guide to GIS and image processing Clark University; Worcester; MA 01610 - 1477 USA.
16. Echeverria C.; Coomes D.A.; Hall M.; & Newton A.C. 2008. Spatially explicit models to analyze forest loss and fragmentation between 1976 and 2020 in southern Chile. Ecological Modelling 212:439-449

17. Eftekhar B.; Mohammad K.; Ardebili H.; Ghodsi M.; & Ketabchi E. 2005. Comparison of artificial neural network and logistic regression models for prediction of mortality in head trauma based on initial clinical data. BMC Medical Informatics and Decision Making 5:3
18. ERSDAC. 2003. ASTER User's Guide. Earth Remote Sensing Data Analyis Center (ERSDAC). http://www.science.aster.ersdac.or.jp/en/index.html. Accessed January 2009
19. Etter A.; McAlpine C.; Phinn S.; Pullar D.; & Possingham H. 2006. Unplanned land clearing of Colombian rainforests: Spreading like disease? Landscape and Urban Planning 77:240-254
20. Etter A.; McAlpine C.; Pullar D.; & Possingham H. 2006. Modelling the conversion of Colombian lowland ecosystems since 1940: Drivers; patterns and rates. Journal of Environmental Management 79:74-87
21. Etter A.; McAlpine C.; Wilson K.; Phinn S.; & Possingham H. 2006. Regional patterns of agricultural land use and deforestation in Colombia. Agriculture, Ecosystems & Environment 114:369-386
22. FAO. & CETEFOR. 2002. Análisis multi-temporal de imágenes satelitales para la estimación de pérdida de cobertura forestal primaria y evaluación del cambio de uso en el bosque de uso múltiple del trópico de Cochabamba.
23. FAO. & FRAa. 2005. Global Tables. Food and Agriculture Organization of the United Nations. FAO; Global Forest Resources Assessment. FRA. http://www.fao.org/forestry/32040/en/, accessed November 2008
24. FAO. & FRAb. 2004. Global forest resources assessment (update 2005). Terms and definitions.(Final version). Food and Agriculture Organization of the United Nations. FAO; Global Forest Resources Assessment. FRA. http://www.fao.org/docrep/007/ae156e/AE156E00.HTM, accessed December 2008
25. FAO. 1996. Forest resources assessment. Survey of tropical forest cover and study of change processes. Food and Agriculture Organization of the United Nations. Number 130; Rome: FAO
26. FAO. 2002. Second expert meeting on harmonizing forest related definitions for use by various stakeholders. FAO Corporate Document Repository. Food and Agriculture Organization of the United Nations. http://www.fao.org/docrep/005/y4171e/Y4171E00.HTM, accessed December 2008
27. Geist H.J. & Lambin E.F. 2001. What drives tropical deforestation? A meta-analysis of proximate and underlying causes based on sub-national case study evidence. LUCC Report Series No. 4. CIACO; Louvain-la-Neuve.
28. Gong P. 1997. Remote sensing and image analysis. University of California at Berkeley
29. Granger C.W.J. 2007. Modelling Amazon Deforestation for Policy Purposes. Insights. Melbourne Economics and Commerce 2
30. Hu Z. & Lo C.P. 2007. Modeling urban growth in Atlanta using logistic regression. Computers; Environment and Urban Systems 31:667-688
31. Hubert-Moy L.; Cotonnec A.; Le Du L.; Chardin A.; & Perez P. 2001. A Comparison of Parametric Classification Procedures of Remotely Sensed Data Applied on Different Landscape Units. Remote Sensing of Environment 75:174-187
32. IDRISI. 2006. IDRISI 15 Andes Edition. Clark Labs; Clark University 950 Main Street; Worcester MA 01610-1477 USA
33. ILWIS. 2007. ILWIS Help: Principal Component Analysis. ITC. International Institute for Geo-Information Science and Earth Observation. The Netherlands

