Understanding the drivers affecting land use change in Ecuador: An application of the Land Change Modeler software

Dhruva Rajan
School of GeoSciences
The University of Edinburgh

Statement of Originality

"I declare that this dissertation represents my own work, and that where the work of others has been used it has been duly accredited. I further declare that the length of the components of this dissertation is **5507** words for the Research and **10,458** words for the Technical Report."

Dhruva Rajan

11th August

Acknowledgements

Part I
Research Paper

Understanding the drivers affecting land use change in Ecuador: an application of the Land Change Modeler software

DHRUVA RAJAN

School of GeoSciences, Drummond Street, University of Edinburgh, Edinburgh EH8 9XP, UK.

Abstract

Deforestation modelling is a relatively new field of study. The importance of this science has been advanced with the emergence of deforestation as one of the leading causes of global climate change. The advent of REDD (reduced emissions from deforestation and degradation) and related policy mechanisms has accelerated the need for modelling deforestation. This project looks at developing a methodology for modelling deforestation using the Land Change Modeler software. To generate the model of change in forest cover, satellite images (years 1996 and 2001) were used to produce land cover maps which were then used with the software to estimate probabilities of pixels changing from forests to other land use types. Various drivers of deforestation input into the model were proximity to roads, proximity to towns and slope. The model was developed by analysing change in forest cover between 1996 and 2001 and computing the probability of each cell undergoing change by deforestation. The success was measured by validating the different models (generated using various drivers with the LCM) for the year 2006 with a classified image of actual deforestation in 2006. The predicted rate of deforestation (based on the model generated) was 1.5% per year. This compares to an 'actual' rate calculated by a differencing of forest pixels in two satellite images from 2001 and 2006 of 1.8%. The results show that the software was able to accurately model 61% of the pixels that underwent deforestation when using all of the three drivers mentioned above. Testing the various drivers revealed that proximity to towns was the single driver most strongly influencing the predicted pattern and that predicted deforestation using a finer 20m resolution slope data was more similar to the actual pattern of deforestation observed in the 2006 imager data than the prediction based on a coarser 90m slope.

1. Introduction

Given the extent of the areas involved and the need for repeated monitoring over time, measuring and analysing deforestation is now predominantly accomplished through the use of remote sensing. The advantage of using remote sensing has already been widely discussed (see Gilruth et al. 1990; Kerr & Ostrovsky 2003; Liverman 1998). Remote sensing incorporates a technique called change detection, which is used to analyse the rate of deforestation. However, simple change detection techniques consider only two time periods, and hence yield a single estimate of deforestation between two dates. It does not itself help us to understand how or why the deforestation may have taken place in certain places rather than others.

This project attempts to understand the rate of deforestation and analyse drivers of change. Instead of limiting the study to just two land cover maps and identifying areas of change (change detection), we are more interested in looking at drivers of land use change that alter the places where deforestation occurs over time. This project tries to answer the following questions –

- Can we find geographical data sets that can help to explain the actual patterns of deforestation?
- Does the Land Change Modeler generate models of deforestation that are reliable?
- Are the models of deforestation sensitive to the resolution of the input data used?

These questions are answered by generating a series of models and comparing patterns of deforestation at two dates in Ecuador using the Land Change Modeler software. Although it is not meant to be a comprehensive methodology, it is an attempt to develop a simple land use change model which incorporates various geographical data sets as possible explanatory variables (as 'drivers of change'). Reflecting on the usefulness of this approach contributes towards a critique of land change models; something that is lacking in the scientific community (Pontius 2002).

2. Background

According to According to estimates by the International Panel on Climate Change (IPCC), over the last decade, 1.6 billion tons of carbon has been released annually by land-use change activities (Bolin et al., 2000). The Stern Report (Stern, 2007) points out that nearly one-fifth of today's total annual greenhouse gas emissions come from land-use change, most of which can be traced back to tropical deforestation. Deforestation is caused by several 'drivers'. Sometimes these drivers may be physical and measurable or social and political (in which case they are considerably more difficult to measure). In the context of REDD (Reduced Emissions from Deforestation and Degradation) and Carbon Trading, modelling these changes in land use has assumed great importance (Kohl et al., 2009).

The premise of REDD is that tropical forest countries would be compensated if they reduce their rates of deforestation and thus their emissions of greenhouse gases (Bali Action Plan, 2007). This compensation would be given in monetary terms, which would be a part of the global carbon market. Several tropical, developing countries have become

interested in developing forest conservation programmes which leads to a reduction in deforestation rates, thereby generating a sizeable amount of carbon credit.

Once we are able to predict how the current land use might affect the state of the forest after a period of time, we arrive at what is known as a "Baseline scenario" (Madeira, 2008). This describes the future emission pathway without any climate protection measures or what is referred to as a *business as usual scenario*. Such baselines are crucial to measuring the emission reduction performance (Angelsen, 2008) as well as ensuring accurate crediting of the carbon credit account (OECD/IEA, 2003). For an avoided deforestation project to generate a sizeable amount of carbon credit, the baseline needs to indicate that the forest was under sufficient threat of deforestation (Brown et al., 2007).

When initial REDD mechanisms were conceived for project level implementation, it became obvious that a project specific system could be compromised due to the phenomenon of "leakage" (Brown et al., 2000), where deforestation occurs outside the project area, hence negating any conservation measure taken within the project area. Thus, as part of the REDD mandate, there has been a debate as to whether participating countries must implement the REDD mechanism at a national scale or project level scale (Angelsen et al., 2008, Verchot et al., 2009, Madeira, 2008). This debate is significant because it has serious implications on the way carbon emissions are accounted for. The scale of assessment may also influence the approach that is taken to monitor deforestation.

2.1 Models of Land use Change

Modelling land use change helps us unravel the dynamics of land use systems. Sensitivity of land-use patterns to changes in key variables can be tested. "Sensitivity analysis can help to identify the most important mechanisms of change in a certain area that could not be identified from field observation" (Verburg et al., 2006).

Although several types of models exist, Verburg et al. (2006) give us a common structure for a large number of land use change models (see Figure 1). A distinction is made between drivers that influence the magnitude of change and those that influence the location of change. Based on the drivers that influence the change, we can come up with what is known as a suitability map, which indicates the suitability of a location for a specific land-use type relative to the suitability of other locations. This is the main component of any land-use change model (Clarke et al., 1997; Zamyatin and Markov, 2005).

The bottom-up approach is chosen when the spatial dynamics and allocation rules determine the aggregated quantity of land-use or land-cover change. This would be used when a set of predefined characteristics are expected to shape the land use (e.g. an industrial estate or a town). A top-down approach is chosen when the quantity of change is to be estimated or explained based on a set of driving factors (e.g. roads, rivers, proximity to towns etc). This would be used when only driving factors are quantifiable. In most cases these geographical data are proxies for the underlying processes which they may represent but are much harder to measure.
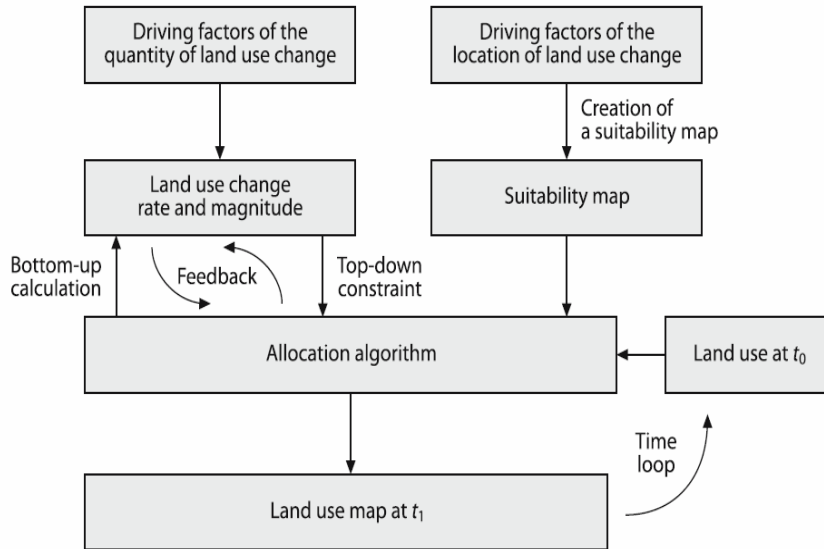
Figure 1 – Flow chart showing process of creating a land use change model (Source: Land use and Land Cover Change (the IGBP series – edited by E.F Lambin and H.J Geist)
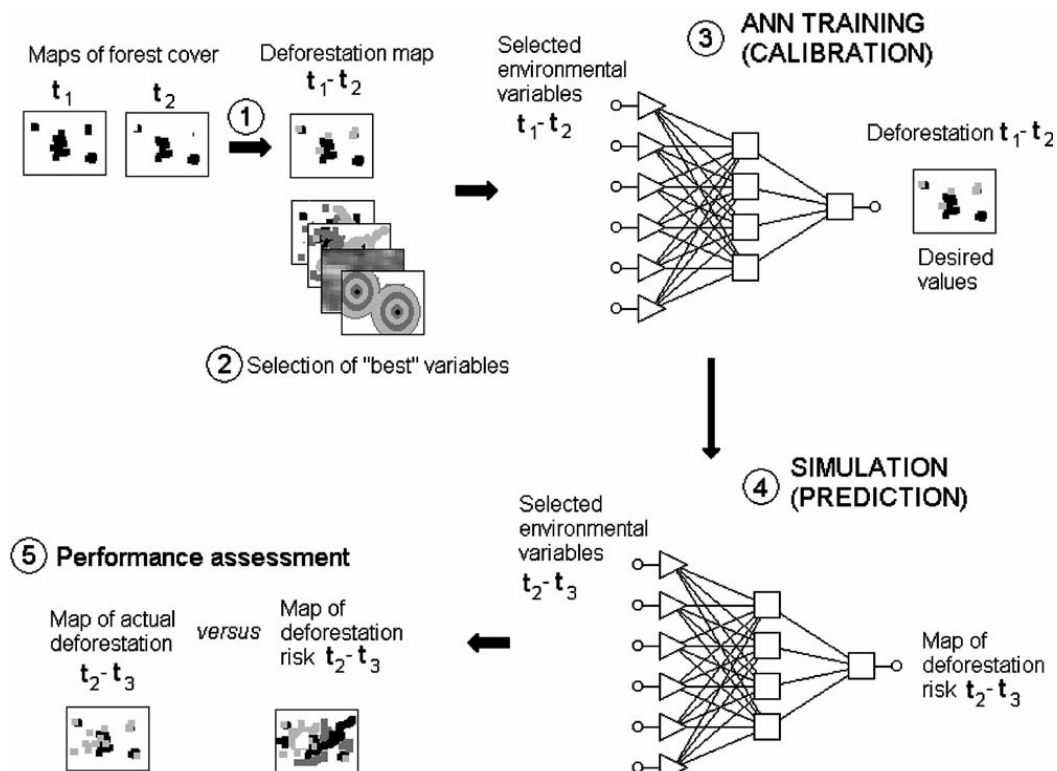

Figure 2 - Flowchart of the neural-network based model for simulating deforestation (J.F. Mas et al.(2004))

The allocation algorithm is usually a neural network, which functions as described in the Figure(2) above. The main steps for creating models of deforestation according to the diagram are described below:

(1) Creating maps of deforestation by comparing land cover maps at two or more time periods.

(2) Analysis of relationships between deforestation and drivers in order to select "best" explanatory variables (so called 'driver' maps)

(3) Calibration of the model (training of the ANN)

(4) Simulation (running the model to obtain a map of deforestation which predicts deforestation for the given target date)

(5) Assessment of the model performance (comparison between actual and predicted deforestation).

This paper presents a top-down type of methodology based on the steps described above, using an artificial neural network called Multi Layer Perceptron (part of the Land Change Modeler software developed by Clark Labs) to model changes in land use, specifically deforestation. In addition to producing a map of predicted deforestation, it also explores the effect of using different drivers of deforestation to create the predicted deforestation through a sensitivity analysis. To conduct the study, a case study area was chosen for which there was previous work to map land cover change and some understanding of the drivers leading to deforestation of this area.

3. Material and Methods

3.1 Study Area:

The study area includes a part of the province Imbabura in North Ecuador. This part of Ecuador contains the inter-Andean central highlands (sierra), one of the most rugged terrains in the world. The elevation changes from about 40m above sea level to almost 4500m above sea level. The slope ranges from 2% to about 60% in some places. The area under study contains mainly cloud forest on the lower slopes of the mountains, and high altitude grasslands towards the top of the mountains. Forest area covers about 42% of the study area.

Agriculture is the main land use in the countryside, leading to deforestation (Southgate and Whitaker, 1994; Bebbington, 1993). The other biggest land use is for cattle rearing (i.e. Pasture). This has increased over the last 20 years due to the rising demand in meat and milk products (Zevallos, 1989; Commander and Peek, 1986).

With the increase in agricultural activity, there was an automatic urbanisation of the rural areas. Quito, the capital of Ecuador (located about 60km south of the study area) underwent major expansion, leading to growth in urban area, leading to further deforestation. The capital of the Imbabura province, Ibarra, is the biggest town in the study area with a population of 153,256 (http://www.statoids.com/yec.html). Due to the rapid urbanisation and agricultural expansion, the rate of deforestation has increased enormously. A baseline study conducted by Peck et.al., (2010) in association with Rainforest Concern has identified the potential for a REDD project in the area with a success rate (meaning level of reduction in emission) estimated at 96.3% and 95.1% for a 30 year and 40 year project, respectively, based on the baseline calculated.

However, a standard methodology for generating baselines presently does not exist (Brown et al., 2007). People have turned to modelling land use change based on a set of geographical driving factors as one approach to create a baseline, as it allows consideration of regional factors that affect land use change. It also means that the host country or state can identify factors contributing to deforestation quickly. Ecuador represents a good example of an area where baselines need to be produced to estimate extent of deforestation and available carbon stock (due to the rapid rate of deforestation as described by Peck et.al, 2010). Thus we use a model for land use change to produce deforestation models for the generation of baselines in the study area described below.

The highlighted part of Imbabura province is the county of Cotacachi. A part of the county (highlighted in pink in the inset) is the area under study for this project. The area was chosen based on the extent of the driver data available.
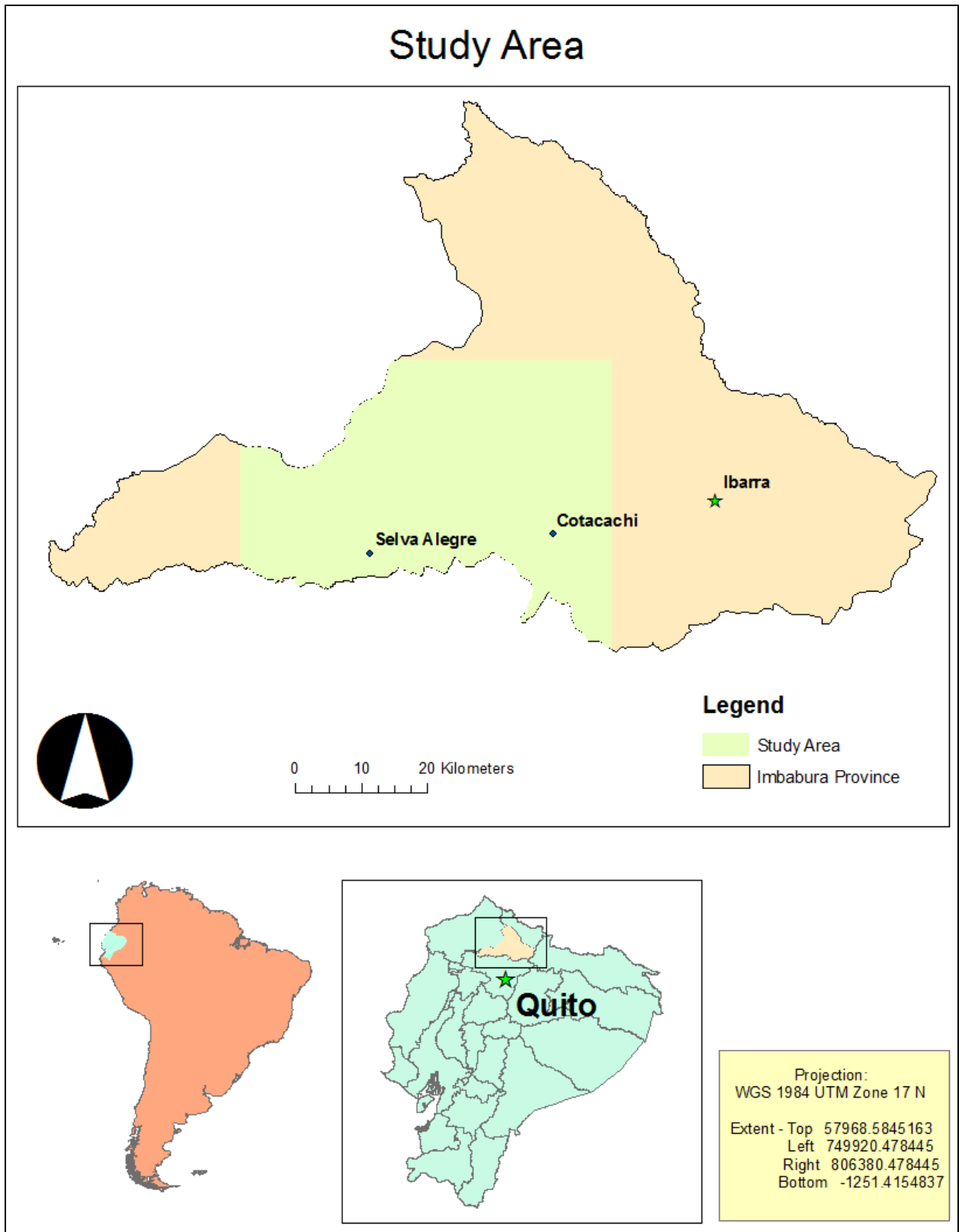
Figure 3 – Project study area within Ecuador, in South America

3.2 Data Sets

Data was obtained from Rainforest Concern's (an NGO working in the Ecuadorian rainforest) scientist, Dr. Mika Peck. Peck et al. (2008) have done studies in the area to identify the various land use types and produced classified land cover maps (based on LANDSAT imagery) for a habitat study of the fauna in the province of Imbabura. In addition to using the classified images (for 2001 and 2006) from Peck et al. (2008) another LANDSAT image (for 1996) was classified for this study (Rajan D.K, 2010) in order to generate a model of change from 1996-2001 based on the land cover maps of the area from classified LANDSAT images and drivers of deforestation present in the area. Slope (elevation was not used because the change in elevation, rather than just elevation has a greater impact on rate of deforestation), proximity to towns and protected areas were chosen as the drivers for this study taking data constraint into consideration. The presence of these drivers has also been correlated with the high rate of deforestation in Ecuador (Southgate et al. (1991), Rudel (1993), C.F. Mena (2006), S. Walsh et al. (2002), Chomitz, K.M. & Gray, D.A. (1995), Soares-Filho, B (2004)).

The model was then used to predict change from 2001 to 2006 and the result compared against the actual land cover from 2006 (as explained in the next section).

Table 1 – A summary of the LANDSAT images used

Path	Row	Date	Extra Comments
10	**60**	**24 07 1996**	Source for this data set was the Global Land Cover Facility (www.landcover.org)
10	**60**	**03 11 2001**	Source for this data set was Dr. Mika Peck (2008)
10	**60**	**13 08 2006**	Gap filled with 09 07 2005 SLC – Off and 03 11 2001 SLC- ON Source for this data set was Dr. Mika Peck (2008)

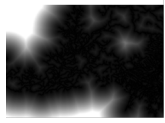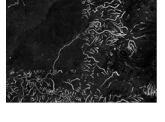***Table 2 – A summary of the data sets***

Data Set	Description	Thumbnail	Projection	Resolution	Source
Proximity to Roads	Vector converted to Raster; Euclidian distance to the roads.		WGS 1984 UTM Zone 17 N	30m	Instituto Geografico Militar, Ecuador
Proximity to Towns	Vector converted to Raster; Euclidian distance to the towns.		WGS 1984 UTM Zone 17 N	30m	Instituto Geografico Militar, Ecuador
Slope	Slope calculated from Digital Elevation Model		WGS 1984 UTM Zone 17 N	20m	Instituto Geografico Militar, Ecuador
Slope	Slope calculated from Digital Elevation Model		WGS 1984 UTM Zone 17 N	90m	SRTM (2000) website
Protected area/ constraints map	Vector converted to Raster; Reclassified to have 0 value for protected area and 1 for all other areas		WGS 1984 UTM Zone 17 N	30m	Instituto Geografico Militar, Ecuador

Table 2a - Data Format

Extent *- Top*	*57968.5845163*
Left	*749920.478445*
Right	*806380.478445*
Bottom	*-1251.4154837*
Number of columns	*1882*
Number of rows	*1974*

3.3 Land Change Modeler

The Land Change Modeler (http://www.clarklabs.org/products/Land-Change-Modeler-Overview.cfm) is a software program produced with the intention of visualizing change, and producing models. It also allows one to analyze measure and project the impacts of land use change on habitat and biodiversity. It allows the user to create models of land use change with a suite of tools that address the complexities of change analysis, resource management and habitat assessment with relatively low level of difficulty and data requirements. The Land Change Modeler is included within the IDRISI GIS and Image Processing software and is available as a software extension for use with ESRI's ArcGIS product.

Land Change Modeler provides:

- A certain set of tools to perform land cover change analysis

- A relatively simple modelling environment, in order to create predictions for future scenarios, integrating drivers of change as well as constraints, such as protected areas.

- Specific tools for the implementation of REDD projects (such as estimation of baselines and modelling deforestation)

(From the clarklabs website - http://www.clarklabs.org)

Models of deforestation were generated for year 2006 using the Land Change Modeler, with a variety of different driver inputs. The overall methodology is given in the flow chart below (Figure 4).

In the first step, the LANDSAT image (of 1996) was classified using ERDAS Imagine. Two other land cover maps (for 2001 and 2006) were obtained already classified by Peck et.al. (2008). All three maps contained the same land cover classes. (Cloud shadow, Water, Urban/Cleared, Pasture, Agriculture and Forest).

Two of the three land cover maps, one dated 1996 and the other 2001, were used (along with the driver maps described above) as input for the LCM. ('Training phase' in Figure 4) Using different combinations of drivers along with a constraint map (which restricted the deforestation in protected areas), the LCM produced a series of models of deforestation for the study area, based on the training data. ('Model generation phase' in Figure 4) These models then had to be validated. The Map Comparison Toolkit (MCK) was used to validate the results by comparing the models generated with a pre-classified 2006 LANDSAT image. ('Comparison phase' in Figure 4)
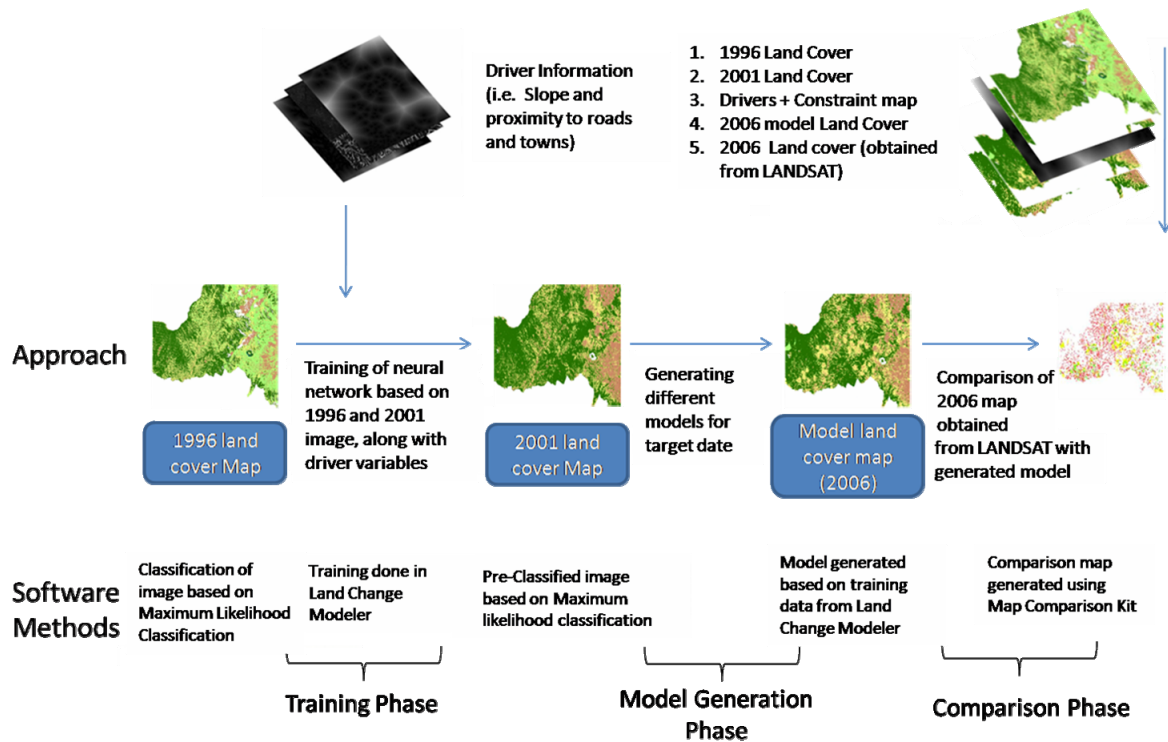
Methodology

Figure 4 – Flowchart describing major steps in the methodology

The MCK generates a comparison map, which identifies areas of similarity and difference between the two maps. It also generates Kappa statistics which measures the amount of coincidence in classes between the two maps.

