UNIVERSITÉ DES ANTILLES ET DE LA GUYANE
FACULTÉ DES SCIENCES EXACTES ET NATURELLES

MASTER EN SCIENCES – TECHNOLOGIES -SANTE

Mention: ECOLOGIE

Spécialité : Écosystèmes Tropicaux Naturels et Exploités

Predicting the deforestation risk in Suriname using a spatially explicit Random Forest model

Kasanpawiro Cindyrella

Directeur (s) de stage :
Camille Dezecache and
Sabrina Coste

12th of June 2015

(Foundation for Forest Management and Production Control, Ds.Martin Luther King weg perc.no.283, Paramaribo Suriname)

Mémoire soutenu le : 22nd of June 2015

UNIVERSITÉ DES ANTILLES ET DE LA GUYANE
FACULTÉ DES SCIENCES EXACTES ET NATURELLES

Summary

The deforestation is increasing worldwide mainly due to human activities. This has a considerable impact on carbon dioxide levels and increases global warming. Suriname is still considered a High Forest cover, Low Deforestation (HFLD) country, but there are areas that undergo high deforestation pressure due to population growth and the gold rush.

In this study the focus will be on predicting the deforestation risk in a shifting cultivation and gold mining site in Suriname using spatial factors. Gold mining is now suspected to be the main driver of deforestation in Suriname and the population growth in some villages in the interior can lead to high deforestation pressure in this area. The spatial risk of deforestation due to shifting cultivation and gold mining will be predicted by creating a model using Random Forest, which is a powerful machine learning classifier tool.

The explicative variables used in the deforestation location model were the distances to former deforestation, villages, rivers and streams. The data were analyzed using open-source software's: GRASSGIS and R.

The generalized error of the model for shifting cultivation and gold mining were 7.43% and 4.46% respectively. Distance to former deforested areas seemed to be the most important predictor for both study areas, which can be explained by former studies showing that locations surrounded by recently deforested land have a higher risk to be deforested. The distances to streams seemed to be more important for deforestation in the shifting cultivation study area than in the gold mining area, whereas the distances to villages were more important in the gold mining site than in the shifting cultivation site.

When the model was used with data from another period in the future, this resulted in a higher error rate meaning that the model cannot predict accurately for future predictions. The ratio of forest gain to forest loss shows that the regeneration of deforested areas due to gold mining is slow compared to the regrowth in shifting cultivation, whereas the forest loss is much greater in the gold mining than in shifting cultivation. Finally, the deforestation risk maps produced by the model give a better idea of the deforestation pressure on the shifting cultivation and gold mining study area.

Keywords: shifting cultivation, gold mining, deforestation pressure, Random Forest tool

Table of Contents

Summary 1
List of abbreviations 3
1 Introduction 4
1.1 Approach of the study 5
2. Materials and Methods 6
2.1 Study area 6
2.2 Data gathering 7
2.2 Data processing 8
3. Results 19
3.1 Modeling using Random Forest 19
3.1.1 Modeling with data 2004-2008 19
3.1.2 Modeling with data 2008-2013 25
3.2 Deforestation potential map 27
3.3 Deforestation trend per year per study area 29
3.4 Ratio forest gain to forest loss 30
4 Discussion 31
5 Conclusions 35
Acknowledgement 36
References 37

List of abbreviations

REDD	Reducing Emissions of Deforestation and forest Degradation
OGS	Commission Regulation Gold Sector
GMD	Geological Mining Services
SBB	Foundation for Forest Management and Production Control
CELOS	Center for Agriculture Research in Suriname
GRASS	Geographic Resources Analysis Support System
GIS	Geographic Information System
FCMU	Forest Cover Monitoring Unit
OOB	Out Of Bag

1 Introduction

Deforestation is an important driver of carbon emissions worldwide. It is responsible for 6-17% of global anthropogenic carbon dioxide emissions in the atmosphere (A. Bacciniet al., 2012). According to the Food and Agriculture Organization of the United Nations (FAO) 7.3 million hectares of forest are lost each year (Alina Bradford, 2015). In many tropical countries forests are degraded and destroyed to expand timber, mining and agricultural industries (Ramirez-Gomez, S. 2011). Deforestation is causing loss of biodiversity, disruption in the water cycle, soil erosion, decrease in life quality and enhancing the greenhouse effect (Alina Bradford, 2015). The deforestation rates in the Guiana Shield are relatively low in comparison with other tropical forests. They contain the largest primary tropical rainforest on earth, which is of great importance in the context of climate change. However, the development of the economy and demography of the countries in the Guiana Shield leads to an increasing deforestation pressure (Baseline Report for NC-IUCN, 2001). Suriname is known as a High Forest cover Low Deforestation country, but there are areas undergoing deforestation pressure due to population growth and the gold fever (Garry D.Peterson & Marieke Heemskerk, 2001). Gold mining is now suspected to be the main driver of deforestation in Suriname and the population growth in the interior could lead to a deforestation pressure. Within REDD+, Reducing Emissions from Deforestation and forest Degradation, or any public management policy aiming to reduce the effects of human activities on the forest, it is important to know more about the patterns of deforestation for each driver. There are two types of patterns: spatial patterns and temporal patterns. The spatial patterns of deforestation are the structures on Earth and can be recognized by their arrangement such as a line or clustering of points. Temporal pattern of deforestation is the deforestation change in time. Combining the spatial patterns, based on the location model, and the temporal trends, that include the amount of deforestation, a deforestation model for Suriname can be created. In this study the focus will be on the spatial patterns of deforestation due to gold mining and shifting cultivation. The process of deforestation in gold mining and shifting cultivation is different and causes different patterns. The difference in patterns makes the comparison between gold mining and shifting cultivation easier. Shifting cultivation is a method of small-scale farming that includes clearing and burning the land, planting and harvesting crops, and then abandoning the land before moving to a new area (Katherine Lininger, 2011). A very small part carried out in this study focuses on extracting the temporal trends of deforestation due to shifting cultivation and gold mining. By characterizing the ratio of forest loss to forest gain, this will indicate the difference between shifting cultivation and gold mining.

1.1 Approach of the study

Research questions that were considered in this study were:

- Which explicative variables should be taken into account for modeling and why?
- Which period of the data should be used for calibrating the model?
- Is the model valid for future predictions?
- What is the temporal trend of the deforestation in shifting cultivation area and gold mining area from the period 2000-2013?
- What is the ratio of forest gain to forest loss in a shifting cultivation area compared to a gold mining area?

Sub-objectives that were detected from the research questions are:

- Determining the important and relevant explicative variables for the model;
- Determining the difference of deforestation between a shifting cultivation and gold mining area.

The deforestation model was created in order to build deforestation potential maps showing the spatial patterns of deforestation due to gold mining and shifting cultivation.
To create the model, some explicative variables have been taken into account. Global forest change data of the period 2004-2008 were used to calibrate the model. To test if the model can be used for a period in the future, the global forest change data 2008-2013 were used.
The temporal trend of the deforestation in the shifting cultivation study area and gold mining study area has been estimated per year. The ratio of forest gain to forest loss was also estimated by dividing the forest gain with the forest loss.

2. Materials and Methods

In the previous chapter the objective of this study has been described, which is to create a model that predicts the spatial probability of deforestation in Suriname due to shifting cultivation and gold mining. To achieve this objective, the method to carry out this research will be specified and described in this chapter.

2.1 Study area

This study focused on two areas: a shifting cultivation site and a gold mining site. The shifting cultivation study site was chosen in an area where there are a lot of villages and people doing shifting cultivation. The gold mining area was chosen in an area where deforestation has increased due to gold mining during the last ten years. See below figure 2.1 for the shifting cultivation and the gold mining study areas.

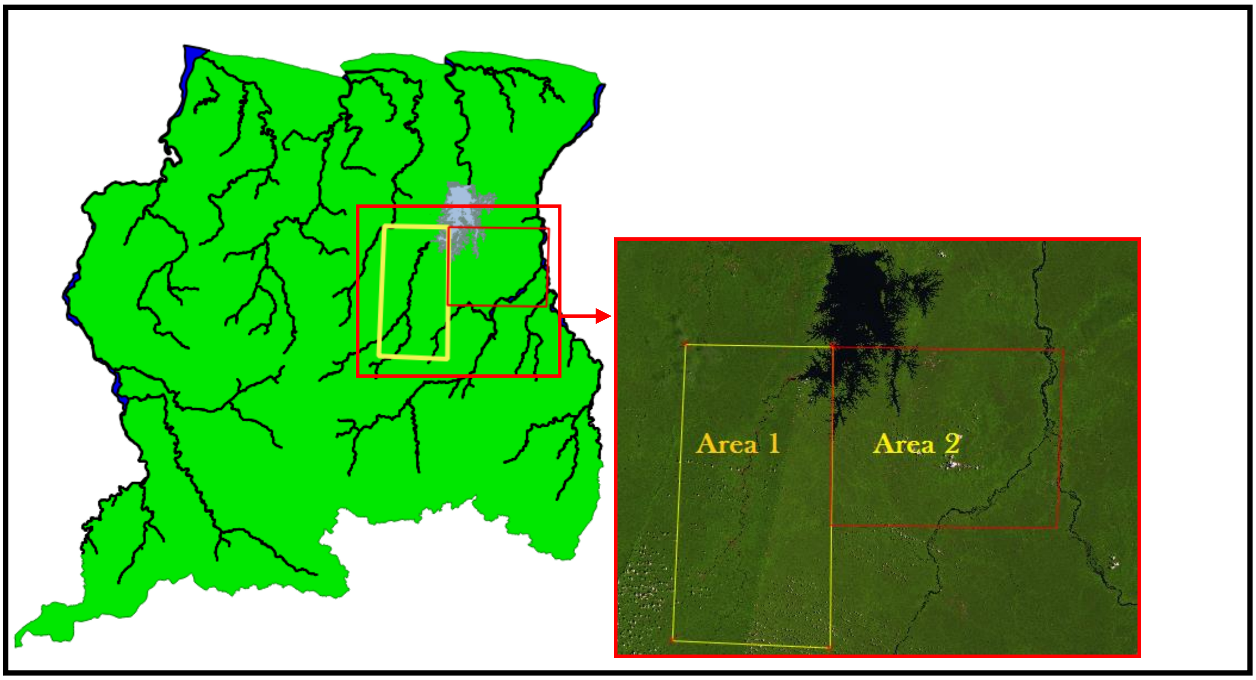


Figure 2.1 The shifting cultivation (Area1) and gold mining (Area2) study site

The shifting cultivation study site (Area 1)

The shifting cultivation study site covers an area of 6186 km^2 and is located upstream of the Suriname River at Boven-Suriname in district Sipaliwini. Along this part of the river there are a lot of maroon villages. The villages can only be accessed by the river, but not by road. The villagers rely on the forest as a source of food, fuel, medicine and land for agriculture (Katherine Lininger, 2011). Agriculture is performed using the shifting cultivation method. Shifting cultivation is a type of small-scale farming that involves clearing the land, burning the plant material, planting and harvesting the crops, and then abandoning the land to go fallow. Fallow periods in a shifting cultivation system vary and can be long enough to let forests regenerate in abandoned land (Katherine Lininger, 2011).

