2023

Driving forces behind deforestation in Upper West, Ghana

BACHELORTHESIS EARTH SCIENCES, ECONOMICS AND SUSTAINABILITY

MARIJN VAN T HOF (2707083)

FOREWORD

This report is the outcome of my bachelor's thesis earth sciences, economics and sustainability at the Vrije Universiteit Amsterdam. I would like to thank Dr. Eric Koomen for his guidance and support throughout the research. I would also like to thank Renato Stopic and Eduardo Dias for their support and for the chance to take seat in their office. Finally, I would like to thank Sommalife, and more specifically Joost Westerhout, for his feedback and expert knowledge on the study area.

This report summarizes the findings of my study on the forest cover changes and driving forces behind them. This report could be used to improve forest management in Upper West and could also be used to assist future studies in Northern Ghana on the topic.

ABSTRACT

Forests are a key source of the world's carbon and without them life on earth would be essentially unsustainable. Recent global studies show that Upper West, Ghana has not subject to deforestation the past decade. However satellite image analysis have shown severe forest cover changes in the region in the 2014-2022 period. The aim of this study was to explain deforestation in Upper West using a logistic regression and predict future locations for deforestation based on expected driving forces, and has done so successfully. Forest density is strongly correlated with deforestation, with denser forests having higher probability of deforestation. Population also showed strong positive correlation with deforestation, indicating that population growth leads to more deforestation. Poor government policies also lead to more deforestation in the region as was confirmed by a positive correlation between protected areas and deforestation. The logistic regression was used to create a prediction for future deforestation with satisfactory level of confidence and can be used to help policy makers prevent deforestation in the future.

TABLE OF CONTENTS

1. INTRODUCTION

With 294 Gt of carbon stored in forest cover, 42 Gt in dead wood, and 324 Gt in soils and litter, forests are a key source of the world's carbon (FAO, 2010). According to studies, the estimated 658 Gt total carbon content of the world's forests in 2009 was higher than the amount of carbon present in the entire atmosphere (FAO, 2011). This shows that without forests, life on earth would be essentially unsustainable. Many economic activities such as mining, wood logging and farming continuously contribute to the degradation of these forests (Fagariba, Song, & Soule, 2018). Global deforestation has a huge effect on the climate and was estimated to cause 20% of the annual greenhouse gas emissions between 1990 and 2005 (Gupta, 2012).

Approximately 90% of African nations rely on agriculture and forest products directly or indirectly for economic and food security (Nielsen & Reenberg, 2010). One of these countries is Ghana, which is the country this report will focus on. About two thirds of Ghana is covered by savannah (Vedeld, Angelsen, Bojö, Sjaastad, & Berg, 2007). Roughly 6% of Ghana's total savanna zone is permanently forested, and roughly 50% is unreserved savanna woodland (Adusei & Dunyah, 2016). A study by Etwire et al. has revealed that Ghana has already lost about 80% of it's forest cover in a century (2013). With an average annual deforestation rate of 75,000 hectares, the original forest cover of 8.2 million hectares has been drastically reduced to about 1.7 million hectares (Acquah & Onumah, 2011). The direct causes for deforestation in Ghana are charcoal burning, hunting, agriculture, overgrazing and wood logging (Fagariba, Song, & Soule, 2018). The biggest indirect causes for deforestation in northern Ghana are population growth and a lack of an alternative livelihood (Fagariba, Song, & Soule, 2018).

The focus in this report will be on the Upper West region in Ghana, an area which is mostly covered by savannah areas. A lot of indigenous communities in this region rely on shea trees as an economic income. These trees produce shea nuts, which can be used to make shea butter. Shea butter has gained popularity recently, especially in western nations where it is primarily utilized in cosmetics for hair and skincare goods (Grand View Research, 2022). The communities that harvest these nuts mostly live in unprotected savannah areas, which means the trees in these areas can be cut down without any consequences (Sommalife, 2023). Sommalife, a non-profit organization in Ghana that helps communities get fair prices for shea butter, states that this is a threat to the livelihood of many of these communities (2023).

Previous global studies show that there is no forest left in Upper West and that this has been the case for the past decade (Global Forest Watch, 2023). However forest cover maps that were generated from Landsat 8 imagery show that forest cover has changed in the past decade (van 't Hof, 2023). As this forest cover loss was previously unknown there are no previous studies about the driving forces behind this deforestation. This study aims to explain Northern Ghana's deforestation using spatially explicit driving forces. This information can then be used to predict future deforestation, which is very helpful in creating a more targeted approach at protecting Northern Ghana's savannah forests.

The following research question and sub-questions are attempted to be answered in this research: 'How has deforestation in Northern Ghana changed over the past decade, as revealed through analysis of satellite imagery, and how can this deforestation be explained using spatially explicit driving forces?

Sub questions:

- 'Which spatially explicit driving forces are expected to drive forest loss in Northern Ghana?'
- 'Where does Northern Ghana lose forest cover?
- 'How can local forest loss be explained using spatially explicit driving forces?'

2. DATA & METHODOLOGY

This study aims to predict the spatial deforestation probability and explain the relationship between this process and driving factors. For this a statistical analysis using a logistic regression will be used.. The dependent variable in this regression will be deforestation and the independent variables are the expected driving forces. The data & methodology that was used for the variables is explained in detail in this chapter.

2.1 DEFORESTATION

Forest cover maps from the years 2014, 2018 and 2022 were used to generate deforestation maps. These forest cover maps have been generated previously with a machine learning algorithm using Landsat 8 and Google Earth Pro imagery. This process has been discussed in detail in a technical report (van 't Hof, 2023). In short, Landsat 8 images from 2014, 2018 and 2022 have been classified by a random forests classification model into four different forest cover classes: no forest (<5% tree cover), sparse forest (5-25% tree cover), semi-dense forest (25-50% tree cover) and dense forest (>50% tree cover). High resolution images from Google Earth Pro were used as reference data to train the random forests model. The model classified the Landsat 8 images with accuracies of 72%, 80% and 77% for 2014, 2018 and 2022 respectively. These accuracies are comparable to similar studies (Borges, Higginbottom, Symeonakis, & Jones, 2020).

To create a landcover change map from the forest cover maps the raster calculator was used with the following expression:

$$10 * Forestcover\ 2014 + Forestcover\ 2022$$

The forest cover maps include discrete values between 1 and 4 that each correspond to a forest cover class. In this a 1 corresponds to ‘dense forest’, a 2 to ‘semi-dense forest’, a 3 to ‘sparse forest’ and a 4 to ‘no forest’. The expression generates a new raster that shows a number between 11 and 44, with the first number being the forest cover class of 2014 and the second number being the forest cover class of 2022. This means that a pixel with the value 13 has gone down 2 classes, from ‘dense forest’ to ‘sparse forest’. This raster was then reclassified to 7 new classes to create a map of the forest cover change as can be seen in the Table 1.