3.4 Evaluation of Drivers of Deforestation

In this project, the sensitivity of the model was tested by using different combinations of input drivers. These results were then compared against a reference map to compare kappa values. The Comparison Phase in Figure 4 shows the layers of maps used in the final step. The two land cover maps (1996 and 2001) are used to generate a model of deforestation for 2006, which is then compared to the land cover map for 2006 obtained from classification of a Landsat image, thereby yielding a comparison map, showing areas where the model correctly predicted deforestation (hits), areas where the model predicted deforestation which did not actually occur (false alarms), and areas where the model did not predict deforestation that actually occurred (misses).

4. Results

4.1 Sensitivity Analysis of model

To evaluate the effect of the different drivers of deforestation on the study area, LCM was run using different combinations of drivers (listed in table 2). The Kappa statistics for each are given in Table 3. (*For a detailed discussion on Kappa refer to Section 4.1 in the Technical Report (Rajan D.K., 2010)* Higher Kappa values indicate greater agreement between the model and the 2006 LANDSAT image (see Fig. 6)

Table 3 – Kappa statistics of the different drivers

	Road	Slope	Town	Road + Town	Slope + Town	Road + Slope	All
Kappa	0.62511	0.58932	0.62742	0.62451	0.6207	0.61831	0.61751
Fraction correct	0.81317	0.79531	0.8143	0.81286	0.81097	0.80979	0.80936

Note – Numbers highlighted show highest value in respective category

The table shows us how the different drivers affect the final model produced. In fact, the model built with proximity to towns as the only driver has the highest Kappa value. Significantly, we can see that all the models have similar Kappa Values. The graph below gives us a quick summary of the comparative Kappa values according to different combinations of drivers.
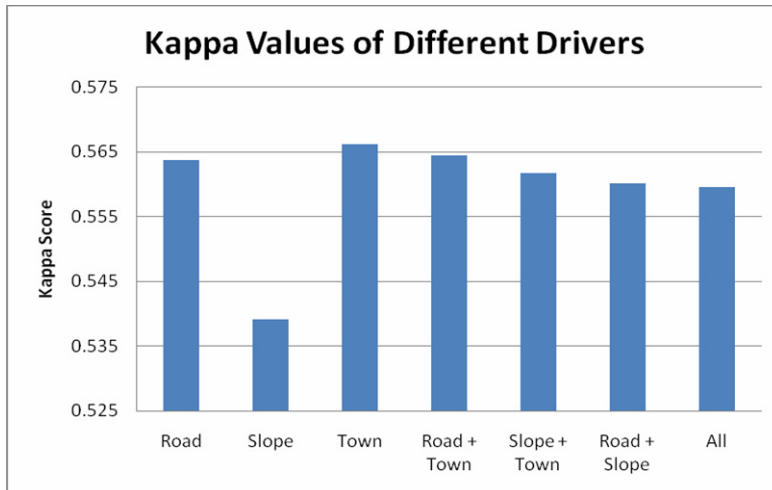

Figure 5 – comparison of performance of drivers

The graph illustrates that the model of deforestation driven by only proximity to towns has the highest Kappa values when compared to the 2006 reference land cover map. One other notable feature from the graph is that the difference between the lowest and highest Kappa value is not high (i.e. only a 0.04 difference). Therefore we can conclude

that it does not make a huge difference whether using a single or a combination of these drivers in this area.
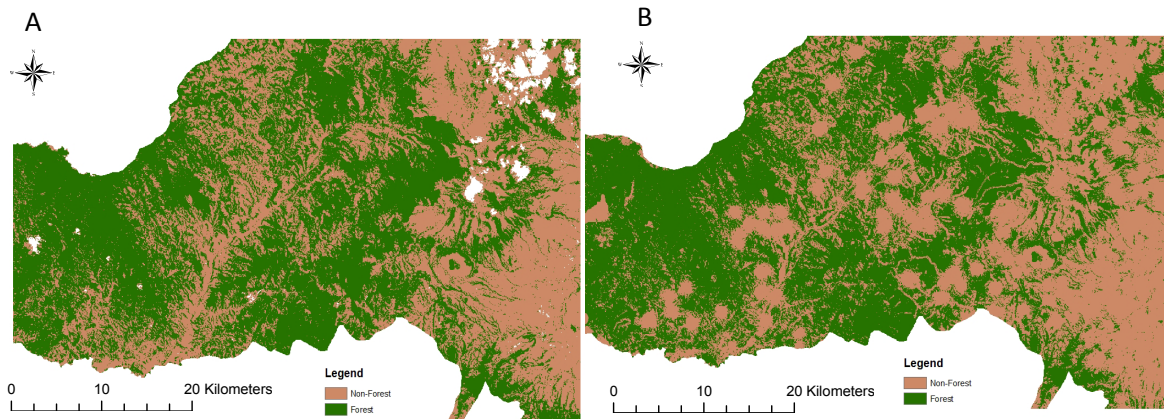

Figure 6 – a) Original Land cover map of 2006; b) Model of land cover in 2006 produced using all drivers

The map above (Fig 6b) shows the model of land cover produced using all three drivers (proximity to roads, proximity to towns and slope). The comparison map (Fig 7a below) shows the difference in areas of predicted remaining forest cover with areas of actual remaining forest cover. We are interested in forest cover rather than areas of non-forest in this comparison, so areas of non-forest stand out. Hits refer to areas where forest pixels exist in both maps. Misses refer to areas where forest pixels are found in original land cover map but are not forest in the model (therefore the model has missed areas of forest that exist). False alarms refer to areas where the forest pixels are found in the model image but these are actually non-forest pixels in the original image (therefore the model has falsely predicted forest to have existed in those areas).
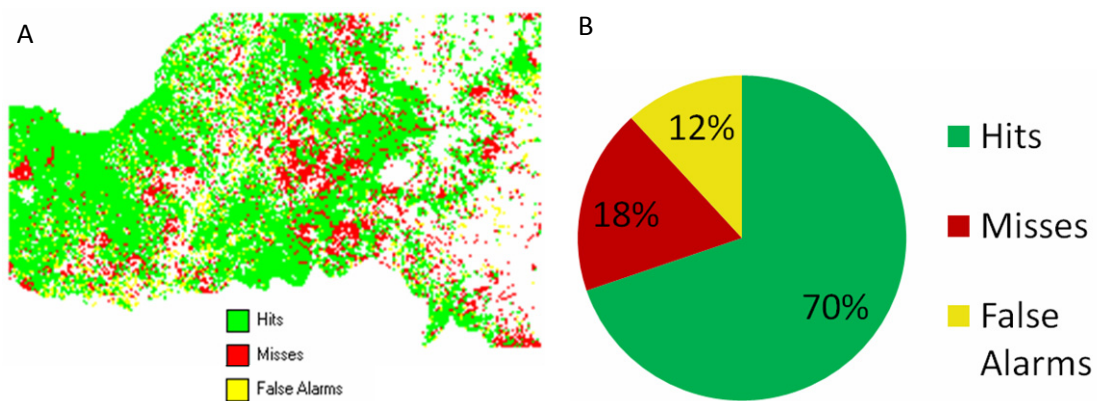

Figure 7 –a) Comparison map produced by comparing 6A with 6B; b) Pie chart showing percentage of areas correctly predicted (hits), areas missed and false alarms

The percentage of hits are high (70%), which is quite close to the Kappa value of 0.62 (as described in table 3).

The Kappa values (from table 3) show that the land cover model (using all drivers) produced for 2006 is moderately similar (almost 0.62) with the land cover map obtained from LANDSAT in 2006 (Kappa score of 1 = exactly the same; Kappa score of 0 = completely different). The fraction correct statistic shows that the model produces an 80% match in deforested areas with reference to the 2006 image. However we must be critical of this value because the fraction correct method is biased, due to the fact that it does a cell by cell comparison. Hence even if NoData pixels match up well, it returns a high value. Therefore, a more accurate value of comparison is the Kappa score.

Effect of spatial resolution

One of the key questions in REDD and land use change modelling is that of resolution. If deforestation needs to be modelled at a global scale, then data will be coarse. The map below shows a model generated using 90m slope instead of 20m slope (other driving factors remaining the same)
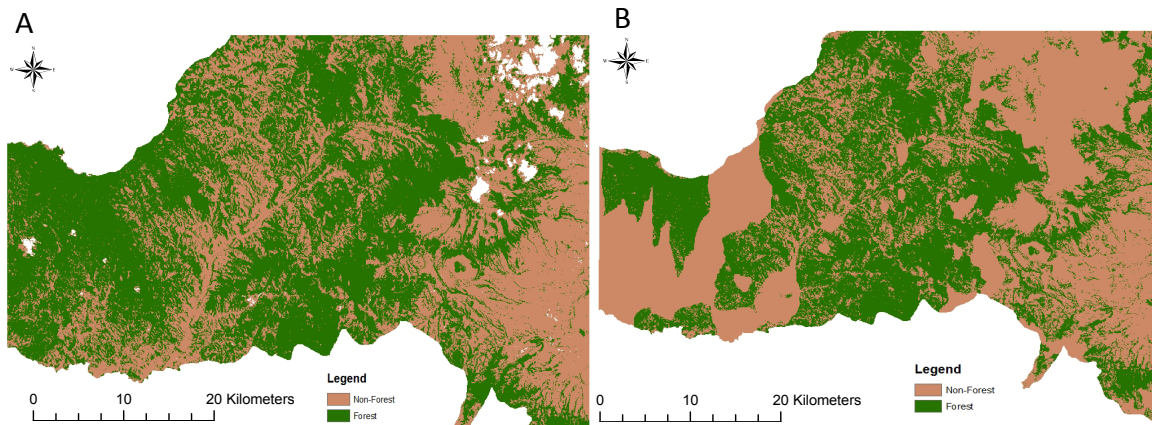

Figure 8 – a) Original Land cover map of 2006; b) Model of land cover in 2006 produced using all drivers with 90m slope
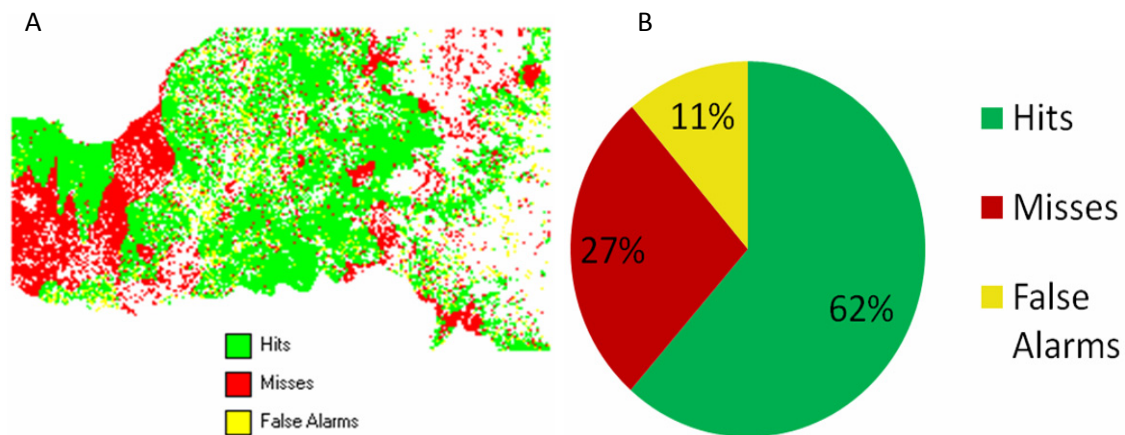

Figure 9 – a) Comparison map produced by comparing 8A with 8B; b) Pie chart showing percentage of areas correctly predicted (hits), areas missed and false alarms

It is immediately clear that the areas predicted by the model produced with 90m Slope have less hits and false alarms than the model produced with 20m slope. The number of misses is more than the model created with all drivers and the 20m slope.

Table 5 – Comparison between 20m slope and 90m slope

	Map 20m Slope	**Map 90m Slope**
Kappa	0.61751	0.52164
Fraction correct	0.80936	0.7585

The Kappa values confirm that the 90m slope does not perform so well in terms of overall accuracy. This reflects the fact that finer resolution results in greater detail, hence more accurate modelling.

The graphs below (Figure 10a and 10b) also indicate that there is greater variability and a higher rate of estimated/predicted deforestation depending on the combination of drivers used when using the 90m slope. This could be due to the fact that a 90m slope is coarse, leading to a smoothing effect of the slope therefore allowing more areas of a low degree of slope to be potentially 'vulnerable' to deforestation. When comparing the Kappa values between the 20m slope and 90m slope, we can observe that all the models produced using 20m slope have higher Kappa values than the 90m slope models. This is important to note, as 90m datasets yield predictions that are an overestimation of actual level of deforestation, and any results from a model using coarse resolution slope data must be treated with caution.
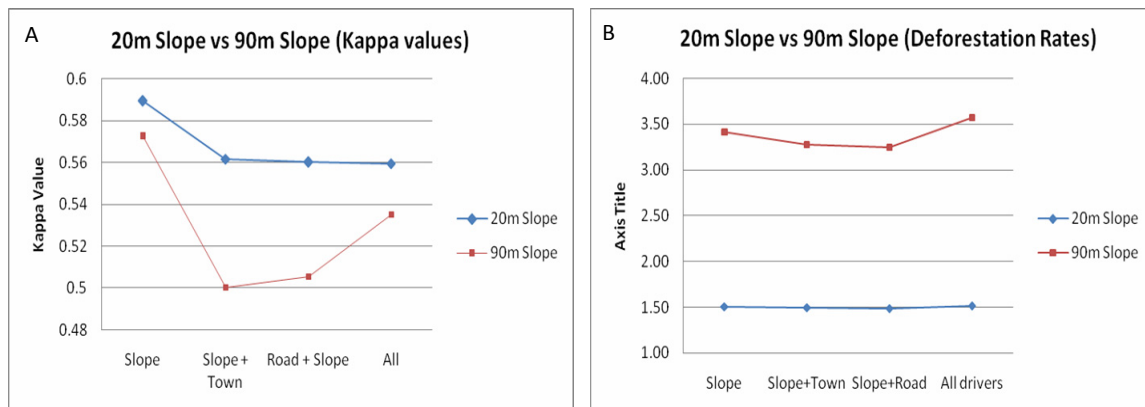

Figure 10- showing difference between using slope (20m resolution) and slope (90m resolution); a) difference between Kappa values, b) difference in predicted deforestation rates

4.2 Deforestation Rate

The rate of deforestation was calculated by subtracting the model land cover map for 2006 (generated by LCM) from the land cover map of 2001 (classified from LANDSAT). This yielded the pixels that were predicted by the model to have undergone change from 2001 to 2006 (i.e. from forest to non-forest). These were then divided by the number of years (i.e. 5) to obtain average number of pixels that changed per year. The predicted rate of deforestation obtained for the years 2001-2006 based on the LCM model built from 1996 and 2001 training data (based on all drivers) was 1.51 % per annum.

Rate of deforestation for the County of Cotacachi as computed by a previous study using the same imagery (Peck M. 2010) was 1.92% of forest loss per annum for the period 2001-2006, using change detection techniques. Using the Map Comparison Toolkit to measure the area deforested between 2001 and 2006 using the same land cover maps the result was 1.83% per annum. This is an underestimation of the area of deforestation when compared with that of Peck et al.(2010) This minor difference could possibly be attributed to differences in the area under study (our study area slightly smaller than the one studied by Dr.Peck).

The rate of deforestation measured by change detection is higher (1.8-1.9%) than the one measured by the model (1.5%). Therefore the software predicts less deforestation than is actually observed (i.e. it underestimates deforestation). This is a significant point to note. When measuring rates of deforestation, different drivers yielded different results.

Table 6 - describing the sensitivity to the presence of different drivers affecting deforestation rates

Driver present	% Loss per year (2001 to 2006)
Road	1.49
Slope	1.51
Town	1.51
Road+Town	1.50
Slope+Road	1.49
Slope+Town	1.50
All drivers	1.51

Although the values of loss are similar, there is an indication that (value circled in red in Table 6) all the drivers together estimate a higher rate of deforestation than a single driver. This value is also closer to the value (1.92%) that was estimated by Dr. Peck. However, as described by Figure 5, when using towns as the single driver of deforestation, the model returns a higher Kappa value when compared with the reference 2006 map. This means that towns show better predictions of geographical locations of deforestation. Whereas when using all drivers, the predicted rate of deforestation is closer to reality (although not by much). As the values of all the drivers are location within a very small range of values, it does not matter much what driver (or combination of drivers) we use for generating the model.

Modelling future change – The comparison between the model produced with all three drivers and the original map shows that there is an 80% match between the generated model of land cover for 2006 and the actual LANDSAT land cover image for 2006. The Kappa value for this comparison is 0.61. This means, we can have 61% confidence in the models generated by the software. Using the same settings as used to generate the model with all the drivers, the LCM was run to generate a prediction for 2050, with intermediate output stages at every 5 years. The assumption here is that no changes have taken place to the existing roads and towns, which is very rarely true. However such a predictive model allows for roughly estimating the extent and level of deforestation into the future in order to plan conservation measures effectively. (*See discussion for more*)

The predicted level of deforestation by 2050 is nearly 61% loss from 2001 (which is equal to a loss of 1.25% per year). The result is shown below:
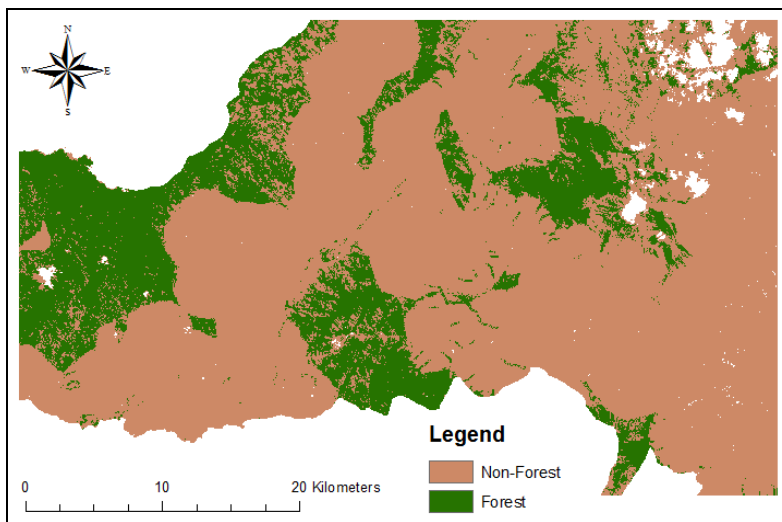
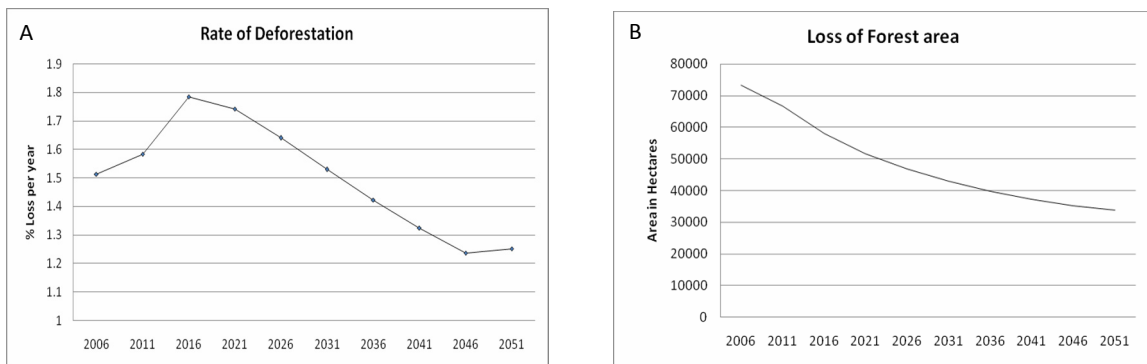

Figure 11 – Model of deforestation for 2050

Figure 12 - modelling future scenarios; a) describing change in rate of deforestation, b) describing loss in forest area over time.

The graph above (Figure 12a) shows the changing rate of deforestation over a period of nearly 50 years (from 2006 to 2050). The interesting thing to note is that rate of

deforestation begins to increase up to 2016, and then gradually drops down over the next 30 years. This drop is characteristic of unplanned deforestation, because with a decrease in forest cover over time, the rate of deforestation becomes controlled by the lack of further easy access to forest area. (If we assume that no new roads are being constructed)

Figure 12b shows the reduction in area of forest. There is a steady decrease in the area of forest over the 50 years. The loss is forest are corresponds to the rate of deforestation described in 12a.

5. Discussion

With the advent of global climate change, the need for understanding and mitigating deforestation has never been greater. Without a succinct understanding of the processes and patterns of deforestation however, change prediction is difficult. Deforestation is affected by several anthropogenic related activities, and in order to reduce the rate of deforestation, the primary objective then becomes understanding drivers of deforestation.

A lot of study has been done on the explanatory value of different drivers of deforestation. The table below describes briefly, some of the studies done previously and their results:

Table 7 - Conclusions from previous Spatial Regression Models about the effects of different variables on Deforestation (Table Source: Kaimowitz et.al. 2002)

Study	Country	Presence of More roads	Closer to markets	Better soils &/or drier	Nearer forest edge
Brown et al. (1993)	Malaysia	NA	NA	NA	Increase
Chomitz & Gray (1995)	Belize	Increase	Increase	Increase	NA
Deininger and Minten (1996)	Mexico	Increase	NA	Increase	NA
Gastellu-Etchegorry & Sinulingga (1988)	Indonesia	NA	NA	Increase	NA
Liu et al. (1993)	Philippines	Increase	NA	NA	
Ludeke et al. (1990)	Honduras	Increase	Increase	NA	Increase
Mamingi et. al (1996)	Cameroon and Zaire	Increase	Increase*	Increase	NA
Mertens and Lambin (1997)	Cameroon	Increase	Increase	NA	Increase
Nelson and Hellerstein (1995)	Mexico	Increase	Increase	NA	NA
Rosero-Bixby and Palloni (1996)	Costa Rica	Increase	NA	Increase	Increase
Sader and Joyce (1988)	Costa Rica	Increase	NA	Increase	NA

* Only in Cameroon. No effect in Zaire

From the table, we can draw a few main observations. Almost all the studies indicate that the likelihood of forests being cleared increases when they are closer to roads (From table: Chomitz and Gray, 1995; Deininger and Minten 1996; Liu et al.,1993; Ludeke et al.,1990; Mamingi et. al. 1996; Mertens and Lambin, 1997; Nelson and Hellerstein, 1995; Sader and Joyce, 1988; Rosero-Bixby and Palloni, 1996).

Proximity to markets is the next most important variable that seems to increase the rate of deforestation. This has been noted by several of the authors above (Chomitz & Gray, 1995, Ludeke et al., 1990, Mamingi et. al, 1996, Mertens and Lambin, 1997, Nelson and Hellerstein, 1995). This indicates that urban areas (in which these markets occur) have an effect on the areas of deforestation. As seen from our results, the proximity to towns had the highest Kappa value, which follows the results of previous studies conducted on the effect of driver variables on deforestation.

Some of the other variables that have been studied in other projects include soil type and edge of the forest exposed to anthropogenic activities. Kaimowitz et al. sum up that forest fragments have a higher risk of being lost than forests in large compact areas, with those close to the forest edge especially likely to be cleared (Brown et al., 1993; Liu et al., 1993; Ludeke et al., 1990; Mertens and Lambin, 1997; Rosero-Bixby and Palloni, 1996).

