Factors Influencing Above Ground Biomass of Forest Plots in the Congo

https://github.com/nikki-egna/ENV872_Final_Project_Congo_Carbon

Nikki Egna

Contents

List of Figures

1 Rationale and Research Questions

Above ground biomass (AGB) is all biomass above the soil, including woody and herbaceous vegetation, stems, stumps, branches, bark, and foliage. Above ground biomass can be a predictor of overall health of a forest ecosystem. It is also important for looking at carbon storage potential of forests, which has significant implications for climate change mitigation. (Ducanson et al., 2019) Researchers in forestry, ecology, and agriculture often adopt the "plot method" for studying AGB by estimating and monitoring production or changes in carbon stocks. (Ducanson et al., 2019) The study conducted by the Poulsen Tropical Ecology Lab utilizes this "plot method" to estimate AGB from plots in the Republic of Congo, Africa. The Poulsen Lab estimated AGB using methodology from Chave et al. (2014). This study utilizes these AGB estimates to look at various environmental factors (covariates), and how they might contribute to differences in AGB at each forest plot. The environmental factors included in this study are Human Footprint Index (HFI), GlobCover, annual precipitation, soil type, distance from the nearest road, distance from the nearest village, distance from the nearest river, distance from the nearest saw mill, and distance from the nearest protected area. This study also quantifies the change in AGB between the measurement collection years of 2005, 2009, and 2013.

Question 1:

How has above ground biomass changed over time at the plot locations?

Question 2:

Which environmental factors are signicant predictors of above ground biomass?

2 Dataset Information

This repository uses data from the Poulsen Tropical Ecology Lab, and contains information on tree diameter and species from plots in the Congo. This data is being collected in order to model the amount of above ground biomass (AGB) within each plot across the study sites. Utilizing remote sensing techniques, the Poulsen Lab performed analysis to determine several covariate values at each plot, including distance to nearest road, distance to nearest village, distance to nearest river, distance to nearest protected area, Human Footprint Index, and Globcover value. This project will look at which covariates have a significant impact on AGB at the study sites.

The CongoCarbon_Raw_Data.csv was collected by the Poulsen Tropical Ecology Lab, and their on-ground team based in the Congo. This data is not public facing, and access was granted through Dr. John Poulsen. The CongoCarbon_AGB_by_Plot.csv was created by John Poulsen, Anna Nordseth, and Nikki Egna based on the raw data. The covariate data (CongoCarbon_Plot_Covariates.csv) was created by Nikki Egna in March of 2020, and the data it utilizes was obtained from the following sources: Human Footprint Index was derived from Wildlife Conservation Society (WCS), Center for International Earth Science Information Network (CIESIN), and Columbia University and was accessed in December of 2019. The GlobCover raster was downloaded from the ESA Globcover 2005 Project, led by MEDIAS-France/POSTEL, in December of 2019. Precipitation data was obtained from the Climate Hazards group Infrared Precipitation with Stations (CHIRPS) in December of 2019. The roads, river, soil, saw mill, and villages data comes from the Ministère de l'Économie Forestière CNIAF_MEFDD (https://cog.forest-atlas.org), accessed in April, 2020.

2.1 Metadata

2.1.1 CongoCarbon_Raw_data.csv

Column Name	Description
Tree ID	Identification number assigned to the unique tree (Numeric)
Tag No	Secondary tree identification number (Numeric)
Plot	Study plot number (Numeric)
Subplot	Subplot number within each plot (Numeric)
Family	Tree family name (Character)
Species	Tree species name (Character)
WD	Wood density (Numeric)
DBH0_05	First diameter measurement in 2005 (Numeric; Centimeters)
DBH1_05	Second diameter measurement in 2005 (Numeric; Centimeters)
DBH2_05	Third diameter measurement in 2005 (Numeric; Centimeters)
DBH3_05	Fourth diameter measurement in 2005 (Numeric; Centimeters)
DBH4_05	Fifth diameter measurement in 2005 (Numeric; Centimeters)
DBH0_09	First diameter measurement in 2009 (Numeric; Centimeters)