As one of the goals of this study is to predict the locations for deforestation the regression only included the areas that have experienced a decrease in forest cover. To do this the raster was reclassified to 0’s and 1’s, with 0 being no deforestation and 1 being deforestation.

Table 1: Forest cover change classes and their corresponding meaning

Forest cover change class	Corresponding number & meaning
Heavy afforestation	41 (3 classes higher in forest density)
Medium afforestation	31, 42 (2 classes higher in forest density)
Light afforestation	21, 32, 43, (1 class higher in forest density)
No change	11, 22, 33, 44 (No change in class)
Light deforestation	12, 23, 34 (1 class lower in forest density)
Medium deforestation	13, 24 (2 classes lower in forest density)
Heavy deforestation	14 (3 classes lower in forest density)

2.2 EXPLANATORY VARIABLES

2.2.1 Literature review

As briefly mentioned in the introduction, the most important direct causes for deforestation in Northern Ghana are agriculture, wood logging and charcoal burns (Fagariba, Song, & Soule, 2018). And the biggest indirect causes are population growth, poor government policies and the lack of an alternative livelihood (Fagariba, Song, & Soule, 2018). With this knowledge and a literature review on studies that also use spatially explicit driving forces to explain deforestation a list of expected driving forces has been constructed in Table 2. Figure 1 also gives an indication of the spatial dispersion of some of the expected driving forces

Slope and elevation are expected explanatory variables for deforestation in this study and can be linked to the direct causes of deforestation. A steeper slope for example makes it less likely that agricultural land use will occur (Oğuz, 2020). Multiple previous studies have shown significant correlation between slope, elevation and deforestation (Bavaghar, 2015; Gayen & Saha, 2017), which make them interesting variables to include in the regression.

Forest density and distance to forest edge are used as potential explanatory variables for deforestation in this study. Distance to forest edge is a variable that is often used in similar studies that look at driving forces behind deforestation (Gayen & Saha, 2017; Kucsicsa & Dumitrica, 2019; Saha, et al., 2020). Forest density is not used that often in similar studies. But forest density, just like the distance to a forest edge, can be linked to the direct causes of deforestation in Northern Ghana. Mainly for wood-logging and charcoal burning a higher distance to forest edge and/or a lower forest density potentially lead to more travel costs. The study by Reddy et al. states for example that deforestation rates in denser forests are higher compared to open forests in Bangladesh (Reddy , Pasha, Jha, Diwakar, & Dadhwal, 2016). Forest density is expected to have a big effect on deforestation in Upper West as most 'semi-dense forest' and 'dense forest' areas have been deforested in the 2014-2022 period.

The argument of higher travel costs can also be used for the distance to the nearest river and road. This is why the distance to the nearest road and stream will be used as an explanatory variable for deforestation in this study. The distance to the nearest river and road are used often in similar regressions (Saha, et al., 2020; Kucsicsa & Dumitrica, 2019).

The distance to nearest settlement and building and population density are all used as explanatory variables for deforestation in this study and have been used before in similar studies about deforestation (Ludeke, Maggio, & Reid, 1990; Saha, et al., 2020). Population growth, an indirect cause of deforestation in Northern Ghana, can be linked to these variables (Fagariba, Song, & Soule, 2018). It can be linked directly to the population density as the population density will increase as population grows in an area. Population also indirectly leads to more buildings and larger/more settlements as more people need housing.

Poor government policies is an indirect cause of deforestation in Northern Ghana. The spatially explicit driving force that can be linked to this cause is the current protected forest areas. In theory these protected areas should protect forest areas and prevent deforestation. However government policies, like creating designated areas where forest is protected, are often more effective in democratic countries with higher levels of corruption control and protection of property right (Abman, 2018). Ghana is a flawed democracy with a democracy score of 6,43 in 2022, which makes it one of the most democratic African countries (Oluwole, 2023). However Ghana still shows signs of private media actively contributing to political corruption (Asomah, 2020). This makes the presence of protected areas an interesting driving force for Northern Ghana.

Table 2: Expected (spatially explicit) driving forces for Northern Ghana's deforestation, including their unit and range in the study area.

Driving forces	Unit	Range	Mean	Std. Dev.
Elevation	meter	10 – 502	258	67
Slope	degrees	0 – 53	3,6	2,5
Protected areas	yes/no	0-1	-	-
Distance to forest edge	kilometre	0 – 82,2	14,8	18,6
Population density	1000 inhabitants /10km2	0 – 107,2	0,06	1,18
Distance to nearest settlement	kilometre	0 – 28,9	4,3	4,4
Distance to nearest building	kilometre	0 – 12,4	1,0	1,5
Distance to nearest (major) road	kilometre	0 – 64,8	10,9	10,5
Distance to nearest river	kilometre	0 - 18,7	4,4	3,2
Forest density	Forest density classes	2-4	-	-