The gold mining study site (Area 2)

The gold mining study site covers an area of 5513 km^2 and includes the Lelygebergte, a part of the Brokopondo Lake, the Marowijne River and the Tapanahony River. The gold mining sites are accessible through the rivers and the Brokopondo Lake. There is one gold mining company with mining authorization and a lot of small-scale and medium-scale gold miners in this study site (John Johans (OGS), personal communication, February 2015). Along the Marowijne River and the Tapanahony River there are also villages where shifting cultivation is likely to occur.

Before starting to mine, an area is being deforested to put camps and a working space. The average measured size of a mining pit can be 80mx80m. The most small-scale goldminers mine land-based secondary gold deposits that are near forest creeks and rivers (Garry D. Peterson & Marieke Heemskerk, 2001). The intensity of gold mining sites is assumed to be dependent on a combination of the gold price, the presence of gold and the accessibility of the area (Pansa (GMD), personal communication, February 2015).

2.2 Data gathering

Table 2.1 gives an overview of how the data was gathered and table 2.2 gives an overview of all the data that has been used in this study.

Table2.1 Overview of the approach for data gathering

Interviews	**Internet**
The institutions where data was gathered: • SBB (Foundation for Forest Management and Production Control) • CELOS (Center for Agriculture Research in Suriname) • Tropenbos International Suriname • GMD (Geological Mining Services) • OGS (Commission Regulation Gold Sector)	Global forest change 2000-2013 data by Hansen: • Tree canopy cover for year 2000 • Global forest cover loss 2000–2013 • Global forest cover gain 2000–2012 downloaded from the website: http://earthenginepartners.appspot.com/science-2013-global-forest/download_v1.1.html

Table2.2 Overview of the purpose and source of the data used

Data	Purpose	Source
Tree cover 2000	Forest cover 2000 map	http://earthenginepartners.appspot.com/science-2013-global-forest/download_v1.1.html
Forest cover loss 2000-2013	Deforestation maps for 2000-2004, 2004-2008, 2008-2013	http://earthenginepartners.appspot.com/science-2013-global-forest/download_v1.1.html
Forest cover gain 2000-2012	Forest gain map 2000-2012	http://earthenginepartners.appspot.com/science-2013-global-forest/download_v1.1.html
Villages	Raster map for distance to villages as predictor	Foundation for Forest Management and Production Control
Streams	Raster map for distance to streams as predictor	Foundation for Forest Management and Production Control
Rivers	Raster map for distance to rivers as predictor	Foundation for Forest Management and Production Control
Shifting cultivation	The shifting cultivation shapefile was used to exclude the shifting cultivation areas in the gold mining study site	Foundation for Forest Management and Production Control
Borders of Suriname	The shapefile of the borders of Suriname was used to extract the data that covers Suriname	Foundation for Forest Management and Production Control

2.2 Data processing

After gathering these data, they were processed using the software GRASS GIS (Geographic Resources Analysis Support System) and analyzed with R, following the steps below:

- **Step1: Data projection**

  For adjusting the deforestation and the forest cover data to data available in Suriname, the maps were reprojected to EPSG: 32621, WGS 84/UTM zone 21N. This also allows an easier interpretation of the results by using meters as the spatial unit for distances instead of degrees of latitude/longitude.

- **Step2 : Data processing**

  All the necessary data were imported into a GRASS database, where after binary maps were created for:

- Forest cover 2000: The tree cover 2000 data is a raster with values indicating the percentage of canopy closure in every pixel. The very high values correspond to forests and the very low values correspond to non-forested pixels due to human activities or naturally not forested. Forest definition differs among regions (Louis V. Verchot et al., 2007). Using a higher threshold means more exclusion of land area (see figure 2.1 below).

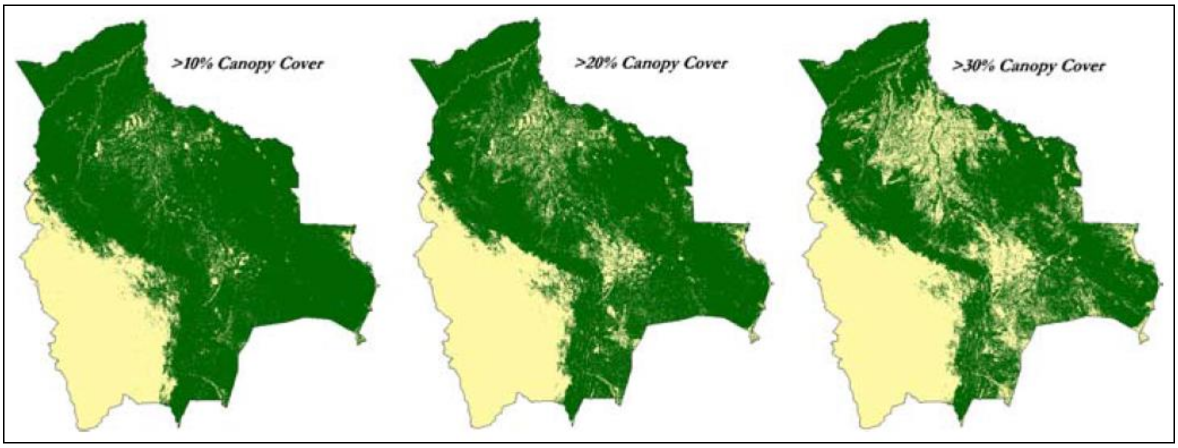


Figure2.2 Exclusion of areas in Bolivia based on crown cover (Louis V. Verchot et al., 2007)

For the forest cover 2000 map of Suriname a threshold of 75% crown cover was chosen, which is an arbitrary choice that was considered as relevant for a tropical rainforest like in Suriname. A threshold that is too high would produce an underestimation of the forest cover and a lower threshold would give the opposite. By taking a lower threshold, naturally degraded forest might be included, which are subject to natural evolution and can disappear naturally.

- Deforestation and no-deforestation (2000-2013): The deforestation and no-deforestation maps were created for three periods: 2000-2004, 2004-2008 and 2008-2013. The no-deforestation maps are the maps that indicate areas covered by forest in 2000 and which are still forested in the next period. The data were divided into three periods, because: (1) Data from 2000-2004 gave information about former deforested areas, (2) data from 2004-2008 was used to create the model, and (3) data from 2008-2013 was used to test if the model is still valid for future predictions.

- Forest gain (2000-2012): The forest gain map was available for the period 2000-2012. The forest gain is defined as the opposite of forest loss, or a non-forest to forest change between 2000-2012 (Hansen et al., 2013).

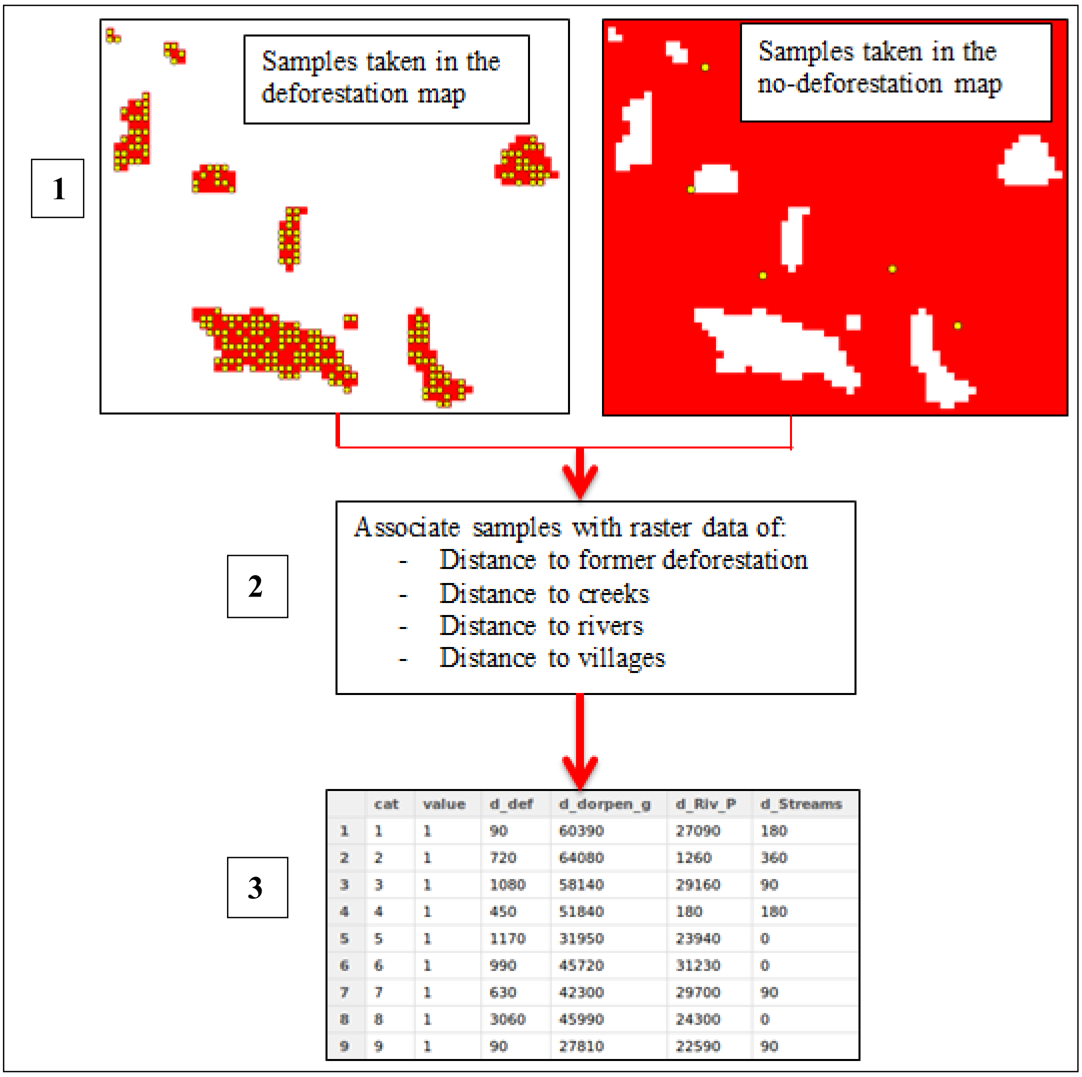
For creating the deforestation model, some explicative variables must be considered. The shapefiles of these explicative variables were rasterized, where after the distances raster's were computed. The explicative variables that have been taken into account in the model are given below:

- Distance to former deforested areas: Locations surrounded by recently deforested land are likely to suffer deforestation themselves. This deforestation increases the probability of other nearby locations (Isabel M.D. Rosa et al, 2013). In this study the distance to former deforestation was estimated from 2000-2004 deforestation map.
- Distance to villages: People in villages surrounded by forests rely on the forest as a source of food, fuel, medicine and land for agriculture. This can lead to replacing forests with small-scale agricultural fields (Katherine Lininger, 2011) using the shifting cultivation method. A study from 2009 showed that 90 percent of the families living along the Tapanahony, Lawa and Marowijne River are mostly dependent on small-scale gold mining (Leontien Cremers et al., 2013). This means that the presence of villages is also an important factor to predict locations of gold mining.
- Distance to streams: Most small-scale gold miners in Suriname mine land-based secondary gold deposits that are near forest creeks (Garry D. Peterson & Marieke Heemskerk, 2001). Alluvial gold-mining is then closely associated with smalls streams and creeks.
- Distance to rivers: Rivers can be used as a transport medium to shifting cultivation lands. Rivers shapefile overlaps with streams shapefile, but only for main rivers. The hypothesis is that main rivers are used for transportation, so are necessary for people to settle and practice shifting cultivation. Small creeks are more associated with alluvial gold-mining.