Column Name	Description
DBH1_09	Second diameter measurement in 2009 (Numeric; Centimeters)
DBH2_09	Third diameter measurement in 2009 (Numeric; Centimeters)
DBH3_09	Fourth diameter measurement in 2009 (Numeric; Centimeters)
DBH4_09	Fifth diameter measurement in 2009 (Numeric; Centimeters)
DBH0_13	First diameter measurement in 2013 (Numeric; Centimeters)
DBH1_13	Second diameter measurement in 2013 (Numeric; Centimeters)
DBH2_13	Third diameter measurement in 2013 (Numeric; Centimeters)
DBH3_13	Fourth diameter measurement in 2013 (Numeric; Centimeters)
DBH4_13	Fifth diameter measurement in 2013 (Numeric; Centimeters)
AGB05.MgE	AGB estimate 2005 (Numeric; milligrams)
AGB09.MgE	AGB estimate 2009 (Numeric; milligrams)
AGB13.MgE	AGB estimate 2013 (Numeric; milligrams)

2.1.2 CongoCarbon_Plot_Covariates.csv

Column Name	Description
Plot	Tree plot ID number (Factor)
Latitude	Latitude of the plot (Numeric)
Longitude	Longitude of the plot (Numeric)
HFI	Human Footprint Index value at the plot location (Factor)
GlobCover	GlobCover vegetation index value at the plot location (Factor)
Precip_sum_2013	Annual precipitation sum for 2013 at plot location (Numeric; millimeters)
Soil	Soil type index at plot location (Factor)
Dist_Road_m	Distance from plot to the nearest road (Numeric; Meters)
Dist_Village_m	Distance from plot to the nearest village (Numeric; Meters)
Dist_River_m	Distance from plot to the nearest river (Numeric; Meters)
Dist_PA_m	Distance from plot to the nearest protected area (Numeric; Meters)
Dist_Saw_Mills_m	Distance from plot to the nearest saw mill (Numeric; Meters)

3 Exploratory Analysis

3.1 Raw Data

To explore the raw data, the first few rows and the structure of the dataframe were analyzed. The class of each of the columns was also assesed, and any incorrect column classes were corrected. For this analysis, only the 'Plot' column needed to be changed from numeric to a factor variable. The number of empty rows within the AGB measurement columns were also checked and quantified. It is critical to know if there were any missing measurements. Approximately 10% of the data had missing values. The number of missing measurements, was reported back to The Poulsen Tropical Ecology Lab. It should be noted that the presence of missing data may alter the accuracy of the analysis presented here.

```
#Check first few rows of data and the structure of the data
head(Raw_CongoCarbon_Data)
str(Raw_CongoCarbon_Data)

#Check class of necessary columns
lapply(Raw_CongoCarbon_Data, class)

#Change necessary classes
Raw_CongoCarbon_Data$Plot <- as.factor(Raw_CongoCarbon_Data$Plot)

#Check for NAs
anyNA(Raw_CongoCarbon_Data$AGB05.MgE)
anyNA(Raw_CongoCarbon_Data$AGB09.MgE)
anyNA(Raw_CongoCarbon_Data$AGB13.MgE)

#Look at number of NA values
nrow(subset(Raw_CongoCarbon_Data, is.na(AGB05.MgE)))
nrow(subset(Raw_CongoCarbon_Data, is.na(AGB09.MgE)))
nrow(subset(Raw_CongoCarbon_Data, is.na(AGB13.MgE)))

#Proportion of NAs
(nrow(subset(Raw_CongoCarbon_Data, is.na(AGB05.MgE)))/nrow(Raw_CongoCarbon_Data))*100
```

3.2 Covariate Data

The dataset containing the covariate values at each plot was analyzed similarly to the raw data. The beginning rows of the dataset, as well as the structure, and then the class of each column were assessed.

```r
head(CongoCarbon_Covariates)
str(CongoCarbon_Covariates)

#Check the class for all columns in CongoCarbon_Covariates
lapply(CongoCarbon_Covariates, class)
```