34. ITC. 2005. Exercises instruction book. Natural Resources Management RS and GIS. International Institute for Geo-Information Science and Earth Observation. The Netherlands
35. ITC. 2007. Exercises instruction book. Natural Resources Management. Modules 3&4 RS and RS/GIS 2007 Geo referencing exercise, Westinga E. ITC. International Institute for Geo-Information Science and Earth Observation. The Netherlands
36. Jennes J. 2006. Topographic Position Index. tpi_jen.avx, extension for ArcView 3.x; v. 1.3a. Jenness Enterprises. Available at: http://www.jennessent.com/arcview/tpi.htm, accessed December 2008
37. Kaimowitz D. & Angelsen A. 1998. Economic Models of Tropical Deforestation; a review. Center for 1nternational Forestry Research. C1FOR. Bogor Indonesia
38. Lambin E.F. 1994. Modelling deforestation processes; a review. EUR 15744 EN; TREES series B: Research Report No. 1. Joint Research Centre; Institute for Remote Sensing Applications; European Space Agency; Luxembourg; Office for Official Publications of the European Community. p128
39. Liebetrau A.M. 1983. Measures of association Newbury Park; CA: Sage Publications. Quantitative Applications in the Social Sciences Series No. 32.
40. Lippitt C.D.; Rogan J; Toledano J; Sangermano F; Eastman JR; Mastro V; Sawyer A. 2008. Incorporating anthropogenic variables into a species distribution model to map gypsy moth risk. Ecological Modelling 210:339-350
41. Loza A.V. 2004. Spatial logistic model for tropical forest conversion : a case study of Carrasco province; 1986 - 2002; Bolivia. In: ITC; Enschede; p 74
42. Ludeke A.K.; Maggio R.C.; & Reid L.M. 1990. An analysis of anthropogenic deforestation using logistic regression and GIS. Journal of Environmental Management 31:247-259
43. Mahiny A.S. & Turner B.J. 2003. Modeling past vegetation change through remote sensing and G.I.S: A Comparison of Neural Networks and Logistic Regression Methods. School of Resources; Environment and Society; The Australian National University; Australia.
44. Manzano N. 2007. Impacto ambiental y sostenibilidad de la producción de coca y cocaína en la región del Chapare; Bolivia. Instituto de Estudios Sociales y Económicos. IESE. Universidad Mayor de San Simón. Cochabamba; Bolivia.
45. Maria de Almeida C.; Batty M.; Vieira Monteiro A.M.; Câmara G.; Soares-Filho B.S.; Cerqueira G.C.; & Pennachin C.L. 2003. Stochastic cellular automata modeling of urban land use dynamics: empirical development and estimation. Computers; Environment and Urban Systems 27:481-509
46. Mas J.F. 2004. Mapping land use/cover in a tropical coastal area using satellite sensor data, GIS and artificial neural networks. Estuarine, Coastal and Shelf Science 59:219-230
47. Mas J.F.; Puig H.; Palacio J.L.; & Sosa-López A. 2004. Modelling deforestation using GIS and artificial neural networks. Environmental Modelling & Software 19:461-471
48. McConnell W.J.; Sweeney S.P.; & Mulley B. 2004. Physical and social access to land: spatio-temporal patterns of agricultural expansion in Madagascar. Agriculture; Ecosystems & Environment 101:171-184
49. McDonald R.I. & Urban D.L. 2006. Spatially varying rules of landscape change: lessons from a case study. Landscape and Urban Planning 74:7-20
50. Mcgarigal K. & Marks B.J. 1995. FRAGSTATS: spatial pattern analysis program for quantifying landscape structure. Gen. Tech. Re PNW-GTR-351. U.S. Department of Agriculture; Forest Service; Pacific Northwest Research Station; Portland; OR; USA.

51. Mertens B. & Lambin E.F. 2000. Land-Cover-Change trajectories in Southern Cameroon Annals of the Association of American Geographers 90: 467-494.
52. Millington A.C. & Jepson W. 2008. Land Change Science in the Tropics. Agricultural land-use trajectories in a cocaine source region: Chapare; Bolivia.
53. MindSites Group L. 2008. http://imaging.geocomm.com/features/sensor/landsat7/ GeoCommunity
54. Moreno N.; Quintero R.; Ablan M.; Barros R.; Dávila J.; Ramírez H.; Tonella G.; & Acevedo M.F. 2007. Biocomplexity of deforestation in the Caparo tropical forest reserve in Venezuela: An integrated multi-agent and cellular automata model. Environmental Modelling & Software 22:664-673
55. Müller D. & Mburu J. 2008. Forecasting hotspots of forest clearing in Kakamega Forest; Western Kenya. Forest Ecology and Management 257:968-977
56. Nagendra H.; Munroe D.K.; & Southworth J. 2004. From pattern to process: landscape fragmentation and the analysis of land use/land cover change. Agriculture; Ecosystems & Environment 101:111-115
57. Nefeslioglu H.A.; Gokceoglu C.; & Sonmez H. 2008. An assessment on the use of logistic regression and artificial neural networks with different sampling strategies for the preparation of landslide susceptibility maps. Engineering Geology 97:171-191
58. CEOYale. 2007. ASTER. The Center for Earth Observation, Yale University. http://www.yale.edu/ceo/Documentation/ASTER.pdf, accessed September 2008
59. Orzolini V. 2008. Isiboro Sécure National Park and Indigenous Territory. TIPNIS. Manager; June September 2008. Personal communication.
60. Parkswatch. 2004. http://www.parkswatch.org/main.php?l=eng Access date: June 2008
61. Pijanowski B.C.; Brown D.G.; Shellito B.A.; & Manik G.A. 2002. Using neural networks and GIS to forecast land use changes: a Land Transformation Model. Computers; Environment and Urban Systems 26:553-575
62. Pontius R.G. & Schneider L.C. 2001. Land-cover change model validation by an ROC method for the Ipswich watershed; Massachusetts; USA. Agriculture; Ecosystems & Environment 85:239-248
63. Pontius R.G.; Shusas E.; & McEachern M. 2004. Detecting important categorical land changes while accounting for persistence. Agriculture; Ecosystems & Environment 101:251-268
64. Pontius R.G. & Spencer J. 2005. Uncertainty in extrapolations of predictive land-change models. Environment and Planning B-Planning & Design 32:211-230
65. Puyravaud J.P. 2003. Standardizing the calculation of the annual rate of deforestation. Forest Ecology and Management 177:593-596
66. Ribando C.M. 2007. Bolivia: Political and Economic Developments. Analyst in Latin American Affairs. Foreign Affairs. Defence; and Trade Division. USA
67. Ritter N. 2006. Biodiversity and Phytogeography of Bolivia's Wetland Flora. Doctoral dissertation. University of New Hampshire. United States: pp 56
68. Rogan J.; Franklin J.; Stow D.; Miller J.; Woodcock C.; & Roberts D. 2008. Mapping land-cover modifications over large areas: A comparison of machine learning algorithms. Remote Sensing of Environment 112:2272-2283
69. Rossiter D.G. & Loza A. 2008. Technical Note: Analyzing land cover change with logistic regression in R. Technical Report ITC; Enschede; NL. Version 2.01; 19-July-2008. 69 pp. First version April 2004.
70. Rudel T. & Roper J. 1996. Regional patterns and historical trends in tropical deforestation; 1976–1990: a qualitative comparative analysis. Ambio 25; pp. 160–166.