- **<u>Step3: Dataset building</u>**

Figure 2.3 illustrates the three steps used for building the datasets:

1. Pixels were sampled in the deforestation and no-deforestation maps (same number of pixels for both layers);
2. The values of the corresponding explicative variables were then associated to each sampled pixel;
3. The sampled pixels in the deforestation and no-deforestation maps and the corresponding values of the explicative variables were brought together in one table which formed the dataset.



	cat	value	d_def	d_dorpen_g	d_Riv_P	d_Streams
1	1	1	90	60390	27090	180
2	2	1	720	64080	1260	360
3	3	1	1080	58140	29160	90
4	4	1	450	51840	180	180
5	5	1	1170	31950	23940	0
6	6	1	990	45720	31230	0
7	7	1	630	42300	29700	90
8	8	1	3060	45990	24300	0
9	9	1	90	27810	22590	90

Figure 2.3 Overview of the dataset building in three steps: (1) Pixel sampling in deforestation and no-deforestation maps, (2) Associate samples with corresponding explicative variable, and (3) Merging the samples in one table

Five datasets following the method explained above:

- A dataset for shifting cultivation and another for gold mining in 2004-2008: pixels were sampled in the deforestation and no-deforestation maps of 2004-2008 at step1. This dataset was created to calibrate the model.
- A dataset for shifting cultivation and another for gold mining in 2008-2013: pixels were sampled in the deforestation and no-deforestation maps of 2008-2013 at step 1. This dataset was created to validate the model for future predictions.
- A cleaned dataset for gold mining in 2004-2008: this dataset is from the gold mining study area, but excludes samples in the shifting cultivation area occurring in this region. The shifting cultivation shapefile created by the Forest Cover Monitoring Unit (FCMU) of Suriname was used to exclude the shifting cultivation derived deforestation in the gold mining study area. The deforestation and no-deforestation pixels were then sampled in this new region, where no shifting cultivation occurs. The reason for sampling in the gold mining study site again, was to see if the shifting cultivation occurring in the gold mining study site would influence the model.

➢ **Step 4: Modeling with Random Forests**

Random Forest (Breiman, 2001) is a method that can be used for regression, but also for the classification of multisource remote sensing and geographic data. In this study it has been used as a classifier. This Random Forest classifier includes a combination of tree classifiers which are generated using random vector sampled independently from the input vector (see figure 2.5). Each tree will then cast a vote for the most popular class to classify an input vector of explained variable (M. Pal, 2005). Figure 2.4 shows an example of a “decision tree”. An input is entered at the top and as it traverses down the tree the data gets bucketed into smaller and smaller sets (Dan Benyamin, 2012).

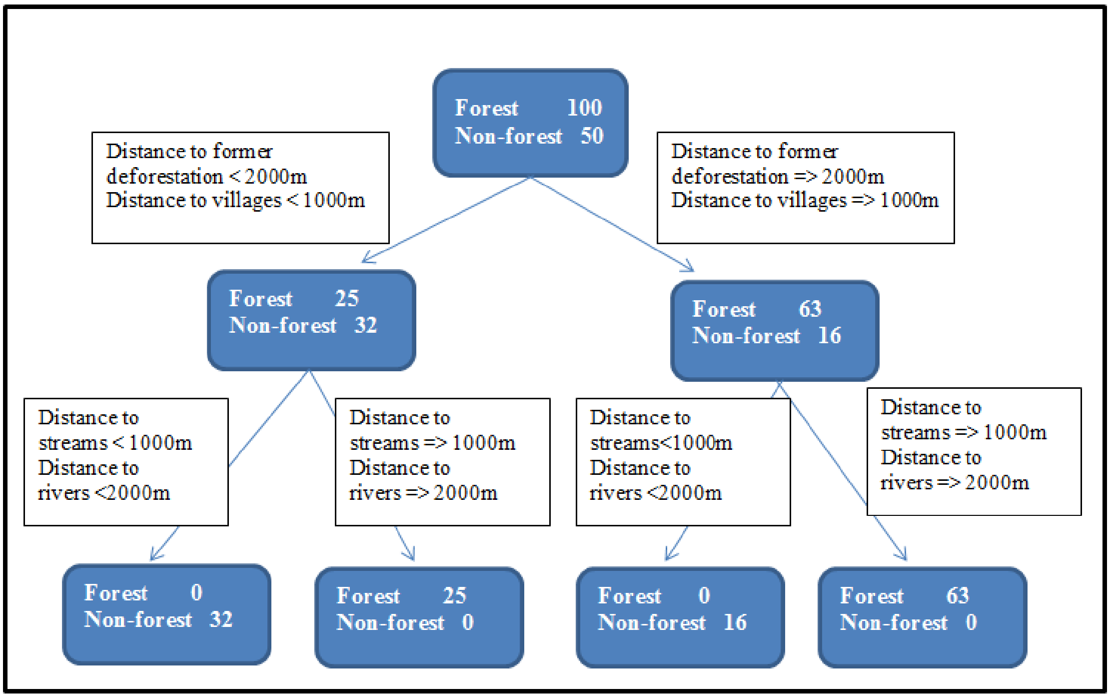


Figure2.4 An illustration of a decision tree where an input is entered at the top and the data is getting bucketed into smaller sets as it traverses down the tree

In figure 2.5 an illustration is given of the process in Random Forest. This process will be explained below:

First, training datasets were randomly selected with replacement from the total data to create individual trees. Each training dataset contained 2/3 of the total data. The rest of the data (1/3) was included in another dataset, called Out of Bag (OOB). The individual trees choose explicative variables randomly and in combination at each node to grow the tree (V.F. Rodriguez-Galiano et al., 2011). In this study there are four explicative variables in total. The number of variables at each node was set on 2, which is the default value. At the first node two variables at random were chosen. Each variable will then find a value which optimizes the split. At the next node, two variables will be chosen again at random from all explicative variables and follow the same procedure. Thus, one variable can be used more than one time in the nodes of the tree. The OOB dataset was formed for every individual tree and is not considered as training data, but for classification by the tree of the training dataset. The proportion between the misclassifications and the total number of OOB dataset gives an unbiased estimation of the generalization error of the model (V.F. Rodriguez-Galiano et al., 2011). It is not necessary to use a test subset independent from the training set to do cross validation.

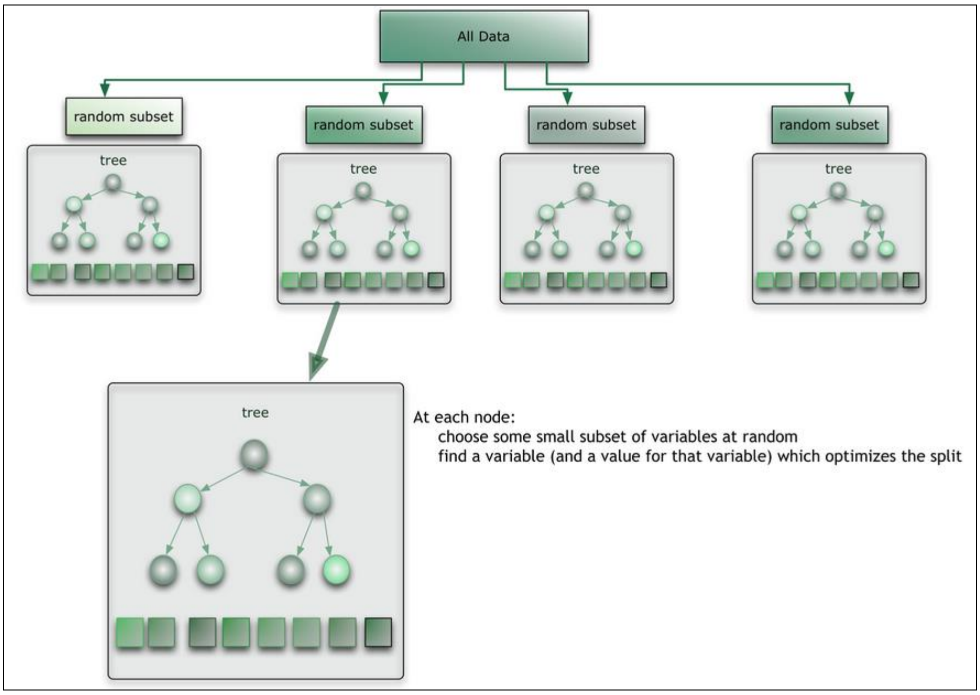


Figure 2.5 An illustration of the Random Forest method where training datasets are randomly chosen and individual trees are grown by the nodes (Dan Benyamin, 2012).

The number of training datasets to be selected can be determined by the Out of Bag (OOB) error. In this study 100 trees were chosen, because after plotting the 100 cases in a graph it was discovered that the OOB error rate is already constant reaching the 100 samples (see figure 2.6). This means that there is no need to take more than 100 sample cases. The generalization error converges as the number of trees increases, so there is no over-fitting the data (V.F. Rodriguez-Galiano et al., 2011).

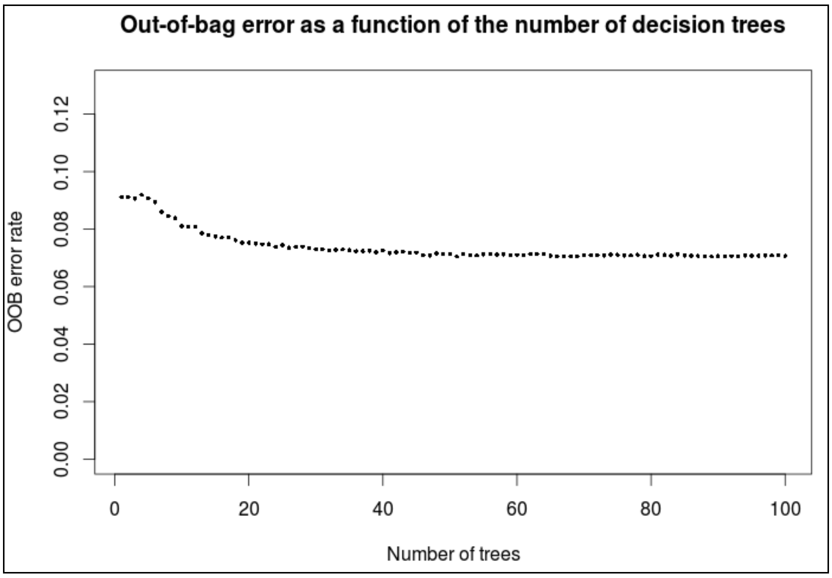


Figure2.6 Example of the constancy of OOB error rate when reaching 100 trees

In this study there was still a verification conducted on the OOB error rate. This was done by manually dividing the samples between calibration and validation samples: 2/3 for the calibration and 1/3 for the validation of the model 2004-2008. After calibrating the model, it was validated using the validation sample. A confusion matrix was then created and out of the confusion matrix the error rate was estimated and compared with the OOB error rate. This will give an indication of the reliability of the OOB error rate from Random Forest. The false positive error rate and false negative error rate were also calculated out of the confusion matrices. The false positive error rate is the incorrect classification to deforestation and the false negative error rate is the incorrect classification to forest.