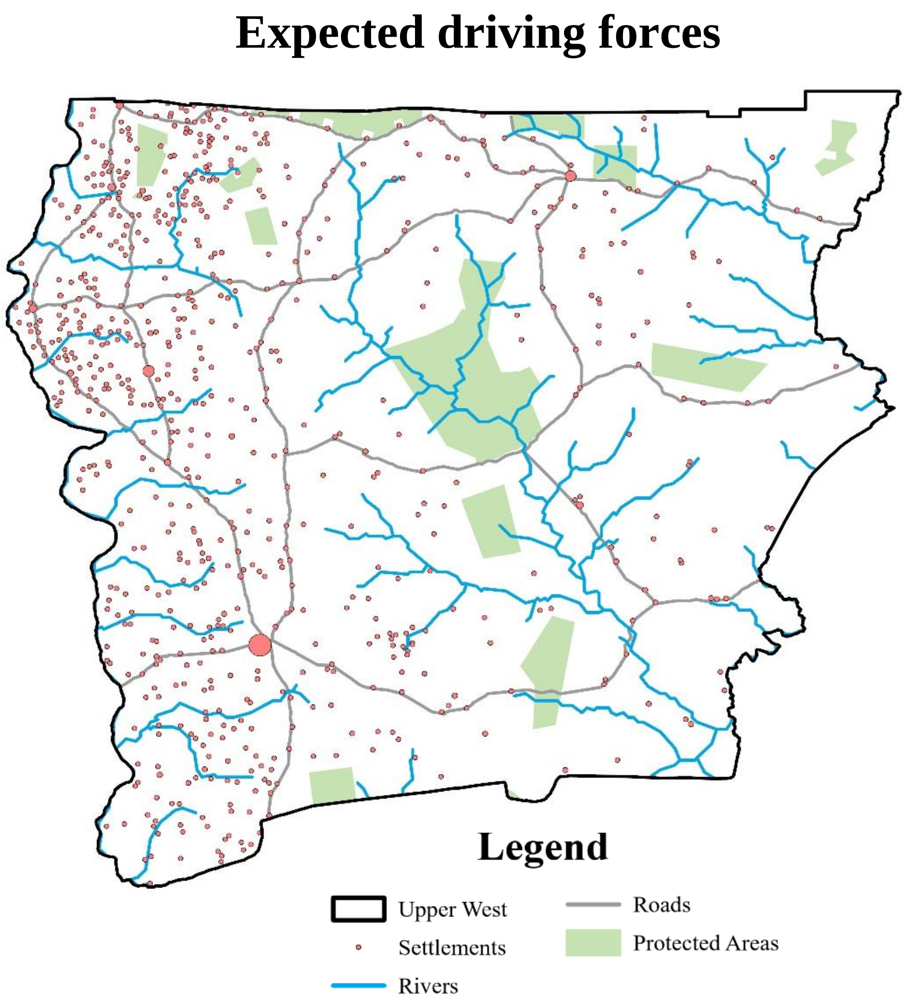

Figure 1: indication of spatial dispersion of (expected) driving forces

2.2.2 Data & Methodology

Elevation and slope

ASTER data was used to generate elevation and slope maps. ASTER is a Japanese instrument on the Terra satellite that among other things produces detailed terrain hight models (NASA, 2023). These detailed terrain hight models, also called global digital elevation models or GDEM, were downloaded from NASA Earthdata. The GDEM images from ASTER have a spatial resolution of 30 meters and are created from the automatic processing of the entire ASTER level 1a archive of scenes between the 1st of march 2000 and 30th of November 2013 (Japan ASTER Science Team, 2023). To cover the entire study area six GDEM images were downloaded and combined into a single raster using ArcGIS' 'mosaic to new raster' tool.

The elevation raster that is used in the regression consists of the raw data from the GDEM raster from ASTER. The slope raster was computed from the GDEM raster using ArcGIS' 'Slope' tool. Both the slope and elevation raster layers were converted to integers to minimize the necessary computing power for the regression.

Distance to nearest road

To calculate the distance to the nearest road for each pixel the Ghana Road shapefile from the world bank data catalogue was used. The shapefile contains the roads in vector format and also contains the surface type and condition for each road (The World Bank, 2023). The dataset has been updated last the 25th of July 2017.

To generate a raster that contains the distance from each cell to the nearest road and river/stream the 'Euclidian distance' tool in ArcGIS was used. This tool take a vector data layer as input and creates a raster with values for the distance to the closest vector feature for each cell. The deforestation map from paragraph 3.1.1 was used to define the cell size and extent of the images to make sure the whole study area is covered.

Forest density and distance to forest edge

The forest cover maps from 2014 and 2022, previously discussed in paragraph 2.1, were to determine the forest density and the distance to forest edge. The raw data from the forest cover maps will be used for the forest density. This raster consists of three classes (1-3): 'dense forest', 'semi-dense forest' and 'sparse forest'. It only contains three classes instead of four because only forested areas will be included in the regression analysis.

The forest edge in Savannah forests is not very clear. A single 'no forest' cell of 30x30m does not indicate the edge of a forest zone but is often surrounded by forested cells. For this reason three 'forest edge' rasters were created. To do this the forest cover maps were firstly converted to polygons using ArcGIS' 'raster to polygon' tool. These polygons were then filtered to only contain areas in the 'no forest' class. After that these polygons were filtered into three layers with all a different minimum area: 1 cell (900 m²), 1 ha and 1 km². Finally, using the 'Euclidian distance' tool in ArcGIS three distance rasters were computed with the same parameters as before. Which of the three rasters explains deforestation will be tested in the regression.

Distance to nearest river

To calculate the distance to the nearest stream the Rivers of Africa shapefile from the Food and Agriculture Organization of the United Nations was used. The shapefile contains Africa's rivers and streams including some other information about each stream/river (FAO, 2023). The dataset has last been updated on the 13th of May 2022. This dataset was then converted to a distance raster using the 'Euclidian distance' tool in ArcGIS with the same parameters as before.

Population density and distance to nearest settlement

Global Human Settlement Layer data will be used for the population density and the distance to the nearest settlement and building, more specifically the population grid from the years 2015 and 2020. The grids get updated every 5 years which means that the grids do not match the exact years from the deforestation maps. The grid from 2015 will be used with the 2014 deforestation map and the grid from 2020 will be used with the 2022 deforestation map. The population grids were downloaded from the joint research centre data catalogue of the European commission and depict the distribution of human population, expressed as the number of people per cell (Joint Research Centre, 2023).

The population density was then computed from the population grid using the 'Focal statistics' tool in ArcGIS. Using this tool the sum of all people in the surrounding 1 km^2, 5 km^2 and 10 km^2 was computed for each cell of the deforestation raster. Which of these different areas, 1 km^2, 5 km^2 or 10 km^2 explains deforestation best will be tested in the regression.

For the distance to the nearest settlement the population grid was firstly converted to polygons using ArcGIS' 'raster to polygon' tool. This polygon layer was then filtered to only contain polygons that have more than zero people living in them. The polygons in that layer that are within 100 meter of each other were then aggregated into one polygon using the 'Aggregate polygons' tool. This layer shows all urban areas including the amount of people living in each urban area. This layer was then filtered into four different layers containing the settlements with a minimum of 100, 500, 1000 and 5000 inhabitants. Finally, using the 'Euclidian distance' tool in ArcGIS four distance rasters were computed with the same parameters as before. Which of the four rasters explains deforestation will be tested in the regression.

Protected areas

The data on protected areas in Upper West has been retrieved from the Food and Agriculture Organization of the United Nations. The World Database on Protected Areas from 2020 was used. This dataset contains all protected areas in the world containing the year the areas started being in a protected status. All protected areas in Upper West have been protected for multiple decennia, which means the same dataset can be used for both 2014 and 2022. The polygon dataset was then converted to a raster containing values of 0's and 1's. A 1 corresponds to a cell that is in a protected area and a 0 corresponds to a cell that is unprotected.