71. Schneider L.C. & Pontius R.G. 2001. Modeling land-use change in the Ipswich watershed; Massachusetts; USA. Agriculture; Ecosystems & Environment 85:83-94
72. SERNAP. 2004. Diagnóstico Integral Preliminar. Territorio Indígena y Parque Nacional Isiboro Sécure; TIPNIS ; Servicio Nacional de Areas Protegidas. Cochabamba; Bolivia.
73. SERNAP. 2006. Servicio Nacional de Áreas Protegidas. SERNAP. Protected areas national service Bolivian Government. http://www.sernap.gov.bo/
74. Steininger M.K.; Tucker C.J.; Townshend J.R.G.; Killeen T.J.; Desch A.; Bell V.; & Ersts P. 2001. Tropical deforestation in the Bolivian Amazon. Environmental Conservation 28:127-134
75. USGS. 2008. Aster Images years 2001; 2004; 2006. United States Geological Survey http://glovis.usgs.gov/, accessed August 2008
76. USGS. 2008. U.S. Geological Survey http://landsat.usgs.gov/tools_faq.php?id=6, accessed August 2008
77. Van Den Eeckhaut M.; Vanwalleghem T.; Poesen J.; Govers G.; Verstraeten G.; & Vandekerckhove L. 2006. Prediction of landslide susceptibility using rare events logistic regression: A case-study in the Flemish Ardennes. Belgium. Geomorphology 76:392-410
78. Van Gils H.; Batsukh O.; Rossiter D.; Munthali W.; & Liberatoscioli E. 2009. Forecasting the pattern and pace of Fagus forest expansion in Majella National Park, Italy. Applied Vegetation Science 11:539-546
79. Van Gils H. & Loza A.V. 2006. What drives conversion of tropical forest in Carrasco province, Bolivia?. Ambio: A journal of the human environment 35: 81-85
80. Vanclay J.K.; Kaimowitz D.; Puntodewo A.; & Mendez P. 1999. Spatially explicit model of deforestation in Bolivia. Data Management and Modelling using Remote Sensing and GIS for Tropical Forest Land Inventory. Rodeo; Jakarta; 82: 371-373
81. Westburn. 2002. Westburn Publishers Ltd; The Westburn Dictionary of Marketing edited by Michael J Baker. www.themarketingdictionary.com. accessed August 2008

Appendices

Appendix I. Field Work

Ia. Field Data Sheet Form

The Field Data Sheet Form was developed to document land use and land cover information and identify their GPS coordinates in a systematic way.

LAND USE/COVER AND GPS SURVEY			
Place:	ID point:		Transect:
DATE: September / / 2008			
UTM COORDINATES	X:		Y:
LAND FORM	Plain		Observations:
	Hill		
	Piedmont		
	Terrace		
LAND USE/COVER Line interceptor: field measurements	Closed Forest		
	Open Forest:		
	Natural open forest		
	Anthropic open forest:		Signs of crops
			Signs of logging
			Signs of slash or burn
			Other indicator:
	Crops		Type of crop
	Village		Observations:
	Bare soil		
	Lagoon or river		
	Grassland		
	Pasture:		
	Fences		
	Cattle prints		
	Cattle dung		
	Cattle presence		
	Other:		
COVER (%)			
FOREST USES	Logging		Type of wood:
	Secondary products		Type of product:

m		m		m		m	
0		0		0		0	
1		1		1		1	
2		2		2		2	
3		3		3		3	
4		4		4		4	
5		5		5		5	
6		6		6		6	
7		7		7		7	
8		8		8		8	
9		9		9		9	
10		10		10		10	
11		11		11		11	
12		12		12		12	
13		13		13		13	
14		14		14		14	
15		15		15		15	
16		16		16		16	
17		17		17		17	
18		18		18		18	
19		19		19		19	
20		20		20		20	
21		21		21		21	
22		22		22		22	
23		23		23		23	
24		24		24		24	
25		25		25		25	
	25		50		75		100

Ib. Ground Control Points: Collected Data Table

This table contains descriptions of the Ground Control Points measured during the field work.