4 Analysis

The raw data was wrangled in order to obtain the AGB sum of all the trees within each plot. Columns for the change in AGB between the years were created by subtracting the sum AGB of the following measurement year from the sum AGB from the prior measurement year. This created a data frame (AGB_by_Plot) with the plot number and the total sum AGB at each plot for the years 2005, 2009, and 2013, as well as the change in AGB from 2005-2009, 2009-2013, and 2005-2013. The AGB_by_Plot data from was then merged with the raw covariates data to create one final dataframe. This final data frame includes plot number, all covariate values at each plot, the sum AGB for each year at each plot, and the change in AGB from one measurement year to another for each plot. This data frame is saved as CongoCarbon_AGB_and_Covariates_by_Plot.csv in the Data/Processed directory of the repository.

```r
#Create columns with the sum AGB by plot for each year '05, '09, '13
AGB_by_Plot <- Raw_CongoCarbon_Data %>%
  group_by(Plot) %>%
  summarize(sum_AGB05 = sum(AGB05.MgE, na.rm = TRUE),
            sum_AGB09 = sum(AGB09.MgE, na.rm = TRUE),
            sum_AGB13 = sum(AGB13.MgE, na.rm = TRUE))

#Create columns for change in AGB from '05-'09, '09-'13, & '05-'13
AGB_by_Plot$change0509 <- with(AGB_by_Plot, sum_AGB09 - sum_AGB05)
AGB_by_Plot$change0913 <- with(AGB_by_Plot, sum_AGB13 - sum_AGB09)
AGB_by_Plot$change0513 <- with(AGB_by_Plot, sum_AGB13 - sum_AGB05)

#Combine AGB_by_Plot and CongoCarbon_Covariates by plot
AGB_and_Covariates_by_Plot <-
  merge(CongoCarbon_Covariates, AGB_by_Plot, by="Plot")
```

Further data wrangling was necessary in order to graph this data. The dataframe created above (AGB_and_Covariates_by_Plot) is in horizontal form. In order to move forward with this analysis, the data must be in vertical form, so the melt() function was implemented to complete this task. Through use of this tool, a new dataframe (melted_AGB_by_Plot) was created, containing columns for plot, year, the variable (either sum AGB or change in AGB) and the AGB value. Rather than having multiple columns for each measurement year of

AGB and multiple columns for the change in AGB measurements, there is now one column for AGB value, one column for year, and one column for variable. The melted dataframe was then subsetted to make two new sets of data, one for the sum values of AGB, and one for the values of change in AGB. The same process was applied to the original raw data (Raw_CongoCarbon_data) by selecting the necessary columns and applying the melt function.

```r
#Change dataframe from horizontal to vertical
melted_AGB_by_Plot <- melt(AGB_by_Plot, id=c("Plot"))
#Create a column for year
melted_AGB_by_Plot$Year <- melted_AGB_by_Plot$variable
melted_AGB_by_Plot$Year <- gsub("sum_AGB05","2005",
                                melted_AGB_by_Plot$Year, fixed=T)
melted_AGB_by_Plot$Year <- gsub("sum_AGB09","2009",
                                melted_AGB_by_Plot$Year, fixed=T)
melted_AGB_by_Plot$Year <- gsub("sum_AGB13","2013",
                                melted_AGB_by_Plot$Year, fixed=T)

#Create dataframe for the sum of AGB by plot
sum_AGB_by_plot <- melted_AGB_by_Plot[melted_AGB_by_Plot$variable
                                      %like% "sum", ]
#Create dataframe for the change in AGB by plot
change_AGB_by_plot <- melted_AGB_by_Plot[melted_AGB_by_Plot$variable
                                         %like% "change", ]
#Create dataframe for the change in AGB by year
AGB_change_by_year <- change_AGB_by_plot %>%
  group_by(variable) %>%
  summarise(sum=sum(value))

#Subset necessary data
Raw_data_subset <- Raw_CongoCarbon_Data %>%
  dplyr::select(Tree.ID, Plot, AGB05.MgE, AGB09.MgE, AGB13.MgE)

#Melt data
melted_raw_data <- melt(Raw_data_subset, id=c("Tree.ID","Plot"))

#Create a column for year
melted_raw_data$Year <- melted_raw_data$variable
melted_raw_data$Year <- gsub("AGB05.MgE","2005",
                             melted_raw_data$Year, fixed=T)
melted_raw_data$Year <- gsub("AGB09.MgE","2009",
                             melted_raw_data$Year, fixed=T)
melted_raw_data$Year <- gsub("AGB13.MgE","2013",
                             melted_raw_data$Year, fixed=T)
```

4.1 Data visualization

Figure 1 shows histograms of the AGB values of the raw data for each year. This provides general insight about the data, such as the normality, whether there are outliers present, and also compares the values of AGB measurements for each of the years. We can see that, for all years, a vast majority of the measurements are very low (between 0 and 10 Mg), but there are a few significant outliers (Figure 1).