Some authors have explained that there is greater forest clearing in areas with good soil (Chomitz & Gray, 1995, Deininger and Minten, 1996, Gastellu-Etchegorry & Sinulingga, 1988, Mamingi et. al, 1996, Rosero-Bixby and Palloni, 1996, Sader and Joyce, 1988.) In this project, we have not looked at variables like soil type and fragmentation due to the lack of available data for the study area.

The results of our project indicate that the driver 'proximity to towns' had the highest Cramer's V value (Rajan D.K, *Technical Report,* 2010) meaning it was the 'most explanatory' variable for the change of land use. This means that the presence of towns coincided with the areas of deforestation more often than any other driver. This supports evidence that deforestation occurs in areas where towns and congregations of people are present. This follows the trend described by authors above.

The value of modelling land use change

This exercise has shown us the value of modelling land use change. The results give us an understanding of the drivers of deforestation in the area, as well as the rates of expected deforestation if development activities ***stayed the same***. This is important to consider because there was no data input for the new roads or towns that would almost certainly be constructed. The LCM allows one to model expected infrastructural changes like roads or buildings. This is treated as a dynamic variable, hence reflected the reality of situations where there is change in infrastructure over time. However, as this information was not available for this project, a scenario described by this model would be a conservative estimate of expected deforestation. Brown et.al. (2007) suggest that a 10 year period is the furthest projection into the future one should consider. Having said that, the model would provide the user with a good understanding of the qualitative and quantitative levels of change in the study area (i.e – spatial and non-spatial information about the rate of deforestation). With this information, the processes leading to deforestation can be worked out, allowing for a well-planned initiative towards reducing the rate of deforestation.

Visual versus automated map comparison

The figure below shows how visual comparison might be done. For example, the reference map (D) is used to compare how the different models have predicted deforestation (A, B and C). As one might immediately conclude, it is dependent on each individual's interpretation. This will therefore change from one individual to another. For this reason, we have considered an automated process for the purpose of comparing maps.
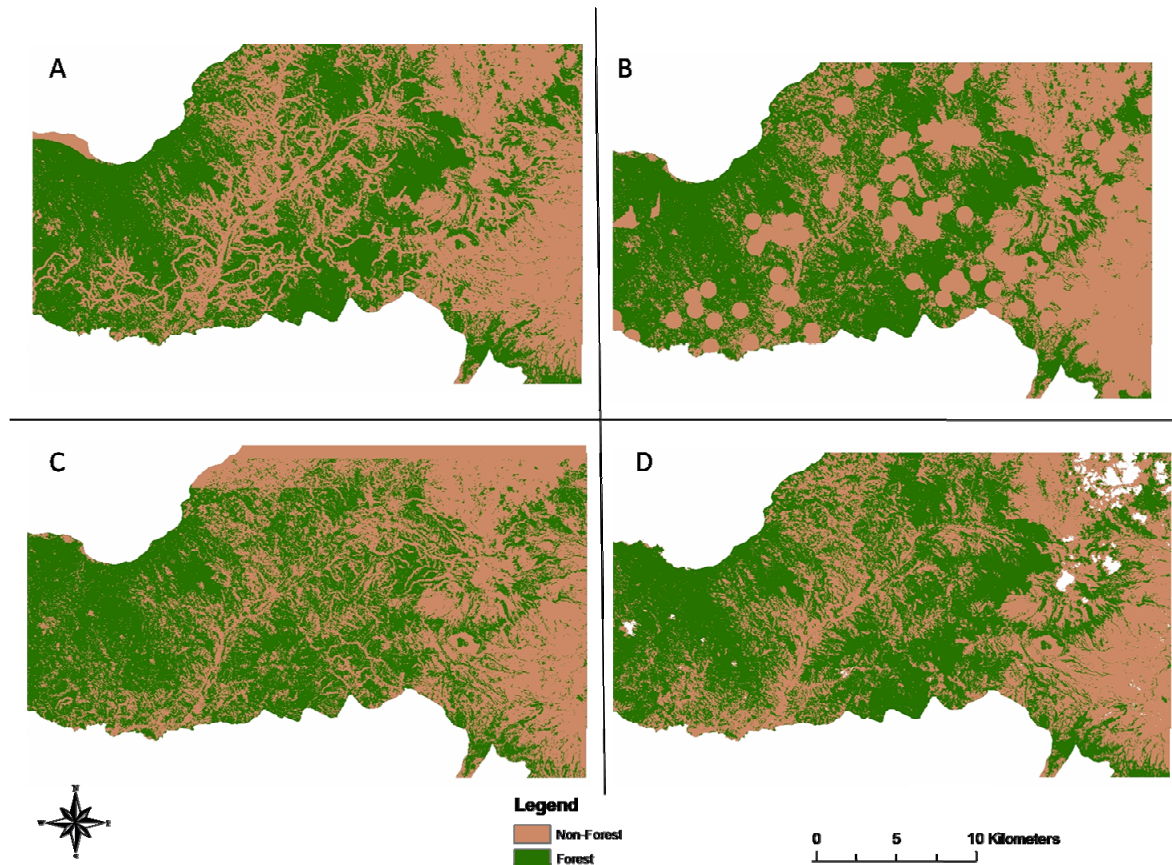

Figure 13 – Model land cover generated for 2006 using A) only roads; B) only towns; C) only slope. D) Land cover map based on LANDSAT image for 2006

In this project, the Map comparison toolkit represents an important stage of the methodology. The validation was made possible by the fact that one could perform automated comparisons between maps in very little time. The toolkit also offered the ability to generate Kappa statistics to compare how different models had performed. The applications of automated map comparison are growing, and with better development, these will prove to be very useful tools in the near future.

An assessment of the LCM as a modeller for deforestation

One of the results of the project is to assess the accuracy and reliability of the LCM as a tool for modelling change. Brown S. (2003) gives four criteria that can be used to assess a model of land use change: (1) Ease of operation and repeatability of process; (2) Accuracy and precision of model calibration and model validation; (3) Ability to deal with

multiple scales and multiple land uses; (4) Cost effective with data needs, time and skill required.

(1) The biggest advantage of the software is the simplicity of the use interface. It is very easy to operate, and efficient in producing models. The process is easily repeated without having to go through too many pre-model steps over and over again.

(2) The results of this project show that the LCM is quite accurate at predicting change (80% for the fraction correct and Kappa value of 0.62 when using all drivers). When using towns as the single driver, it resulted in model with a higher fraction correct and Kappa value than the combination of drivers (when compared with a reference 2006 land cover map).

(3) The software was tested using different resolutions (20m slope and 90m slope). This demonstrates the flexibility of the model in dealing with different resolutions. Although we were only interested in deforestation, the model allows the user to analyse change in multiple land use classes.

(4) One of the primary difficulties with the Land Change Modeler is that data has to be corrected and formatted in a particular fashion. This formatting was one of the most time-consuming processes in the software program. The lack of easily available help for any problems encountered during the modelling process made the task even more time consuming. However, once data was in the right format, the time period for model generation is very low compared to other models (GEOMOD, SLEUTH etc.) Pontius gives a comprehensive assessment of the various models in his paper. (Pontius R.G 2004)

One of the unique features of the Land Change Modeler is that it allows for the addition of constraints maps. These maps are used to define areas where land use change will not occur (e.g. protected areas). The constraints map restricts the change in land use in demarcated areas, therefore reflecting real situations where one would not expect deforestation to occur in protected areas.

The Land Change Modeler also allows the user to allow for infrastructural changes like the development of roads and towns. This feature is important when modelling long term changes. This represents the dynamic nature of land use change that is usually observed in reality.

6. Limitations of this study

As with any study, this one has its limitations. One must identify and accept these limitations so as to understand the results in the right context. The following are the significant limitations of this study:

1. Data

There is a limitation of driver data in this study. We have only considered three patterns that may influence deforestation (proximity to roads, proximity to towns and

slope). More drivers might have produced different results. However, previous studies have found that these three drivers are often influential on the patterns of deforestation.

2. Drivers

In many cases of land use change, the major driver of change is often one that is not quantifiable. For example, land ownership or change in forest policy is a driver that cannot necessarily be quantified and fed into a model. In these cases, history is not a good indicator of the future, and hence generating a model becomes difficult. This is because changes in land ownership do not follow a set pattern, wherein past trends act as indicators of future trends. Existing models do not incorporate this information in the generation of the model simply because it is difficult to quantify such variables. Lambin (1994) describes the plethora of factors which one may not necessarily be able to quantify, that are responsible for land use change. For instance, political instability, change in land ownership, war, environmental factors, change in law (e.g. decommissioning of a reserved area) are all factors that impact land use change. In this project the emphasis is on quantifying factors to model change. As a result, the factors mentioned above have not been taken into consideration in the study.

3. Image processing

This study was not about finding the most optimal image processing technique for land cover classification. Nor was it to generate the most accurate land use classification. On the contrary, it was about using these data to generate predictive models to aid research into rates of deforestation (why and how it occurs). Errors due to misclassified pixels will always exist. In this project these errors have not been validated and explicitly acknowledged (through a confusion matrix). The only image processing undertaken by this author was to classify the 1996 LANDSAT 5 image. This was only to allow for a prediction that could be validated (as the 1996-2001 images were used to produce a model for 2006).

7. Conclusion

With a better understanding of the drivers that affect land use change, we can develop better models that give us a credible picture of the levels of deforestation in different areas of the world. By analysing factors that influence the rates of deforestation, conservation methods can be made more efficient. This project was an attempt at exploring the factors affecting deforestation in Ecuador and generating such a model using the Land Change Modeler software.

The main findings were that a model was developed that estimated rate of deforestation for the area of Ecuador to be around 1.5% per annum (when compared to 1.9% by change detection in a previous study). Proximity to towns was the most important driver with a 62% agreement with the reference map for deforestation in 2006. Adding further drivers did not improve the agreement. On comparing different resolutions of slope data, it was found that the finer 20m resolution DEM yielded a higher percentage of agreement than the coarser 90m DEM data.

References

Abcli, W.S., Nsolomo, V.R. & Morogoro, T., 1998. Causes of Deforestation in Tanzania and Possible Measures to Take. In *Proceedings of the first annual Forestry Research Workshop*. p. 249.

Agarwal, Chetan; Green, Glen M.; Grove, J. Morgan; Evans, Tom P.; Schweik, Charles M. 2002. A review and assessment of land-use change models: dynamics of space, time, and human choice. Gen. Tech. Rep. NE-297. Newtown Square, PA: U.S.

Andrienko, G.L. & Andrienko, N.V., 1999. Interactive maps for visual data exploration. *International Journal of Geographical Information Science*, 13(4), 355–374.

Angelsen A. and Kaimowitz D., "Rethinking the causes of deforestation: lessons from economic models," The *World Bank Research Observer* 14, no. 1 (1999): 73.

Angelsen, A., 2008. REDD models and baselines. *International Forestry Review*, 10(3), 465–475.

Angelsen, A. et al., 2008. What is the right scale for REDD. *Moving Ahead with REDD: Issues, Options, and Implications*.

Bebbington, A., 1993. Sustainable livelihood development in the Andes: Local institutions and regional resource use in Ecuador. *Development Policy Review*, 11(1), 5–30.

Bolin, B. et al., 2000. Global perspective. *Land Use, Land Use Change and Forestry*, 23–51.

Bromley, R., 1981. The colonization of humid tropical areas in Ecuador. *Singapore Journal of Tropical Geography*, 2(1), 15–26.

Brown S, Masera O, et al (2000) Project-based activities. In Watson RT, Noble I.R., et al (eds) Land use, land-use change, and forestry; Special Report to the Intergovernmental Panel on Climate Change, Cambridge University Press (UK), Ch. 5: 283-338

Brown, S., 2003. Finalizing Avoided-Deforestation Project Baselines. *Final Report for the US Agency for International Development*.

Brown, S. et al., 2007. Baselines for land-use change in the tropics: application to avoided deforestation projects. *Mitigation and Adaptation Strategies for Global Change*, 12(6), 1001–1026.

Cañadas-Cruz, L., 1983. El mapa bioclimático y ecológico del Ecuador. *Quito; Banco Central del Ecuador, 1983, 210 p.*

Commander, S., Peek, P., 1986. Oil exports, agrarian change and the rural labor process: the Ecuadorian Sierra in the 1970s. *World Development* 14(1), 79±96.

Clarke, K.C., Hoppen, S. & Gaydos, L., 1997. A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area. *Environment and planning B*, 24, 247–262.

Chomitz, K.M. & Gray, D.A., 1995. Roads, lands, markets, and deforestation: a spatial model of land use in Belize. *World Bank Economic review,* 1996.

Dale, V.H. et al., 1993. Causes and effects of land-use change in central Rondônia, Brazil. *Photogrammetric Engineering and Remote Sensing*, 59(6), 997–1005.

Deininger, K. & Minten, B., 1997. Determinants of forest cover and the economics of protection: An application to Mexico. *American Journal of Agricultural Economics*, 84(4), 943–960.

Denman KL: Couplings between changes the climate system and biogeochemistry Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge; 2007.

Dickinson R. E and Henderson-Sellers A., "Modelling tropical deforestation: A study of GCM land-surface parameterizations," Quart. J. Roy. Meteor. Soc 114 (1988): 439– 462.

Dushku A. and Brown S., "Spatial Modeling of Baselines for LULUCF Carbon Projects: The GEOMOD modeling approach," in 2003 International Conference on Topical Forests and Climate Change:" Carbon Sequestration and the Clean Development Mechanism, 2003.

Gastellu-Etchegorry, J.P. & Sinulingga, A.B., 1988. Designing a GIS for the study of forest evolution in Central Java. *Tijdschrift voor economische en sociale geografie*, 79(2), 93–103.

Geist, Helmut J. & Lambin, Eric F., 2006. What Drives Tropical Deforestation?LUCC Series Report 4. LUCC International Project Office University of Louvain Department of Geography.

Gilruth, P.T., Hutchinson, C.F. & Barry, B., 1990. Assessing deforestation in the Guinea Highlands of West Africa using remote sensing. *Photogrammetric engineering and remote sensing*, 56(10), 1375–1382.

Godfrey, B.J., 1990. Boom towns of the Amazon. *Geographical Review*, 80(2), 103–117.

GOFC-GOLD: Reducing greenhouse gas emissions from deforestation and degradation in developing countries: a sourcebook of methods and procedures for monitoring,

measuring and reporting. GOFC-GOLD Report version COP14-2, GOFC-GOLD Project Office, Natural Resources Canada, Alberta, Canada; 2009.

Henderson-Sellers A. et al., "Tropical deforestation: Modeling local-to regional-scale climate change," Journal of Geophysical Research-Atmospheres 98, no. D4.

Imbernon, J. & Branthomme, A., 2001. Characterization of landscape patterns of deforestation in tropical rain forests. *International Journal of Remote Sensing*, 22(9), 1753–1765.

IPCC: Good practice guidance for land use, land-use change and forestry The Intergovernmental Panel on Climate Change. IPCC/IGES, Hayama, Japan; 2003.

Kaimowitz, D. & Angelsen, A., 1998. *Economic models of tropical deforestation: a review*, Cifor.

Kerr, J.T. & Ostrovsky, M., 2003. From space to species: ecological applications for remote sensing. *Trends in Ecology & Evolution*, 18(6), 299–305.

Lambin E. F, Department of Agriculture, Forest Service, Northeastern Research Station. 61p. "Modelling deforestation processes: a review," European Commission, Luxemburg (1994).

Lambin, E.F., 1997. Modelling and monitoring land-cover change processes in tropical regions. *Progress in physical Geography*, 21(3), 375.

Land Change Modeler (2009). Clark Labs, Clark University, Worcester, MA. http://www.clarklabs.org

Land Change Modeler Tutorial

Liu, D.S., Iverson, L.R. & Brown, S., 1993. Rates and patterns of deforestation in the Philippines: application of geographic information system analysis. *Forest Ecology and Management*, 57(1-4), 1–16.

Liverman, D.M., 1998. *People and pixels: Linking remote sensing and social science*, National Academies Press.

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(3), 247–259.

Map Comparison Kit (2007). Netherlands Environmental Assessment Agency and Research Institute for Knowledge Systems, Netherlands. http://www.riks.nl/mck/

Madeira, E.C., 2007. Policies to Reduce Emissions from Deforestation and Degradation (redd) in Developing Countries.

Mena, C.F., Bilsborrow, R.E. & McClain, M.E., 2006. Socioeconomic drivers of deforestation in the northern Ecuadorian Amazon. *Environmental management*, 37(6), 802–815.

Meyer, W.B. & Turner, B.L., 1994. *Changes in land use and land cover: a global perspective*, Cambridge Univ Pr.

Mertens B. and Lambin E. F, "Spatial modelling of deforestation in southern Cameroon Spatial disaggregation of diverse deforestation processes," Applied Geography 17, no. 2 (1997): 143–162.

Nelson, G.C. & Hellerstein, D., 1997. Do roads cause deforestation? Using satellite images in econometric analysis of land use. *American Journal of Agricultural Economics*, 79(1), 80.

OECD/IEA (2003) Forestry projects: lessons learned and implications for CDM modalities. *OECD/IEA Information Paper*

Pahari K. and Murai S., "Modelling for prediction of global deforestation based on the growth of human population," ISPRS Journal of Photogrammetry and Remote Sensing 54, no. 5-6 (1999): 317–324.

Peck M., 2008. Developing a habitat action plan for *Ateles fusciceps,* University of Sussex

Peck M., (2010) "Establishing a baseline: Determining deforestation rates in NW Ecuador using LANDSAT satellite imagery". A report prepared for Rainforest Concern.

Pontius R.G and Neeti Neeti (2009), Uncertainty in the difference between maps of future land change scenarios. Sustainability Science Volume 5, Number 1 / January, 2010 39-50

Pontius R.G.(2008), Comparing the input, output, and validation maps for several models of land change. The Annals of Regional Science. Volume 42, Number 1. 13-35

Rudel, T.K. & Horowitz, B., 1993. *Tropical deforestation: small farmers and land clearing in the Ecuadorian Amazon*, Columbia Univ Pr.

Rajan D.K, 2010. *Understanding drivers affecting land use change in Ecuador: A case study in Ecuador using the Land Change Modeler.* Technical Report

Sader, S.A. & Joyce, A.T., 1988. Deforestation rates and trends in Costa Rica, 1940 to 1983. *Biotropica*, 20(1), 11–19.

Sánchez-Azofeifa, G.A., Harriss, R.C. & Skole, D.L., 2001. Deforestation in Costa Rica: A Quantitative Analysis Using Remote Sensing Imagery1. *Biotropica*, 33(3), 378–384.

Soares-Filho, B. et al., 2004. Simulating the response of land-cover changes to road paving and governance along a major Amazon highway: the Santarém–Cuiabá corridor. *Global Change Biology*, 10(5), 745–764.

Southgate, D., Sierra, R. & Brown, L., 1991. The causes of tropical deforestation in Ecuador: a statistical analysis. *World Development*, 19(9), 1145–1151.

Southgate, D. & Whitaker, M., 1994. *Economic progress and the environment: one developing country's policy crisis.*, Oxford University Press, New York, 150pp.

Stern N: The economics of climate change The Stern review. Cambridge University Press, Cambridge; 2007.

Sunderlin, W.D. et al., 2002. Economic crisis, small-scale agriculture, and forest cover change in southern Cameroon. *Environmental Conservation*, 27(03), 284–290.

Vanclay, J.K. et al., 1999. *Spatially explicit model of deforestation in Bolivia*, Citeseer.

Veldkamp, A. & Lambin, E.F., 2001. Predicting land-use change. *Agriculture, Ecosystems and Environment*, 85(1-3), 1–6.

Verburg, P.H. et al., 2006. Modeling land-use and land-cover change. *Land-use and Land-cover Change: Local Processes and Global Impacts Eds EF Lambin, HJ Geist (Springer, Berlin) pp*, 117–137.

Verchot, L., Petkova, E. & BOGOR, I., 2009. The state of REDD negotiations: consensus points, options for moving forward and research needs to support the process. *Unpublished manuscript. CIFOR, Bogor, Indonesia*.

Walsh, S. et al., 2002. Characterizing and modeling patterns of deforestation and agricultural extensification in the Ecuadorian Amazon. *Linking people, place, and policy: a GIScience approach*, 187–214.

Zamyatin, A. & Mikhaylov, P., Technology of suitability maps designing for land use and land cover change modelling.

Zamyatin, A. & Markov, N., 2005. Approach to Land Cover Change Modelling Using Cellular Automata. In *Proceedings of 8th Conference on Geographic Information Science.–Estoril.–Portugal.–AGILE*. pp. 587–592.

Zevallos J.V., 1989. Agrarian reform and structural change: Ecuador since 1964. In: hiesenhusen, W.C. (Ed.), Searching for Agrarian Reform in Latin America. Unwin Hyman, Boston, pp. 42-69.

Part II

Technical Report

Understanding the drivers affecting land use change in Ecuador: an application of the Land Change Modeler software

Technical Report

DHRUVA RAJAN

School of GeoSciences, Drummond Street, University of Edinburgh, Edinburgh EH8 9XP, UK

Summary

This document complements the research paper submitted by Dhruva Rajan, 2010. The information provided here supports and details the methods used. The following are included:

- Detailed description of datasets used

- Detailed description of data processing

- An introduction and working of the Land Change Modeler

- Working of the Map Comparison Kit

- Extensive review of results

- Suggestions for future work

TABLE OF CONTENTS

List of Figures

List of Tables

1. Introduction

This report is intended for those who wish to understand more about the processes undertaken to achieve the results described in the research paper.

The flow diagram below (Figure 1) describes the overall methodology.
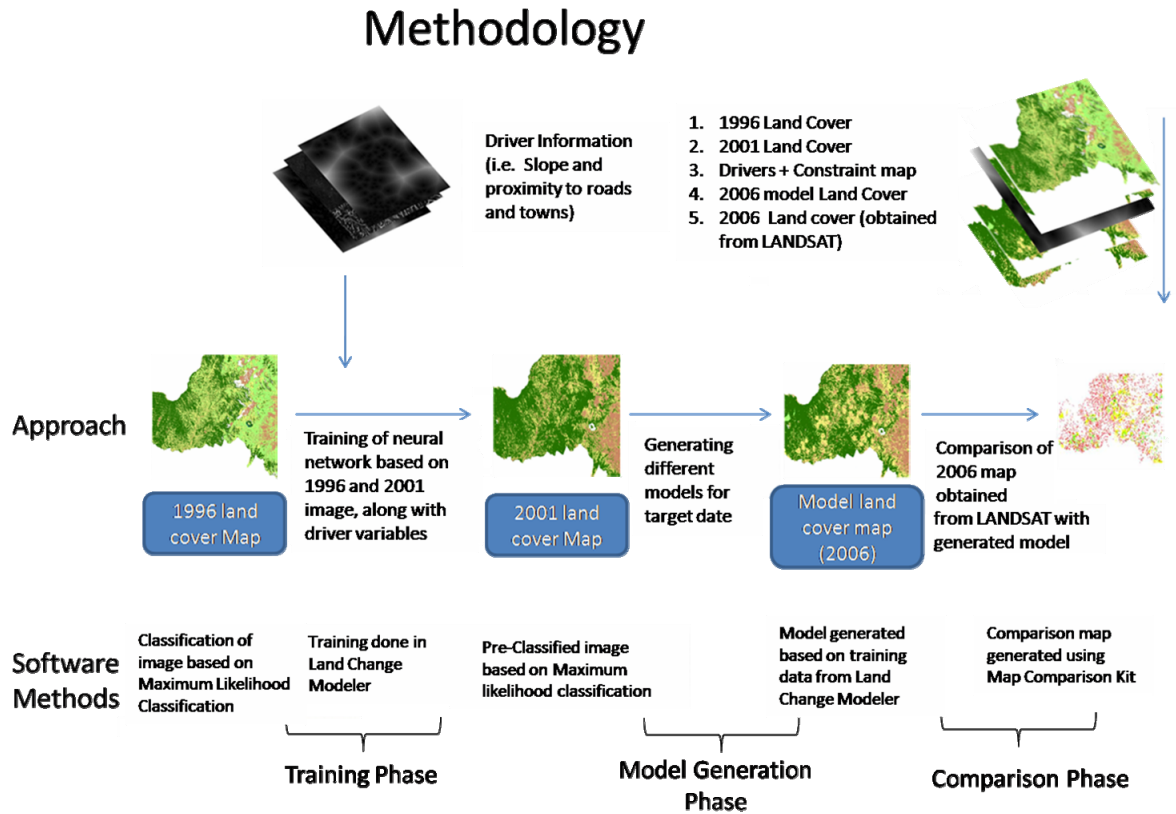
Figure 1 – Flowchart describing major steps in the methodology

1.1 Overall Approach

The overall methodology as outlined above (Figure 1) consisted of 3 major steps. First, the LANSAT 5 image for 1996 was classified using Maximum Likelihood Classification to obtain a land cover map. This map and one for 2001(which was already classified by Dr.Mika Peck (of Rainforest Concern)), along with driver data (raster maps of proximity to road, towns and slope calculation) and a constraint map (preventing deforestation in protected areas) were used in the Land Change Modeler to train the artificial neural network, in order to aid the production of a model for the later target date. (Training phase in Figure 1)

Next, the training data was used by the LCM to generate a model of deforestation for the desired date. This involved using the LCM to create transition potentials, which were then used for the final model generation. (Model Generation phase in Figure 1)

Finally, the Map Comparison Kit was used to validate the model produced. Using a classified land cover map for 2006, the level of deforestation was compared with the model produced for 2006. (Comparison phase in Figure 1)

1.2 Data Sets

The data sets and sources for the drivers are outlined in Table 1. The data set and sources for the LANDSAT images are given in Table 2.