ID	X	Y	Landscape	Data cover field work	Classification	Interceptor %
C1	212520.1	8156186	Plain	Pastos, Arbustos, arboles dispersos	Grassland	< 10
C2	213270.1	8157176	Plain	Pastos, Arbustos, arboles dispersos	Grassland	< 10
C3	212760.1	8157251	Plain	Pastos, Arbustos, arboles dispersos	Grassland	< 10
C4	213615.1	8157296	Plain	Pastos, Arbustos, arboles dispersos	Grassland	< 10
C5	216516.6	8158031	Plain	Ambaibo Macho, Ambaibo Hembra,	Secondary Forest	44
C6	210300.1	8159636	Plain	Pastos, Arbustos, arboles dispersos	Grassland	< 10
C7	214039.8	8160247	Plain	Ambaibo Hembra, Balsa, Pacay	Secondary Forest	55
C8	219900.1	8161181	Plain	Balsa, Pacay, gomerilla	Secondary Forest	51
C9	202575.1	8162306	Plain	Mandioca, Citricos, Coca ; Ganaderia	Crops and Cattle pastures	
C10	203970.1	8162321	Plain	Ambaibo Hembra, Balsa, Pacay	Secondary Forest	50
C11	206190.1	8162396	Plain	Citricos, Mandioca, Coca ; Ganaderia	Crops and Cattle pastures	
C12	209475.1	8164826	Plain	Ambaibo Hembra, Balsa, Pacay	Secondary Forest	34
C13	206940.1	8164961	Hill slope	Verdolago, Almendrillo, Pachiuba, Gabetillo, Cedro, Sangre de Toro, Bibosi, Ochoo, Blanquillo,	Forest	85
C14	202215.1	8166311	Hill slope	Mapajo, Mara o Caoba , Ochoo, Blanquillo, Coquino, Bibosi, Almendrillo, Pachiuba Chonta Ocoro	Forest	72
C15	214410.1	8166326	Plain	Mandioca, Arroz, Coca; Ganaderia	Crops and Cattle pastures	
C16	203850.1	8167286	Hill slope	Verdolago, Sangre de Toro Ocoro Pachiuba, Coquino, Bibosi, Almendrillo Peloto	Forest	68
C17	207270.1	8167301	Plain	Plátano, Citricos, Mandioca, Coca	Crops	
C18	195300.1	8167346	Plain	Plátano, Cítricos, Mandioca, Coca; Ganadería	Crops and Cattle pastures	
C19	218670.1	8167826	Plain	Mandioca, Arroz, Coca	Crops	
C20	216960.1	8167886	Plain	Gomerilla Ambaibo Hembra, Balsa, Pacay	Secondary Forest	41
C21	203940.1	8168186	Hill slope	Verdolago, Ocoro, Cedro Peloto, Piraquina, Pachiuba, Ochoo, Blanquillo, Coquino, Gabetillo, Bibosi	Forest	85
C22	200355.1	8169236	Plain	Ambaibo Macho, Ambaibo Hembra,gomerillo	Secondary Forest	42
C23	215490.1	8170706	Plain	Pacay, Ambaibo Macho, Ambaibo Hembra,	Secondary Forest	46
C24	212400.1	8171336	Plain	Verdolago, Bibosi, Sangre de Toro, Coloradillo, Peloto Mapajo, Palo María, Pachiuba, Ochoo	Forest	90
C25	206520.1	8171846	Plain	Citricos, Mandioca, Coca; Ganaderia	Crops and Cattle pastures	
C26	213645.1	8172191	Plain	Verdolago, Cedro, Chonta, Yesquero, Mapajo, Palo María, Almendrillo, Pachiuba, Gabetillo, Cedro	Forest	81
C27	204135.1	8172401	Plain	Pastos, Arbustos, arboles dispersos	Grassland	< 10
C28	213360.1	8173901	Plain	Cedro, Chonta, Yesquero, Mapajo, Palo María, Almendrillo, Pachiuba, Gabetillo, Bibosi, Piraquina	Forest	86
C29	204525.1	8173976	Plain	Ochoo, Mapajo, , Blanquillo, Coquino, Bibosi, Almendrillo, Pachiuba, Chonta, Ocoro, Gabetillo,	Forest	79
C30	211200.1	8173991	Plain	Plátano, Mandioca, Maiz, Coca	Crops	
C31	201780.1	8174306	Plain	Mandioca, Citricos, Coca	Crops	
C32	209940.1	8174846	Plain	Bibosi, Coloradillo, Peloto Mapajo, Palo María, Pachiuba, Ochoo, Blanquillo, Coquino, Verdolago	Forest	81
C33	194010.1	8175461	Hill slope	Mandioca, Arroz, Cítricos, Coca: Ganadería	Crops and Cattle pastures	
C34	209640.1	8176061	Plain	Ambaibo Macho, Ambaibo Hembra, Balsa, Pacay	Secondary Forest	41
C35	203025.1	8176361	Plain	Citricos, Mandioca, Maiz, Coca	Crops	
C36	202080.1	8176481	Plain	Pastos, Arbustos, arboles dispersos	Grassland	< 10
C37	197948.9	8176786	Hill slope	Ochoo, Verdolago, Cedro, Mapajo, , Blanquillo, Coquino, Bibosi, Almendrillo, Pachiuba, Ocoro	Forest	80
C38	198165.7	8176810	Plain	Bibosi, Coloradillo, Peloto Mapajo, Palo María, Pachiuba, Ochoo, Blanquillo, Coquino, Verdolago	Forest	83
C39	204090.1	8177996	Plain	Mandioca, Maíz, Plátano, Coca	Crops	

Ic. Field Work Pictures

Pictures of land cover categories in the TIPNIS

P1. Primary Forest	P2 Secondary Forest
	
P3 Crops (Coca)	**P4 Cattle Pastures**
	
P5 Line interceptor1	**Line interceptor2**
	

Appendix II. CHI SQUARE "Landscape Position" and "Type of Settlement"

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Forest conversion * Landscape position	6389926	100.0%	0	.0%	6389926	100.0%