Relative importance of explicative variables

Random Forest (Breiman, 2001) can also assess variable importance by estimating the mean decrease accuracy. The mean decrease accuracy is determined during the calculation phase of the OOB error. The calculations are carried out tree by tree as the random forest is build. The more the accuracy of Random Forest decreases due to a variable, the more important this variable is. The higher the mean decrease accuracy, the most important the variable is (Andy Liaw & Matthew Wiener, 2002).

Advantages and disadvantages of Random Forest

Some advantages of using Random Forest in remote sensing are:

- It can still run efficiently with large data bases
- The important variables in the classification are being estimated.
- It performs an internal unbiased estimate of the generalization error (Out of Bag error)
- It is robust to noise (V.F. Rodriguez-Galiano et al., 2011).

The main disadvantage of Random Forest is that it can be difficult to understand the split rules for the final classification, because of the multiple classification trees generated from resampling the same dataset. Therefore Random Forest can be considered to be a black box type classifier (V.F. Rodriguez-Galiano et al., 2011).

➢ **Step 5: Partial dependence plots**

Partial dependence plots were created to show the relationship between each explicative variable with the predicted probabilities of deforestation obtained from Random Forest. This was executed using the R software. Figure 2.7 shows an example of a partial plot. The y-axis shows the logit scale of the probability of deforestation and the x-axis shows the explicative variable (D. Richard Cutler et al., 2007). The probability of deforestation is estimated by the amount of votes for a deforested pixel during the process within Random Forest.

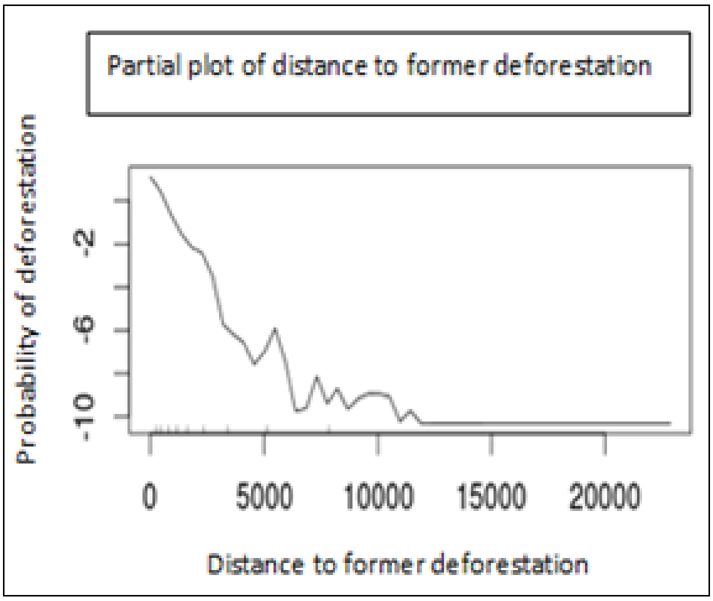


Figure 2.7 An example of a partial plot where the relationship of the probability of deforestation is shown with the distance to former deforestation

- **Step 6: Create deforestation potential map**

  The raster images of the explicative variables were used to create the deforestation potential map. The potential of deforestation for each pixel was calculated based on the random forest model previously calculated, associated with the corresponding value of the same pixel for all explanatory variables. The term "probability" of deforestation is not appropriate to use in this step, because of its definition in statistics. The term "probability" in statistics is equal to the number of ways that something can happen divided by the total amount of outcomes. As the amount of pixels in the model is known, the term "probability" can still be used. However, in the deforestation potential map the amount of forested and deforested pixels is unknown. This is why the term "probability" will not be used anymore, when the map is created.

The packages used in R were:

- Rgdal (Roger Bivand et al., 2015)
- Spgrass6 (Roger Bivand et al., 2014)
- RandomForest (Leo Breiman et al., 2014)
- Raster (Robert J. Hijmans et al., 2015)
- SDMTools (Jeremy Van DerWal et al., 2014)

Deforestation per year per study area

Binary maps were created for the deforestation per year per study area. This was executed in GRASSGIS using R. With the r.report tool of GRASSGIS the amount of deforested pixels could be calculated for each year and this was then converted into hectares. The deforestation trend was then illustrated by creating a graph. This graph will be able to give a better idea on the temporal patterns of deforestation due to gold mining and shifting cultivation whereby comparison could be made.

Forest gain and forest loss

Besides estimating the deforestation per year per study area, the forest gain and forest loss data by Hansen were also compared with each other within the two study sites. The pixels in the forest gain map with a value that corresponds to regrowth were summed. The total of these pixels were then converted into hectares. The same principle was used for the forest loss data. The ratio was estimated by dividing the amount of gained forest with the amount of forest loss. This will give an idea on the intensity of shifting cultivation and gold mining on the forest. The hypothesis is that in gold mining the areas being deforested are much greater than in shifting cultivation. As gold miners also use excavators and other machines, it is presumed that the regrowth of forest will be very slow in contrary to shifting cultivation.

3. Results

After describing the methods in the previous chapter, the results of the analyzed data will be given in this chapter. The first results are outputs from the models calibrated for the shifting cultivation and gold mining areas respectively, using Random Forest. The deforestation potential maps for both study sites, the deforestation trends from 2000 to 2013 and the ratio of forest loss and forest gain per study area will also be given.

3.1 Modeling using Random Forest

Three datasets have been used for modeling with Random Forest: The shifting cultivation sample, the gold mining sample and a “cleaned” gold mining sample excluding areas of shifting cultivation within this region. Table 3.1 gives an overview of the information of each dataset used in the modeling process.

Table3.1 Overview of the information of each dataset used in the modeling process

Study site	**Data**	**Explicative variables**	**Number of trees**	**Number of variables tried at each split**	**Out of bag (OOB) estimate of error rate**
Shifting cultivation (Area1)	Calibration sample (2004-2008)	• distance to former deforestation • distance to villages • distance to rivers • distance to streams	100	2	7.43%
Gold mining (Area2)	Calibration sample (2004-2008)				4.46%
Gold mining (Area2)	“Cleaned” gold mining calibration sample (2004-2008) without shifting cultivation				5.09%

3.1.1 Modeling with data 2004-2008

Error rate

For the shifting cultivation study site (area1) the calibrated random forest model produced an Out Of Bag (OOB) error rate of 7.43% and for the gold mining study site (area2) the OOB error rate was 4.46%. For the gold mining study site another dataset was created, which includes no shifting cultivation occurring in the gold mining region. A new model was calibrated using this new dataset with the same explicative variables as the former model. The OOB error rate with this dataset is 5.09%. To verify the reliability of the OOB error rate, the models were validated using the validation sample for prediction. With the confusion matrix the error rate was estimated and compared with the OOB error rate.

Table 3.2 gives an overview of the confusion matrices and the error rates estimated for every study area. For the shifting cultivation site (2004-2008) the OOB error rate is approximately the same as given in table 3.1. The false negative error rate is 1.8% and the false positive error rate is 12.6%. The confusion matrix with the calibration samples within Random Forest, give a false negative error rate of 2.3% and a false positive error rate of 12.7%.

The error rate of the gold mining dataset (2004-2008) given in table 3.2 is not the same as the OOB error rate of this dataset in table 3.1. The false negative error rate for this dataset is 1.6% and the false positive error rate is 3.2%. Out of the confusion matrix with the calibration samples estimated within Random Forest, the false negative error rate is 1.6% and the false positive error rate is 7.3%.

The error rate of the cleaned gold mining dataset is approximately the same as the OOB error rate of this dataset in table 3.1. Out of the confusion matrix of the validation sample of this new dataset, the false negative error rate is 2% and the false positive error rate is 8.7%. Out of the confusion matrix with the calibration samples estimated within Random Forest, the false negative error is 2.3% and the false positive error rate is 7.9%.

Table 3.2 Overview of the confusion matrices and error rates of the validation samples for each dataset of 2004-2008

			Observation		**False negative error rate**	**False positive error rate**	**Error rate (incorrect prediction/ total cases)**
Shifting cultivation (2004-2008)	**Prediction**		**0**	**1**	1.8 %	12.6%	7.2%
		0	3923	83			
		1	565	4405			
Gold mining (2004-2008)		**0**	6254	106	1.6%	3.2%	2.4%
		1	207	6361			
Cleaned gold mining (2004-2008)		**0**	3044	67	2%	8.7%	5.4%
		1	290	3267			

Relative importance of explicative variables

The relative importance of the explicative variables was also extracted from Random Forest model. Figure 3.1 shows the relative importance of the explicative variables in the shifting cultivation study site, where the distance to former deforested areas is the most important variable, followed by distance to streams, distance to villages and at least distance to rivers.

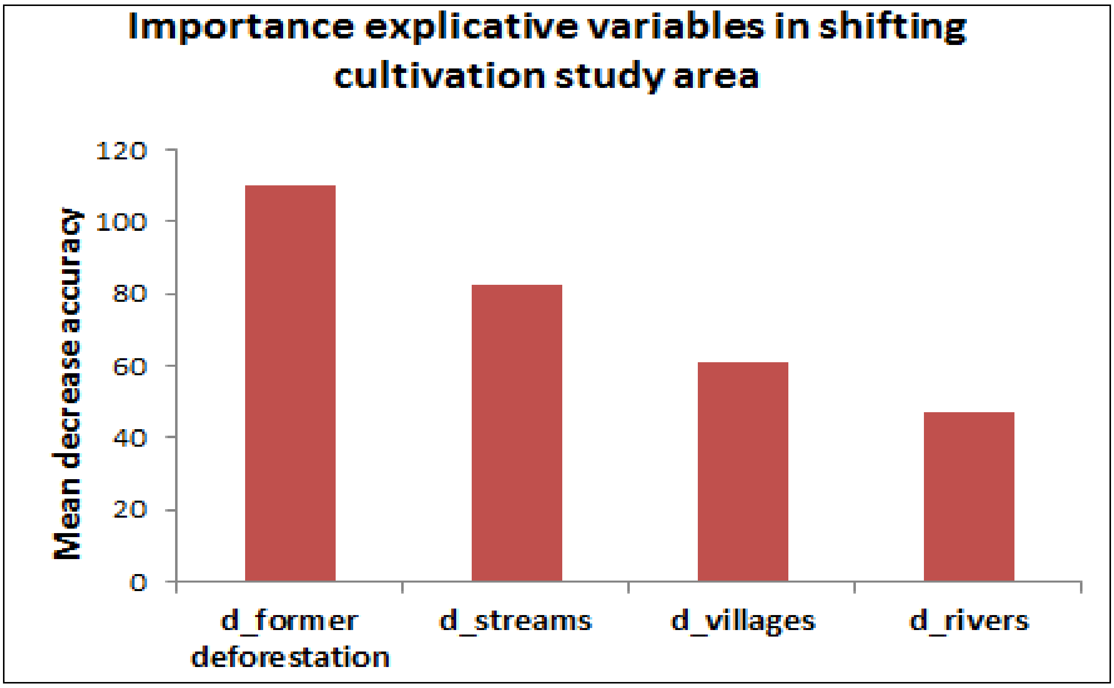


Figure 3.1 The relative importance of explicative variables in the shifting cultivation site

Figure 3.2 shows the relative importance of the explicative variables in the gold mining study site, where the distance to former deforestation areas was the most important variable, followed by distance to villages, distance to rivers and at least distance to streams.