Table 1 – A summary of the data sets used in the modelling software

Data Set	Description	Thumbnail	Projection	Resolution	Source
Proximity to Roads	Vector converted to Raster; Euclidian distance to the roads.		WGS 1984 UTM Zone 17 N	30m	Instituto Geografico Militar, Ecuador
Proximity to Towns	Vector converted to Raster; Euclidian distance to the towns.		WGS 1984 UTM Zone 17 N	30m	Instituto Geografico Militar, Ecuador
Slope	Slope calculated from Digital Elevation Model		WGS 1984 UTM Zone 17 N	20m	Instituto Geografico Militar, Ecuador
Slope	Slope calculated from Digital Elevation Model		WGS 1984 UTM Zone 17 N	90m	SRTM (2000) website
Protected area/ constraints map	Vector converted to Raster; Reclassified to have 0 value for protected area and 1 for all other areas		WGS 1984 UTM Zone 17 N	30m	Instituto Geografico Militar, Ecuador

Table 1a - Data Format

***Extent** - Top*	*57968.5845163*
Left	*749920.478445*
Right	*806380.478445*
Bottom	*-1251.4154837*
Number of columns	*1882*
Number of rows	*1974*

Table 2 – A summary of the LANDSAT images used

Path	Row	Date	Extra Comments
10	**60**	**24 07 1996**	Source for this data set was the Global Land Cover Facility (www.landcover.org)
10	**60**	**03 11 2001**	Source for this data set was Dr. Mika Peck (2009)
10	**60**	**13 08 2006**	Gap filled with 09 07 2005 SLC – Off and 03 11 2001 SLC- ON Source for this data set was Dr. Mika Peck (2009)

NOTE – Of the above images, the 2001 and 2006 images were obtained as pre-classified images (using MLC)

1.3 Choice of software

In the course of producing and testing the model, different software packages were used. These are outlined in the flow chart below:
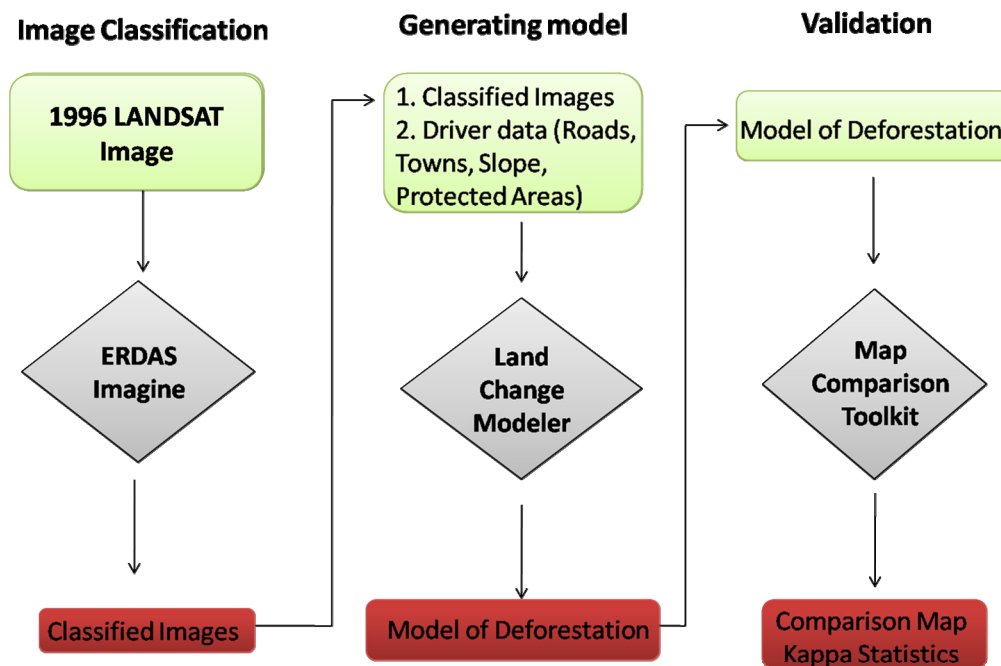
Figure 2 – Flowchart showing software packages used

In the first step, LANDSAT images were classified using ERDAS Imagine (ver. 9.1). Although two images were already classified (Peck.M., 2008), a third image had to be classified in order to validate the performance of the model. This image was classified using the same land cover classes used in the first two images.

Two images, the 1996 and the 2001 (classified as land cover maps), were used (along with the driver maps) as input for the Land Change Modeler. The LCM is available as a plug-in for Arc GIS ver. 9.2 or higher.

The LCM uses the data to produce a model of deforestation for the study area. In our case it was used to generate a model of deforestation for 2006. The Map Comparison Toolkit (MCK ver. 3.2.0) was used to validate the results by comparing the model generated with the pre-classified 2006 LANDSAT image. The tool was developed by the Research Institute for Knowledge Systems.

The MCK generates a comparison map, which identifies similar and dissimilar areas between the two maps. It also generates Kappa statistics for the comparison, which allows statistical validation in addition to a visual one. The MCK is also used in this project to specifically look at only the changes to the forest class. This is done by using the per category algorithm (explained in section 4).

2. Image Classification

The aim of the classification process is to categorize all pixels in an image into one of several land cover classes. Normally, multispectral data are used to perform the classification and the spectral pattern present within the data for each pixel is used as the numerical basis for categorization (Lillesand and Kiefer, 1994). The objective of image classification is to identify and portray the features occurring in an image in terms of the type of land use these features actually represent on the ground.

The process of classification consists of two stages. The first is the recognition of 'classes' of real-world objects. In the context of remote sensing of the land surface, these categories could include, for example, forest, water, pasture etc. The second stage in the classification process is the labelling of the entities (normally pixels) to be classified. These labels are normally numerical, so that a pixel that is recognised as belonging to the class 'water' may be given the label '1', 'forest' may be labelled '2', and so on. The process of image classification requires the user to perform the following steps:

"(i) Determine beforehand, the number and nature of the categories in terms of which the land cover is to be described

(ii) Assign numerical labels to the pixels on the basis of their properties using a decision-making procedure, usually termed a classification rule or a decision rule."

(Mather, P.M., 2004)

Two main classification methods are: Supervised Classification and Unsupervised Classification. In this project, the supervised classification method has been used to classify the image.

2.1 Supervised Classification

With supervised classification, examples of the land cover type of interest in the image are identified. These are called "training sites". The image processing software is then used to develop a characteristic reflectance value for each information class (i.e. land use class). Once this has been achieved for each land use class, the image is then classified by examining the reflectance for each pixel and making a decision about which of the signatures it resembles most. (Eastman, 1995)

2.1.1 Maximum likelihood Classification

"Maximum likelihood Classification is a statistical decision criterion to assist in the classification of overlapping signatures; pixels are assigned to the class of highest probability. The maximum likelihood classifier is considered to give more accurate results than parallelepiped classification however it is much slower due to extra computations." (Mather, P.M., 2004) In the classification of the 1996 LANDSAT image, Maximum Likelihood Classification was used to classify the different land use classes in the image.

2.2 Identification of areas of known land use classes

For the 1996 image, supervised classification was carried out. Training areas for the supervised classification were identified using previous field knowledge. This was done using aerial photography and field observations. The data was obtained from Rainforest Concern's scientist, Dr.Mika Peck (mikapeck@yahoo.com)

The training sets for the image were identified with advice from Dr. Peck. First, the different types of land cover were identified and then selected using the AOI polygon tool in ERDAS Imagine. The sample pixels were selected using polygons to obtain multiple pixels within a single area. Then, a snapshot of the sample training pixels was sent to Dr. Peck as an image. He then verified the accuracy of the training classes and suggested the changes that had to be made. These were taken into consideration before the final classification was done.
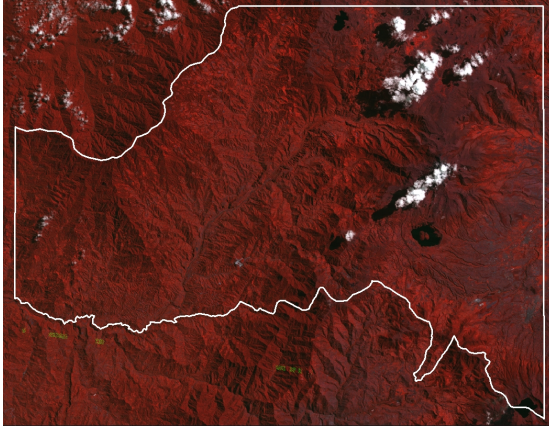
Figure 3 - Unclassified LANDSAT image (1996) with the outline of study area displayed as a 4, 2, 3 RGB false colour composite.

The specific land use classes were classified based on this field knowledge. Supervised classification was performed using parametric rule based decision with a maximum likelihood classification. The maximum likelihood rule was chosen as it uses surrounding pixels to determine the value for the pixel in question. (Mather P., 2004)

Table 3 – A summary of the land cover classes identified for 1996 image

Land Use Class	Category Code	% Study Area
Cloud Shadow	1	0.26
Water	2	0.06
Urban/Cleared	3	7.5
Pasture	4	28.2
Agriculture	5	21.7
Forest	6	42.1

The image below shows the classified image, clipped to the extent of our study area.
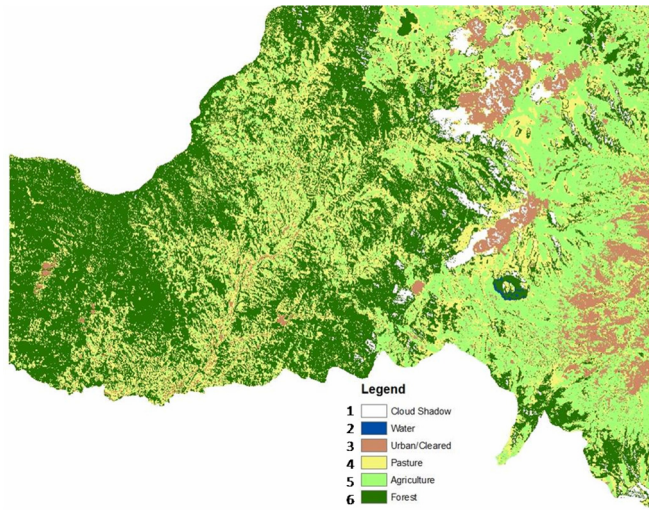
Figure 4 - Classified image of study area 1996 (Land Cover Map)

3. Land Change Modeler

The Land Change Modeler (www.clarklabs.org – Student 1 year licence) was used to model changes in land use, using two maps in chronological order. In our case, land cover maps produced based on LANDSAT 5 (24 07 1996) and LANDSAT 7 (03 11 2001) are used as the training maps. Once the Artificial Neural Network was trained with the two images, a prediction for 2006 was generated. The third LANDSAT 7 image (13 08 2006) was used to validate the model generated. The steps below outline the process of modelling change in land use based on two LANDSAT images.

3.1 Load start and end images

The first step involved identifying the input images for the modeller. The input images used were land cover maps based on classified LANDSAT 5 (24 07 1996) and LANDSAT 7 (03 11 2001) image. Before using the data with the Land Change Modeler, one must ensure that both images have the following characteristics:

- Classes

- Projection

- Extent

- Resolution

The Land Change modeler only accepts raster datasets which are either 8-bit unsigned, 6-bit signed or 32-bit float. Thus the Copy Raster function in ArcGIS was used to convert the LANDSAT Images to 8-bit unsigned format.

Clipping LANDSAT Image – The image taken from the Glovis (www.landcover.com) server was much larger than required. In order to clip the raster to the exact shape required, the Spatial Analyst was used.

First, an analysis mask was set for the Spatial Analyst. This can be done by going to options and general tab in the Spatial Analyst. The mask used was a shape file of the shape of the part of Imbabura (Imbabura_Clip.shp) district that was under study. By using a mask, one can cut raster images to the exact shape as that of a vector.

Once the mask has been set, the cell size was set to 30, and then using the Raster Calculator, the raster is evaluated. The evaluate function just multiplies every cell with 1. Since an analysis mask had been imposed, the resulting analysis was carried out only within the area that corresponded to the extent of the Imbabura shapefile.

The remaining raster images were also clipped this way to ensure uniform extent and shape.

The images also had to be in the same projection (WGS 1984 UTM Zone 17 N) and contain the same number of rows (1882) and columns (1974) in the raster data set. ArcGIS was used to manipulate the raster data to arrive at same size and projection.

Data Input   **Data Output**
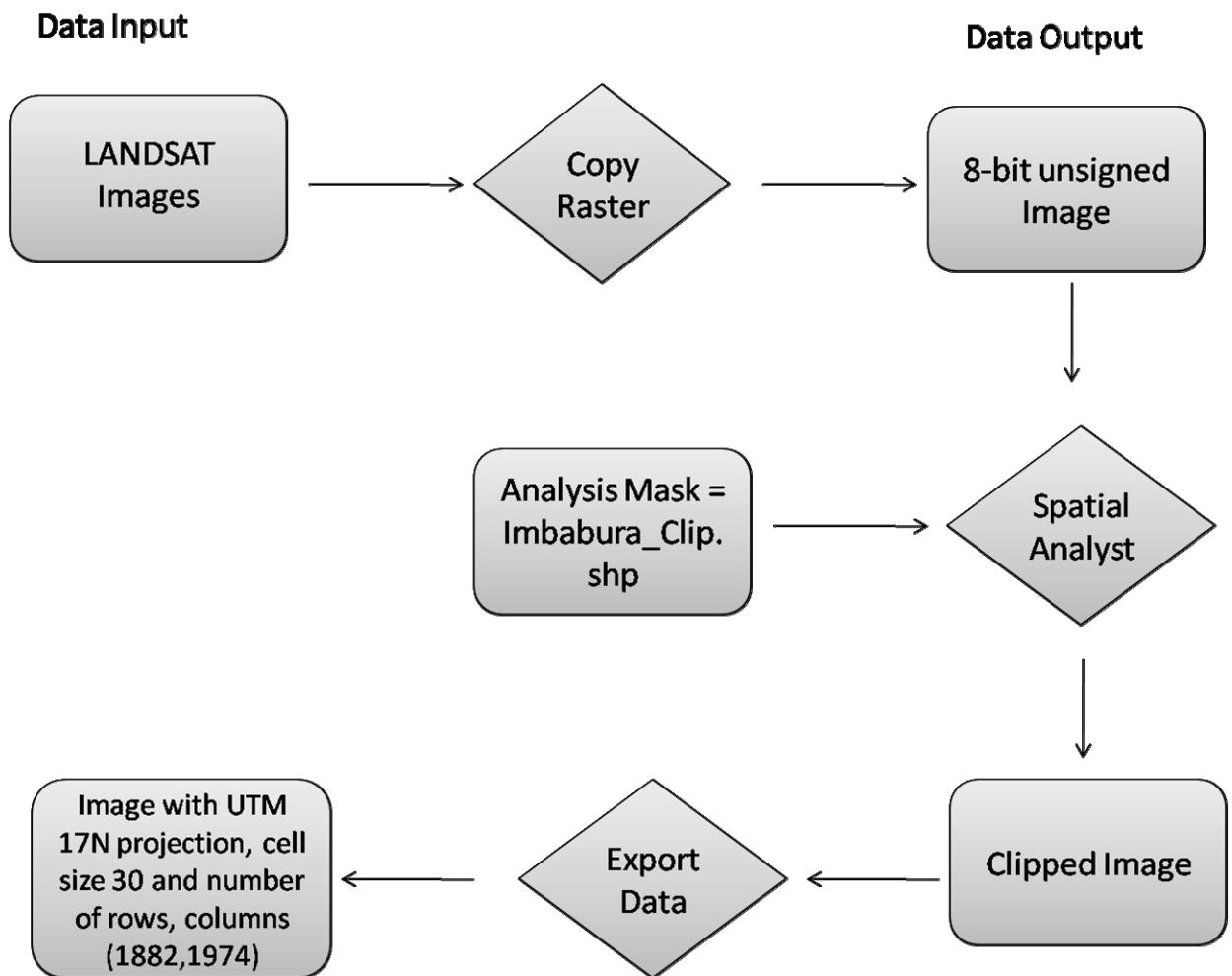
Figure 5 - Flowchart describing the steps in preparation of the data

3.2 Change Analysis

The next step is to quantify change that has occurred in the images from 1996 to 2001. This helps understand areas of change, and therefore build the change prediction. The Land Change Modeler does this by automatically identifying specific pixels that have undergone a change from the first image to the second one. The Land Change Modeler also can identify specific land cover classes that have undergone change. For example, one can analyse only the forest class to identify which land cover classes increased with a corresponding decrease in the forest class. This way, one can identify possible causes for deforestation.

This functionality was used to identify the land use classes that contributed most towards deforestation. By analyzing the different land classes, it became clear that deforestation did not result in any direct increase in the urban or cleared areas. Rather, the process was more staged. Forest was first converted to Pasture, and then Pasture to Agriculture. Finally, Agriculture became urban land. This process was an important finding to enable conclusions about the processes leading to deforestation in the area.

Once the amount of change has been identified, the locations undergoing this change can be mapped. This change map is then used by the Land Change Modeler in its prediction

model. This information goes into the training of the neural network, and eventually affects the resulting model. The areas of change indicate where further change might occur. This is because the neural network works in such a way that, the areas of past change influence areas of future change. Therefore the generation of the change map is an important step in the modelling process.
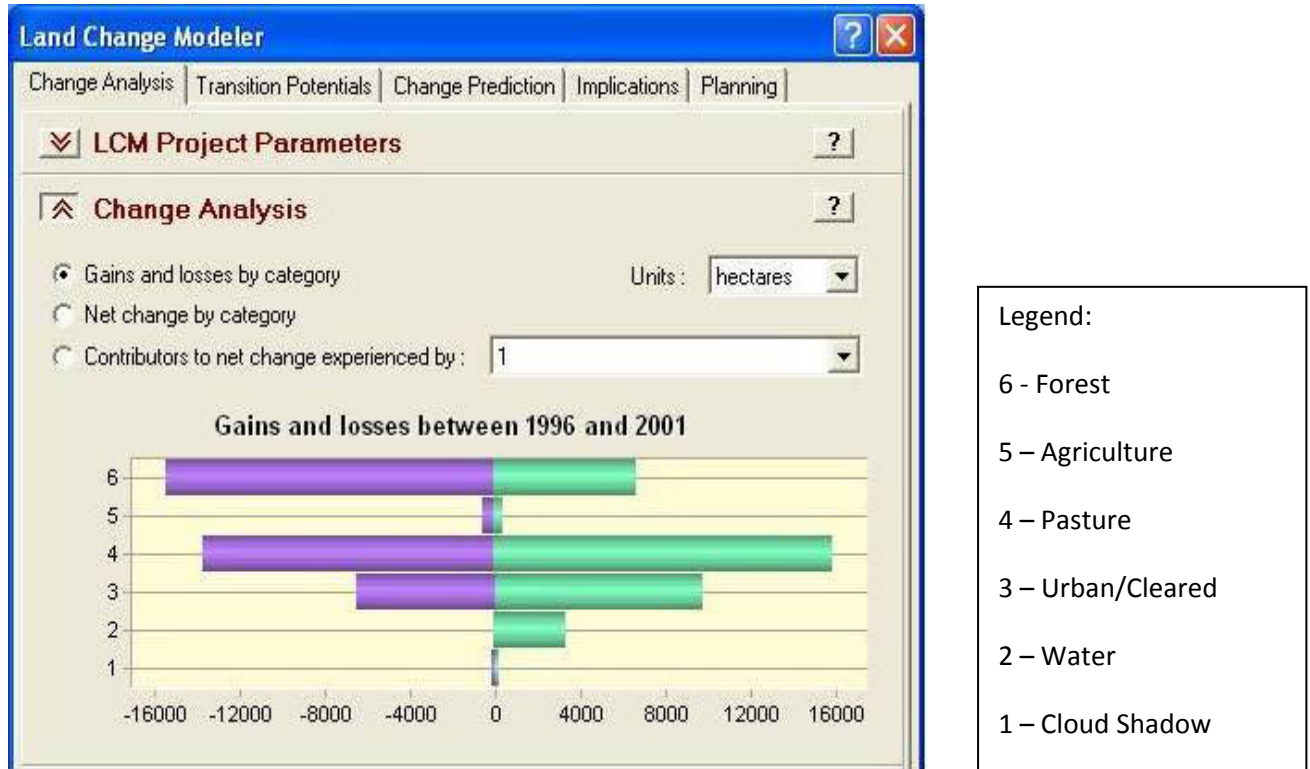
Figure 6 – Change analysis showing gains and losses

The graphic above (Figure 6) produced by the LCM indicates change of land use measured from the two LANDSAT images between 1996 and 2001. The land change modeller allows the user to change these analysis settings to analyse either each land use change individually, or all the land use changes together. The above graphic shows a scenario where we are trying to compare all the changes at once. The Gains/Losses per class are shown. This allows one to summarize the changes in land use to understand whether there have been significant gross changes in the time period or not.
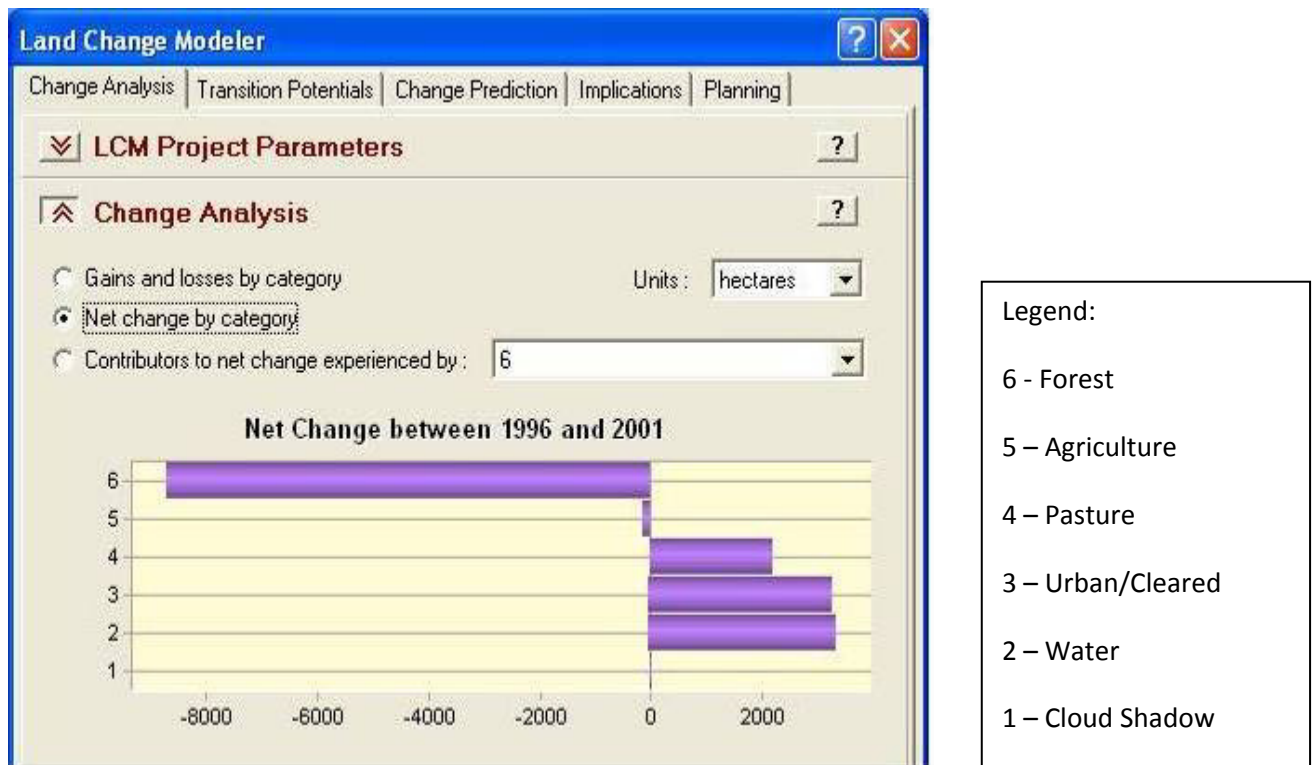
Figure 7 – Change analysis showing net change by category

This graphic (Figure 7) tells us about the change in Forest areas (class 6). This is the most important change in land use that we are interested in modelling as we use it to measuring deforestation. The modeller can measure change in hectares.