Distance to the nearest building

For the distance to the nearest building the GHS built-up surface layer was downloaded from the joint research centre data catalogue of the European commission. The GHS built-up surface layer depicts the distribution of the built up surface (Joint Research Centre, 2023). It shows this as the fraction of a cell that is covered by buildings, with buildings being 'any roofed structure erected above ground for any use' (Joint Research Centre, 2023).

The same methodology of the distance to nearest settlement will be used to compute the distance to the nearest building. The polygon layer the built-up surface layer however will be filtered to only contain polygons with a higher value than zero, which means it contains all cells that are built-up.

For the variables distance to nearest building, settlement, river, forest edge and road the logarithm of the values has also been computed. Distance often decays non-linearly and the distance decay function often shows more resemblance to the logarithmic function (Hammond & Youngs, 2011; Vries, Nijkamp, & Rietveld, 2004). For this reason both the linear and logarithmic effect of all variables related to distance will be tested in the regression to determine what variable explains deforestation best.

2.3 REGRESSION

For the regression all variables were compiled into a .csv file. To do this the deforestation raster was firstly converted into points, each point representing a raster cell. Then all the values from the rasters of the variables are joined to the points using ArcGIS' 'Extract multi values to points tool'. Finally the attribute table of the points was exported to a .csv file.

A logistic regression will be used to explain deforestation. A logistic regression is a regression that is often used for classification and predictive analysis (IBM, 2023). The outcome of a logistic regression is a probability, in the form of a number between 0 and 1. In this case 1 corresponds to deforestation and 0 to no deforestation. In a logistic regression, the probability of deforestation is considered to be a function of the explanatory variables and is defined by the logistic function (Kucsicsa & Dumitrica, 2019):

$$p = E(Y) = \frac{exp^{(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots \beta_n X_n)}}{(1 + exp^{(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots \beta_n X_n)})}$$

In this function p is the probability; E(Y) is the expected value of the dependent variable; β_0 is the constant to be estimated by the regression; β_1, β_2 ... β_n are the coefficients to be estimated for each independent variable (X_1, X_2 ... X_n) (Kucsicsa & Dumitrica, 2019). The logistic regression from the python module statsmodels will be used. Statsmodels has the ability to show a summary of the regression, containing metrics like the R-squared, confidence interval and p values. This is why it is preferred in this study over other popular python modules like sci-kit learn's logistic regression.

As previously discussed, the different options for the variables 'population density', 'distance to nearest settlement' and 'distance to forest edge' have to be tested to prevent multicollinearity in the regression. Each option of these variables will be used in a separate regression with the dependent variable deforestation. The option that best explains the deforestation, indicated by the highest R-squared, will be used in the final regression.

The final regression will be computed and the effects of all variables will be reviewed. To get an easier comparison of the relative effect between variables another regression will be developed using the standardized version of all explanatory variables (Bavaghar, 2015; Etter, McAlpine, Wilson, Phinn, & Possingham, 2006)

Using the logistic function and the coefficients obtained from the regression the probabilities for all raster cells will be calculated using the 'raster calculator' tool in ArcGIS. For this the raster data from 2022 will be used to finally create a probability map for 2022. This probability map will show the probability of deforestation in 2022 based on the spatial patterns of the 2014-2022 period.

3 RESULTS

3.1 FOREST COVER CHANGE

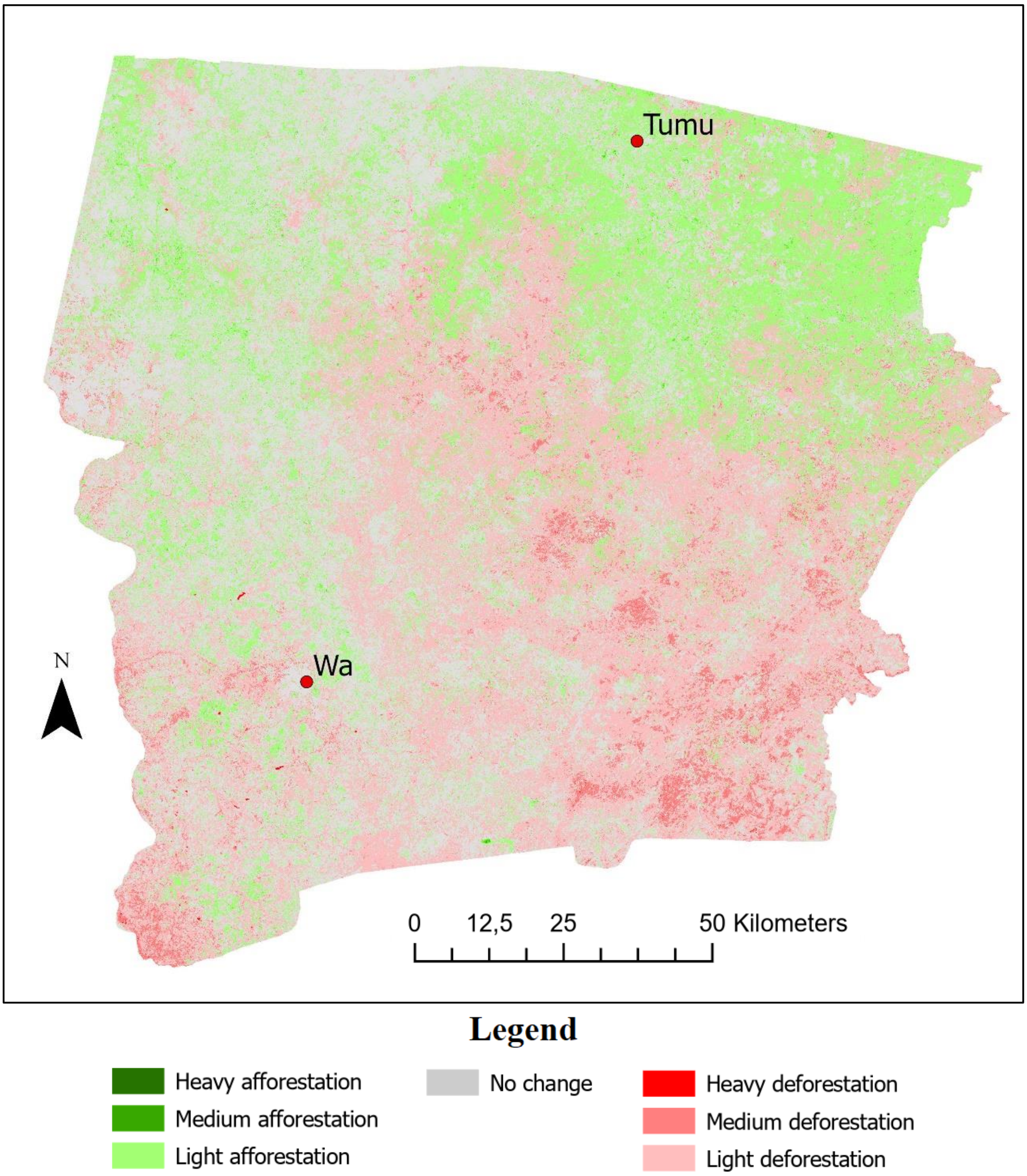

Figure 2: Forest cover change in Upper West between 2014 and 2022.