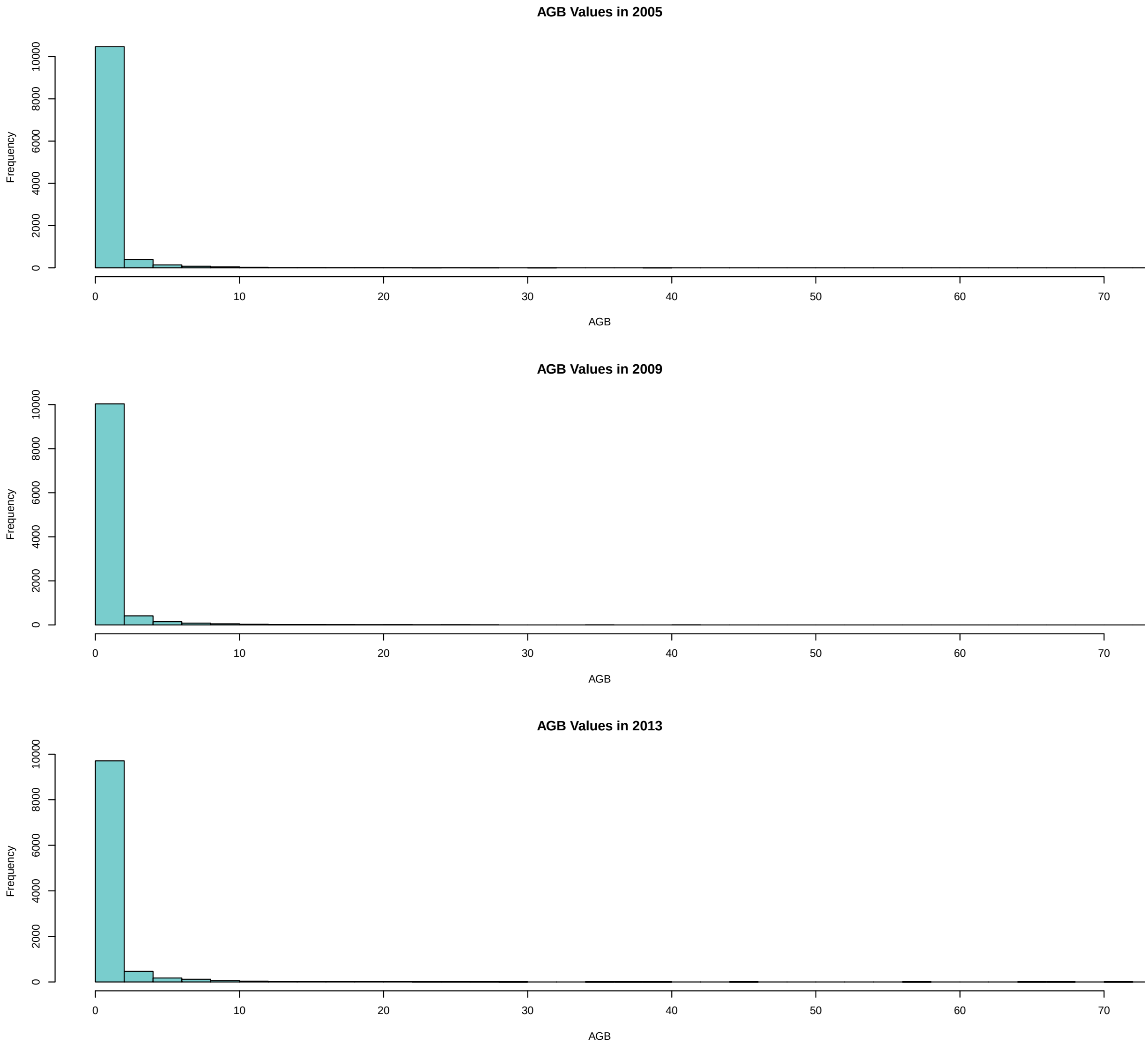

Figure 1: Histogram showing the distribution of AGB measurements of the trees for 2005, 2009, and 2013.

```
## null device
##           1
```

In order to further vizualize how AGB may differ between the measurements years, a line graph represents the change in the sum of AGB of each plot over time (Figure 2). Each line represents a different plot number. Here, we can speculate that there was not much change in AGB from 2005 to 2009, however there was significant change from 2009 to 2013.