Forest conversion * Landscape position Crosstabulation

			Landscape position				Total
			Steep slope	Gullybed	Ridgetop	Gentle slope	
Forest conversion	Change	Count	26582	628	4614	356710	388534
		% within Forest conversion	6.8%	.2%	1.2%	91.8%	100.0%
		% within Landscape position	13.2%	7.7%	14.2%	5.8%	6.1%
		% of Total	.4%	.0%	.1%	5.6%	6.1%
	No change	Count	174305	7562	27968	5791557	6001392
		% within Forest conversion	2.9%	.1%	.5%	96.5%	100.0%
		% within Landscape position	86.8%	92.3%	85.8%	94.2%	93.9%
		% of Total	2.7%	.1%	.4%	90.6%	93.9%
Total		Count	200887	8190	32582	6148267	6389926
		% within Forest conversion	3.1%	.1%	.5%	96.2%	100.0%
		% within Landscape position	100.0%	100.0%	100.0%	100.0%	100.0%
		% of Total	3.1%	.1%	.5%	96.2%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	22590.557[a]	3	.000
Likelihood Ratio	17378.926	3	.000
Linear-by-Linear Association	20502.017	1	.000
N of Valid Cases	6389926		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 497.99.

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Forest conversion * Type of settlement	6389847	100.0%	0	.0%	6389847	100.0%

Forest conversion * Type of settlement Crosstabulation

			Type of settlement			Total
			Yuracare	Encroachment	Trinitarios	
Forest conversion	Change	Count	31191	289656	67973	388820
		% within Forest conversion	8.0%	74.5%	17.5%	100.0%
		% within Type of settlement	.9%	11.1%	18.2%	6.1%
		% of Total	.5%	4.5%	1.1%	6.1%
	No change	Count	3368258	2327597	305172	6001027
		% within Forest conversion	56.1%	38.8%	5.1%	100.0%
		% within Type of settlement	99.1%	88.9%	81.8%	93.9%
		% of Total	52.7%	36.4%	4.8%	93.9%
Total		Count	3399449	2617253	373145	6389847
		% within Forest conversion	53.2%	41.0%	5.8%	100.0%
		% within Type of settlement	100.0%	100.0%	100.0%	100.0%
		% of Total	53.2%	41.0%	5.8%	100.0%

Chi-Square Tests

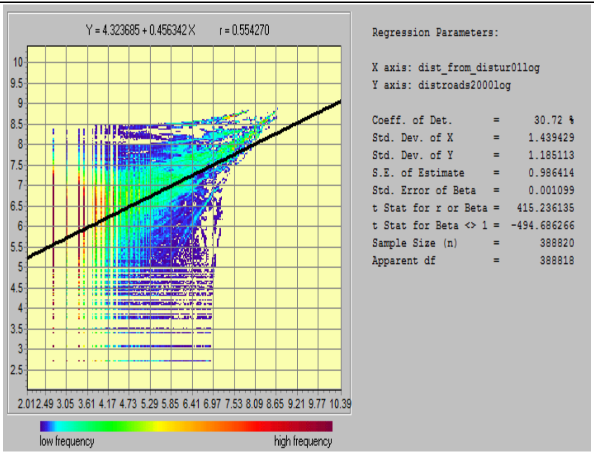
	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	368619.502[a]	2	.000
Likelihood Ratio	400219.542	2	.000
Linear-by-Linear Association	365120.221	1	.000
N of Valid Cases	6389847		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 22705.75.

Appendix III Correlation Analysis between Explanatory Variables Results

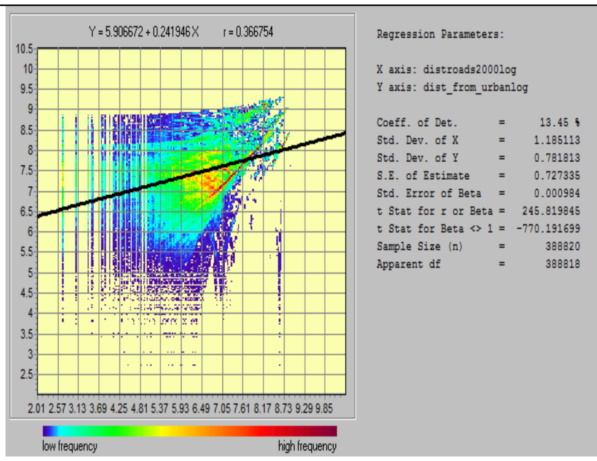
Distance from forest edge and Distance from roads

Correlation Coefficient = 0.55



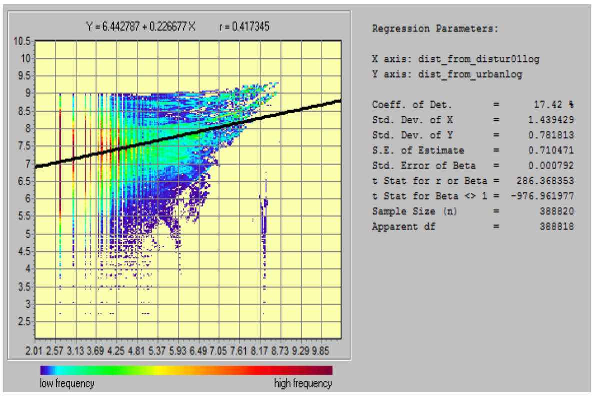
Distance from forest edge and Distance from roads