The forest cover change map between 2014 and 2022 has been computed from the forest cover and is shown in Figure 2**.** The north of the Upper West Region shows mostly light afforestation. On the other hand a big part of the southeastern part of Upper West has experienced deforestation. Most of this area experienced light deforestation, some medium deforestation is however concentrated in the far southeastern corner and in the middle of Upper West.

The exact amount of areas that experienced afforestation or deforestation are shown in Table 3. 38,9% or 6890,5 km² of Upper West did not experience change, which means that 61,1% did experience change. Most of these areas experienced deforestation, 6570 km² or 37,1% in total. Heavy deforestation hardly happened in Upper West, while 33,7% or 5967 km² of Upper West experienced light deforestation and 3,4% or 597 km² experienced medium deforestation. A total of 23,9% of Upper West experienced afforestation. This was mostly light afforestation, as hardly any areas experienced medium and heavy afforestation with a total of 0,52% or 90 km². A total of 4142 km² or 23,4% of Upper West experienced light afforestation. Both afforestation and deforestation happened abundantly in Upper West between 2014 and 2022, however deforestation was more common than afforestation.

Table 3: Absolute and relative forest cover changes in Upper West

Forest cover change type	Area (km²)	Percentage (%)
Heavy afforestation	0,5	0,002
Medium afforestation	89,5	0,5
Light afforestation	4141,8	23,4
No change	6890,5	38,9
Light deforestation	5967,4	33,7
Medium deforestation	596,9	3,4
Heavy deforestation	5,1	0,02

3.2 REGRESSION

The logistic regressions on the ‘population density’ variables concluded that the population within the surrounding 10 km² of a cell explained deforestation best. The regressions of the variables showed an R^2 of 0.003, 0.007 and 0.010 for 1 km², 5 km² and 10 km² respectively. The logistic regressions on the ‘distance to nearest settlement’ variables concluded that the distance to a settlement with a minimum of a 100 inhabitants best explained deforestation. The regressions of the variables showed an R^2 of 0.046, 0.004, 0.005 and 0.0017 for a minimum of 100, 500, 1000 and 5000 inhabitants respectively. Finally, the logistic regressions on ‘the distance to forest edge’ variables concluded that the distance to a forest edge, where a forest edge is defined as an area of at least 1 km², best explains deforestation. The different thresholds for forest edge showed an R^2 of 0.061, 0.083 and 0.526 for 1 cell, 1 ha and 1 km² respectively. This means that only the population within the surrounding 10 km² of a cell, the distance to the nearest settlement with a minimum of 100 inhabitants and the distance to the nearest forest edge, being an unforested area of at least 1 km², are included in the final regression. None of the variables related to distance showed a significant improvement in R^2 as a logarithmic correlation compared to the linear correlation. Which means the linear variant of the variables related to distance have been included in the final regression. This final regression is shown in Table 4.

The coefficient in Table 4 indicates how much the probability of deforestation changes when the corresponding explanatory variable increases one unit. This value is the one that is used in the probability map in the next chapter. All of these coefficients and all of the standardized coefficients are significant. The standardized coefficient shows the relevant effect of each of the explanatory variables. Forest density has the biggest effect on deforestation (standardized β = -2.249) and an increase in the forest density causes an increase in deforestation probability. This correlation seems positive, but a lower forest cover class indicates a higher forest density, which makes it a negative correlation. Elevation has the second largest relevant effect (standardized β = -0.247). The effect is negative indicating that a higher elevation negatively impacts deforestation probability. Distance to forest edge (standardized β = -0.176) and population density (standardized β = 0.178) have a similar effect on deforestation probability. However, distance to forest edge has a negative effect on deforestation probability, indicating that cells that are further away from a forest edge have lower chances of deforestation. While population density has a positive effect on deforestation, indicating that a higher population density in the surrounding 10 km² of a cell leads to a higher deforestation probability. The distance to the nearest road (standardized β = 0.097) and the distance to the nearest river (standardized β = -0.016) both have a positive effect on deforestation probability, indicating that a cell that is further from a road/river has a higher chance of deforestation. However, the effect of the distance to nearest road is around 6 times bigger than the effect of the distance to the nearest river, which has the smallest relative effect of the explanatory variables. Distance to the nearest building (standardized β = -0.085) and the distance to the nearest settlement (standardized β = -0.027) both have negative effects on deforestation, indicating that cells that are further away from a settlement/building have lower chances of deforestation. Slope (standardized β = -0.051) has a negative effect on deforestation probability, indicating that deforestation is less likely to occur on steeper slopes. Finally, protected areas (standardized β = 0.063) have a positive effect on deforestation, indicating that a protected area a higher chance of deforestation than a non-protected area.

Table 4: Logistic regression results. S.E. = standard error

Explanatory variables	Unit	Coefficient (β)	S.E.	Standardized Coefficient (β)	S.E.
Constant	-	12.770*	0.010	-0.046*	0.001
Elevation	1 m	-0.006*	3.1e-05	-0.247*	0.001
Slope	1 deg	-0.021*	0.000	-0.051*	0.001
Protected areas	0-1	0.232*	0.004	0.063*	0.001
Distance to forest edge	1 km	-0.014*	7.7e-05	-0.176*	0.001
Population density (10 km²)	1000 people	0.025*	0.000	0.178*	0.001
Distance to nearest settlement	1 km	-0.010*	0.000	-0.027*	0.001
Distance to nearest building	1 km	-0.138*	0.002	-0.085*	0.001
Distance to nearest (major) road	1 km	0.017*	0.000	0.097*	0.001
Distance to nearest river	1 km	0.005*	0.000	0.016*	0.001
Forest density	Forest cover classes	-4.444*	0.002	-2.249*	0.001
				R^2	0.534

[1]S.E. = Standard error [2]* : $P < 0.001$ [3]S.E. for the standardized coefficients are not all the same, S.E. are all below 0.001. Statsmodels' regression table does not show more than 3 decimals.