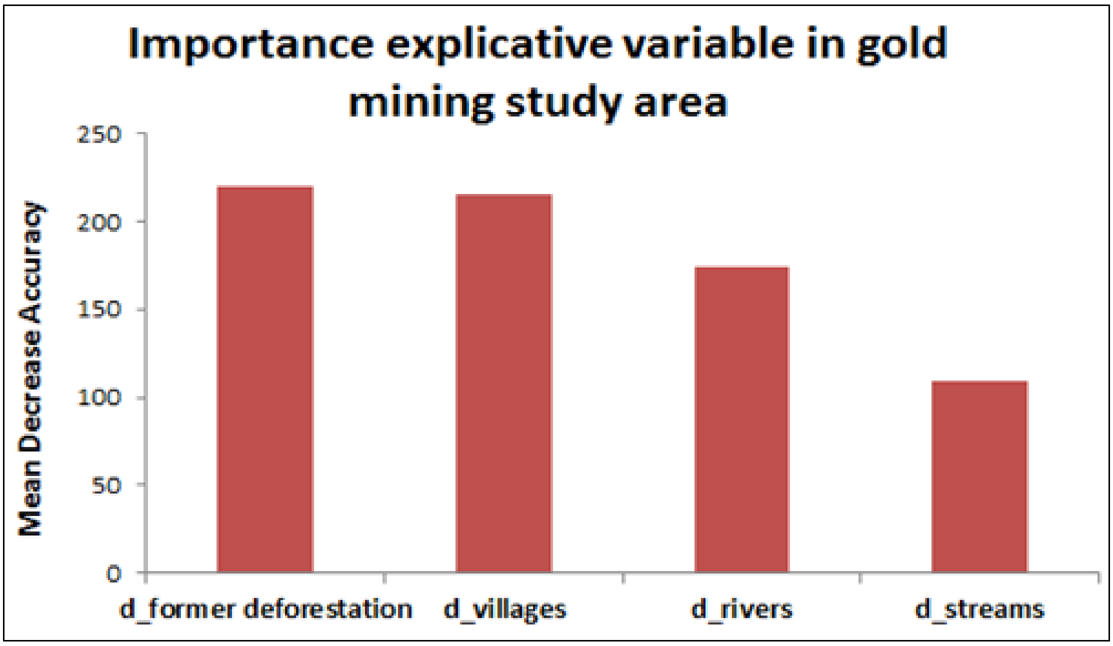


Figure 3.2 The relative importance of explicative variables in the gold mining study site

Figure 3.3 shows the relative importance of the explicative variables in the gold mining study site without shifting cultivation, where the distance to former deforestation areas was the most important variable, followed by distance to villages, distance to rivers and at least distance to streams.

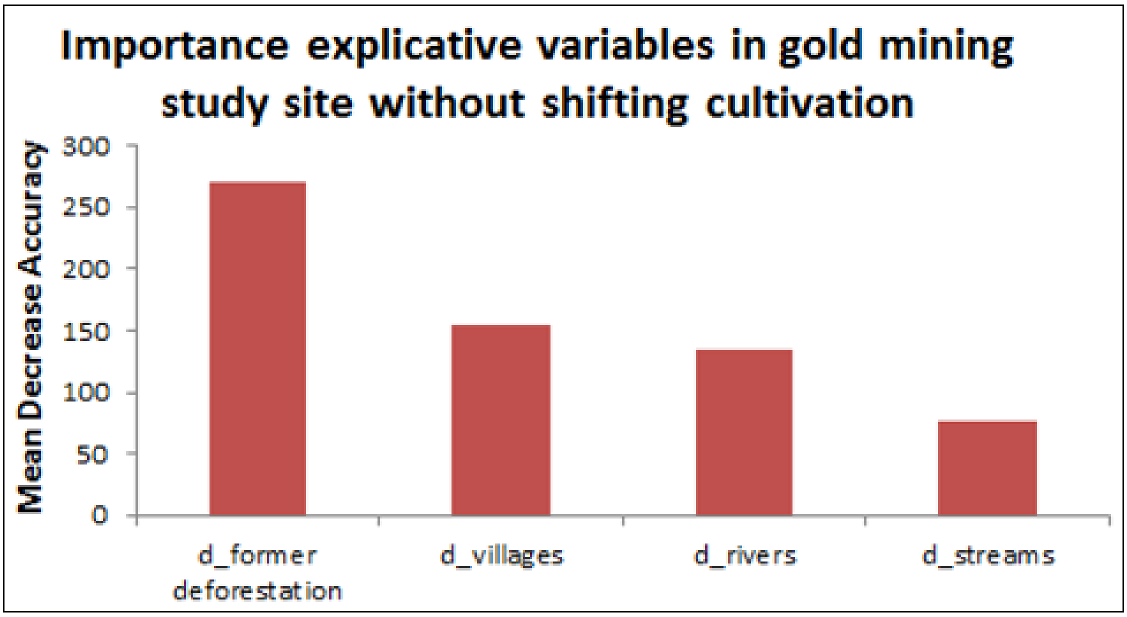


Figure 3.3 Importance explicative variables gold mining without shifting cultivation area

Partial plots

Shifting cultivation study site (Area1)

Partial plots show the relationship with the probability of deforestation with each explicative variable. The x-axis on the partial plots indicates the distances in meters and on the y-axis is the logit scale of the probability of deforestation.

Figure 3.4 shows the partial plot of the distance to former deforestation, distance to streams, distance to villages and distance to rivers related with the probability of deforestation.

According to the partial plot of former deforestation, the probability to be deforested decreases, when the distance to former deforestation increases.

The partial plot of the distance to streams shows that the deforestation is likely to increase between approximately 0 and 100m from streams, but at a distance of more than 200m the deforestation decreases and remains low.

The partial plot of the distance to villages has a clear trend of a decrease of the probability of deforestation when the distance to villages increases.

The partial plot of the distance to rivers shows an unstable trend which is hard to interpret. At approximately 13km from rivers there is suddenly an increase of the probability of deforestation and then remains constant.

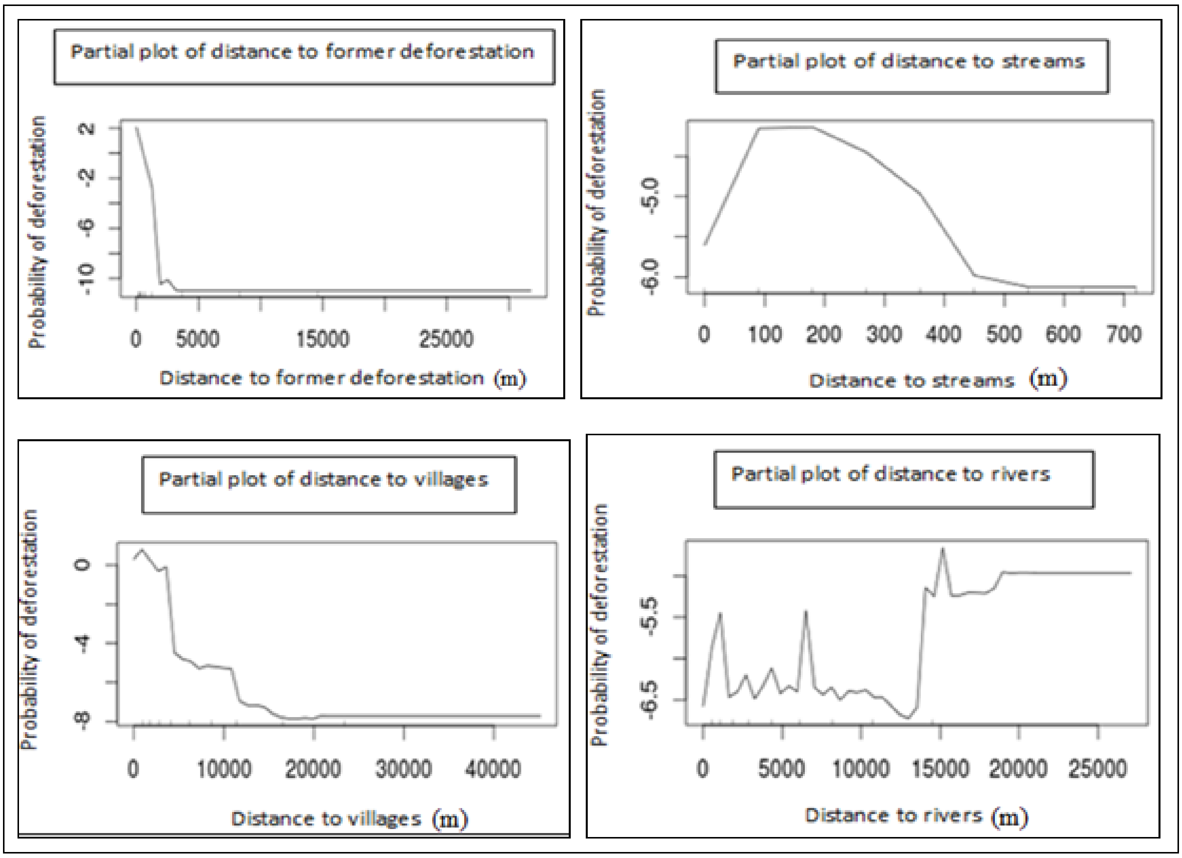


Figure 3.4 Partial dependence plots of the probability of deforestation with the explicative variables in the shifting cultivation area

<u>Gold mining study site (Area 2)</u>

Figure 3.5 shows the partial plots of the variables for the gold mining study site.

In the partial plot of distance to former deforestation, the probability of deforestation decreases.

The partial plot of distance to villages shows an unstable trend, where the probability of being deforested increases and decreases constantly while the distance to villages increases.

The partial plot of distance to rivers shows a decrease of the probability of deforestation, but at a distance of approximately 10km from rivers it begins to increase more and more.

Finally, the partial plot of distance to streams shows that the probability of deforestation will decrease when the distance to streams increases.