Correlation Coefficient = 0.37



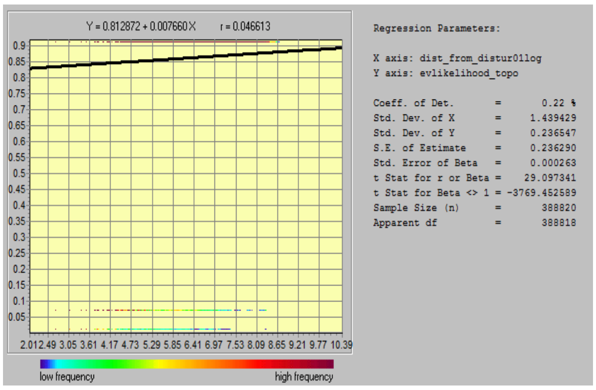
Distance from forest edge and Distance from settlement

Correlation Coefficient = 0.42



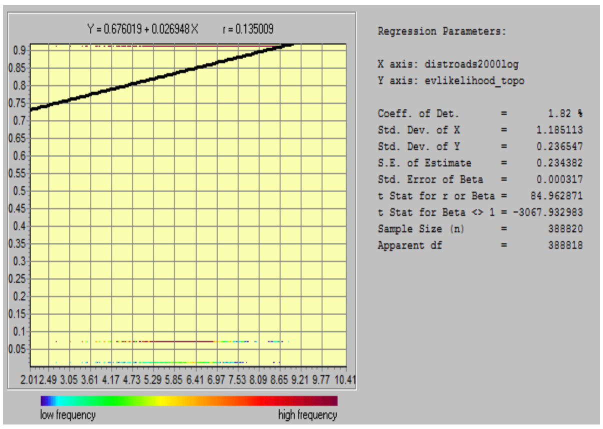
Distance from forest edge and Landscape Position

Correlation Coefficient = 0.05



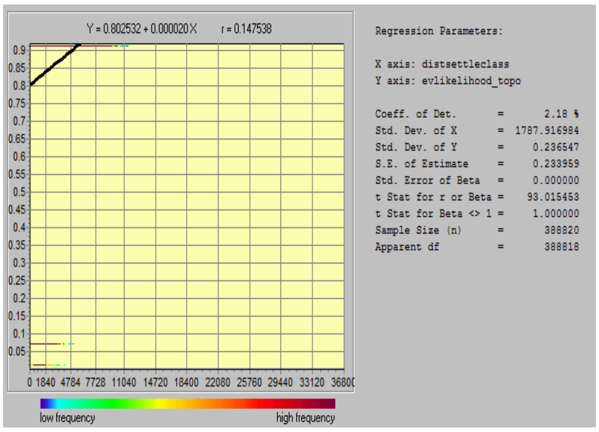
Distance from forest roads and Landscape Position

Correlation Coefficient = 0.13



Distance from settlements and Landscape Position

Correlation Coefficient= 0.15



Appendix IV Logistic Regression Statistics (best fitted predictor set M4)

```
Logistic Regression Results:

    Regression Equation :

     logit(change0104) = 1.5483 - 0.341876*dist_from_disturlog - 0.211048*dist_from_roadslog - 0.275174*dist_from_urbanlog + 1.269546*evlikelihood_topo

    Individual Regression Coefficient
--------------------------------------------------
Variables             |    Coefficient
--------------------------------------------------
Intercept             |    1.54825200
dist_from_disturlog   |    -0.34187598
dist_from_roadslog    |    -0.21104795
dist_from_urbanlog    |    -0.27517425
evlikelihood_topo     |    1.26954628
--------------------------------------------------
    Regression Statistics :
    Number of total observations     = 11202660
    Number of 0s in study area       = 10813840
    Number of 1s in study area       = 388820
    Percentage of 0s in study area   = 96.5292
    Percentage of 1s in study area   = 3.4708

    Number of auto-sampled observations = 1068664
    Number of 0s in sampled area     = 1031730
    Number of 1s in sampled area     = 36934
    Percentage of 0s in sampled area = 96.5439
    Percentage of 1s in sampled area = 3.4561

    -2logL0                  = 321144.7472
    -2log(likelihood)        = 242638.8269
    Pseudo R_square          = 0.2445
    Goodness of Fit          = 506280.2109
    ChiSquare( 4)            = 78505.9204


    Means and Standard Deviations
------------------------------------------------------------
                     Mean        Standard Deviation
------------------------------------------------------------
dist_from_disturlog   7.731486      2.223464
dist_from_roadslog    8.157789      1.520242
dist_from_urbanlog    8.768269      1.091226
evlikelihood_topo     0.505964      0.454772
change0104            0.034561      0.182665
------------------------------------------------------------

    Classification of cases & odds ratio
----------------------------------------------------------------
 Observed |    Fitted_0  |    Fitted_1  | Percent Correct
----------------------------------------------------------------
    0     |  1031687        43          99.9958
    1     |    36932         2           0.0054
----------------------------------------------------------------
    Odds Ratio   =   1.2993
    Reclassification of cases & ROC (Sample-based computation when applicable):
(1) Select a new threshold value such that, after reclassification,  the number
   of fitted 1s matches the number of observed 1s in the dependent variable
    New cutting threshold =   0.1983

 Classification of cases & odds ratio by using the new threshold
-------------------------------------------------------------------
 Observed |    Fitted_0  |    Fitted_1  | Percent Correct
-------------------------------------------------------------------
    0     |   999261      32469         96.8530
    1     |    32469       4465         12.0891
-------------------------------------------------------------------
  Adjusted Odds Ratio =   4.2322
      True Positive =  99.5319%
     False Positive =   3.1470%
(2) ROC* Result with 100 thresholds (Sample-based computation when applicable):
      ROC =   0.8877
 * ROC=1 indicates a perfect fit; and ROC=0.5 indicates a random fit.
```

Appendix V AUC / ROC Logistic Regression Model

(AUC/ROC) IDRISI Results for the Logistic Regression Prediction 2006 compared to Forest Change 2006

ROC = 0.852

****

The following section list detailed statistics for each threshold.