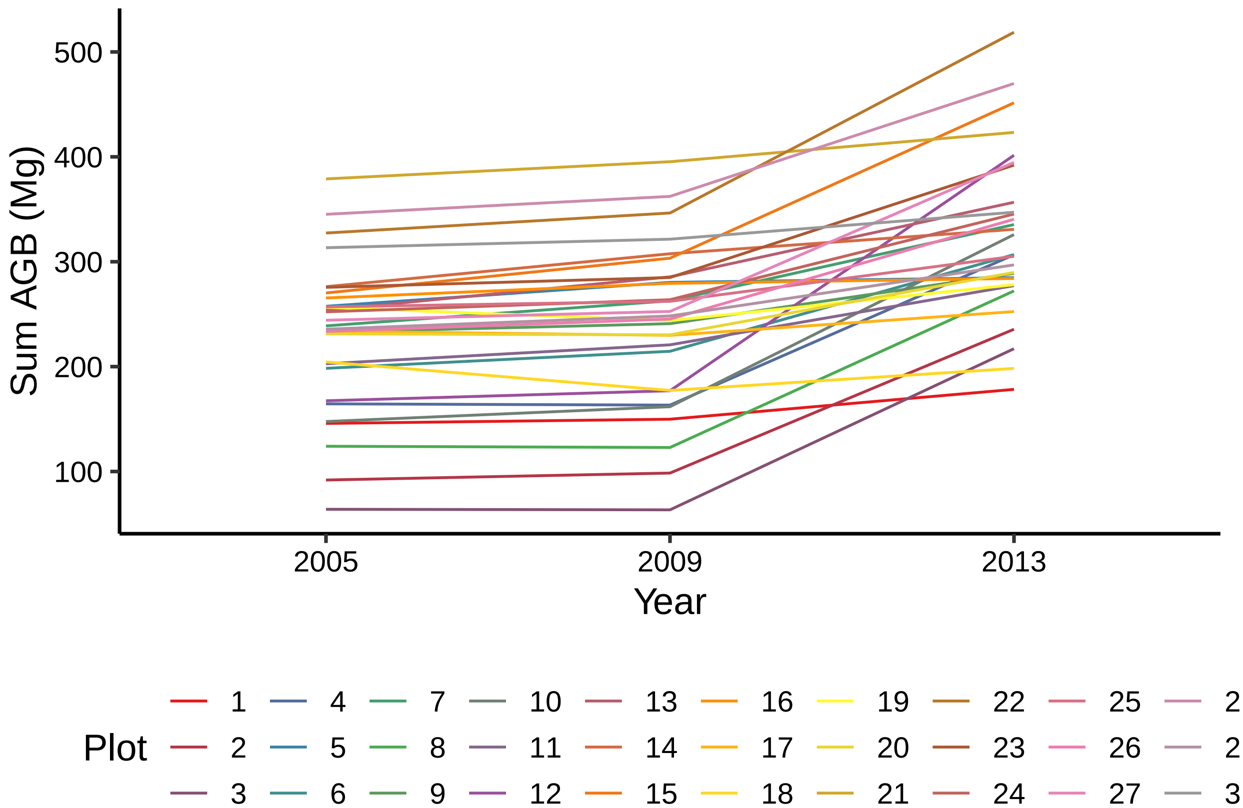

Figure 2: Relationship of the changes in total above ground biomass at each plot between 2005, 2009, and 2013.

The following violin plot shows the distribution of AGB measurements within each plot for 2005, 2009, and 2013 (Figure 3). Similar to the histogram, we can see that a majority of the values are low measurements, with the presence of significant outliers. We can see here the differences in AGB values not only between the different trees within each plots (shown by the long tails), however we also see that there are differences between the overall sum AGB from one plot to another.