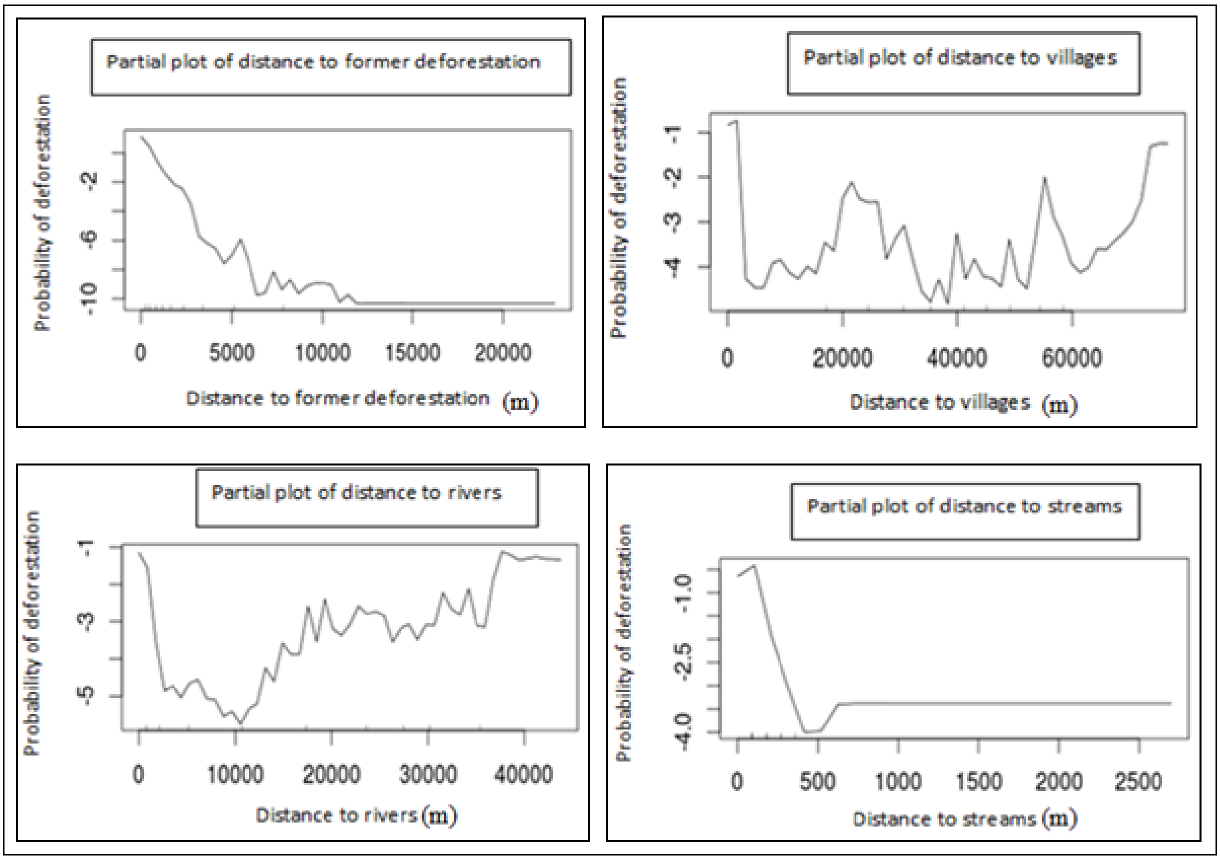


Figure 3.5 Partial dependence plots of the probability of deforestation with the explicative variables in the gold mining area

3.1.2 Modeling with data 2008-2013

After calibrating the model with the dataset of 2004-2008, the dataset of 2008-2013 was used for prediction with the same model build for 2004-2008. This was performed to know if the model is still valid for the period 2008-2013. Two confusion matrices were created for shifting cultivation and gold mining to estimate the false negative and false positive error rates and compare these with the error rates of the dataset from 2004-2008.

Table 3.3 Confusion matrix shifting cultivation sample (2008-2013)

		Observation		**Error rate (incorrect prediction/ total cases)**
		0	**1**	
Prediction	**0**	17838	5283	18.7%
	1	2245	14800	

The confusion matrix in table 3.3 gives a false positive error rate of 11.2% and a false negative error rate of 26.3%. The confusion matrix in table 3.4 of the gold mining sample gives a false positive error rate of 7% and a false negative error rate of 74.7%. According to table 3.3 and table 3.4, the error rates of 2008-2013 are much higher than the error rates from the dataset 2004-2008.

Table 3.4 Confusion matrix gold mining sample (2008-2013)

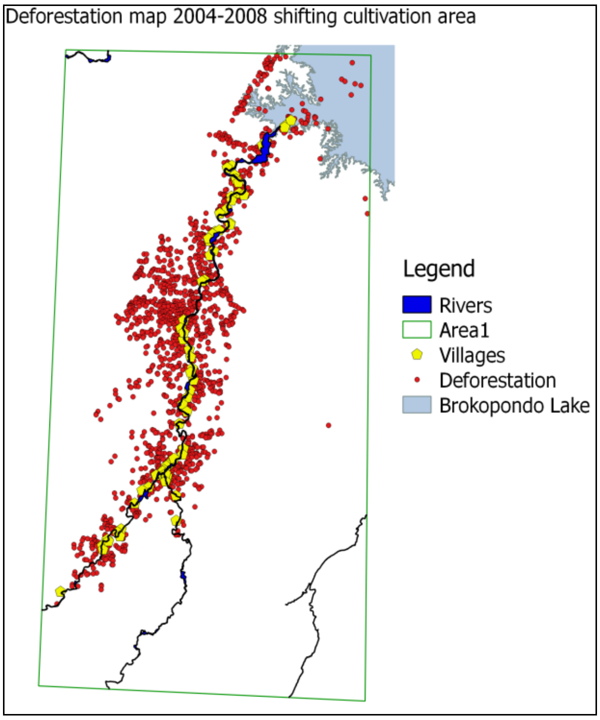
		Observation		**Error rate (incorrect prediction/ total cases)**
		0	**1**	
Prediction	**0**	64605	51887	40.8%
	1	4877	17595	

3.2 Deforestation potential map

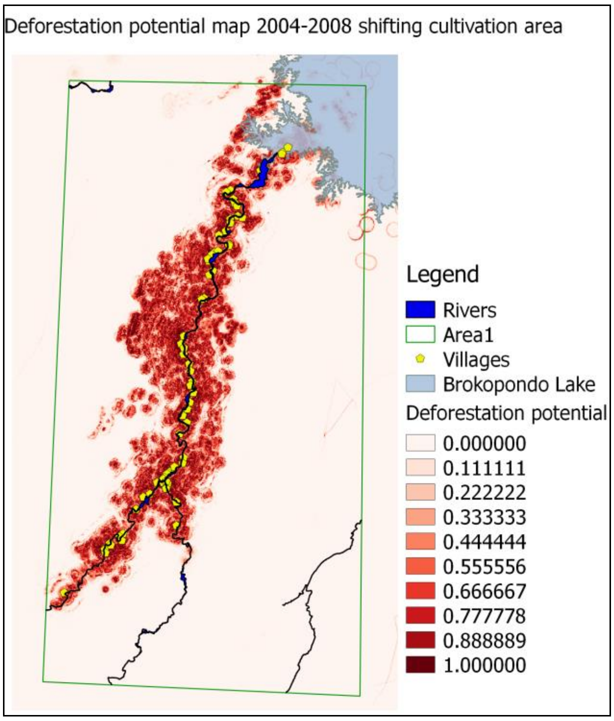
After creating the deforestation models for shifting cultivation and gold mining, deforestation potential maps were made for both study sites.

Deforestation potential map for Area 1 (shifting cultivation area)

In map 3.1 the observations of deforestation from Hansen data is given for the period 2004 to 2008 for the shifting cultivation study area. In map 3.2 the potential of deforestation is shown. The darker the area, the more risk it has to be deforested. The area close to the rivers and villages has the highest potential to be deforested. This map is created based on the dataset of 2004 -2008. The two maps overlap quite well.



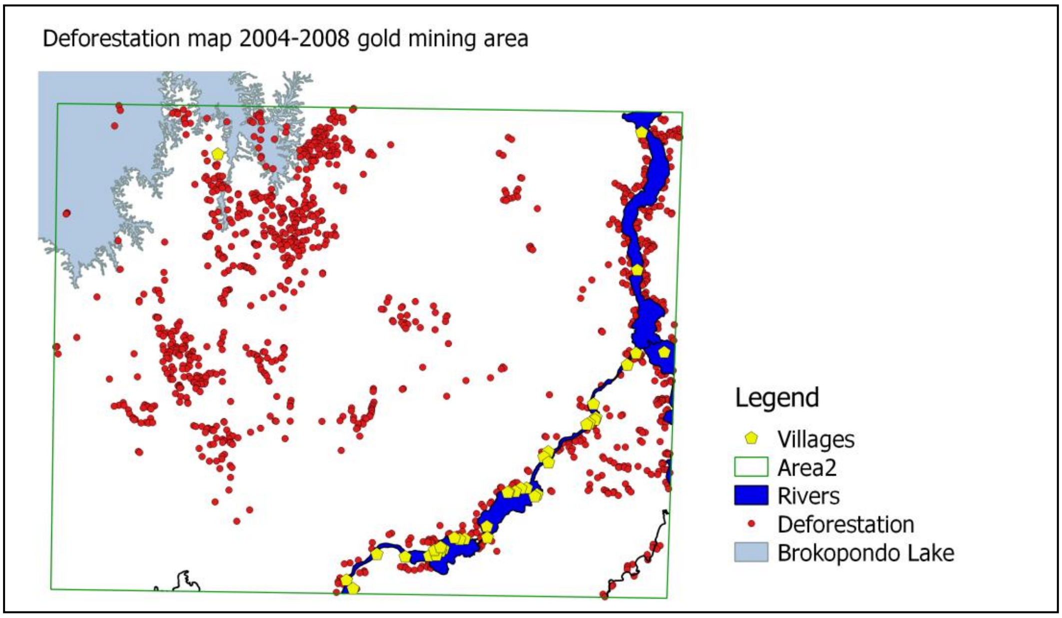
Map3.1. Deforestation map in 2004-2008 in shifting cultivation study area



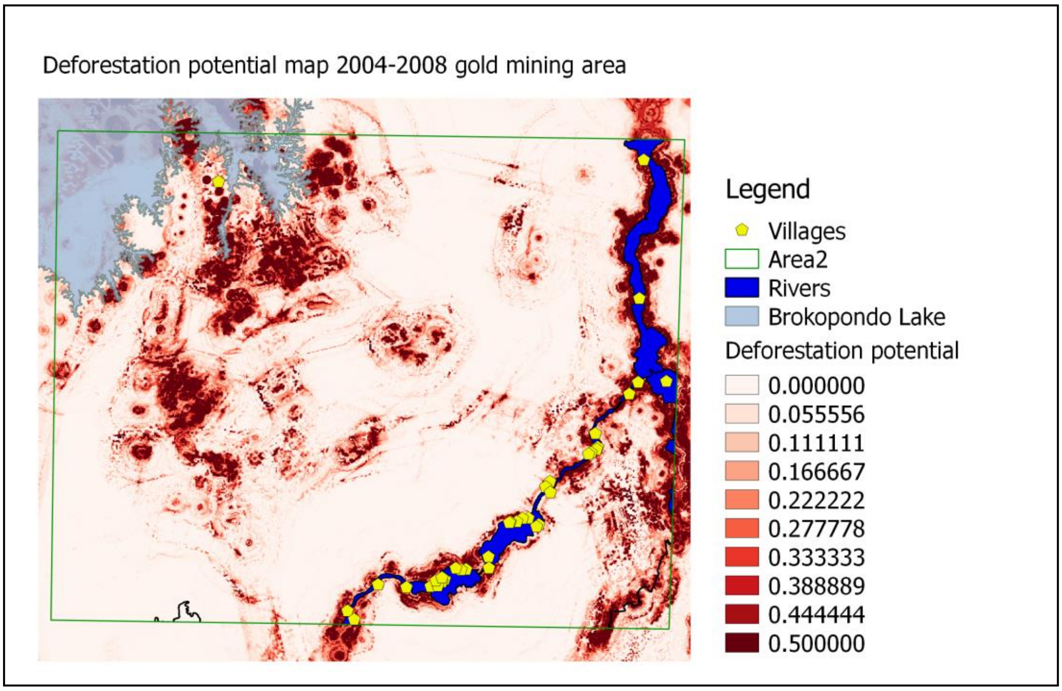
Map 3.2 Deforestation potential map 2004-2008 in shifting cultivation area

Deforestation potential map for Area2 (gold mining area)

In map 3.3 the deforestation of the period 2000 to 2004 is given for the gold mining study area. In map 3.4 the deforestation potential map is shown. The darker the area, the more potential it has to be deforested. This map is created based on the dataset of 2004 -2008.



