2023

Modelling forest cover Northern Ghana

EXTENDED BACHELORTHESIS EARTH SCIENCES, ECONOMICS AND SUSTAINABILITY

MARIJN VAN 'T HOF

2707083

1. Goal

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 its 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, over grazing and wood logging (Fagariba, Song, & Soule, 2018). Indirect causes in are mostly 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 of 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). According to Sommalife this is a threat to the livelihood of many of these communities (2023). According to Global Forest Watch however there has been no deforestation in the Upper West region in Ghana since 2010 (2023). This is due to the fact that this study was mainly focused on more dense forests. Global Forest Watch's study even indicates that there is no forest left in the Upper West Region, likely because the area is very sparsely covered with trees (2023). For many indigenous communities these areas are of huge importance. This study will attempt to map forest cover in the Upper West Region using a different approach, using satellite imagery and machine learning algorithms. Which aims to answer the following question: 'Can satellite data and machine-learning algorithms be used to map the current and historical location of savannah forests in Northern Ghana?'

The results of this report can then be used to identify the spatially explicit drivers of deforestation. This allows NGOs such as Sommalife to protect forests against deforestation in a more targeted way.

2. Methodology

To determine the deforestation in Northern Ghana, or more specifically the deforestation of the Upper West region, 3 satellite images from Landsat 8 were downloaded for the years 2014, 2018 and 2022 from the USGS Earth Explorer with a maximum cloud percentage of 10%. An indication of the location and size of the satellite images can be seen in Figure 1. Landsat has been used often for land cover classifications in Savannah regions and has yielded good results (Boakye, Adjei, Odai, & Annor, 2008; Higginbottom, Symeonakis, Meyer, & van der Linden, 2018). In the report by Higginbottom et al. Landsat data was used to detect woody vegetation in savannah area. Woody vegetation is also of particular importance in this report. After all, many grasses grow in the savannah, which is why this analysis requires a clear distinction between woody vegetation and grasses. To clarify this distinction, the selected satellite images were all from the month of February, which is at the end of the dry season (Lieberman, 1982). It is easiest to distinguish between grasses and woody vegetation during these months, and there is little cloud cover (Boakye, Adjei, Odai, & Annor, 2008; Haro-Carrión & Southworth, 2018)

Figure 1: Location and size of Landsat 8 images

To get an idea of the deforestation, the satellite images will be classified according to forest density. For this, 4 different land cover classes have been established based on tree cover: Dense forests (>50% tree cover), Semi-dense forests (25-50% tree cover), sparse forests (5-25% tree cover), no forests (<5 % tree cover). A similar classification has been used in the report by Eskandari et. al, which also examines tree cover in savannah areas (2020).

In this study a supervised machine leaning algorithm will be used to classify each pixel of the satellite images into one of the four landcover classes. The specific algorithm that will be used is a random forests classifier model. The random forests model is a machine learning classifier that combines decision trees with bootstrapping and aggregation. It has been proven to be highly accurate and more time effective when compared to other machine learning algorithms like Support vector machine (SVM) and maximum likelihood. (Rodriguez-Galiano, Ghimire, Rogan, Chica-Olmo, & Rigol-Sanchez, 2012; Li, Cheng, Chen, Chen, & Liu, 2015). The random forests model has been successfully applied before to African savannahs, specifically in southern Africa (Symeonakis, Petroulaki, & Higginbottom, 2016; Higginbottom, Symeonakis, Meyer, & van der Linden, 2018) and eastern Africa (Ng, et al., 2017).

To train the random forests model there is a need for reference or 'ground-truth' data. Because there is little online reference data about Ghana, this study uses high resolution images from Google Earth Pro to generate the reference data. In a study by Eskandari et al., which was conducted in a comparable area to Northern Ghana of sparse tree cover in the Zagros forest in Iran, Google Earth data has been proven to be reliable as reference data (2020).

Unfortunately, not all areas within our study area are covered with high resolution images. For this reason the areas that are covered with high resolution images were identified within Google Earth Pro for each of the years. Then, for each year 300 random points were selected within these areas using the 'create random points' tool in ArcGIS. These points were all at least 500 meters apart to cover as much of the study area as possible. To get the actual landcover of these points the polygon of the point's pixels were extracted from ArcGIS and imported to Google Earth. For each of these polygons the tree cover was estimated manually. An indication of what each landcover looks like in Google Earth Pro can be seen in Figure 2. The estimated landcover were then added to Excel files for each of the years. Finally, around 100 extra points were manually gathered in the landcover classes that did not have many training points. This was mostly in the classes 'dense forest' and 'no forest'. On top of that some points were gathered in pixels that have a relatively rare landcover, like a river, road, houses or burned areas to make sure these were classified in the right class. For these cases the right class being 'no forest'.