Figure 7 shows us that the forest class has reduced by almost 8500 hectares from 1996 to 2001. This is consistent with the high rate of deforestation that has been observed by authors (Peck M., 2010) in the area. There also seems to be an almost corresponding gain of nearly 7000 hectares in classes 4, 3 and 2, which are Agriculture, Urban and Water respectively. In order to explain if these changes are related to land use, one can use a simple tool within the software.

The graphic below (Figure 8) shows the results of a tool that allows one to compare the contributors to net change in a specific land use class. This allows us to pin point the conversion of land use, and hence begin to suggest some of the factors during the change.
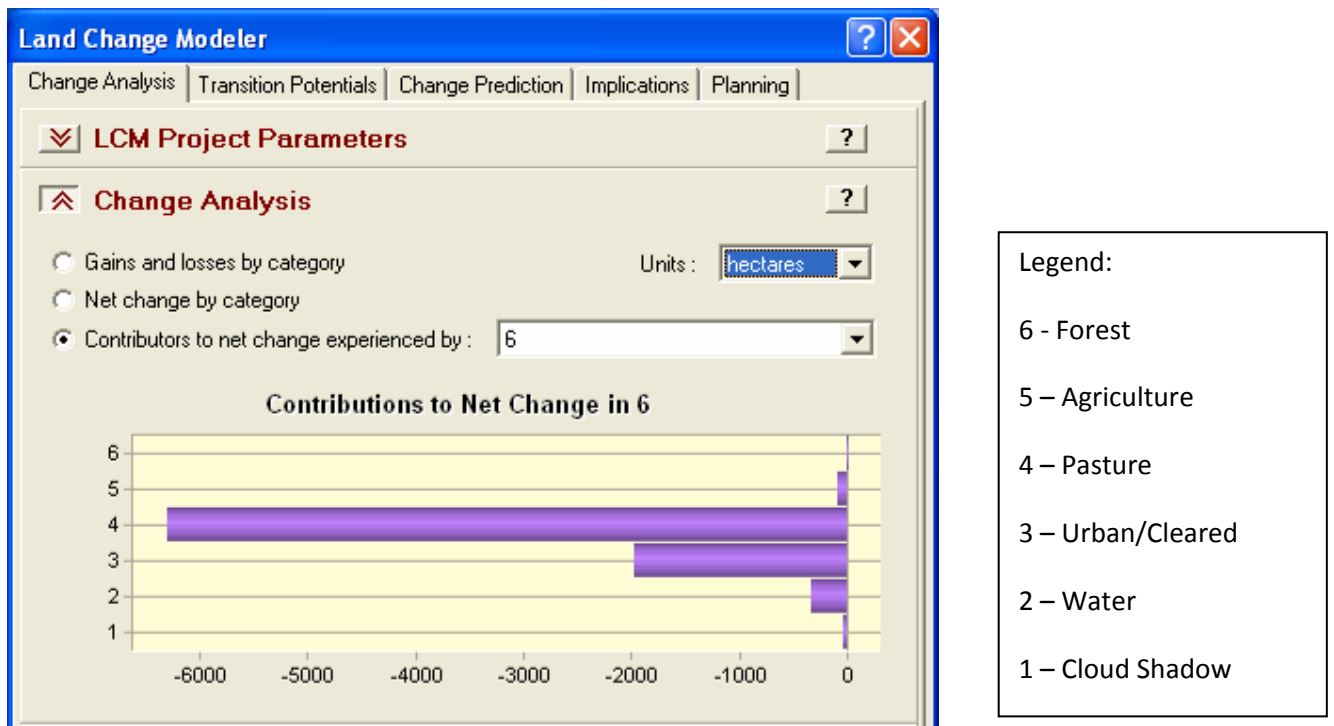
Figure 8 – Change analysis showing contributors to net change

From the graph, it becomes apparent that most of the pixels that were forest (6) in 1996 have changed to pasture in 2001. The next largest change from forest class has been to class 3 (Urban). Most forest areas that have experienced a change have changed to pasture land between 1996 and 2001.

3.3 Creating Change Maps and Maps of Trends

The LCM allows one to create maps of changes to specific classes. For instance, the map below (Figure 9) shows the change in all land use classes from 1996 to 2001. This change map was produced using a tool in the LCM.
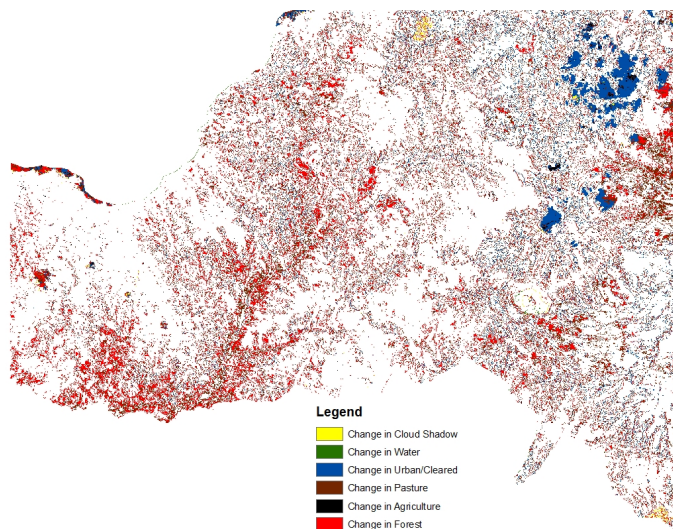
Figure 9 – Map showing change in all land use classes between 1996 and 2001

The change map shows change in land use classes. The areas of Red are the ones we are interested in, as it indicates areas of deforestation. This however, provides only a 'snapshot' of the areas of change.

To assess 'directionality' of change, so that one may predict future trends, the modeller allows one to create maps of trends. These maps indicate trends of change that have occurred from the earlier map to the later one. The map below shows an example of a trend map. These are created to explore the geography of whether certain changes are occurring in particular regions. As one can see, this representation allows interpretation of directionality of the change, as well as the magnitude of that change. This can be useful to answer questions like "in which areas do we expect to find the greatest amount of deforestation?"
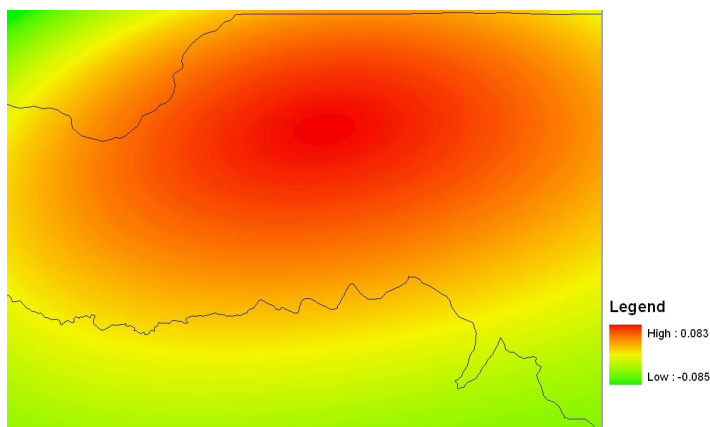
Figure 10 – Trend map showing change from Forest to Pasture
This map (Figure 10) was produced by selecting only changes in pixels from Forest class to Pasture class. It is a smoothed surface of change produced.

Dark Red areas indicate areas where there is generally a high likelihood of Forest changing to Pasture. Greener areas indicate areas where there is a lower likelihood of the change from pasture to forest.

3.4 Selecting transitions to model

The LCM allows one to select specific transitions to model. The maximum number of transitions is defined by the number of land use classes present in the image. That would be all the transitions from each class to every other class.

In this project we are interested only in deforestation, which is the transition from the forest class to other classes. Thus we have to model 5 transitions, as we have 6 land use classes. However, the transitions to cloud shadow and water have been excluded, these are artefacts of the Image Processing and neither are transitions that occur in reality.

This also reduces the errors due to misclassified pixels. This is because with fewer transitions, there is less chance of attributing change to error from wrongly classified pixels.

3.5 Adding in drivers of Deforestation

Once there is evidence that there has been some degree of change in the area, the next step is to try and map the change, in order to project it into the future for our study. The change

is predicted based on coincidences between pixels that undergo change during the training period 1996-2001 and the relationship (e.g proximity or average) of these change areas with other data sets – which may be causes or drivers of this change. These drivers are added as raster datasets. Different data were treated with different calculations before being added to the model. The drivers were chosen based on the availability of data and their expected relative importance to the rate of deforestation.

3.5.1 Road

Vector data was first converted to raster using ToRaster tool. Once the raster was obtained, a distance calculation using Spatial Analyst Tools on ArcGIS was used to calculate straight line distance from the road.

The resulting raster was exported to the desired location, with the same projection and number of rows and columns as the original landsat image. This was achieved by using the export data function in ArcGIS.

Once the data was obtained in the desired format, it could be added to the modeller.

3.5.2 Slope (0-4300m)

The slope was calculated using ArcGIS's Spatial Analyst tool. The Digital Elevation Model (Instituto Geografico Militar) for the study site was obtained at 20m resolution. The slope was calculated using the DEM. A similar calculation was performed for the 90m resolution DEM, to obtain slope at 90m resolution. The raster datasets were then exported as above, to match the resolution (in terms of number of rows and columns) of the original images and then added to the LCM.

3.5.3 Small towns (181 towns)

Small towns were obtained as point data. They were first converted to raster format, and then a distance calculation was run to find the gridded distances away from these towns. The raster was once again exported to the same projection and size as the other images.
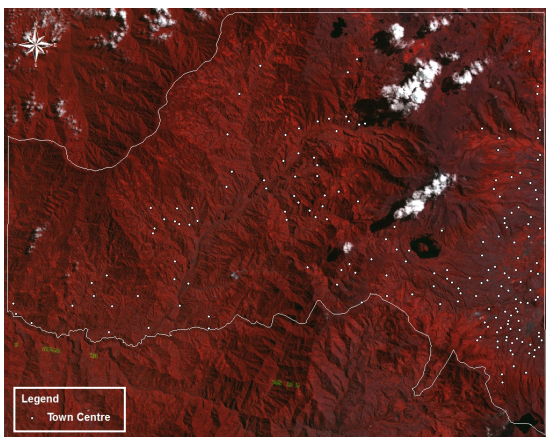
Figure 11 – Distribution of towns within study area

3.5.4 Protected areas (6)

The protected areas were added as constraints to the model after the training of the neural network. This factor is important to acknowledge because the presence of protected areas

should mean that no deforestation can take place within those areas. In order for a raster layer to be added as a protected area, it had to be re-classified.

Existing protected areas were available as polygons. They were first converted into raster using the convert tool. Once the raster was generated, it was reclassified into values of 0-1. The classified values are important here because they act as multipliers. So values of 0 on the map are treated as absolute constraints while values of 1 are unconstrained and consequently have no impact. Values less than 1 but above 0 act as disincentives while values greater than 1 act as incentives. Once reclassified, the raster was then exported with the same number of rows and columns as the other images (and the same projection).

The land change modeller models constraint maps in a different way to the way the maps of drivers are used. This is because drivers only influence the rate of change of land use. On the other hand, the constraints, such as protected areas, act to completely restrict the change in land use.
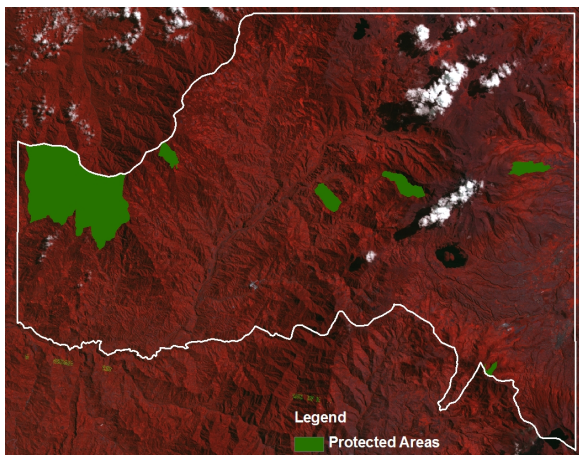
Figure 12 – Distribution of Protected Areas within study area

3.6 Cramer's V

Before the drivers can be added to the model, they are checked for the Cramer's V and p values. These values check the degree of explanatory value for each of the drivers to be added to the LCM. This driver variable test procedure is based on a contingency table analysis. For qualitative variables, it uses the originally defined categories of the variable to test association with the distribution of land cover types in the later land cover map. Quantitative variables, on the other hand, are divided to 256 categories in order to conduct this test. The Cramer's test does not give conclusive proof that a particular variable explains the change in land use. It is rather a more intuitive tool that can be used to understand the significance (statistical) that a particular variable has in influencing change.
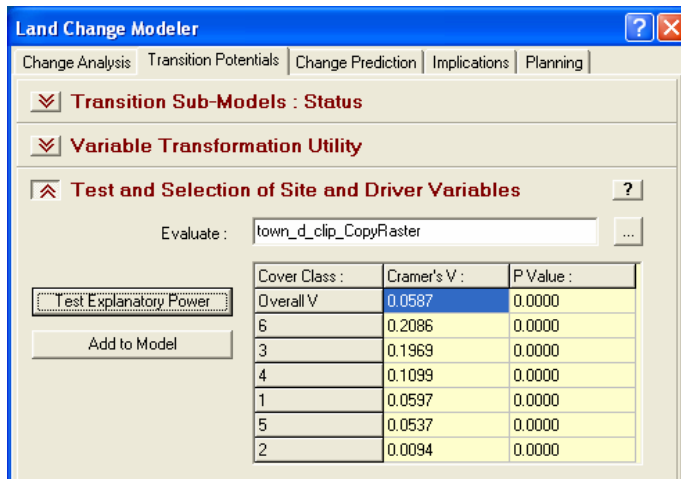
Figure 13 – Table showing Cramer's V and P values for Towns

A high Cramer's V indicates that the potential explanatory power of the variable is good, but does not guarantee a strong performance since it cannot account for the mathematical requirements of the modelling approach used and the complexity of the relationship (LCM tutorial and Help files). However, it is a good indication that a variable can be discarded if the Cramer's V is low. The p value expresses the probability that the Cramer's V is not significantly different from 0 (Note that this assumes that all pixels are independently sampled and have no spatial dependence in their values). A low value of p is not a good indicator of a variable's worth (in terms of its explanatory value), but a high value is a sure sign that it can be rejected (LCM Tutorial and Help Files). In our case, all the P values are very low. This is because the Cramer's V values are also relatively low (the overall score is 0.05 – in the case of proximity to towns). However, the variable (proximity to towns) was not discarded because further detailed data was not available for testing.

The figure below (Figure 14) shows the Cramer's V value for slopes as the explanatory driver. In this case, two p values are very high (for classes 1 and 3). This means that slope as an explanatory driver may not be appropriate to model change to urban land (because category 3 is Urban) or Cloud Shadow (which is category 1).
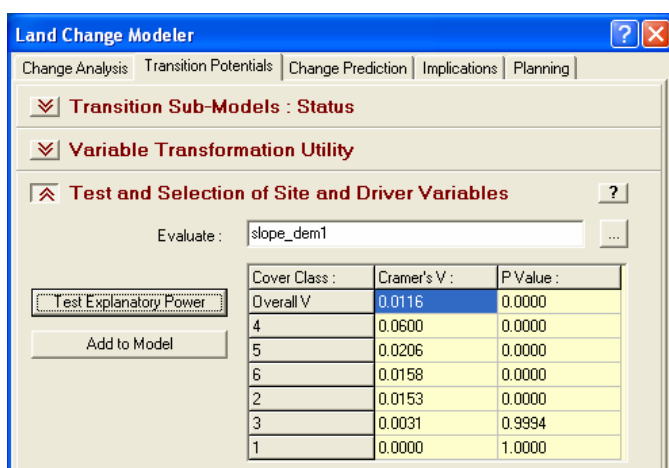
Figure 14 – Table showing Cramer's V and P values for Slope

3.7 Running the Modeller

Once the drivers with a high level of significance (based on the Cramer's V value) have been selected, the Multi – Layer Perceptron uses half the pixels that underwent the transition for training and the other half for testing of the network.

The default modelling procedure is the MLP neural network (the other option is simple linear regression) which is what is used in this analysis. The model identifies the number of cells that transitioned during the training period (in this case from 1996-2001) for the smallest transition (least number of pixels change) in the group of classes that are being modelled. It also identifies cells that could have undergone the transition but did not (i.e. persisted). This allows one to gauge the sample size to be used for the training of the MLP. Although one can select a smaller sample size, a larger sample usually yields better results (LCM tutorial).

The MLP performs the training by identifying two sets of classes. The first one consists of the pixels that underwent transition from the first land cover map to the next. The second is a set of pixels that persisted from the first land cover map to the next.

It then automatically starts the training process. The MLP uses examples obtained from the training classes and develops a function that predicts the potential for transition based on the values at any location for the driver variables. It does this by taking half the samples it was given to train on and reserving the other half for testing success.

The MLP constructs a network of neurons between the driver variables, the transitions and a web of connections between the neurons that are applied as a set of (initially random) weights. With each pixel MLP looks at from the training data, it gauges its error and adjusts the wrights. As the weights are adjusted and more actual changes in the test set are correctly predicted, the accuracy improves, and the precision improves (i.e. the RMS gets lower). When the MLP has completed training, a decision can be taken as to whether it has performed well enough or not. If the accuracy is not up to the mark (the LCM tutorial suggests 80% or higher), then the training can be re-run. Once an acceptable level of accuracy has been achieved, a map showing the Transition Potential is created. (LCM Tutorial)

While running the MLP, there are several options that can be used to change the way the neural network runs the model.

3.7.1 Automatic Training mode

When calculating transition potentials, the Land Change Modeler launches MLP in a special automatic training mode. This automatic mode monitors the start and end learning rate of the neural network. The learning procedure starts with an initial learning rate and reduces it progressively over the iterations until the end learning rate is reached when the maximum number of iterations is reached. "If significant oscillations in the RMS error are detected after the first 100 iterations, the learning rates (start and end) are reduced by half and the process is started again." (LCM Tutorial)

All other parameters of the MLP are used by the Land Change Modeler at their default values. However, Land Change Modeler applies some special modifications to the outputs. When specific transitions are being modelled, Land Change Modeler masks out the transition potentials of all cases that do not match the case of any specific transition (LCM

Tutorial). For example, if the transition being modelled is from forest to agriculture, values will only exist in pixels that were forest to start with.

3.7.2 Learning Rate

The critical factor in the use of MLP is the learning rate. What is ideally required is a smooth descent of the RMS error curve. If it is flat over a large number of iterations (more than 2000), the training should be stopped and the start and end learning rates be halved. (LCM Tutorial)

"If the RMS error curve has descended and flattens out over a large number of iterations (>1000), the training should be stopped and one can proceed to the next step. If however one experiences a slow but progressive increase in accuracy and decrease in the RMS errors, then the MLP must run until the end of its iterations. If it reaches the end of its iterations and it still appears to be learning (the accuracy is increasing and the RMS is dropping), it should be re-run with a larger number of iterations (e.g., an additional 25%)." (LCM Tutorial)

Optional parameters are also included in the MLP dialog that may be useful. The authors of the LCM software suggest that the default settings are usually optimal for most uses. (LCM Tutorial) However, two parameters are occasionally helpful in achieving the best solution.

3.7.3 Sigmoid Constant

The first is the Sigmoid Constant. This function is often referred to as the activator function in a neural network. Its purpose in an artificial neural network is to generate a degree of non-linearity between the neuron's input and output. An easy way to understand this function is to think of it as a sigmoid curve. It is common knowledge that a sigmoid curve has a slow lag phase, followed by an exponential log phase and then terminated by a stationary phase. The sigmoid function does exactly this with a neural network. It works to reduce the slope between the correctly predicted pixels and wrongly predicted pixels. A value greater than 1 (generally not more than 10) will make the decision boundary between correct and wrong locations less steep. However, upon testing of the model with different sigmoid constants, it was concluded that a value of 1 gives the best result.

3.7.4 Layer Nodes

The second is the number of hidden layer nodes. Each hidden unit sums the signal with different weights, then applies its activation function to compute its output signal, and sends this signal to the unit in the output layer.

With a small numbers of layer nodes relative to the number of input layers, the hidden layers act like Canonical Components, reflecting the common underlying themes in the driver variables. With large numbers relative to the number of driver variables, the nodes capture very specific characteristics. In general, it has been found that two layers perform well. (LCM Tutorial)

3.8 Create Map of Transition Potential

The Transition Potential is a map estimating suitability for change in each pixel. The word 'suitability' in this case may sound counter intuitive, and an easier way of explaining this is

by calling it a vulnerability to change. This is then used further by the software to generate the predictive model for the desired future target date.

The software uses the trained data from the MLP to generate a transition potential. The potential is shown in varying degrees of transition estimated. 0 indicates no potential to transition. A value of 0.46 indicates a high transition potential.
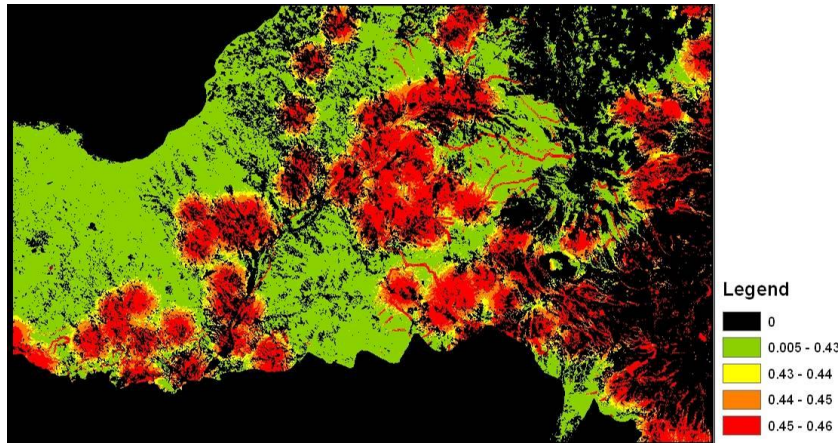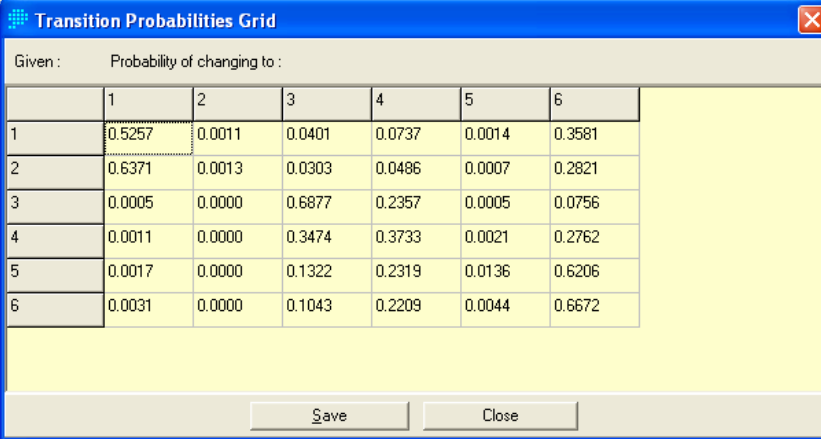
Figure 15 – Map showing transition potential from forest to pasture (to be redone)
Transition potential – Areas in Red indicate greater probability of transition. Areas in Green indicate lower probability of transition. Most of the values of transition potential lay in the region between 0.43 and 0.46. As a result, the image has been displayed to pick out the difference in values at the higher end of the scale.

3.9 Generate model for desired date

The Land Change Modeler provides two basic types of models to predict change: a hard prediction model, and a soft prediction model. The hard prediction model is based on a competitive land allocation model similar to a multi-objective decision process. The soft prediction yields a map of vulnerability to change for the selected set of transitions. In general the results of the soft prediction are preferred for habitat and biodiversity assessment. (LCM Tutorial) The hard prediction yields only a single possibility while the soft prediction is an assessment of change potential.