Map3.3 Deforestation map 2004-2008 gold mining study area



Map3.4 Deforestation potential gold mining study area

3.3 Deforestation trend per year per study area

The deforestation per year per study site is given in this paragraph. Also the trend of forest gain and forest loss in the two study areas is shown. In figure 3.6 the deforestation trend fluctuates. The highest deforestation in the shifting cultivation study site occurred in the year 2008 and 2012.

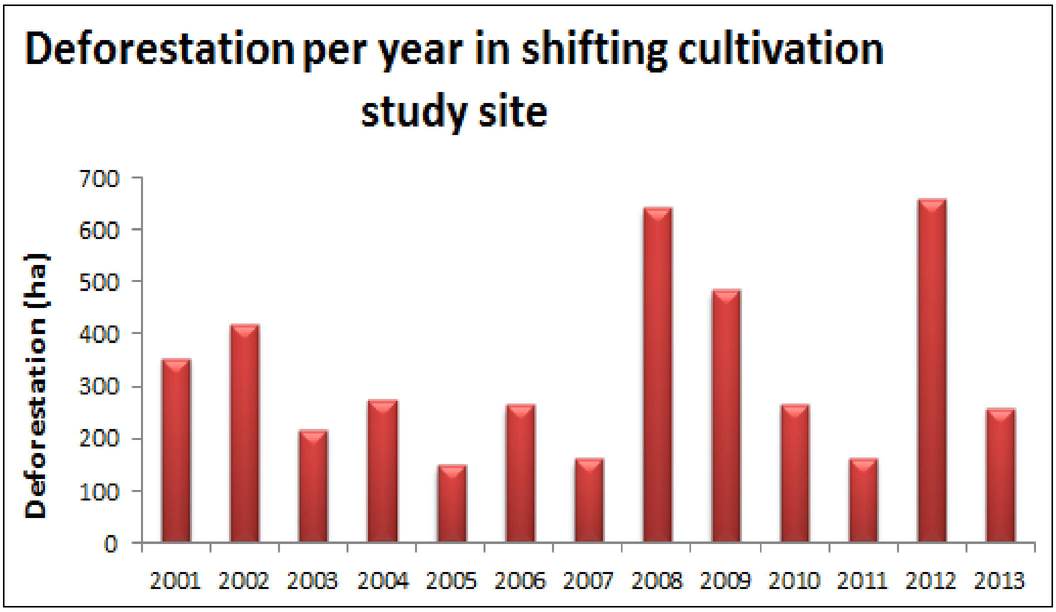


Figure 3.6 Deforestation trend 2001-2013 for shifting cultivation

In figure 3.7 the deforestation in the gold mining study site has an increasing trend, but decreases in 2013.

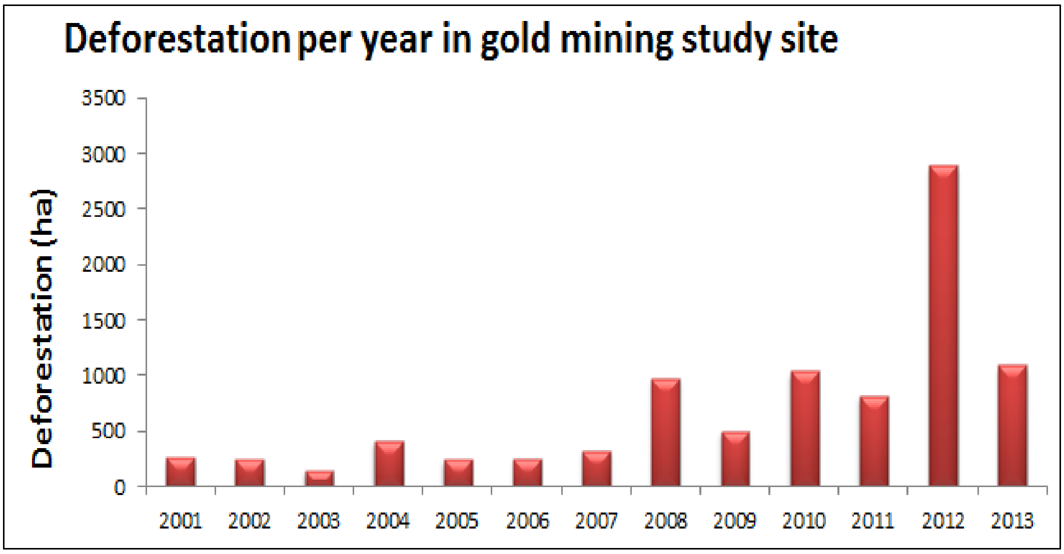


Figure 3.7 Deforestation trend 2001-2013 for gold mining

3.4 Ratio forest gain to forest loss

According to table 3.5 the ratio of forest gain to forest loss is much higher in the shifting cultivation study site than in the gold mining study site. Figure 3.8 shows that the forest gain in the shifting cultivation study site is higher than in the gold mining site. On the contrary, the forest loss is less in the shifting cultivation area than in the gold mining area.

Table 3.5 Overview of forest gain and forest loss in both study sites and their ratio's

	Shifting cultivation study site (ha)	**Gold mining including the shifting cultivation area (ha)**
Forest gain	2457.9	643.32
Forest loss	4870.8	9722.61
Ratio	0.5046	0.066

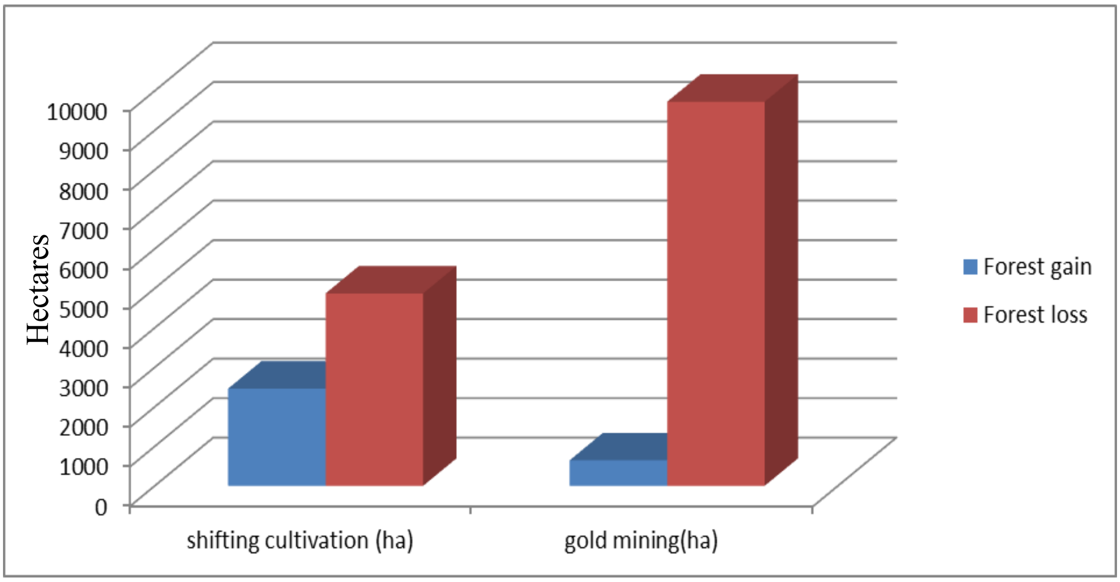


Figure 3.8 Forest gain and forest loss in the two study areas

4 Discussion

The selection of explicative variables

For the modeling of deforestation in shifting cultivation and gold mining area there were four explicative variables used: distance to former deforestation, distance to rivers, distance to streams and distance to villages. There were also other variables considered, but not taken into account, because of the following:

- The greenstone belt: The greenstone belt is characterized by metavolcanics, which consists of three distinct formations: Rosebel, Armina and Paramaka formations. These formations may be gold bearing and determines the gold mining activities in the greenstone belt (Heemskerk M, 2009). A part of the greenstone belt covers a large portion of the gold mining study site, but was not taken into account as an explicative variable. As the greenstone belt covers a big part of the study area, it will not be able to give information about specific gold mining sites. If the study area was partially located outside from the greenstone belt, this could give a better result about the specific gold mining sites.
- The Brokopondo lake: The Brokopondo lake covers a small part of the study area and deforestation is distributed upon the whole study area. This makes it difficult to presume that the lake can be considered an explicative variable, as it has no link with the deforested areas.
- Distance to roads: In the two study areas there are barely no roads detected by 30meter resolution satellite images. The FCMU in Suriname is working on a gold mining study at the moment. Roads due to gold mining will also be mapped. For the moment there is no complete road map available of these study areas.
- Nature reserves: The two study areas are not covered by any nature reserve.
- Slope: During the interview at the Geological Mining Services and the Management Gold mining Sector, it was said that gold mining also occurs on the slope of the mountains. Slope could be considered an explicative variable in the gold mining areas. At first, it was included in the model for gold mining, but the model seemed to have higher error rate with the slope then when it was excluded. This could be explained by a higher amount of alluvial gold mining (near creeks) than on the slopes of mountains.

Deforestation caused by shifting cultivation

The difference between the false positive error rate and the false negative error rate is 10.4%. This big difference can be explained by the difficulties of predicting deforestation correctly. A forest patch is most likely surrounded with forest, whereas a deforested patch can be deforested or forested in the surrounding area.

The contagious character of deforested pixels is the reason why the distance to former deforestation is the most important variable in the shifting cultivation study site.