Dense forest (>50% tree cover)

Semi-dense forest (25-50% tree cover)

Sparse forest (5-25% tree cover)

No forest (<5% tree cover)

Figure 2: Landcover classes in Google Earth Pro

Then, bands 2, 3, 4 and 5 were downloaded from the Landsat 8 images. These are the bands of visible and near-infrared light. These bands are most useful at detecting vegetation, as they can be used to calculate the most popular vegetation index, the NDVI. (Auravant, 2023). Of course a lot more vegetation indices exist, but NDVI is most widely used to detect vegetation greenness and has been shown to be related to the tree cover (Xue & Su, 2017)**.** Two variants of the NDVI are the green NDVI and the MSAIV2. The green NDVI is more sensitive to chlorophyll variations and has a higher saturation point, which makes it better at detecting vegetation in more advanced stages (Auravant, 2023). The MSAIV2 index however is better at detecting vegetation in the early stages of development (Auravant, 2023). The NDVI, GNDVI and MSAIV2 were all computed by using the formulas shown below.

$$NDVI = \frac{NIR - RED}{NIR + RED}$$

$$GNDVI = \frac{NIR - GREEN}{NDVI + GREEN}$$

$$MSAIV2 = \frac{2 * NIR + 1 - 1\sqrt{(2 * NIR + 1) - 8(NIR - RED)}}{2}$$

Together with the bands and the reference data these were then added into 3 different excel files, one for each year. These are our training data sets.

To validate the model 20% of the training points in the datasets were left out. The model will estimate the forest cover at these points after it is trained on the remaining 80% of the training points. This estimation can then be used to make a validation of the model by computing the confusion matrix and calculating the accuracy and F1 score, which is an accuracy metric that is good at dealing with datasets that have class imbalance (Korstanje, 2023). After the model was trained using the datasets it became clear that only the NDVI and GNDVI showed an increase in the model's accuracy. For this reason the MSAIV2 was left out of the model.