3.3 PROBABILITY MAP

Figure 3 shows the probability of deforestation for every raster cell in Upper West. The probability is shown as a value between 0 and 1. With values closer to 0 having a lower probability of deforestation and values closer to 1 having a higher probability. The highest probabilities do not show a clear pattern but they are mostly located in the south and east of Upper West. It is remarkable however that these areas are mostly areas that fall into the 'semi-dense forest' or 'dense forest' classes. Furthermore the areas around cities, like Wa and Tumu for example, show medium probabilities. Medium probabilities also appear relatively much in the southeastern corner of Upper West.

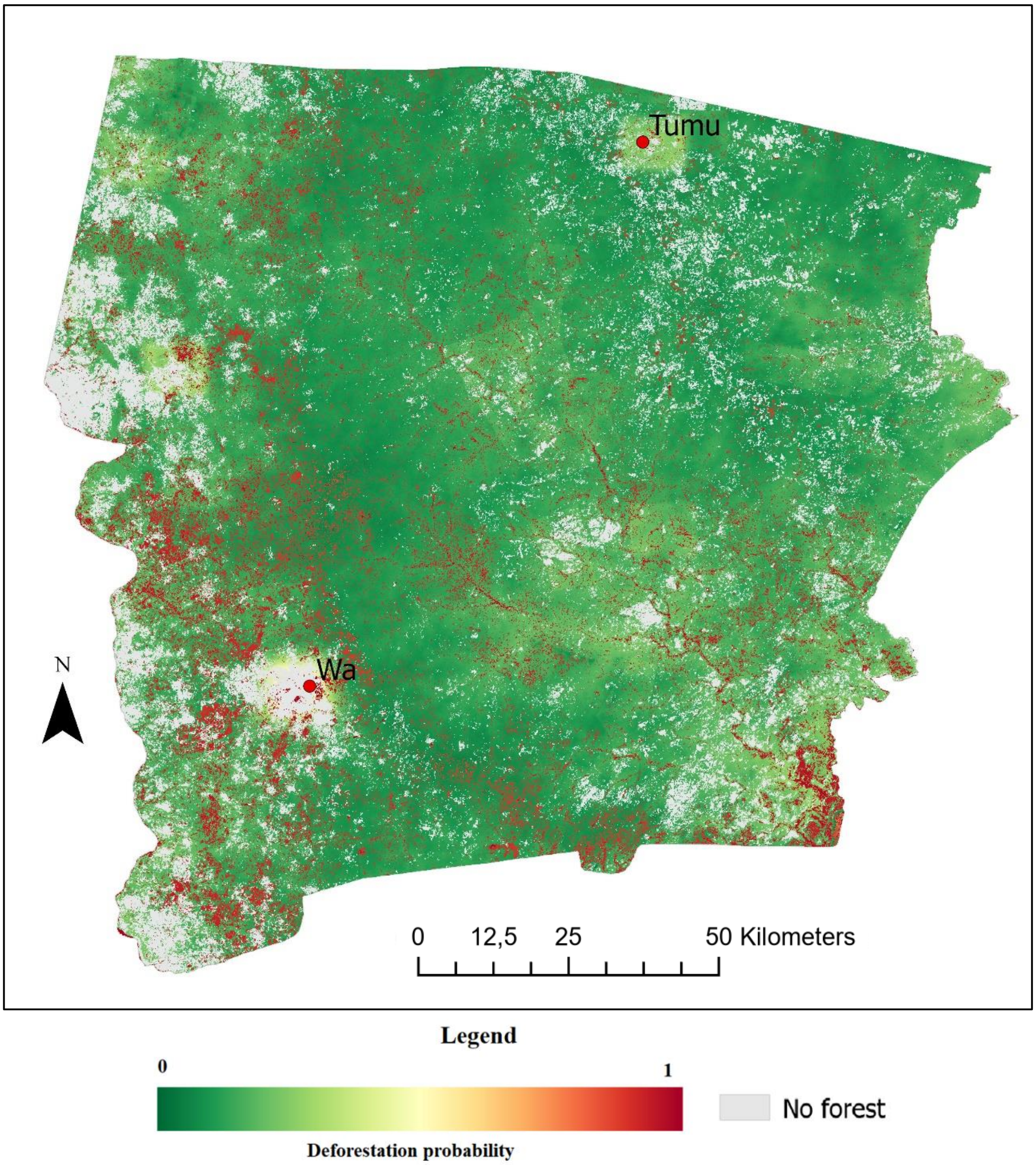

Figure 3: Probabilty of deforestation for Upper West in 2022

4 DISCUSSION

This study has shown that a lot of areas in Upper West have experienced a change of forest cover. The logistic regression has shown the value of producing location-specific logistic regression models that can be used for probability mapping. This is important, as probability mapping can become a very important tool to prevent future deforestation (Lambin, 1994). This is especially the case for Upper West as no similar studies have been previously conducted in this area.

As expected forest density has a big effect on deforestation. Not many studies have previously studied the effect of forest density on deforestation. A study about deforestation in Bangladesh does however show higher deforestation rates in denser forests compared to open forests, confirming the findings in our study (Reddy , Pasha, Jha, Diwakar, & Dadhwal, 2016). The fact that not many studies have studied the effect of forest density on deforestation is likely because many studies have used the definition of a forest as was used by Global Forest Watch (2023). This definition has a higher forest density threshold, which is why Upper West has no forest according to this definition. Our study shows that the difference between 'sparse forest (5-25% forest cover)' and 'semi-dense forest (25-50% forest cover)' leads to a big difference in deforestation probability. These are the two most common forest cover classes in Upper West, and would both not be classified as forest in the definition by Global Forest Watch. A reason why most other studies do not include forest density into their logistic regression could be that the difference in density in forests that are overall denser does not have a significant effect on deforestation.

A higher elevation and a sleeper stope cause lower chances of deforestation. This negative effect of slope is confirmed by a number of other studies (Kucsicsa & Dumitrica, 2019; Bavaghar, 2015; Gayen & Saha, 2017). The effect of elevation is not always negative in other studies. The study by Arekhi for example does show a negative effect of elevation on deforestation (2011). However, a number of other studies also show a positive effect of elevation on deforestation (Kucsicsa & Dumitrica, 2019; Gayen & Saha, 2017). The effect of elevation on deforestation is therefore very location dependent.