The change is modelled using a Markov Chain, which determines the amount of change using the earlier and later land cover maps along with the target date specified. A Markov chain is a discrete random process with the assumption that the next state depends only on the current state (Eastman, 2005). This means that previous states do not influence the outcome. This is important because it means that the model is only influenced by the current training set received from the MLP. Thus the matrix (as shown in fig 16 below) records the probability that each land cover category will transition to every other category, and then applies the specific weights to the model. (Eastman, J.R., and Toledano, J., (2000))

Figure 16 – Matrix showing transition probabilities for change in land use

Here, the classes 1-6 indicate the land use classes. The matrix gives the transition probabilities associated with each class based on the training set of actual changes fed into the MLP.

3.10 Validation

The resulting model needs to be validated. In this project only the hard (scenario) was validated. This validation step involved the need to classify an earlier LANDSAT image, so that one could generate a model for the time at which a later image was available. Hence, a 1996 image was classified, and this was used along with the 2001 image, to generate a model for 2006 (using all the drivers or a combination of drivers and the constraints map), for which a land cover classification exits. The model was then validated using this 2006 land cover map as described below.

Using the Map Comparison Kit (described in section 4), the predicted model for 2006 (Map 2) was compared against the classified LANDSAT image of 2006 (Map 1). The resulting comparison yields two results. The first is a map of comparison. This map has the following attribute table which is used for display.

Table 4 – Attribute table used for display in the comparison map

S .no	Actual Land Cover | Model Land Cover	Description
1	0|0	
2	1|0	Misses – Forest that is present in Map 1, but not in Map 2
3	0|1	False alarm – Forest that is present in Map 2, but not in Map 1
4	1|1	Hits – Forest that is present in both Map 1 and 2
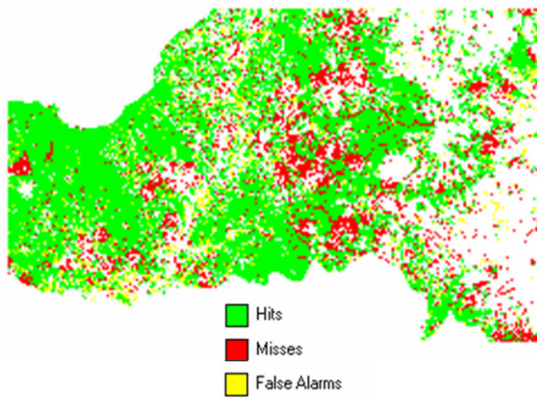
Figure 17 – Map showing areas correctly predicted (hits), false alarms and areas missed

The second result is the number of pixels in the respective categories, which yields a pie chart as shown below.
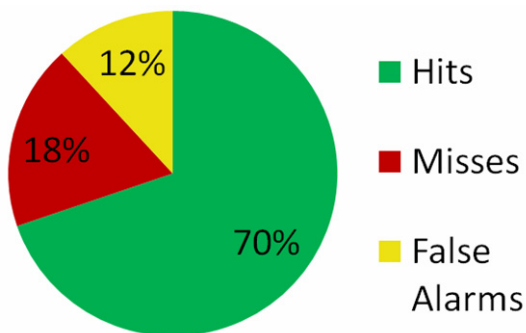
Figure 18 – Pie Chart describing distribution of the hits and misses

We look at comparing maps in terms of area of forest remaining rather than area of non-forest remaining, because Kappa does a pixel wise comparison, which results in values in terms of number of forest pixels, and not non-forest (or deforested pixels).

One must also remember that a hard prediction is just one of the many possible ways in which a particular scenario might be represented. The LCM also generates soft models, if one recalls, that give a map of vulnerability, rather than a pre-defined model.

At this stage, we can use the soft prediction to get a more relevant picture of how well the model has performed.

We can use the soft model to ascertain where future changes are expected in a study area.
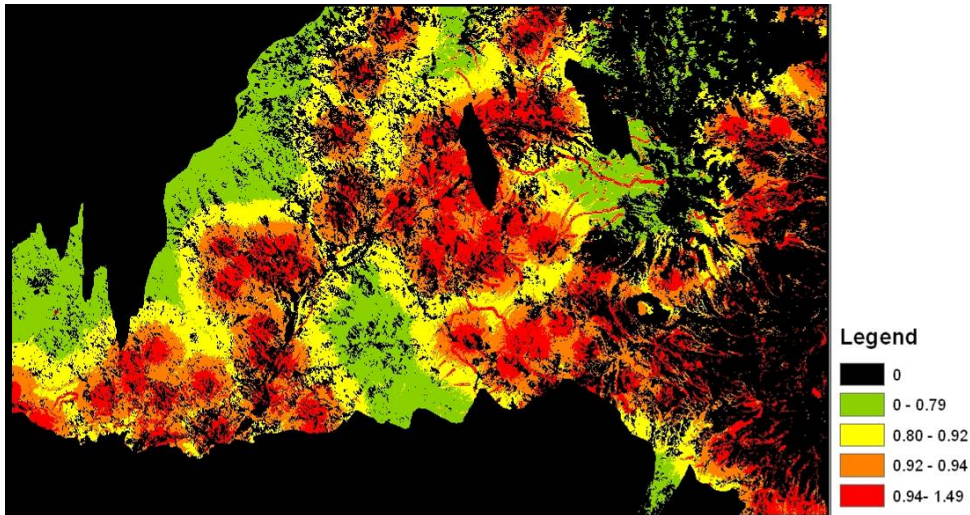
Figure 19 – Soft map showing vulnerability to change
The scale of vulnerability is measured from 0 to 1.50 (0 being no vulnerability and 1.50 being very high vulnerability)

Areas in Red and Orange indicate a high vulnerability (greater than 60%) to change in land use. In this case we are looking at change from forest to any other land use (i.e. deforestation). The areas in yellow and green indicate lower vulnerability to change, reducing from yellow (50 - 60% vulnerability) to green (less than 50%). This scale was chosen to fit the values present.

Sometimes a soft model is a better one to use in cases where there may more than a few factors driving the deforestation. For example, past history might indicate that the deforestation did not occur in a particular pattern. If the user thinks that the deforestation does not have a pattern, then by using a soft model, the places vulnerable to deforestation can be identified, leaving room for some level of ambiguity. Using a map like this would involve understanding the fact that it is a model which describes vulnerability to deforestation, and not like the previous hard land cover map, which shows a single scenario out of many possible ones.

4. Map Comparison toolkit (MCK)

The Map Comparison Kit (MCK) (http://www.riks.nl/mck/) is a software tool for the comparison of raster maps developed by the Research Institute for Knowledge Systems, originally with the only purpose of evaluating output of dynamic spatial models. The software was known as 'Analyse tool'.

Since 2001, major extensions to the software have been made by account of the Netherlands Environmental Assessment Agency (Milieu- en Natuurplanbureau, (MNP). This is the version (version no. 3.2.0) that is popularly used for comparing maps. The program allows users to obtain a good understanding of the differences between pairs of maps, concerning:

- the overall area of the differences

- the location (spatial) of these differences

 (from- http://www.riks.nl/mck/)

This is a valuable tool to use for this particular project, as it allows statistical quantification of the difference between two maps. Such a statistical analysis allows us to quantify the difference between the predicted model and the actual image.
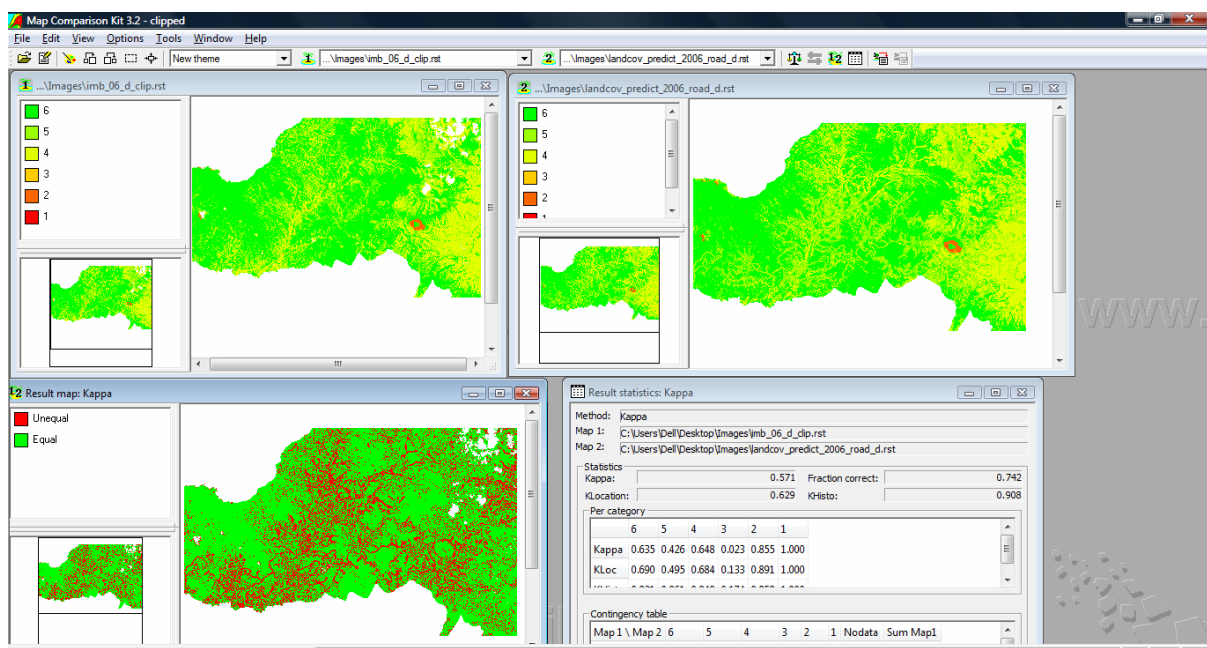
Figure 20 – Map Comparison Kit
The software has a simple user interface that allows for quick comparison of maps. The MCK software only allows the user to add raster or ASCII datasets. The raster format was used in this project, as the LCM also operates with raster datasets.

The first step in the MCK is to add the desired raster datasets and to create a log file. This log file is the place where the information about the data in a particular project is stored. Once the files have been selected, they can be imported to the software.

Once the files have been imported, the user can use the drop down menu to select the reference map and the test map. These are referred to as Map 1 and Map 2. If one wants to see the map, the display map button next to the drop down menu can be used to display the respective maps.

Once the two maps have been selected, one needs to choose a comparison algorithm. The comparison algorithm can be selected using a tab on the top right-hand corner. There are several options for the comparison method, but in this project we only use per-category and the Kappa algorithm.

Finally, the compare map button can be used to create the comparison map, which is displayed in the same window (as shown above). The statistics for the comparison is displayed using the display statistics button adjacent to the compare map button.

In the analysis of the model and the original map in this project, the Kappa statistic has been used; as it is one of the most widely used statistic to assess similarities between observed and predicted results (Pontius, 2000).

4.1 Kappa statistic

The Kappa statistic for maps was introduced by Cohen (1960). Often used to assess the similarities between observed and predicted results, it is not only applied to geographical problems (e.g., Monserud and Leemans, 1992 and Pontius, 2000) but many other disciplines such as medical and social sciences. This has led to a lot of information on the Kappa statistic, including an extensive discussion on its functionality (Carletta, 1996, Foody, 2002 and Foody, 2004).

Kappa is a measure of similarity between two maps based on a contingency table (figure below), which is commonly also referred to as the 'confusion' matrix. In essence, Kappa is based on the percentage of agreement between two maps, corrected for the fraction of agreement that can be expected by pure chance.

Contingency Table:

Table 5 – Contingency table showing different categories in the two maps

Map B – Reference Map

Map 1 \ Map 2	6	5	4	3	2	1	Nodata	Sum Map1
6	709054	144071	33315	2	383	0	0	886825
5	96979	247216	39341	2	101	0	0	383639
4	17806	79594	211084	1	66	0	0	308551
3	1	0	0	2	12	0	0	15
2	608	180	62	150	4623	0	0	5623
1	0	0	0	0	0	0	0	0
Nodata	5069	3260	23655	1	301	0	2098129	2130415
Sum Map 2	829517	474321	307457	158	5486	0	2098129	3715068

(Map A – Test Map)

The table above was created from a pixel wise comparison of all cells in map A and map B

The contingency table above details how the distribution of categories in map A relates to that of map B. It tells us about the user and producer accuracy of the two maps. The cell values represent the number of cells of a particular class in map A and map B. For map A this corresponds to the categories specified in the rows. For map B, the columns specify the different categories.

The central diagonal tells us the number of cells that were correctly predicted. The rows indicate a situation where a particular class was predicted as one thing, whereas the actual land use class is another. Thus, if we look at row 1, land use class 6 is identified as some other class (177771/886825) = 20% of the time. Hence the accuracy is 80% for this class.

The last row and column give the column and row totals. Each row total represents the total number of cells of the given category in map A. Similarly, each column total represents the total number of cells of the given category in map B.

The Kappa statistic is defined according to the following equation:

$$K = \frac{P(A) - P(E)}{1 - P(E)}$$

Where:

P(A) is the fraction of agreement.
P(E) is a random location of subject to the observed distribution (meaning a randomly selected location based on the observed distribution)

Pontius (2000) clarified that the basic Kappa statistic combines the similarity of quantity with similarity of location. In this sense 'quantity' means the total presence as a fraction of all cells in a category taken over the whole map. 'Location' means the spatial distribution of this quantity over the map. Pontius introduces two statistics to separately consider similarity of location and similarity of quantity. The statistic for similarity of quantity is called K-quantity (or referred to here as KHisto). The statistic for similarity of location is called KLocation.

KLocation is calculated by the following equation:

$$KLocation = \frac{P(A) - P(E)}{P(max) - P(E)}$$

KHisto can be calculated by the following equation:

$$KHisto = \frac{P(max) - P(E)}{1 - P(E)}$$

Kappa, KLocation and KHisto are connected through the multiplicative relation

Kappa=KLoc×KHisto

Thus, given a fixed value for Kappa, KLoc will increase if KHisto decreases, and vice versa. Furthermore, if categories in both maps lie at identical locations, we have KLoc = 1.0, and Kappa = KHisto. If the frequency of categories in both maps is equal, we

have KHisto = 1.0 and Kappa = KLoc. Examples from practice have been given in Visser (2004).

According to the authors of the MCK, the reasons for choosing the Kappa statistic are threefold:

1. It has been used widely in research fields such as land-use modelling, remote sensing and ecological modelling.

2. Obtaining the chance correction in Kappa is good, therefore allowing us to compare Kappa values across pairs of maps.

(H Visser, T De Nijs, 2006)

However, there are other measures for map similarity than Kappa, and some authors doubt the practical value of Kappa. Indeed, a measure with chance correction may not be advantageous under all circumstances. Pontius (2002) suggests choosing 'disagreement due to quantity' and 'disagreement due to location', as better alternatives for KHisto and KLoc.

As explained above, the rationale for choosing Kappa is not that it is the 'best' measure for all applications; it is not an absolute criterion. It depends on the specific goals of the user. For this project, the Kappa statistic has useful properties, which are good for measuring the difference to a satisfactory extent. Moreover, the Kfuzzy statistics have been proved to score very well if measured in terms of 'similar to human judgement' (H. Visser, and T. de Nijs, 2005). Finally, this approach is unique in having the ability to define both vagueness in location and vagueness in category. This reflects real situations where magnitude of change and location of change can be different.

"By calculating relative measures for Kfuzzy, Khisto and Klocation with the aid of a reference map it is possible to give a founded validation of the similarity between a model map and an observed map." (Hagen A., 2002) In the research paper (Rajan D.K, 2010) we use the overall Kappa to compare a generated model with reality.

4.2 Per category Comparison

The per category comparison yields a comparison of the different land cover categories between the two maps. This allows us to calculate amount of change in a particular land cover category.

This algorithm was used to calculate the difference in areas deforested, between 2001 and 2006, described in section 3.10 earlier as "Validation".

5. **Testing and Validation**

Once a test model was generated, the next step was to test various drivers for their impact on the rate of deforestation. The above mentioned steps were followed in producing the model, the only difference being that the drivers were changed in order to understand the level of impact on the model.

Drivers of deforestation are varied in their presence and effect on the forest. They are therefore hard to pinpoint. They may be simple; For example, Sader and Joyce (1988) examined forest area change associated with transportation networks for Costa Rica. They found that there was a strong relationship between forest clearing and proximity to road. This is an example of the proximity to road being a driver for deforestation.

On the other hand, they also may be hard to quantify; Lambin (1994) describes the plethora of factors that are responsible for land use change. For instance, political instability, change in land ownership, war, are all factors that impact land use change. In most land use change models, these factors are not taken into consideration as they are difficult to model. In this project the emphasis is on quantifying factors to model change. As a result, the factors mentioned above have not been taken into consideration in the study.

Several authors (Lambin (1994), Madeira(2007) , Veburg (2006)) have described common drivers of deforestation that are at least often show correlations with deforestation: (in no particular order of significance)

1. Proximity to Roads

2. Slope

3. Protected areas

4. Soil type

5. Proximity to settlements

6. Edge of Forest exposed to direct contact

Of these drivers of deforestation, proximity to roads, slope, proximity to towns and protected areas were chosen for this study taking data constraint into consideration. The presence of these drivers has also been correlated with the high rate of deforestation in Ecuador (Southgate et al. (1991), Rudel (1993), C.F. Mena (2006), S. Walsh et al. (2002), Chomitz, K.M. & Gray, D.A. (1995), Soares-Filho, B (2004)).

Settings Used for all models:

Layer Nodes – 2, Sigmoid factor -1

Driver data (road, slope, towns) AND Constraints map (with 0 for protected areas and 1 for all other areas)

The box on the left describes the settings used for the generation of all the models below. The only variable changed was the driver data for the different models.
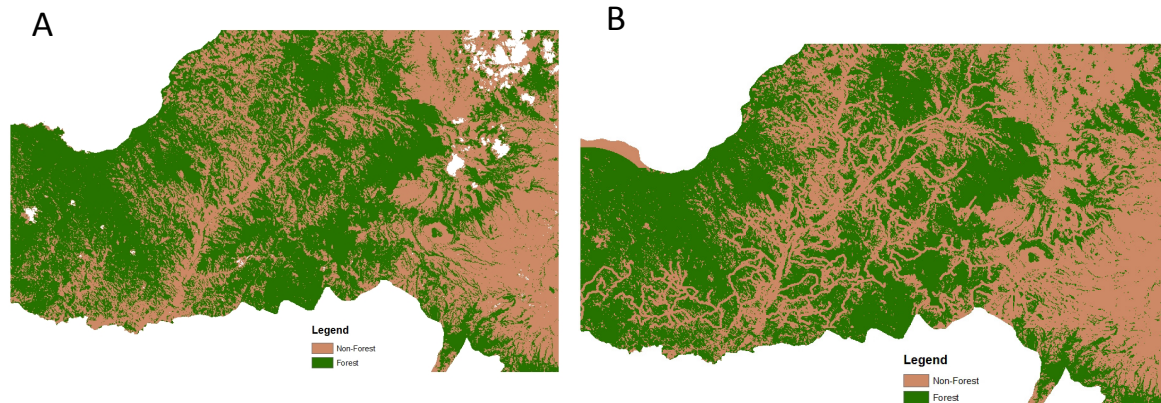
5.1 Testing drivers

Roads

Figure 21 – a) Original 2006 Land Cover Map; b) Model generated using proximity to roads as the only driver of deforestation
In order to assess the validity of a model produced using proximity to roads as the primary driver, it was compared with the original land cover map for 2006.

When the above map was compared with the original 2006 map with MCK, the following output was obtained.

Figure 22 – a) Comparison map between Fig. 21 (roads model) and actual deforested areas in 2006; b) Pie chart summarizing % area predicted correctly (hits), areas missed (misses) and falsely predicted (false alarms)
In this case, when using roads as the main driver of deforestation, we get 70% of all the predicted area correct, but miss 18% of the area that was actually forest in 2006. 12% of the time, forest pixels are predicted where they have not occurred, so these are 'false alarms'.

Slope (20m resolution)
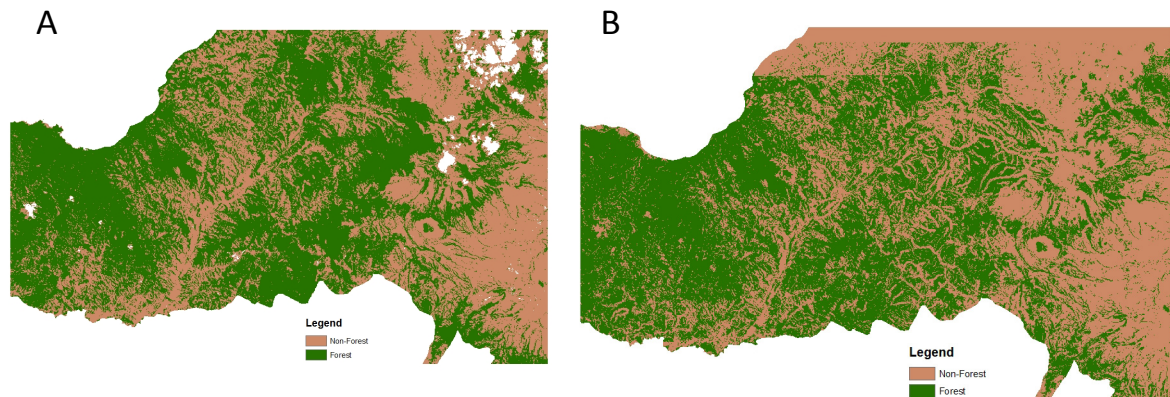
Figure 23 – a) Original 2006 Land Cover Map; b) Model generated using slope as the only driver of deforestation

According to Peck M.(2010), the rate of deforestation in the area is lower at slopes above 45 degrees. In order to take this into consideration, one must consider the effect of slope in the generation of this model.

Slope as the driver of deforestation yields the model illustrated above. The areas of steep slope are not preferred for deforestation, as they are inaccessible to heavy transport vehicles. The comparison map below illustrates the difference between actual deforestation in the area and the deforestation as predicted by slope as the only driver for deforestation.
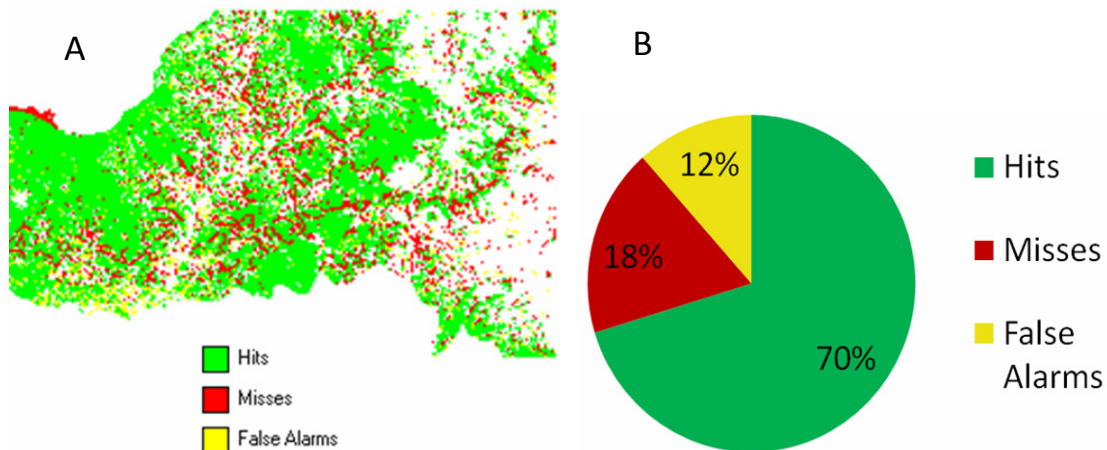

Figure 24 – a) Comparison map between Fig. 23 (slope model) and actual deforested areas in 2006; b) Pie chart summarizing % area predicted correctly (hits), areas missed (misses) and falsely predicted (false alarms)

When using slope as the only driver of deforestation during the years 2001 to 2006, the comparison map reveals the following:

- 68% of the area that was actually forest in 2006 was correctly predicted by the model

- 19% of the area that was forest in 2006 was missed by the model

- 13% of the area that was predicted to have forest by the model was not actually forest in 2006.