The three satellite images were then classified using the three separately trained models for each of the years. After these were classified they were imported into ArcGIS to visualize the raw data in a clear way. The results of these are presented in the next chapter.
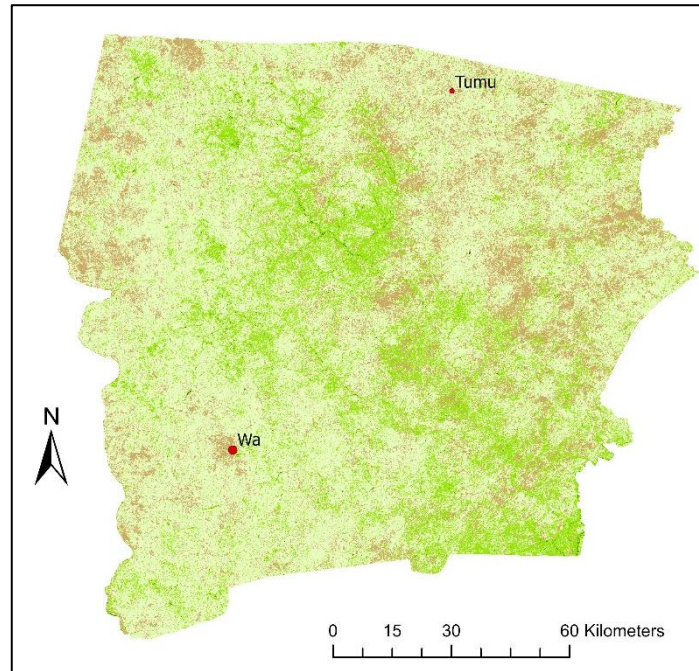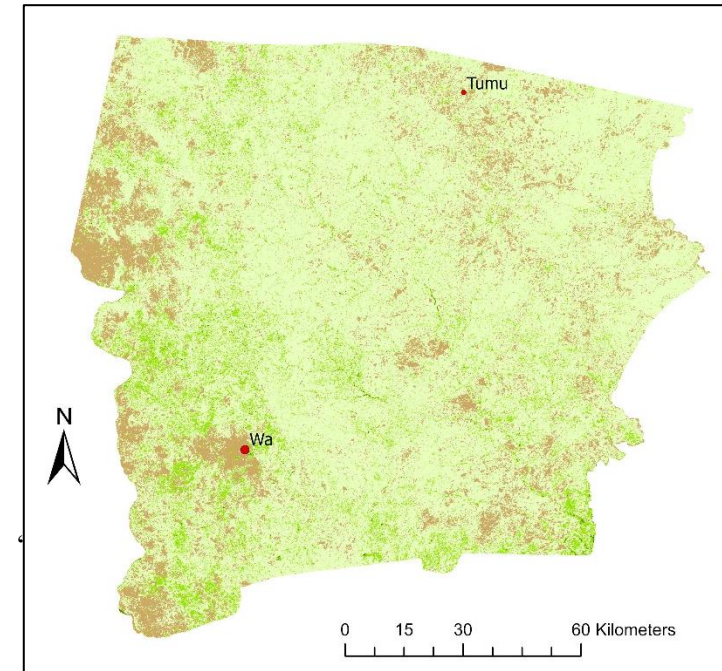
3. Resultaten
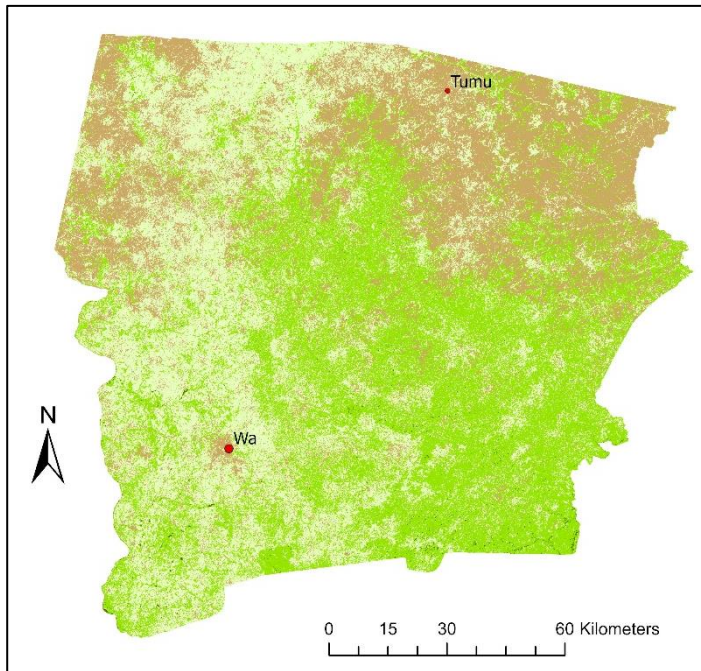
2014   2018   2022

Legend

Dense forests (>50% tree cover)

Semi-dense forests (25-50% tree cover)

Sparse forests (5-25% tree cover)

No forest (<5% tree cover)

Figure 3: Landcover maps 2014, 2018, 2022

In Figure 3 the three classified satellite images are presented. To validate each of the models the accuracy and the f1 score were computed. The accuracy is a number between 0 and 1 that measures number of correctly classified pixels compared to the total number of classified pixels. The F1 score is defined as the harmonic mean of precision (P) and recall (R) and is calculated as follows: $F1 = \frac{2PR}{P+R}$ (Sasaki, 2007).

The F1 score has the advantage of being able to deal with class imbalances (Korstanje, 2023). A simpler accuracy metric like accuracy is not able to do this. The F1 score is able to deal with class imbalances because it takes the type of error into account through the precision and recall. It is expressed in a score between 0 and 1, with a 1 being a perfect classification.

The F1 scores and accuracies are presented in table 1. These accuracies are comparable to similar studies, such as the article by Borges et al. (2020). Borges et al. used random forest models to classify a sentinel 2 image of a savannah region (2020). Their dry season model had an accuracy of 0,773, which is comparable to the accuracy of our models.