In the partial plot of the distance to former deforestation and the distance to villages shown in figure 3.4, there is a clear trend of the probability of deforestation when moving further away from former deforestation and from villages. People tend to deforest in an area where deforestation already occurred. When moving further away from former deforested areas, there are less deforested areas, because people are less inclined to clear land in the middle of the forest. The agriculture lands of people living in the villages decreases when the distance to villages increases. It is presumed that these people prefer having their agricultural lands near to them. In the partial plot of distance to streams the probability of deforestation increases, but at approximately 200meters it decreases. Former studies in the shifting cultivation area have shown that agricultural lands are close to creeks (M.Playfair et al., 2009). When moving further away from creeks, the agricultural lands decrease. This interpretation has to be very carefully done, because of the small range of the distance to streams due to the dense hydro network. Every forested or non-forested pixel can be close to streams, which makes it difficult to identify the patterns. The partial plot of the distance to rivers has an unstable trend, which is hard to interpret. Despite this, the increase of the probability of deforestation at approximately 13km from rivers can be explained by the fact that most of the land close to rivers is already deforested. When people want to deforest more land, they have to move further away from rivers.

The temporal trend of deforestation in the shifting cultivation study site shown in figure 3.6 varies from the period 2000-2013. From 2007-2013 the annual mean of deforestation is higher than from 2001-2007. It is presumed that this increase is due to the population growth in this study area. For the peak in 2008, this can be explained by a project carried out in this area about the establishment of plantation with *Jatropha curcas,* a plant with seeds that contain 37% oil and can be processed to produce biodiesel fuel (Sainik Basti, 2013). The peak in 2012 is presumed to be due to the joint of clearing land because of bad weather in the last years where it was difficult to clear new land for agriculture.

Deforestation caused by gold mining

The validation sample of the gold mining study site gives an error rate which is lower than the OOB error rate in table 3.1. After resampling the validation samples, the error remained the same, which is still lower than the OOB error rate. This was left unexplained. The model with the dataset 2008-2013 shows a higher error rate than modeling with the dataset 2004-2008, which means that the model cannot predict accurately for the future. This can be explained by the classifiers which were generated in the calibrated model with 2004-2008 dataset that cannot correctly predict the classification for the period 2008-2013.

The most important explicative variable is the distance to former deforestation. The contagious character of a deforested pixel is also common in the gold mining study site. Gold miners in Suriname tend to explore sites near the site of another miner who is known or believed to have found a good location (Leontien Cremers et al, 2013). After sampling in the gold mining site again, but excluding the shifting cultivation part, the results of variable importance are the same as before. This means that according to the Random Forest method there is no difference in the importance of the predictors with or without the shifting cultivation part in the gold mining study site. Based on field knowledge and interviews at the Geological Mining Services and Commission Regulation Gold mining sector, deforestation due to gold mining is more explained by the presence of streams. According to Random Forest, the distance to villages is more important than the distance to streams or rivers. It is recommended to carry out an investigation more in depth and consider the situation in reality after a statistical approach.

The partial plot of distance to former deforestation and to streams shown in figure 3.5 has a clear trend of a decrease of the probability of deforestation when moving further from former deforestation and streams. The decrease of the probability of deforestation in the partial plot of distance to streams can be explained by the alluvial gold mining method often used in Suriname which is closely associated with creeks or small streams. In the partial plot of distance to rivers, there is first a decrease and then an increase of the probability of deforestation. It is presumed that at approximately 10km from rivers there are streams which are interesting for alluvial gold mining. The partial plot of distance to villages has an unstable trend which makes it difficult to interpret, even though it is shown to be an important explicative variable for the gold mining study site.

The temporal trend of deforestation in the gold mining study site shown in figure 3.7 indicates an increase of deforestation. The peak in 2012 is presumed to be due to the gold price, but also the technical modernization which allows more gold to be exploited. This results in an acceleration of mining and clearing of more land (Garry D. Peterson & Marieke Heemskerk, 2001). A decrease of the deforestation in 2013 is presumed to be due to the intervention of the Commission Regulation Gold Sector in this area.

Comparison of the deforestation caused by shifting cultivation and gold mining

The distance to former deforestation is the most important explicative variable in the shifting cultivation and gold mining site. The contagious character of forested and deforested pixels is common in both study areas. The distance to villages, streams and rivers have other processes in each study area, which explains the different ranking of importance of the explicative variables.

The deforestation trend per year is different for each study area. The deforestation in the shifting cultivation study area is presumed to be due to the population growth and in the gold mining study area it is presumed to be due to the gold price and technical modernization. An assumption for the peak of deforestation in 2012 is that the satellite images were bad due to bad weather which resulted in an overestimation of deforestation.

The ratio of forest gain to forest loss in the shifting cultivation study site is higher than in the gold mining study site. The ratio in the shifting cultivation study site can decrease if the people in the villages will shift from a shifting cultivation method to permanent agricultural lands. The forest gain in shifting cultivation area is higher than the forest gain in gold mining area. However, the forest loss in shifting cultivation area is lower than in gold mining area. The regeneration of deforested areas due to gold mining seemed to be slower compared to regeneration in a shifting cultivation area. In shifting cultivation, the soil is scarcely disturbed and less area is being cleared compared to gold mining. Within one or two years there will already be regeneration of the cleared area in shifting cultivation (Garry D.Peterson & Marieke Heemskerk, 2001).

5 Conclusions

After describing the results and discussions in the previous chapters, the conclusions of this study will be given here.

This study aimed at predicting the deforestation risk in a shifting cultivation and gold mining site using spatial factors. The deforestation potential maps that were created with the location model give an overview of areas that have high potential of being deforested due to shifting cultivation or gold mining. This kind of information can be useful for policymakers of the country, because it enables them to manage the effects of human activities in the forest.

It can also be concluded that the Random Forest performed well showing low error rates for predicting. The model can also be used for future predictions, but the error rates will be higher. There is still a lot of work to be done in order to improve the model for future predictions. Even though, it can still generate an idea of the deforestation risk in specific areas. Furthermore, the Random Forest algorithm can estimate the importance of variables, which can be useful in the reality knowing which variables can predict deforestation the best.

The ratio of forest gain to forest loss in the shifting cultivation study area is much higher than the ratio of the gold mining study site. The forest loss in the gold mining area is much more than in the shifting cultivation area, whereas the forest gain is much lower in the gold mining site than in the shifting cultivation site. The ratio of the shifting cultivation study site can decrease when people move from shifting cultivation method to a permanent plot. In a permanent plot, there will possibly be no regrowth.

The temporal trend of the deforestation in the gold mining study site and shifting cultivation site shows an increase that could be explained by population growth, the gold price and the technical modernization. Out of the socio-economic data an intensity model can be created. The combination of the intensity model with the location model creates the possibility to build a future deforestation map. In a political perspective this will lead to a better policy to minimize the effects of deforestation due to shifting cultivation and gold mining.

Acknowledgement

I am grateful to ONF international for the financial support during the Master study "Ecology of Tropical Forest". Furthermore, my gratitude goes out to the Foundation for Forest Management and Production Control for giving me the opportunity to enroll in this Master study. Special thanks to Camille Dezecache who supervised me throughout this research with patience. Finally, I want to thank the Center for Agriculture Research in Suriname, the Geological Mining Services, Tropenbos International Suriname, the Foundation for Forest Management and Production Control and the Commission Regulation Gold Sector for their willingness to share field information and geographic data for the progress of this study.

References

Baccini A, Goetz S J, Walker WS, Laporte N T, Sun M, Sulla-Menashe D, Hackler J, Beck PSA., Dubayah R, Fried MA, Samanta S, Houghton RA (2012) Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps. Nature Climate Change (2): 182-185

Baseline Report for NC-IUCN (2001) A baseline analysis for the Guiana Shield region

Bivand R, Keitt T, Rowlingson B, Pebesma E, Sumner M, Hijmans R, Rouault E (2015) Bindings for the Geospatial Data Abstraction Library

Bivand R, Krug R, Neteler M (2014). Interface between GRASS 6+ geographical information system and R

Bradford Alina (2015) Deforestation: Facts, Causes & Effects. Live science. Downloaded on 12-6-2015 at http://www.livescience.com/27692-deforestation.html

Breiman L (2001) Random Forests. Machine Learning, 45: 5–32

Breiman L, Cutler A, Liaw A, Wiener M (2014) Breiman and Cutler's random forests for classification and regression

Cremers L, Kolen J, de Theije M (eds) (2013) Small-scale goldmining in the Amazon. The cases of Bolivia, Brazil, Colombia, Peru and Suriname. 85-101

Cutler DR, Edwards TC Jr., Beard KH, Cutler A, Hess KT, Gibson J, Lawler JJ (2007) Random Forests for classification in ecology. Ecological Society of America. Ecology 88(11): 2783-2792

Dan Benyamin (2015) A Gentle Introduction to Random Forests, Ensembles, and Performance Metrics in a Commercial System. Downloaded on 22-5-2015 at https://citizennet.com/blog/2012/11/10/random-forests-ensembles-and-performance-metrics/

Hansen MC, Potapov PV, Moore R, Hancher M, Turubanova SA, Tyukavina A, Thau, D, Stehman SV, Goetz SJ, Loveland TR, Kommareddy A, Egorov A, Chini L (2013) High-Resolution Global Maps of 21st-Century Forest Cover Change. Science 342: 850-853

Heemskerk M (2009) Small scale gold mining in Suriname: 10
Hijmans RJ, van Etten J, Mattiuzzi M, Sumner M, Greenberg JA, Lamigueiro OP, Bevan A, Racine EB, Shortridge A (2015) Geographic Data Analysis and Modeling

Kimbrough L (2014) Gold mining expanding rapidly along Guiana Shield, threatening forests, water, wildlife.downloaded on 11-6-2015 at http://news.mongabay.com/2014/1022-gfrn-kimbrough-gold-mining-expanding-in-guiana-shield.html

Liaw A, Wiener M (2002) Classification and Regression by randomForest. R News 2/3:18-22

Lininger K (2011) Small-Scale Farming and Shifting Cultivation. Union of concerned scientists:3

Peterson GD, Heemskerk M (2001) Deforestation and forest regeneration following small-scale gold mining in the Amazon: the case of Suriname. Environmental Conservation 28 (2): 117–126

Playfair M, Esseboom M, Helstone A (2009) Exploring Sustainable Agricultural Practices with a view to develop more Sustainable and Productive Land Use. CELOS: 9

Ramirez-Gomez, S. 2011. *Spatial Drivers of Deforestation in Suriname.* Paramaribo : Center for Agriculture research in Suriname; Tropenbos Suriname, 2011.

Rodriguez-Galiano VF, Ghimire B, Rogan J, Chica-Olmo M, Rigol-Sanchez JP (2011) An assessment of the effectiveness of a random forest classifier for land-cover classification. ISPRS Journal of Photogrammetry and Remote Sensing 67:93–104

Rosa IMD, Purves D, Souza C Jr, Ewers RM (2013) Predictive Modelling of Contagious Deforestation in the Brazilian Amazon. PLoS ONE 8(10): e77231.

Sainik Basti (2013) About Jatropha plant. Centre for Jatropha Promotion and biodiesel. Downloaded on 5-6-2015 at http://www.jatrophabiodiesel.org/aboutJatrophaPlant.php

VanDerWal J, Falconi L, Januchowski S, Shoo L, Storlie C (2014) Species Distribution Modelling Tools: Tools for processing data associated with species distribution modelling exercises

Verchot LV, Zomer R, van Straaten O, Muys B (2007) Implications of country-level decisions on the specification of crown cover in the definition of forests for land area eligible for afforestation and reforestation activities in the CDM. Climatic Change 81:415–430