****

With each threshold, the following 2x2 contingency table is calculated

Reality (reference image)

Simulated by threshold 1 0

----

1 A(number of cells) B(number of cells)

0 C D

For the given reference image: A+C=363655 B+D=10839005

No.	Thresholds(%)	A	True positive(%)	B	False positive(%)
1	0	0	0.000	0	0.000
2	1	6198	1.704	105830	0.976
3	2	15635	4.299	208419	1.923
4	3	26347	7.245	309734	2.858
5	4	38007	10.451	410100	3.784
6	5	51149	14.065	508985	4.696
7	6	65547	18.025	606614	5.597
8	7	79953	21.986	704234	6.497
9	8	95183	26.174	801031	7.390
10	9	109935	30.231	898305	8.288
11	10	124041	34.110	996226	9.191
12	11	137972	37.940	1094322	10.096
13	12	151527	41.668	1192793	11.005
14	13	164153	45.140	1292194	11.922
15	14	176827	48.625	1391546	12.838
16	15	188609	51.865	1491791	13.763
17	16	200866	55.235	1591561	14.684
18	17	213210	58.630	1691243	15.603
19	18	225027	61.879	1791453	16.528
20	19	235991	64.894	1892515	17.460
21	20	246710	67.842	1993823	18.395
22	21	257560	70.825	2095000	19.328
23	22	267736	73.624	2196850	20.268
24	23	277433	76.290	2299180	21.212
25	24	286648	78.824	2401991	22.161
26	25	294703	81.039	2505963	23.120
27	26	302271	83.120	2610422	24.084
28	27	309847	85.204	2714872	25.047
29	28	316612	87.064	2820134	26.018
30	29	322878	88.787	2925894	26.994
31	30	328764	90.405	3032035	27.973
32	31	333733	91.772	3139093	28.961
33	32	338527	93.090	3246325	29.950
34	33	342620	94.216	3354259	30.946
35	34	345511	95.011	3463394	31.953
36	35	347681	95.607	3573251	32.967
37	36	349875	96.211	3683084	33.980
38	37	351547	96.670	3793438	34.998
39	38	353643	97.247	3903369	36.012
40	39	355034	97.629	4014004	37.033
41	40	355900	97.867	4125165	38.059
42	41	356569	98.051	4236523	39.086
43	42	357056	98.185	4348062	40.115
44	43	357864	98.408	4459281	41.141
45	44	358647	98.623	4570524	42.167
46	45	359514	98.861	4681684	43.193
47	46	360142	99.034	4793083	44.221
48	47	360787	99.211	4904464	45.248
49	48	361821	99.496	5015457	46.272
50	49	362400	99.655	5126904	47.301
51	50	362825	99.772	5238506	48.330
52	51	362913	99.796	5350445	49.363
53	52	363118	99.852	5462266	50.395
54	53	363180	99.869	5574231	51.428
55	54	363248	99.888	5686189	52.460
56	55	363307	99.904	5798157	53.493
57	56	363375	99.923	5910116	54.526
58	57	363375	99.923	6022142	55.560
59	58	363375	99.923	6134169	56.593
60	59	363375	99.923	6246195	57.627
61	60	363380	99.924	6358217	58.661
62	61	363426	99.937	6470198	59.694
63	62	363460	99.946	6582190	60.727
64	63	363494	99.956	6694183	61.760
65	64	363495	99.956	6806208	62.794
66	65	363496	99.956	6918234	63.827
67	66	363542	99.969	7030215	64.860
68	67	363632	99.994	7142151	65.893
69	68	363655	100.000	7254155	66.926
70	69	363655	100.000	7366181	67.960
71	70	363655	100.000	7478208	68.993
72	71	363655	100.000	7590235	70.027
73	72	363655	100.000	7702261	71.061
74	73	363655	100.000	7814288	72.094
75	74	363655	100.000	7926314	73.128
76	75	363655	100.000	8038341	74.161
77	76	363655	100.000	8150368	75.195
78	77	363655	100.000	8262394	76.228
79	78	363655	100.000	8374421	77.262
80	79	363655	100.000	8486447	78.295
81	80	363655	100.000	8598474	79.329
82	81	363655	100.000	8710501	80.363
83	82	363655	100.000	8822527	81.396
84	83	363655	100.000	8934554	82.430
85	84	363655	100.000	9046580	83.463
86	85	363655	100.000	9158607	84.497
87	86	363655	100.000	9270634	85.530
88	87	363655	100.000	9382660	86.564
89	88	363655	100.000	9494687	87.597
90	89	363655	100.000	9606713	88.631
91	90	363655	100.000	9718740	89.665
92	91	363655	100.000	9830767	90.698
93	92	363655	100.000	9942793	91.732
94	93	363655	100.000	10054820	92.765
95	94	363655	100.000	10166846	93.799
96	95	363655	100.000	10278873	94.832
97	96	363655	100.000	10390900	95.866
98	97	363655	100.000	10502926	96.899
99	98	363655	100.000	10614953	97.933
100	99	363655	100.000	10726979	98.966
101	100	363655	100.000	10839005	100.000

Appendix VI AUC / ROC Multi-Layer Perceptron

(AUC/ROC) IDRISI Results for the Multi Layer Perceptron 2006 compared to Forest Change 2006

ROC = 0.920

****

The following section list detailed statistics for each threshold.