Table 1: Accuracies and F1 scores

Year	F1 score	Accuracy
2014	0,725	0,723
2018	0,801	0,795
2022	0,775	0,773

Another way the models were validated was by computing confusion matrices. A confusion matrix visualizes and summarizes the performance of classification algorithms and makes it easier to see where inaccuracies occur. The confusion matrix of 2022's model can be seen in table 2**.** The confusion matrices for the models of 2014 and 2018 are similar to those of 2022 and can be seen in appendix 1 and 2. Each of the landcover classes are represented by the numbers 1 through 4, with descending tree cover. This means dense forest is represented by 1 and no forest is represented by 4. The confusion matric in table 2 shows the absolute number of pixels. Most inaccuracies occur in the middle two classes, semi-dense forest and sparse forests. This was to be expected as these classes are close together. A pixel that has 20% tree cover is very close to the semi-dense forest class (25-50% tree cover), while it should actually be classified in the sparse forest class (5-25% tree cover). Some inaccuracies could be caused by this. The rest of inaccuracies are minor and could be caused by a variety of things. An example could be a tree in a garden of a house. This pixel could be classified as no forest, because of the highly reflecting roof of the house. But it could also be classified in the sparse- or semi-dense forest class due to the higher value for vegetation indices caused by the tree. The user's and producer's accuracies are also shown in table 2**.** The user's accuracy shows how often the classified landcover pixel is actually landcover on the ground. While the producer's accuracy shows how often the landcover on the ground is actually classified correctly. Both of these again show that the biggest inaccuracies are in the middle two landcover classes.

Table 2: Confusion matrix 2022 model

Reference data

Classified data	1	2	3	4	UA
1	8	1	0	0	89%
2	1	19	5	0	76%
3	0	5	10	2	59%
4	0	0	3	21	88%
PA	89%	76%	56%	91%	

UA = User's accuracy, PA = Producer's accuracy.
1 = dense forest, 2 = semi-dense forest,
3 = sparse forest, 4 = no forest

4. Discussion

In this study a machine learning algorithm was used to classify Landsat 8 images, with a spatial resolution of 30m, based on tree cover. This approach produced models with sufficient accuracies but there are other approaches. Another way to detect tree cover in an area which is sparsely covered with trees could be to identify each tree separately. This is possible by performing an object based classification model. Satellite imagery with higher accuracy is however necessary for this. Most trees in the area have a crown diameter of between 8 and 10 meter, which means separate trees are not visible on Landsat imagery (Kelly, Poudyal, & Bouvet, 2019). Another freely available satellite imagery source is sentinel-2. This imagery has a spatial resolution of 10m, which means it could possibly be used to detect singular trees. Nduji et al. attempted an object based approach using sentinel-2 imagery this in their study, however results showed a detection accuracy of only 68% which is not an improvement to our model (2023). Nduji et al. also attempted Worldview-3 imagery, which has a spatial resolution of 2m, using the same approach of object based detection (2023). Worldview-3 imagery was able to detect trees with a 99% detection accuracy (Nduji, Tolpekin, & Stein, 2023). However, Worldview-3 is not freely available and an order through ESA earth online, the platform Worldview-3 data can be downloaded from, will take around 4-6 weeks. Unfortunately time was a limiting factor in this study, which means the choice for a pixel based approach using Landsat 8 imagery was made.

In the training dataset used for our model burnt areas were classified as 'no forest'. Most of these areas do have some burnt trees still standing however. It is hard to predict whether these trees have survived the fire or not in the Google Earth Pro imagery. Fires in Ghana are mainly used by farmers for numerous reasons. Most of which are related to clearing land for future farms or managing shade on existing farms (Amissah, 2009). Because of this burnt areas were classified as 'no forest'. However for some purposes to use fire, the main goal is not to destroy all trees in an area. Examples of this include fires used for hunting and extracting honey or using fire as a religious and ceremonial tradition (Amissah, 2009). These fires are smaller in size and likely do not kill all trees, while burning most of the grasses. Given that some of the trees survive the fire, our model might misclassify these areas as 'no forest'. These areas are small however and likely do not have a significant influence on the results.

Furthermore some inaccuracies could be caused by the lack of better reference data. Google Earth Pro is built up from a combination of high resolution satellite images and aerial photographs. This means that the reference data was not generated from the same image. The images used to generate reference data can be a couple of months before or after the date of the to be classified Landsat image. This for example means that dry season and wet season images were both used to generate reference data. This could cause inaccuracies as wet season images are much greener which makes it harder to distinguish between woody vegetation and grasses (Boakye, Adjei, Odai, & Annor, 2008). However because of limited reference data for countries like Ghana, Google Earth Pro remains the best option. Inaccuracies were minimize by selecting only reference images from Google Earth Pro that were a maximum of six months from the date of the Landsat image.

Unfortunately, not many landcover maps of Northern Ghana regarding tree cover are available to verify the maps from our model. The only study that created maps similar to our maps is the study by Basommi et al. (2015). Basommi et al. created a landcover map for the years 1991, 2000 and 2014 for the Wa East district (2015). This district is in the southeastern part of the Upper West region, which is the region our maps cover. The 2014 map from the article by Basommi et al. can thus be used to verify

our map. This map confirms the main patterns that are visible in our map. However it is difficult to compare due to the different spatial resolution and landcover classes. This means the best verification comes from expert knowledge.