All variables that relate to population growth have positive effects on deforestation, meaning that more population leads to more deforestation. These variables are distance to nearest settlement and building and population density. This positive effect was expected as population growth is one of the biggest indirect causes for deforestation in Northern Ghana (Fagariba, Song, & Soule, 2018). Other studies have also shown positive effects of population related variables on deforestation (Arekhi, 2011; Kucsicsa & Dumitrica, 2019; Gayen & Saha, 2017). Distance to rivers and roads have positive effects on deforestation. This is unexpected as a higher distance to a river or road increases travel costs, which makes it less profitable. No real explanation for this positive effect was found, but the effect of these variables is only minimal with both standardized coefficients being lower than 0.100. Distance to forest edge does has a negative effect on deforestation, this was expected as deforestation is not likely to randomly happen in the middle of a forest. This negative effect for distance to forest edge was also found by a number of other studies.

A striking result from this study is that protected areas have a higher chance of deforestation. Protected areas not being effective in Upper West does however confirm the findings of Fagariba et al., who stated that poor government policies are an important indirect cause of deforestation in Northern Ghana. Fagariba et al. also state that government policies regarding protecting forests are most effective in democratic countries with higher levels of corruption control and protection of property right. Ghana is one of the most democratic countries in Africa with a democratic index of 6.43 (Oluwole, 2023), but is still quite far behind western countries like the Netherlands and Germany with democratic indeces of 9.00 and 8.80 respectively (Economist intelligence , 2023). On top of that Ghana also shows signs of private media actively contributing to political corruption (Asomah, 2020). Even though poor government policies partially explain why protected areas in Upper West are uneffective, it is unlikely that this results in deforestation to be more likely in protected areas. It is likely that this is caused by something else that was not included in the regression. For example, the thickness or type of trees could make it more profitable to cut down trees in protected areas. After the final regression a cross-variable between forest density and protected areas was added to better inspect the interaction effect between these two variables, as this could possibly give more insight into the positive effect of protected areas. This interaction effect showed that deforestation probability is especially high for the ‘sparse forests’ class within protected areas (standardized β = 0.759*). Which shows that the positive effect of protected areas was mainly caused by the less dense forests within protected areas. Including this cross-variable the ‘semi-dense forests’ class shows a slightly smaller positive effect (standardized β = 0.050*) when in a protected area compared to the regression without the interaction effect. The ‘dense forests’ class even shows a significant negative effect (standardized β = -0.6604*), although there are hardly any areas in Upper West that fall into the ‘dense forests’ class.

The forest cover change map in paragraph 3.1 showed light afforestation in the north of Upper West. Experts on the study area from Sommalife did not expect afforestation in these areas however. The forest cover change maps were generated using satellite imagery from the years 2014, 2018 and 2022, which has been described in detail in a technical report (van 't Hof, 2023). The forest cover maps of 2014 and 2018 showed a sudden change from ‘no forest’ to ‘sparse forest’ in this four year period. The 2018-2022 period did not show much change in this area, which possibly means there might be some inaccuracies in the 2014 forest cover map. These inaccuracies are described in the technical report, but it is important to note that the afforestation in the north could possibly be caused by inaccuracies here as well.

An R-squared of 0.534 of the regression is fairly high when compared to other studies using logistic regressions to predict deforestation. Bavaghar retrieved an R-squared of 0.383 for example in a logistic regression about deforestation in Iran (2015). Gayen & Saha however retrieved an exceptionally high R-squared of 0.959 in logistic regression about deforestation in the Pathro river basin in India (2017). An R-squared of 0.534 indicates that deforestation in Upper West cannot be fully explained by the explanatory factors used in this study. This is expected as lot of other (not spatially explicit) variables also potentially have effects on deforestation. However, the logistic regression can still be used to make predictions on future deforestation with a satisfactory level of confidence.

5 CONCLUSION & RECOMMENDATIONS

Present deforestation is not only a burning issue in Upper West but also in the rest of the world. This study successfully mapped Upper West's forest cover changes for the past decade and showed that deforestation mostly happened in the southeastern corner of the region. A literature review revealed the forces that were expected to drive the deforestation in the area and the expected effect of these forces. A logistic regression confirmed the expectation that forest density plays a big role in Upper West's deforestation, with denser forests having a significantly higher probability of deforestation. The regression also confirmed the positive effect of population growth on deforestation that was found in an earlier study. Finally the regression also confirmed that poor government policies in protecting forests is an indirect cause of deforestation in Northern Ghana by showing that protected areas have a higher probability of deforestation. A probability map of deforestation shows that areas with denser forests are very likely to experience deforestation in the future if current trends resume. The probability map has been generated from the logistic regression with a satisfactory level confidence and can be used as a tool to prevent future deforestation in the study area.

To further improve the confidence behind this study it is recommended that the future forest cover maps of Upper West will be supported by observations in fieldwork to improve accuracies. This could either confirm or deny the afforestation in the northern areas of the region as the confidence in this region was not high. More accurate forest cover change maps also lead to a more confident regression and a better explanation and prediction of deforestation in the area.

It could be very beneficial to further investigate the explanatory variable protected forests. Protected forests are meant to protect forests from deforestation, but it does not seem to work in Upper West. Research could potentially show what the exact issue is and help policymakers prevent deforestation more effectively.

6 References

Abman, R. (2018). Rule of Law and Avoided Deforestation from Protected Areas. *Ecological Economics*, Article ID: 158875970.

Acquah, H., & Onumah, E. (2011). Farmers' Perception and Adaptation to Climate Change: A Willingness to Pay Analysis. *Economics*, Article ID: 67845850.

Adusei, C., & Dunyah, J. Y. (2016). Forest Fringe Communities Participation in Forest Reserve Sustainability in Ghana. *Open Journal of Forestry*, 94-105.

Arekhi, S. (2011). Modeling spatial pattern of deforestation using GIS and logistic regression: A case study of northern Ilam forests, Ilam province, Iran. *African Journal of Biotechnology Vol. 10(72)*, 16236-16249.

Asomah, J. (2020). Democracy, the public sphere, and power elites: examining the Ghanaian private media's role in political corruption. *Political Science*, Article ID: 221117181.

Bavaghar, M. P. (2015). Deforestation modelling using logistic regression and GIS. *Journal of forest science*, 193-199.

Borges, J., Higginbottom, T. P., Symeonakis, E., & Jones, M. (2020). *Sentinel-1 and Sentinel-2 Data for Savannah Land Cover Mapping: Optimising the Combination of Sensors and Seasons.* Manchester: MDPI.

Dai, F. C., & Lee, C. F. (2002). Landslide characteristics and slope instability modeling using GIS, Lantau Island, Hong Kong. *Geomorphology Vol. 42*, 213-228.

Economist intelligence . (2023). *Democracy Index 2022; Frontline democracy and the battle for Ukraine.* London: The Economist Intelligence Unit Limited .