****

With each threshold, the following 2x2 contingency table is calculated

	Reality (reference image)	
Simulated by threshold	1	0
1	A(number of cells)	B(number of cells)
0	C	D

For the given reference image: A+C=363655 B+D=10839005

No.	Thresholds(%)	A	True positive(%)	B	False positive(%)
1	0	0	0.000	0	0.000
2	1	35519	9.767	76509	0.706
3	2	63294	17.405	160760	1.483
4	3	88183	24.249	247898	2.287
5	4	112420	30.914	335687	3.097
6	5	134822	37.074	425312	3.924
7	6	157284	43.251	514877	4.750
8	7	178252	49.017	605935	5.590
9	8	198130	54.483	698084	6.440
10	9	214997	59.121	793243	7.318
11	10	229961	63.236	890306	8.214
12	11	244002	67.097	988292	9.118
13	12	256958	70.660	1087362	10.032
14	13	269074	73.992	1187273	10.954
15	14	279629	76.894	1288744	11.890
16	15	287864	79.159	1392536	12.847
17	16	295261	81.193	1497166	13.813
18	17	301814	82.995	1602639	14.786
19	18	308441	84.817	1708039	15.758
20	19	314678	86.532	1813828	16.734
21	20	320786	88.212	1919747	17.711
22	21	325942	89.629	2026618	18.697
23	22	330554	90.898	2134032	19.688
24	23	335610	92.288	2241003	20.675
25	24	338767	93.156	2349872	21.680
26	25	342237	94.110	2458429	22.681
27	26	344873	94.835	2567820	23.691
28	27	347824	95.647	2676895	24.697
29	28	351205	96.576	2785541	25.699
30	29	353534	97.217	2895238	26.711
31	30	355134	97.657	3005665	27.730
32	31	356538	98.043	3116288	28.751
33	32	357486	98.304	3227366	29.775
34	33	358586	98.606	3338293	30.799
35	34	359271	98.794	3449634	31.826
36	35	360317	99.082	3560615	32.850
37	36	361510	99.410	3671449	33.873
38	37	362192	99.598	3782793	34.900
39	38	363136	99.857	3893876	35.925
40	39	363350	99.916	4005688	36.956
41	40	363495	99.956	4117570	37.988
42	41	363655	100.000	4229437	39.021
43	42	363655	100.000	4341463	40.054
44	43	363655	100.000	4453490	41.088
45	44	363655	100.000	4565516	42.121
46	45	363655	100.000	4677543	43.155
47	46	363655	100.000	4789570	44.188
48	47	363655	100.000	4901596	45.222
49	48	363655	100.000	5013623	46.255
50	49	363655	100.000	5125649	47.289
51	50	363655	100.000	5237676	48.322
52	51	363655	100.000	5349703	49.356
53	52	363655	100.000	5461729	50.390
54	53	363655	100.000	5573756	51.423
55	54	363655	100.000	5685782	52.457
56	55	363655	100.000	5797809	53.490
57	56	363655	100.000	5909836	54.524
58	57	363655	100.000	6021862	55.557
59	58	363655	100.000	6133889	56.591
60	59	363655	100.000	6245915	57.624
61	60	363655	100.000	6357942	58.658
62	61	363655	100.000	6469969	59.692
63	62	363655	100.000	6581995	60.725
64	63	363655	100.000	6694022	61.759
65	64	363655	100.000	6806048	62.792
66	65	363655	100.000	6918075	63.826
67	66	363655	100.000	7030102	64.859
68	67	363655	100.000	7142128	65.893
69	68	363655	100.000	7254155	66.926
70	69	363655	100.000	7366181	67.960
71	70	363655	100.000	7478208	68.993
72	71	363655	100.000	7590235	70.027
73	72	363655	100.000	7702261	71.061
74	73	363655	100.000	7814288	72.094
75	74	363655	100.000	7926314	73.128
76	75	363655	100.000	8038341	74.161
77	76	363655	100.000	8150368	75.195
78	77	363655	100.000	8262394	76.228
79	78	363655	100.000	8374421	77.262
80	79	363655	100.000	8486447	78.295
81	80	363655	100.000	8598474	79.329
82	81	363655	100.000	8710501	80.363
83	82	363655	100.000	8822527	81.396
84	83	363655	100.000	8934554	82.430
85	84	363655	100.000	9046580	83.463
86	85	363655	100.000	9158607	84.497
87	86	363655	100.000	9270634	85.530
88	87	363655	100.000	9382660	86.564
89	88	363655	100.000	9494687	87.597
90	89	363655	100.000	9606713	88.631
91	90	363655	100.000	9718740	89.665
92	91	363655	100.000	9830767	90.698
93	92	363655	100.000	9942793	91.732
94	93	363655	100.000	10054820	92.765
95	94	363655	100.000	10166846	93.799
96	95	363655	100.000	10278873	94.832
97	96	363655	100.000	10390900	95.866
98	97	363655	100.000	10502926	96.899
99	98	363655	100.000	10614953	97.933
100	99	363655	100.000	10726979	98.966
101	100	363655	100.000	10839005	100.000