References

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.

Amissah, L. (2009). Indigenous fire management practices in Ghana. *IUFRO World Serives*, 131-135.

Auravant. (2023, 04 25). *Vegetation indices and their interpretation: NDVI, GNDVI, MSAVI2, NDRE, and NDWI*. Retrieved from auravant.com: https://www.auravant.com/en/articles/precision-agriculture/vegetation-indices-and-their-interpretation-ndvi-gndvi-msavi2-ndre-and-ndwi/

Basommi, P. L., Guan, Q., & Cheng, D. (2015). Exploring Land use and Land cover change in themining areas of Wa East District, Ghana usingSatellite Imagery. *Open Geoscience*, 618-626.

Bennett, L. (2017). *Deforestation and Climate .* Washington: the Climate Institute.

Boakye, E., Adjei, K. A., Odai, N. S., & Annor, F. (2008). Landsat images for assessment of the impact of land use and land cover changes on the Barekese catchment in Ghana. *European Journal of Scientific Research Vol.21 No.4*, 617-626.

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.

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.

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

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.

Global Forest Watch. (2023, 04 25). *Forest change*. Retrieved from Globalforestwatch.org: https://www.globalforestwatch.org/dashboards/country/GHA/8/?category=forest-change&dashboardPrompts=eyJzaG93UHJvbXB0cyI6dHJ1ZSwicHJvbXB0c1ZpZWdlZCI6W10sInNldHRpbmdzIjp7Im9wZW4iOmZhbHNlLCJzdGVwSW5kZXgiOjAsInN0ZXBzS2V5IjoiIn0sIm9wZW4iOnRydWUsInN0ZXBzS2V5Ijo

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.

Haro-Carrión, X., & Southworth, J. (2018). Understanding Land Cover Change in a Fragmented Forest Landscape in a Biodiversity Hotspot of Coastal Ecuador. *Remote sensing 10(12)*.

Higginbottom, T. P., Symeonakis, E., Meyer, H., & van der Linden, S. (2018). Mapping fractional woody cover in semi-arid savannahs using multi-seasonal composites from Landsat data. *ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 139*, 88-102.

IPBES. (2019). *Summary for policymakers of the global assessment report on biodiversity and ecosystem services.* Bonn.

Kelly, B. A., Poudyal, M., & Bouvet, J.-M. (2019). Impact of land use and land use history on fruits production of Vitellariaparadoxa (Shea tree) according to agroclimatic zones in Mali (West Africa). *Current Botany*, 1-7.

Korstanje, J. (2023, 5 18). *The F1 score*. Retrieved from towardsdatascience.com: https://towardsdatascience.com/the-f1-score-bec2bbc38aa6

Kumar, S. a. (2020). Deforestation: Facts, Cause, Effects, And Control . *Journal of Indian Association for Environmental Management (JIAEM) Vol 40, No 1* , 1-5.

Li, X., Cheng, X., Chen, W., Chen, G., & Liu, S. (2015). Identification of Forested Landslides Using LiDar Data, Object-based Image Analysis, and Machine Learning Algorithms. *Remote Sensing, 7(8)*, 9705–9726.

Lieberman, D. (1982). Seasonality and Phenology in a Dry Tropical Forest in Ghana. *Journal of Ecology, Vol. 70, No. 3*, 791-806.

Nduji, N. N., Tolpekin, V. A., & Stein, A. (2023). An object-based image analysis approach for comparing tree detection from satellite imagery at different scales; A case study in Sukumba Mali. *Remote Sensing Applications: Society and Environment, Vol. 30*.

Ng, W.-T., Rima, P., Einzmann, K., Immitzer, M., Atzberger, C., & Eckert, S. (2017). Assessing the Potential of Sentinel-2 and Pléiades Data for the Detection of Prosopis and Vachellia spp. in Kenya. *Remote sensing Vol. 9*, 74.

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.

Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P. (2012). Rodriguez-Galiano, V. F. ; Ghimire, B. ; Rogan, J. ; Chica-Olmo, M. ; Rigol-Sanchez, J. P. *ISPRS Journal of Photogrammetry and Remote Sensing Vol. 67*, 93-104.

Sasaki, Y. (2007). *The truth of the F-measure.* Manchester: University of Manchester.