Towns
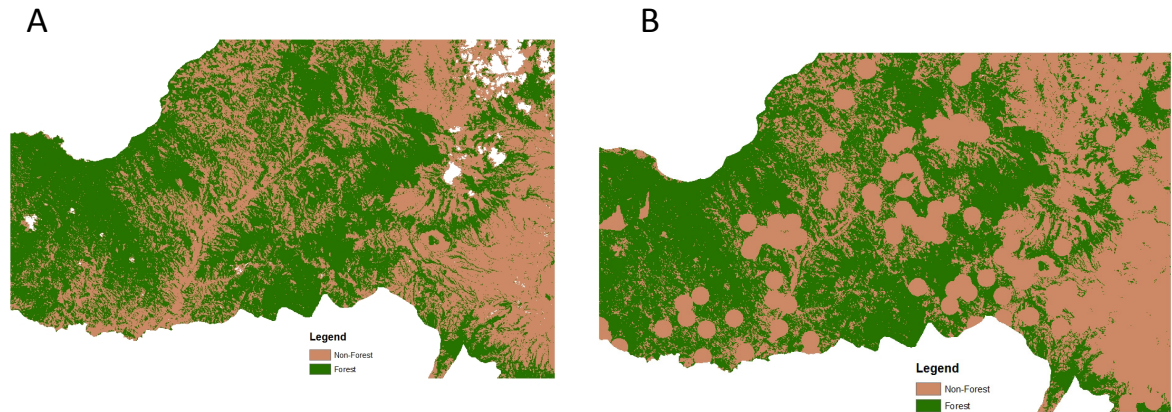
Figure 25 – a) Original 2006 Land Cover Map; b) Model generated using proximity to towns as the only driver of deforestation

With towns as the driver, the model generated does not reflect a plausible situation in reality. The model makes the conversion from forest to urban very high near the towns. This leads to an effect that looks very artificial.
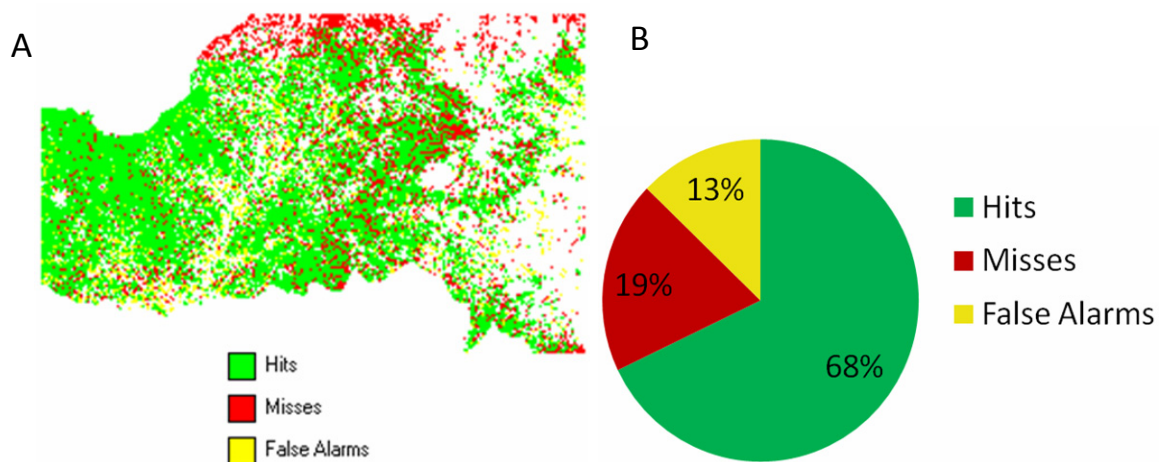

Figure 26 – a) Comparison map between Fig. 25 (towns model) and actual deforested areas in 2006; b) Pie chart summarizing % area predicted correctly (hits), areas missed (misses) and falsely predicted (false alarms)

When comparing the model produced by using towns as the only driver of deforestation with the original 2006 land cover image, the following can be observed:

- 70% of the area was actually forest in 2006 was correctly predicted by the model

- 18% of the area that was forest in 2006 was missed by the model

- 12% of the area which was predicted to be forest by the model, was not actually forest in 2006

These values are better than those returned by the model generated using only slope. It is very similar to the values obtained when comparing the road model. In order to separate the effect each driver, we use another test.

In addition to validating the model by testing the predicted rate of deforestation with the actual rate of deforestation, the MCK also gives Kappa statistics of similarity between two land cover maps. The following table describes the Kappa statistics derived when each model was compared to the original 2006 land cover map.

Table 6- Comparison of Kappa statistics between drivers

	Road	Slope	Town
Kappa	0.62511	0.58932	0.62742
KLocation	0.68179	0.64346	0.68524
KHisto	0.91686	0.91585	0.91563
Fraction correct	0.81317	0.79531	0.8143

The Kappa statistics indicate that the model produced using slope as the principle driver does not have a good fit with the original map. However, the KHisto for all three of the maps are very similar. The KHisto values indicate the quantitative similarity between the two maps in question. (Hagen A., 2002). Thus, a set of similar KHisto values means that the maps are relatively similar in their performance when measured only for quantity of change. As Pontius points out, change can be measured in terms of location and magnitude. This means that quantitative change alone cannot accurately measure change. The KLocation values are similar for the models generated with road and town information. The KLocation values are much lower than the KHisto values, indicating that the model may be better at predicting magnitude of change rather than location of change.

On comparing the three drivers and their Kappa values, we can conclude that the proximity to towns is the best explanatory variable amongst the three drivers. It must also be noted that the difference between the Kappa values is not very high. One conclusion that we can draw, therefore, is that it does not make a huge difference to the Kappa value when changing a single driver. In order to compare the driver impact on the model, one needs to compare how different combinations of the driver data affect the Kappa statistics in the map comparison.

Road and Towns

A   B

Figure 27 – a) Original 2006 Land Cover Map; b) Model generated using both proximity to roads and proximity to towns as the two drivers of deforestation
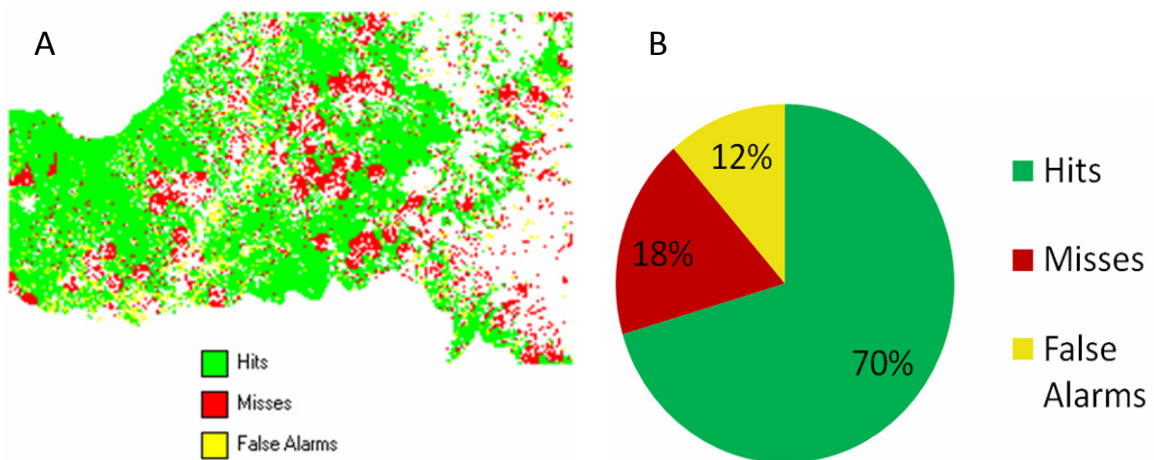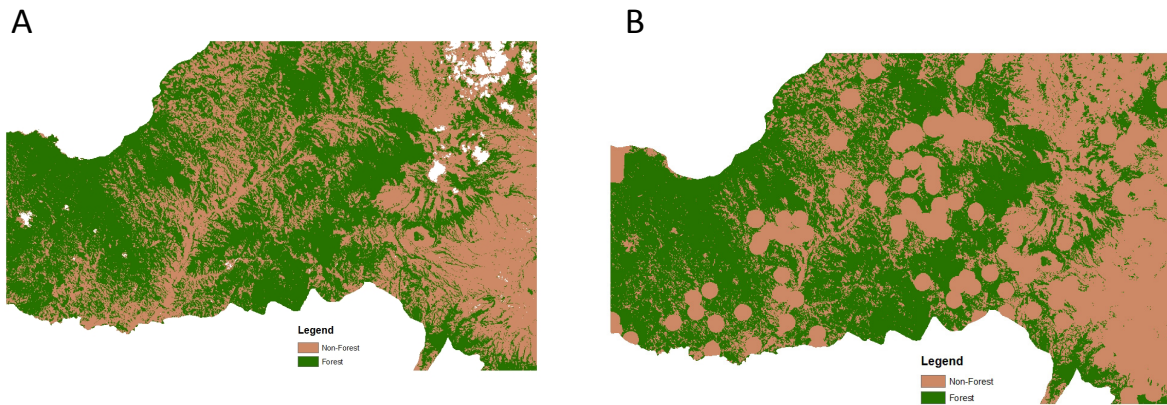

Figure 28 – a) Comparison map between Fig. 27 (roads and towns model) and actual deforested areas in 2006; b) Pie chart summarizing % area predicted correctly (hits), areas missed (misses) and falsely predicted (false alarms)

From the pie chart, one can conclude that roads and towns in combination seem to have the same value for areas of deforestation predicted as the towns model.

Slope (20m resolution) and Towns

A   B

Figure 29 – a) Original 2006 Land Cover Map; b) Model generated using both proximity to towns and slope as the two drivers of deforestation
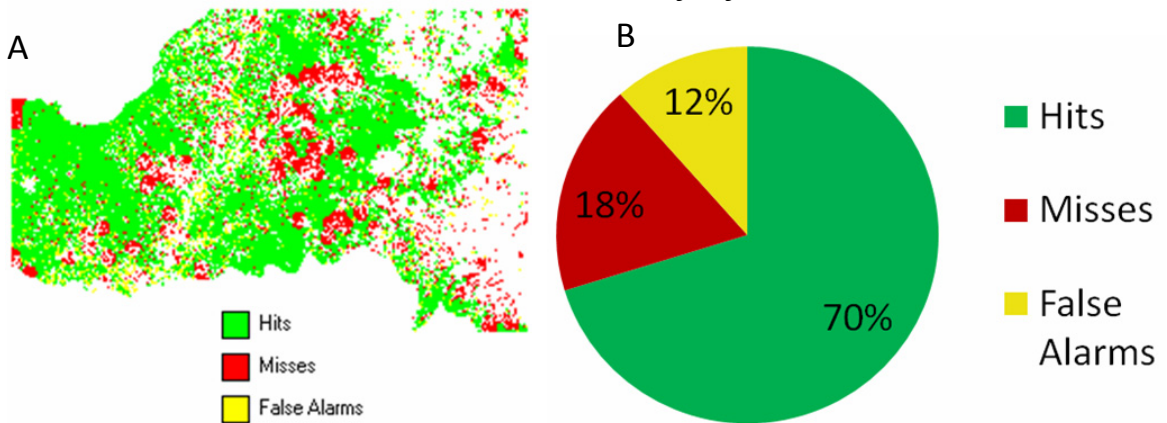
Figure 30 – a) Comparison map between Fig. 29 (towns and slope model) and actual deforested areas in 2006; b) Pie chart summarizing % area predicted correctly (hits), areas missed (misses) and falsely predicted (false alarms)

Once more, the pie chart above (Fig 30 b) indicates a similarity with the towns and the road+town models.
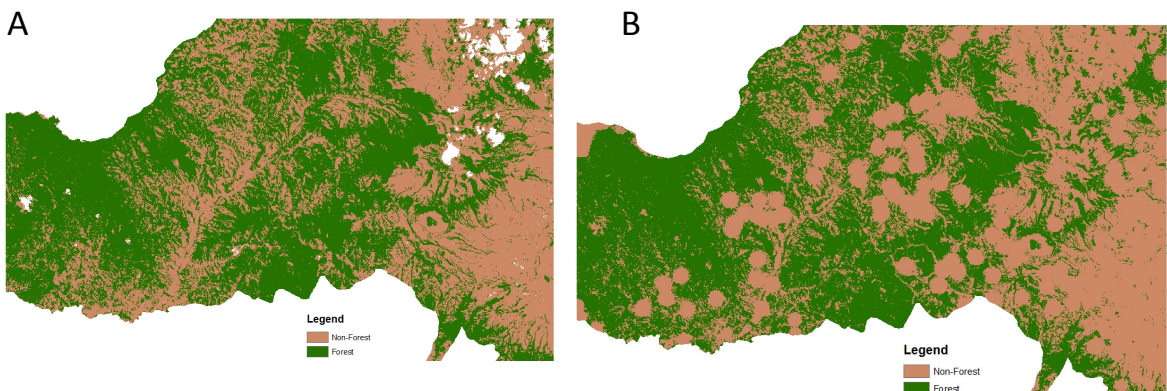
Slope (20m resolution) and Roads

Figure 31 – a) Original 2006 Land Cover Map; b) Model generated using proximity to roads and slope as the two drivers of deforestation

The pie chart below (Figure 32 b) indicates that it is the same percentage of area correctly predicted, areas missed and areas falsely predicted as towns, towns+roads and towns+roads.
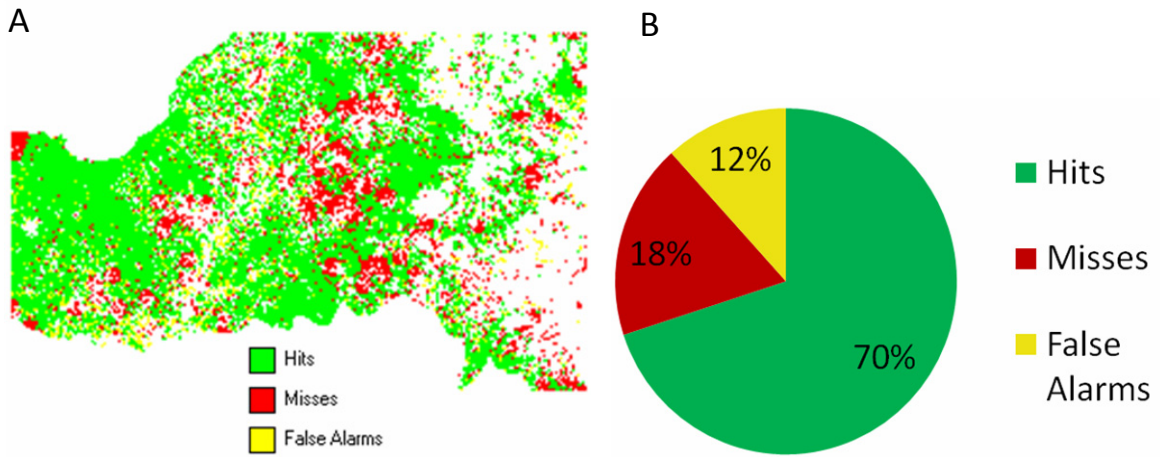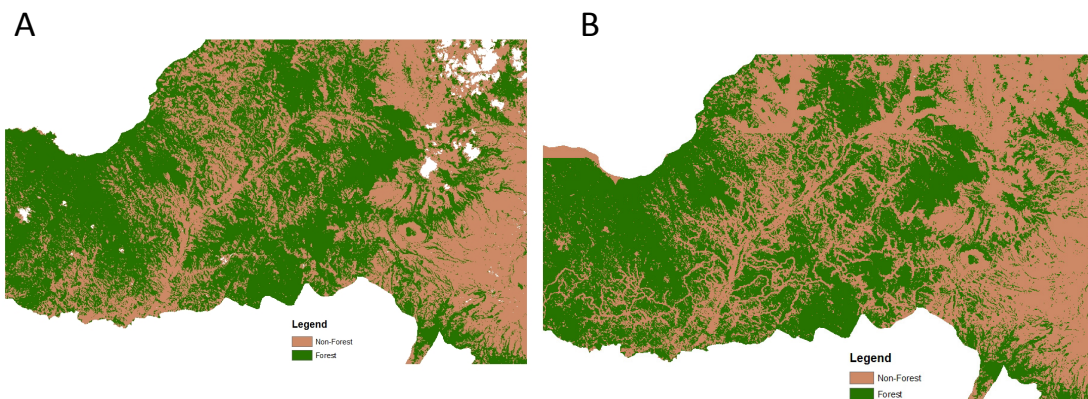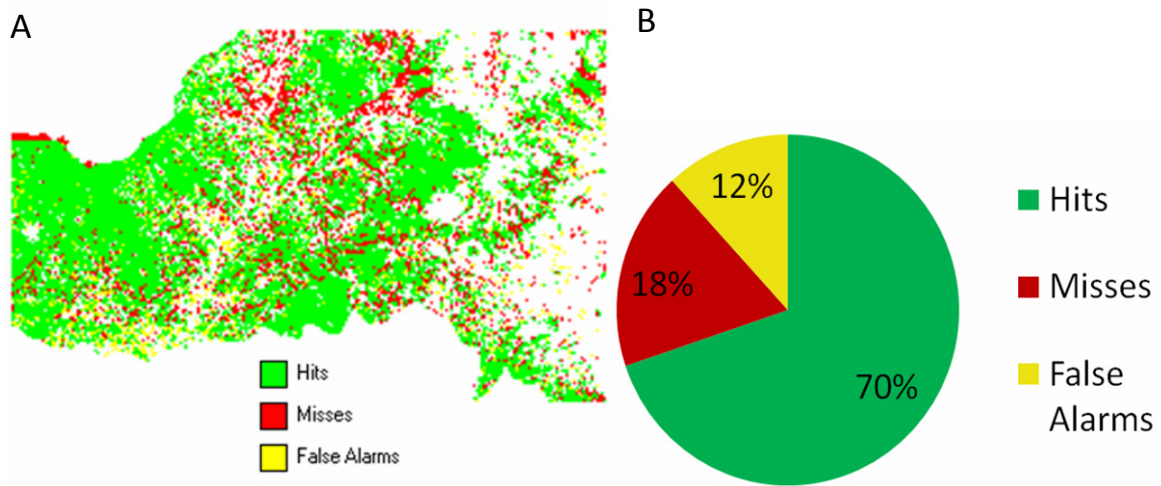
Figure 32 – a) Comparison map between Fig. 31 (roads and slope model) and actual deforested areas in 2006; b) Pie chart summarizing % area predicted correctly (hits), areas missed (misses) and falsely predicted (false alarms)

Although some of the locations of change are different, the overall magnitude (in terms of number of cells) is the same for all four scenarios.

In order to differentiate their individual effects on the model, we can use the Kappa statistic. Different combinations of drivers produce varying results. The best result seems to be achieved by the road and town combination, with a Kappa score of .624 and a fraction of .812. The roads and slopes however, seems to perform comparatively poorly, with a Kappa score of .61 and a fraction of .80. The Kappa values seem to back up the observation that the drivers in combination seem to produce similar results.

Table 7 - Comparison of Kappa statistics between combinations of drivers

	Road + Town	Slope + Town	Road + Slope
Kappa	0.62451	0.6207	0.61831
KLocation	0.6815	0.67721	0.6742
KHisto	0.91636	0.91655	0.9171
Fraction correct	0.81286	0.81097	0.80979

The table below (Table 8) shows the comparison between all the generated models and their kappa values. The model with proximity to towns as the only driver resulted in the best Kappa value. The models produced with slopes (slope, slope+town, slope+road) all have lower Kappa values than the models produced with other values.

Table 8- Comparison of Kappa statistics between all drivers

	Road	Slope	Town	Road + Town	Slope + Town	Road + Slope
Kappa	0.62511	0.58932	0.62742	0.62451	0.6207	0.61831
KLocation	0.68179	0.64346	0.68524	0.6815	0.67721	0.6742
KHisto	0.91686	0.91585	0.91563	0.91636	0.91655	0.9171
Fraction correct	0.81317	0.79531	0.8143	0.81286	0.81097	0.80979

This is because the slope as a driver does not result in a model which had a good Kappa value. Hence the addition of slopes to other drivers, results in a model with lower Kappa values than the other combined models. This is probably because not all areas with a less steep slope will necessarily be deforested (as would be the case when modelling slope as the only driver of deforestation). Hence this results in less correlation with the original 2006 image, and a lower Kappa value.
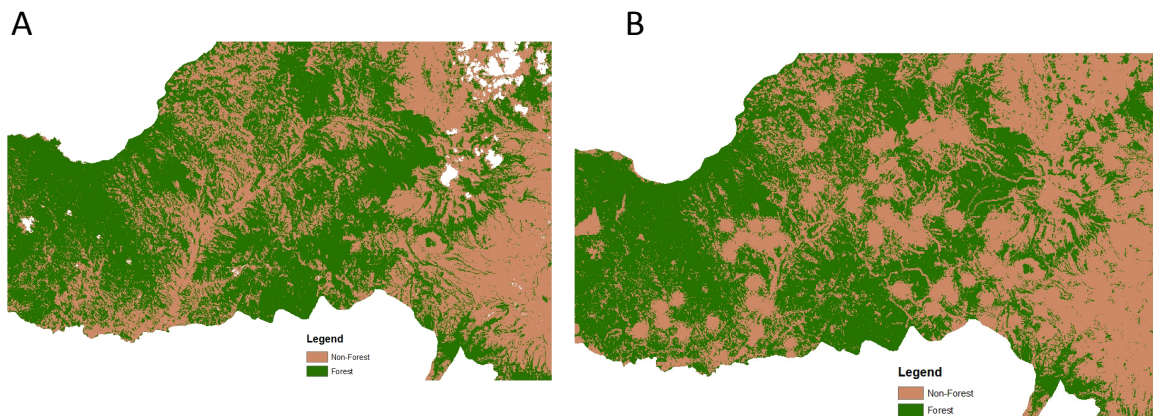
Model with all drivers

Figure 33 – a) Original 2006 Land Cover Map; b) Model generated using all 3 drivers of deforestation
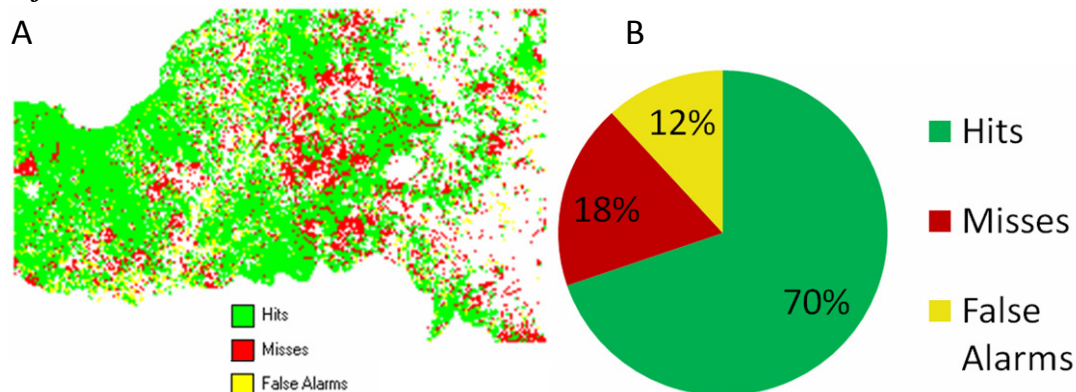

Figure 34 – a) Comparison map between Fig. 33 (all 3 drivers model) and actual deforested areas in 2006; b) Pie chart summarizing % area predicted correctly (hits), areas missed (misses) and falsely predicted (false alarms)

The pie chart above (Figure 34 b) indicates that the final model resembles the models generated with two drivers in terms of the areas correctly predicted, missed and falsely predicted. When the map is compared with the original 2006 map, we get the following statistics

Table 9- Kappa statistics for model generated with all drivers

Kappa	0.61751
KLocation	0.67449
KHisto	0.91553
Fraction correct	0.80936

A look at the Kappa values tells us that it is slightly lower than the models with towns, roads and models with towns+roads, towns+slope as the drivers of deforestation. A reason for this could be that the slope as a driver performs relatively poorly, hence bringing down the overall Kappa value for the model. Thus, a model without slope seems to do better than one with slope. In order to test the effect of slope as the driver, we can explore the models generated using 90m resolution slope to see if it makes a difference.