Eskandari, S., Jaafari, M. R., Oliva, P., Ghorbanzadeh, O., & Blaschke, T. (2020). Mapping Land Cover and Tree Canopy Cover in Zagros Forests of Iran: Application of Sentinel-2, Google Earth, and Field Data. *Remote sensing Vol. 12*, 1912.

Etter, A., McAlpine, C., Wilson, K., Phinn, S., & Possingham, H. (2006). Regional patterns of agricultural land use and. *Agriculture, Ecosystems and Environment*, 369-386.

Etwire, P. M., Al-Hassan, R., & Kowornu, J. (2013). Application of Livelihood Vulnerability Index in Assessing Vulnerability to Climate Change and Variability in Northern Ghana. *Environmental Science*, Article ID: 54508169.

European Comission. (2023). *Quantifying Forest Fragmentation.* Luxembourg: Joint Research Centre European Comission.

Fagariba, C. J., Song, S., & Soule, S. K. (2018). Livelihood Economic Activities Causing Deforestation in Northern Ghana: Evidence of Sissala West District. *Open journal of ecology Vol. 8* , Article ID: 82157.

FAO. (2010). *Global Forest Resources Assessment 2010 main report.* Rome: Food and Agriculture Organization of the United Nations.

FAO. (2011). *State of the world's forests.* Rome: Food and Agriculture Organization of the United Nations.

FAO. (2023, 5 26). *Rivers of Africa*. Retrieved from Aquastat FAO: https://data.apps.fao.org/catalog/dataset/b891ca64-4cd4-4efd-a7ca-b386e98d52e8

Gayen, A., & Saha, S. (2017). Deforestation probable area predicted by logistic regression in Pathro river basin: a tributary of Ajay river. *Spatial information research Vol.26*, 1-9.

Global Forest Watch. (2023, 04 25). *Forest change*. Retrieved from Globalforestwatch.org: https://www.globalforestwatch.org/dashboards/country/GHA/8/?category=forest-change&dashboardPrompts=eyJzaG93UHJvbXB0cyI6dHJ1ZSwicHJvbXB0c1ZpZXdlZCI6W10sInNldHRpbmdzIjp7Im9wZW4iOmZhbHNlLCJzdGVwSW5kZXgiOjAsInN0ZXBzS2V5IjoiIn0sIm9wZW4iOnRydWUsInN0ZXBzS2V5Ijo

Grand View Research. (2022). *Shea Butter Market Size, Share & Trends Analysis Report By Product (Raw & Unrefined, Refined), By Application (Cosmetics & Personal Care, Food), By Region, And Segment Forecasts, 2022 - 2030.* San Francisco.

Gupta, J. (2012). Glocal forest and REDD+ governance: win–win or lose–lose? *Current Opinion in Environmental Sustainability*, 620-627.

Hammond, L., & Youngs, D. (2011). Decay functions and criminal spatial processes: geographical offender profiling of volume crime. *Journal of Investigative Psychology and Offender Profiling*.

IBM. (2023, 6 14). *What is logistic regression?* Retrieved from IBM: https://www.ibm.com/topics/logistic-regression#:~:text=Related%20solutions-,What%20is%20logistic%20regression%3F,given%20dataset%20of%20independent%20variables.

Japan ASTER Science Team. (2023, 5 26). *ASTGTM v003*. Retrieved from USGS.gov: https://lpdaac.usgs.gov/products/astgtmv003/

Joint Research Centre. (2023). *GHSL Data Package 2023.* Luxembourg: Publications Office of the European Union.

Kucsicsa, G., & Dumitrica, C. (2019). Spatial modelling of deforestation in Romanian Carpathian Mountains using GIS and Logistic Regression. *Journal of Mountain Science Vol. 16*, 1005-1022.

Lambin, E. F. (1994). *Modelling deforestation processes.* Luxembourg: European commission.

Lin, M.-L., & Tung, C.-C. (2004). A GIS-based potential analysis of the landslides induced by the Chi-Chi earthquake. *Engineering geology*, 63-77.

Ludeke, A. K., Maggio, R. C., & Reid, L. M. (1990). An analysis of anthropogenic deforestation using logistic regression and GIS. *Journal of Environmental Management*, 247-259.

NASA. (2023, 5 26). *Terra instruments*. Retrieved from terra.nasa.gov: https://terra.nasa.gov/about/terra-instruments/aster

Nielsen, J. O., & Reenberg, A. (2010). Cultural barriers to climate change adaptation: A case study from Northern Burkina Faso. *Global Environmental Change Vol. 20*, 142-152.

Oğuz, H. H. (2020). The Spatial Analysis of The Terrain Impact on Agricultural Land Use by Using Gis And Rs Techniques: A Case Study of Soran District-Erbil-Iraq. *ZANCO Journal of Humanity Sciences*, Article ID: 241581792.

Oluwole, V. (2023, 5 23). Top 10 African nations leading in democracy index rankings. *Business insider Africa*.

Reddy , C., Pasha, S., Jha, C., Diwakar, P., & Dadhwal, V. (2016). Development of national database on long-term deforestation (1930–2014) in Bangladesh. *Global and Planetary Change*, Article ID: 131297381.

Saha, S., Saha, M., Mukherjee, K., Arabameri, A., Ngo, P. T., & Paul, G. C. (2020). Predicting the deforestation probability using the binary logistic regression, random forest, ensemble rotational forest, REPTree: A case study at the Gumani River Basin, India. *Science Of The Total Environment Vol. 730*, Article ID: 139197.

Saha, S., Saha, M., Mukherjee, K., Arabameri, A., Ngo, P. T., & Paul, G. C. (2020). Predicting the deforestation probability using the binary logistic regression, random forest, ensemble rotational forest, REPTree: A case study at the Gumani River Basin, India. *Science of The Total Environment Vol. 730*, Article ID: 139197.

Sommalife. (2023, 04 25). *Impacts*. Retrieved from Sommalife.com: https://sommalife.com/impact/

The World Bank. (2023, 1 19). *Ghana Roads*. Retrieved from Data catalog the world bank: https://datacatalog.worldbank.org/search/dataset/0039327

van 't Hof, M. (2023). *Modelling forest cover Northern Ghana.* Amsterdam: Vrije Universiteit.

Vedeld, P., Angelsen, A., Bojö, J., Sjaastad, E., & Berg, G. K. (2007). Forest environmental incomes and the rural poor. *Forest Policy and Economics*, 869-879.

Vries, J., Nijkamp, P., & Rietveld, P. (2004). Exponential or Power Distance-Decay for Commuting? An Alternative Specification. *Environment and Planning A*, Article ID: 15050089.