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

Symeonakis, E., Petroulaki, K., & Higginbottom, T. (2016). Landsat-based woody vegetation cover monitoring in Southern African savannahs. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences Vol. XLI-B7*, 563-567.

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.

Xue, J., & Su, B. (2017). Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications. *Journal of Sensors, vol. 2017*, 17.

Appendix 1: Confusion matrix 2014

Reference data

Classified data	1	2	3	4	UA
1	2	1	2	0	40%
2	1	17	4	0	77%
3	0	11	24	1	65%
4	0	1	2	17	85%
PA	67%	57%	75%	94%	

Appendix 2: Confusion matrix 2018

Reference data

Classified data	1	2	3	4	UA
1	18	2	0	0	90%
2	1	7	2	2	58%
3	0	4	20	1	80%
4	0	2	2	17	81%
PA	95%	47%	83%	85%	

Appendix 3: Random Forests classificatie model 2022

```python
#importing the necessary modules
import sklearn
import pandas as pd
import numpy as np
import rasterio


#loading in the training and testing data
df_rf = pd.read_csv('D:\Thesis\Reference_data\\train_test_data_2022_manual.csv', sep = ';')
df_rf = df_rf.dropna(how='all')


#replacing comma's in the data for points
df_rf['NDVI'] = df_rf['NDVI'].apply(lambda x: float(x.replace(',', '.')))
df_rf['Green_NDVI'] = df_rf['Green_NDVI'].apply(lambda x: float(x.replace(',', '.')))


#selecting the relevant data and splitting it into train and test data
from sklearn.model_selection import train_test_split
X_2022 = df_rf.iloc[:, 1:-3].values
y_2022 = df_rf.iloc[:, -1].values
X_train, X_test, y_train, y_test = train_test_split(X_2022, y_2022, test_size=0.2,
random_state=42)


#importing the random forest classifier and creating an instance of it
from sklearn.ensemble import RandomForestClassifier
rf = RandomForestClassifier(n_estimators=100, random_state=42, max_depth=7)


#training the random forest classifier using the training data
rf.fit(X_train, y_train)


#using the model to predict the landcover value for the testing data
y_pred2022 = rf.predict(X_test)



#creating a confusion matrix and calculating the accuracy and f1 score for 2022
from sklearn.metrics import confusion_matrix
cf_2022 = confusion_matrix(y_test, y_pred2022)

from sklearn.metrics import accuracy_score
accuracy_2022 = accuracy_score(y_test, y_pred2022)

from sklearn.metrics import f1_score
f1_score2022 = f1_score(y_test, y_pred2014, average='weighted')
```

```
print('Accuracy:', accuracy_2022)
print(cf_2022)
print(f1_score_2022)
```

#opening satellite image(.tiff) using rasterio
```
With rasterio.open('D:\Thesis\Reference_data\\satimage2022.tif') as src:
    #reading the satellite image as a numpy array
    data_sat2022 = src.read()
    #converting the numpy array to a dataframe
    df_sat2022 = pd.DataFrame(data_sat2022.reshape(data_sat2022.shape[0], -1).T,
columns=[f'band{i+1}' for i in range(data_sat2022.shape[0])])
```

#renaming the columns to correspond with the columns of the random forest classifier
```
df_sat2022 = df_sat2022.rename(columns={'band1':'B2_blue', 'band2':'B3_green',
'band3':'B4_red', 'band4':'B5_NIR', 'band5':'NDVI', 'band6':'Green_NDVI'})
```

#predicting the landcover of the satellite image using the random forest classifier
```
sat2022_landcover = rf.predict(df_sat2022)
```

#reshaping sat2022_landcover to the original shape of the satellite image
```
sat2022_landcover_reshaped = np.array(sat2022_landcover).reshape(src.shape)
```

#setting the extent of the image
```
left, bottom, right, top = src.bounds
```

#setting the resolution of the image
```
pixel_size_x, pixel_size_y = src.res
```

#exporting the landcover image using rasterio
```
from rasterio.transform import from_origin
output_file = 'D:\Thesis\Model_results\\2022_landcover.tif'
height, width = sat2022_landcover_reshaped.shape
count = 1
dtype = sat2022_landcover_reshaped.dtype
crs = 'EPSG:4326'
transform = from_origin(left, top, pixel_size_x, pixel_size_y)

with rasterio.open(output_file, 'w', driver='Gtiff', height=height, width=width, count=count,
dtype=dtype, crs=crs, transform=transform) as dst:
    dst.write(sat2022_landcover_reshaped, 1)
```