5.2 Testing models generated using 90m resolution slope

One of the aims of the project was to test the performance of the model when using coarse resolution data. The previous set of results was obtained using 20m resolution DEM (slope). In order to test the level of accuracy of the model with coarse data, models were generated using 90m resolution (slope) data.
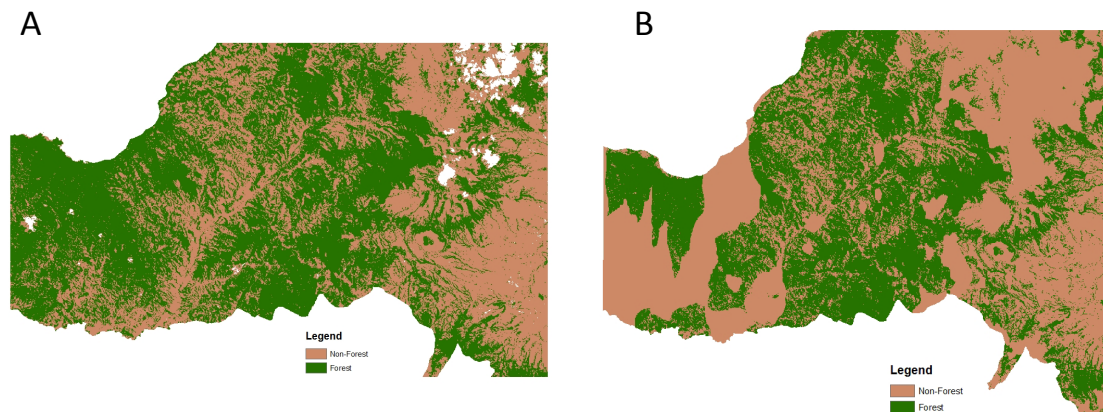
90m Slope

Figure 35 – a) Original 2006 Land Cover Map; b) Model generated using 90m slope as the only driver of deforestation

Figure 36 – a) Comparison map between Fig. 35 (90m slope model) and actual deforested areas in 2006; b) Pie chart summarizing % area predicted correctly (hits), areas missed (misses) and falsely predicted (false alarms)

The pie chart above (Fig 36 b) shows that, while using a 90m slope, 65% of the area that was forest in 2006 was correctly predicted by the model, 25% was missed, while 10% was wrongly predicted to be forest in 2006. This indicates that 20m slope (see Figure 24b) does better than the 90m slope when used as the single driver for deforestation (although only slightly).

Table 10 – Comparison between 20m slope and 90m slope

	Slope 20m	Slope 90m
Kappa	0.58932	0.57231
KLocation	0.64346	0.7051
KHisto	0.91585	0.81167
Fraction correct	0.79531	0.78429

Comparison of Kappa values between 20m slope and 90m slope models shows that the 20m slope model tends to be more similar to the real areas of deforestation than the 90m slope model, confirming what the pie chart above shows.

Slope (90m) and Roads
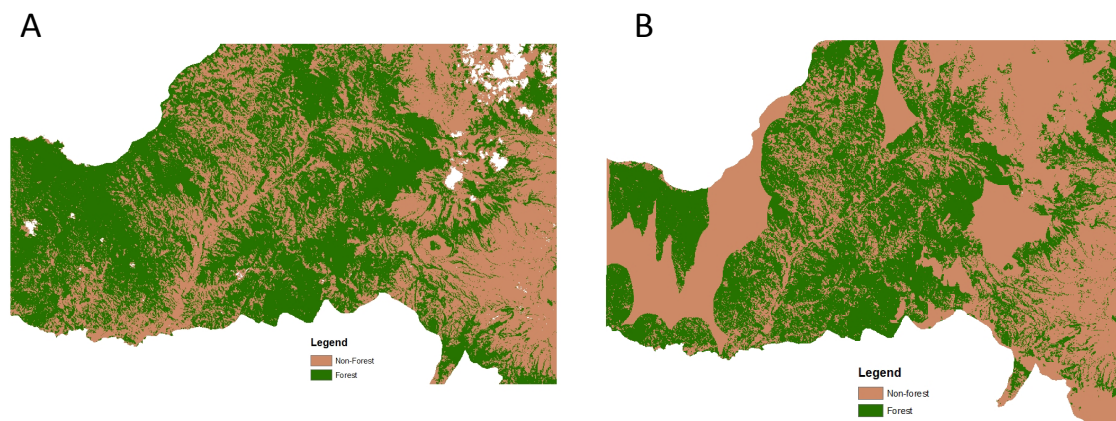
Figure 37 – a) Original 2006 Land Cover Map; b) Model generated using both proximity to roads and 90m slope as the two drivers of deforestation
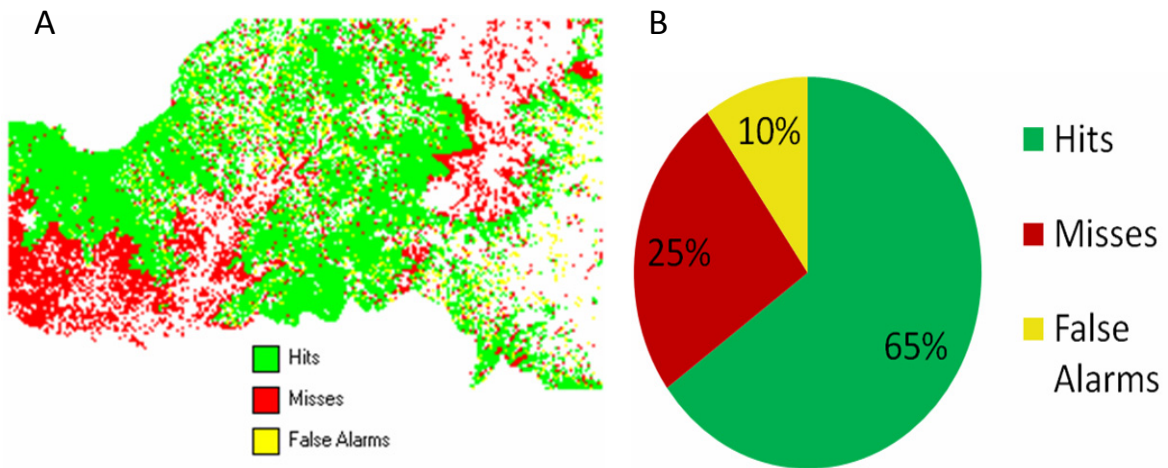

Figure 38 – a) Comparison map between Fig. 37 (roads + 90m slope model) and actual deforested areas in 2006; b) Pie chart summarizing % area predicted correctly (hits), areas missed (misses) and falsely predicted (false alarms)

The comparison map and statistics reveal that when using 90m slope with roads as the drivers for deforestation, it only predicts 61% of the area forested correctly. The area missed is 27%, while the 'false alarms' are 12% of the area.

When Kappa values from Road + 20m slope and Road + 90m slope were compared, we see (from Table 11) that the 20m slope is more similar to the real areas of deforestation than the 90m slope in terms of Kappa.

Table 11 – Comparison between 20m slope and 90m slope

|	Road + Slope (20m)	Road + Slope (90m)
Kappa	0.61831	0.50548
KLocation	0.6742	0.61582
KHisto	0.9171	0.82082
Fraction correct	0.80979	0.75084

Slope (90m) and Towns
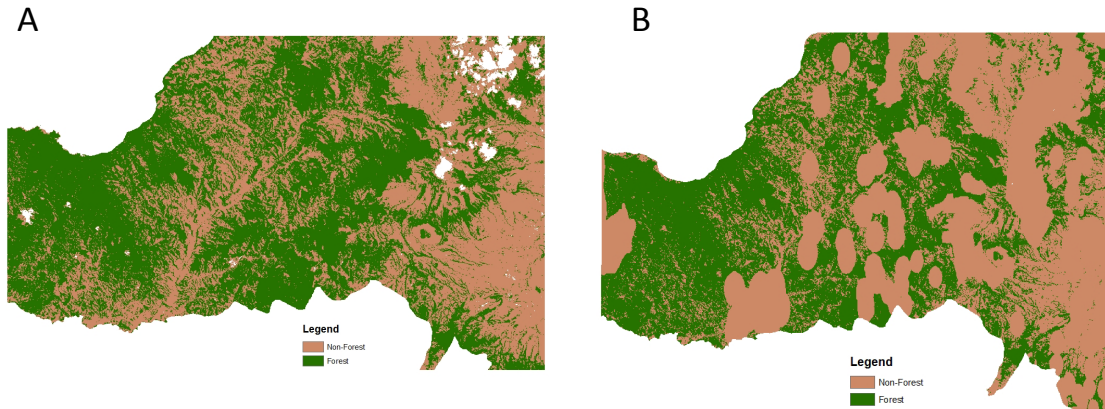
Figure 39 – a) Original 2006 Land Cover Map; b) Model generated using both proximity to towns and 90m slope as the drivers of deforestation
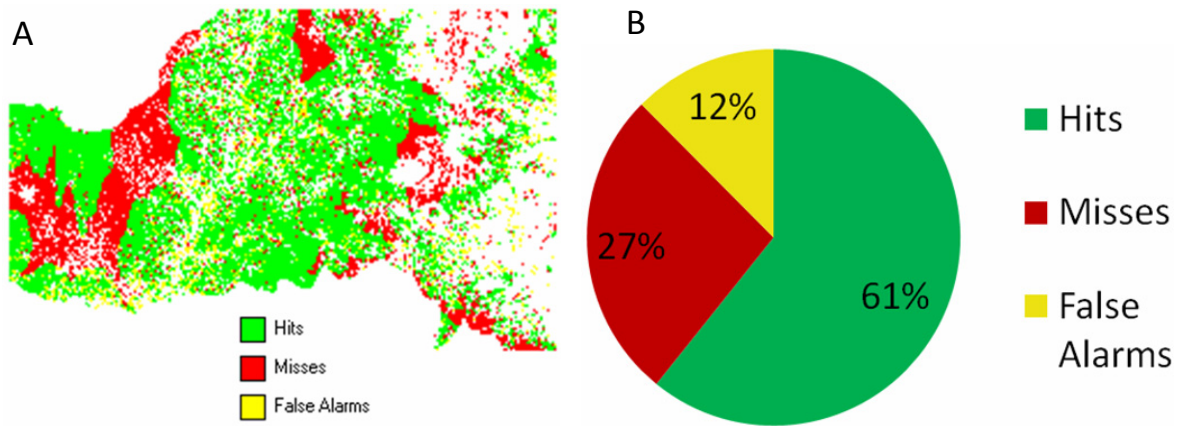

Figure 40 – a) Comparison map between Fig. 39 (towns and 90m slope model) and actual deforested areas in 2006; b) Pie chart summarizing % area predicted correctly (hits), areas missed (misses) and falsely predicted (false alarms)

The statistics from the pie chart above (Fig 40 a) indicate that 63% of the areas that were forest in 2006, have been correctly predicted by the model, while 26% of the areas were missed. 11% of the area which was predicted to be forest was not actually forest in 2006.

A comparison of Kappa values from Table 12 below reveals that the Kappa values for the 20m slope are much better than those for the 90m slope.

Table 12 – Comparison between 20m slope and 90m Slope

|	Town + Slope 20m	Town + Slope 90m
Kappa	0.6207	0.53529
KLocation	0.67721	0.65351
KHisto	0.91655	0.8191
Fraction correct	0.81097	0.76582

All Drivers with 90m resolution slope
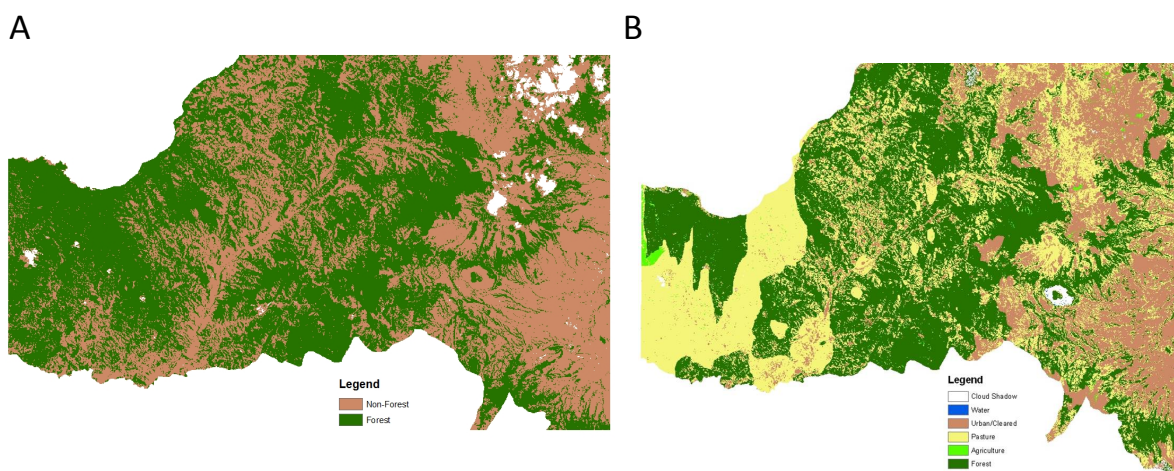
Figure 41 – a) Original 2006 Land Cover Map; b) Model generated using all 3 drivers (road, town and 90m slope) as the drivers of deforestation
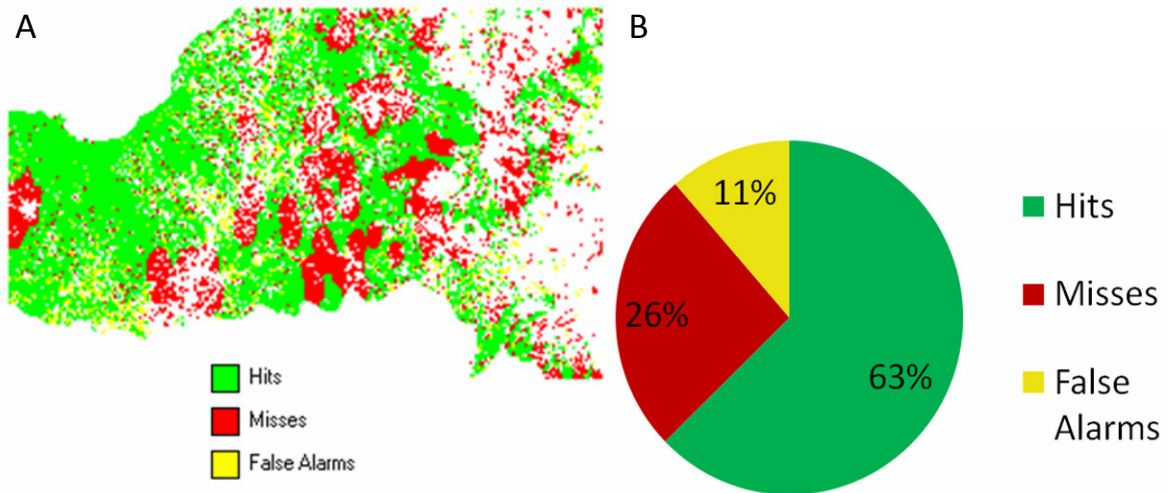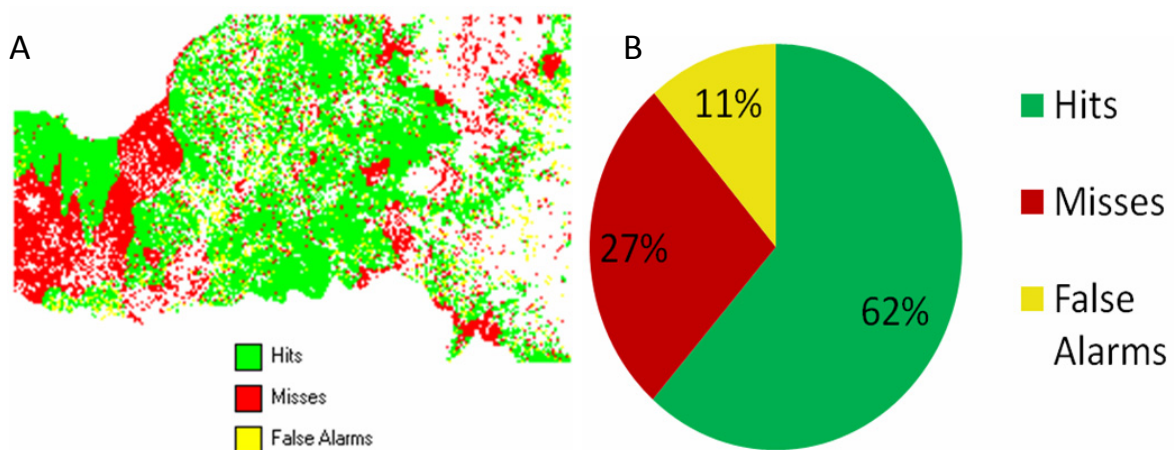

Figure 42 – a) Comparison map between Fig. 41 (all drivers model) and actual deforested areas in 2006; b) Pie chart summarizing % area predicted correctly (hits), areas missed (misses) and falsely predicted (false alarms)

The pie chart above (Fig 42 b) reveals that the model with all the drivers does not predict the areas of deforestation more accurately than the model produced using a combination of

two drivers. The area correctly predicted is only 62% (it was 65% when using 90m slope as the only driver).

A comparison of Kappa values from table 13 below reveals that the model with 20m slope is more similar to the real areas of deforestation than the model with 90m slope. From this it can be inferred that the greater the level of detail, the better the model generated (meaning the closer to reality it is), as all the models with 20m slope have higher Kappa values than the models with 90m slope.

Table 13 – Comparison between 20m slope and 90m slope

|	Map 20m Slope	Map 90m Slope
Kappa	0.61751	0.52164
KLocation	0.67449	0.64959
KHisto	0.91553	0.80304
Fraction correct	0.80936	0.7585

6. Future work

The study towards the identification of a comprehensive methodology for measuring and modelling deforestation has only just begun. This study was just an explorative one, to understand the sensitivity and working of the Land Change Modeler in predicting deforestation in Ecuador based on a set of driving factors. The Land Change Modeler has proved to be robust in its generation of models. The direction for future work would be in terms of exploring different datasets as well as more resolutions. Using MODIS data is one of the options that we only began exploring during the course of the project. A more in-depth look at different satellite imagery will further enhance our understanding of how models of land use change may be generated.

7. References

Atkinson, P.M. et al., 1997. Neural networks in remote sensing. *International Journal of Remote Sensing*, 18(4), 699-709.

Chan, J.C. et al., 2001. Detecting the nature of change in an urban environment: A comparison of machine learning algorithms. *Photogrammetric Engineering and Remote Sensing*, 67(2), 213-225.

Carletta, J., 1996. Assessing agreement on classification tasks: the kappa statistic. *Computational linguistics*, 22(2), 249–254.

Chomitz, K.M. & Gray, D.A., 1995. Roads, lands, markets, and deforestation: a spatial model of land use in Belize. *World*.

Cohen, J., 1968. Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological bulletin*, 70(4), 213–220.

Civco, D.L., 1993. Artificial neural networks for land cover classification and mapping. *International Journal of Geographic Information Systems*, 7(2), 173-186.

Eastman, J.R. et al., 1995. Raster procedures for multi-criteria/multi-objective decisions. *Photogrammetric Engineering and Remote Sensing*, 61(5), 539–547.

Eastman, J.R., and Toledano, J., (2000) "Markov Chain and Cellular Automata Approaches to Land Cover Change Modeling", workshop presentation (invited), *4th International Conference on Integrating GIS and Environmental Modeling*, Banff, Alberta, Sept. 2-8, 2000.

Fleiss, J.L., Cohen, J. & Everitt, B.S., 1969. Large sample standard errors of kappa and weighted kappa. *Psychological Bulletin*, 72(5), 323–327.

Foody, G.M., 2002. Status of land cover classification accuracy assessment. *Remote sensing of environment*, 80(1), 185–201.

Foody, G.M. & Mathur, A., 2004. Toward intelligent training of supervised image classifications: directing training data acquisition for SVM classification. *Remote Sensing of Environment*, 93(1-2), 107–117.

Hagen, A., 2003. Multi-method assessment of map similarity. *International Journal of Geographical Information Science*, 17(3), 235–249.

Hagen-Zanker, A., Straatman, B. & Uljee, I., 2005. Further developments of a fuzzy set map comparison approach. *International Journal of Geographical Information Science*, 19(7), 769–785.

Lambin E. F, Department of Agriculture, Forest Service, Northeastern Research Station. 61p. "Modelling deforestation processes: a review," European Commission, Luxemburg (1994).

Land Change Modeler Tutorial and Help Files

Lillesand, T.M., Kiefer, R.W. & Chipman, J.W., 1994. *Remote sensing and image interpretation*, John Wiley & Sons New York.

Madeira, E.C., 2007. Policies to Reduce Emissions from Deforestation and Degradation (redd) in Developing Countries.

Map Comparison Kit Manual - http://www.riks.nl/RiksGeo/resources/Map%20Comparison%20Kit%20Manual.pdf

Mather, P.M., 2004. *Computer processing of remotely sensed images: an introduction*, Wiley.

Mena, C.F., Bilsborrow, R.E. & McClain, M.E., 2006. Socioeconomic drivers of deforestation in the northern Ecuadorian Amazon. *Environmental management*, 37(6), 802–815.

Peck M., 2008. Developing a habitat action plan for *Ateles fusciceps,* University of Sussex

Peck M., 2010. Establishing a baseline: Determining deforestation rates in NW Ecuador using LANDSAT satellite imagery. A report prepared for Rainforest Concern.

Pontius, R.G., 2000. Quantification error versus location error in comparison of categorical maps. *Photogrammetric Engineering and Remote Sensing*, 66(8), 1011–1016.

Power, C., Simms, A. & White, R., 2001. Hierarchical fuzzy pattern matching for the regional comparison of land use maps. *International Journal of Geographical Information Science*, 15(1), 77–100.

Prentice, I.C. et al., 1992. Special paper: a global biome model based on plant physiology and dominance, soil properties and climate. *Journal of Biogeography*, 19(2), 117–134.

Richards, J.A. and X. Jia, 1999. *Remote Sensing Digital Image Analysis*. (New York: Springer).

Rudel, T.K. & Horowitz, B., 1993. *Tropical deforestation: small farmers and land clearing in the Ecuadorian Amazon*, Columbia Univ Pr.

Sader, S.A. & Joyce, A.T., 1988. Deforestation rates and trends in Costa Rica, 1940 to 1983. *Biotropica*, 20(1), 11–19.

Soares-Filho, B. et al., 2004. Simulating the response of land-cover changes to road paving and governance along a major Amazon highway: the Santarém–Cuiabá corridor. *Global Change Biology*, 10(5), 745–764.

Southgate, D., Sierra, R. & Brown, L., 1991. The causes of tropical deforestation in Ecuador: a statistical analysis. *World Development*, 19(9), 1145–1151.

Tso, B. and P.M. Mather, 2001. Classification Methods for Remotely Sensed Data (New York: Taylor and Francis).

Verburg, P.H. et al., 2006. Modeling land-use and land-cover change. *Land-use and Land-cover Change: Local Processes and Global Impacts Eds EF Lambin, HJ Geist (Springer, Berlin) pp*, 117–137.

Visser, H. & De Nijs, T., 2006. The map comparison kit. *Environmental Modelling & Software*, 21(3), 346–358.

Walsh, S. et al., 2002. Characterizing and modeling patterns of deforestation and agricultural extensification in the Ecuadorian Amazon. *Linking people, place, and policy: a GIScience approach*, 187–214.

8. **Appendix**

List of file names associated with the project:

All the files used in this project can be found deposited with the School of GeoSciences IT department in a folder named *dissfinal*

Folder	Filename	Description
/Base Files	dem.rst	DEM (90m resolution)
	imb_01_d_clip.rst	2001 Land cover map
	imb_06_d_clip.rst	2006 Land cover map
	imb_96_d_clip.rst	1996 Land cover map
	reserve_rc1.rst	Constraints map
	road_d_clip_CopyRaster.rst	Distance from road
	slope_dem1.rst	Slope (20m resolution)
	town_d_clip_CopyRaster.rst	Distance from town
/Results	landcov_predict_all_d.rst	Model generated with all drivers
	landcov_predict_road_d.rst	Model generated with roads
	landcov_predict_roadtown_d.rst	Model generated with road and town
	landcov_predict_slope_d.rst	Model generated with slope
	landcov_predict_sloperoad_d.rst	Model generated with slope and road
	landcov_predict_slopetown_d.rst	Model generated with slope and town
	landcov_predict_soft_d.rst	Soft model with all drivers
	landcov_predict_towns_d.rst	Model generated with towns
	landcov_predict_2011.rst	Model generated for 2011
	landcov_predict_2016.rst	Model generated for 2016
	landcov_predict_2021.rst	Model generated for 2021
	landcov_predict_2026.rst	Model generated for 2026
	landcov_predict_2031.rst	Model generated for 2031
	landcov_predict_2036.rst	Model generated for 2036
	landcov_predict_2041.rst	Model generated for 2041

	landcov_predict_2046.rst	Model generated for 2046
	landcov_predict_2051.rst	Model generated for 2051
	transition_potential3.rst	Transition potential for Forest to Urban
	transition_potential4.rst	Transition potential for Forest to Pasture
	transition_potential5.rst	Transition potential for Forest to Agriculture
	trend_change.rst	Trend Map of change