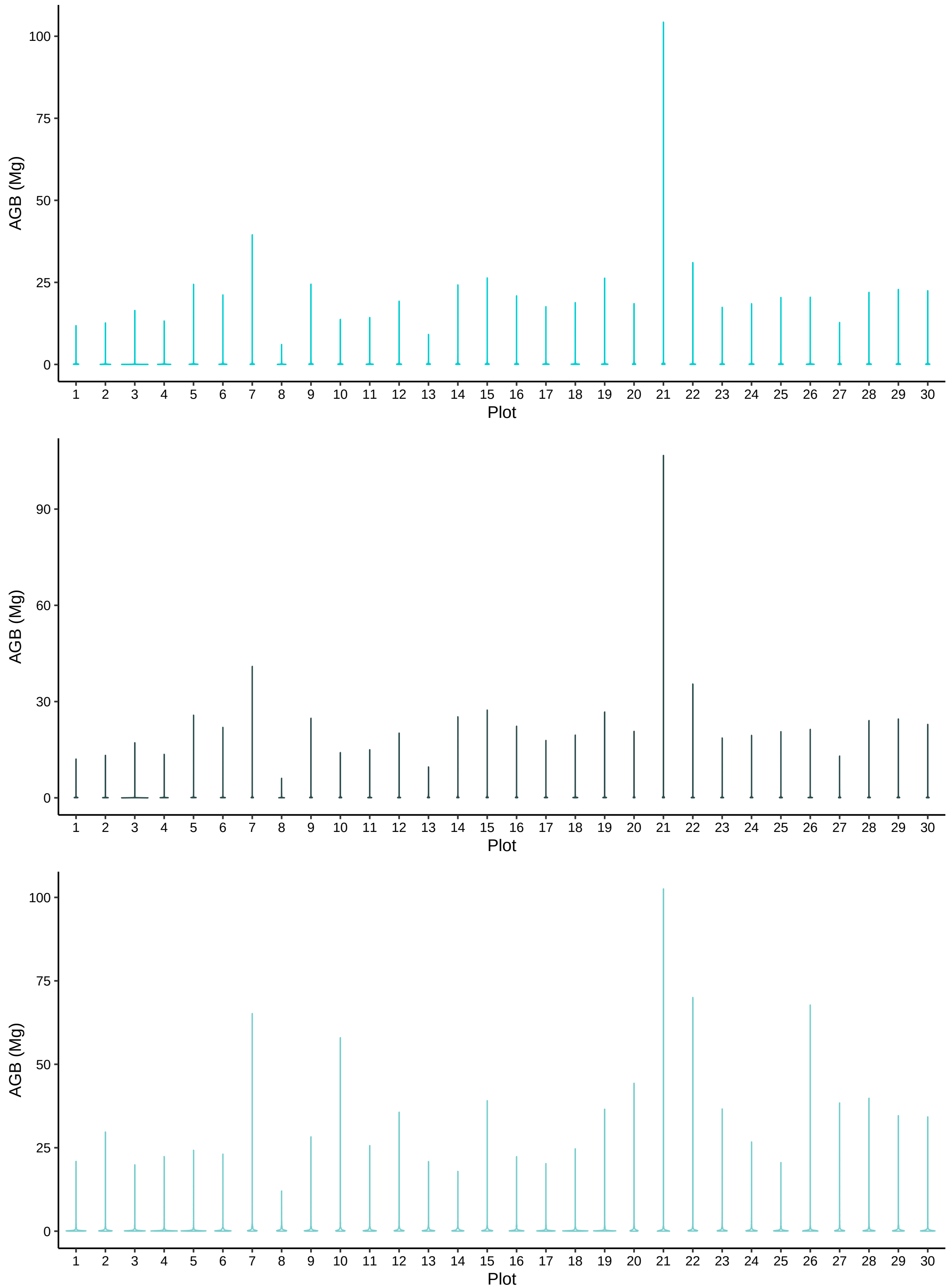

Figure 3: Relationship between above ground biomass at each plot in 2005, 2009, and 2013.

Figure 4 displays the sum AGB of all trees within each plot to vizualize the differences in AGB from one plot to another. For simplicity, the measurements collected in 2013 were the sole focus. Here we can more difinitively see the differences between total AGB at each plot. Where some plots, for example plot 1, had total AGB as low as 100 Mg, plot 22 had a total AGB of over 500 Mg (Figure 4). These differences prompted the second study question: what environmental variables may be affecting the amount of total AGB within a plot.

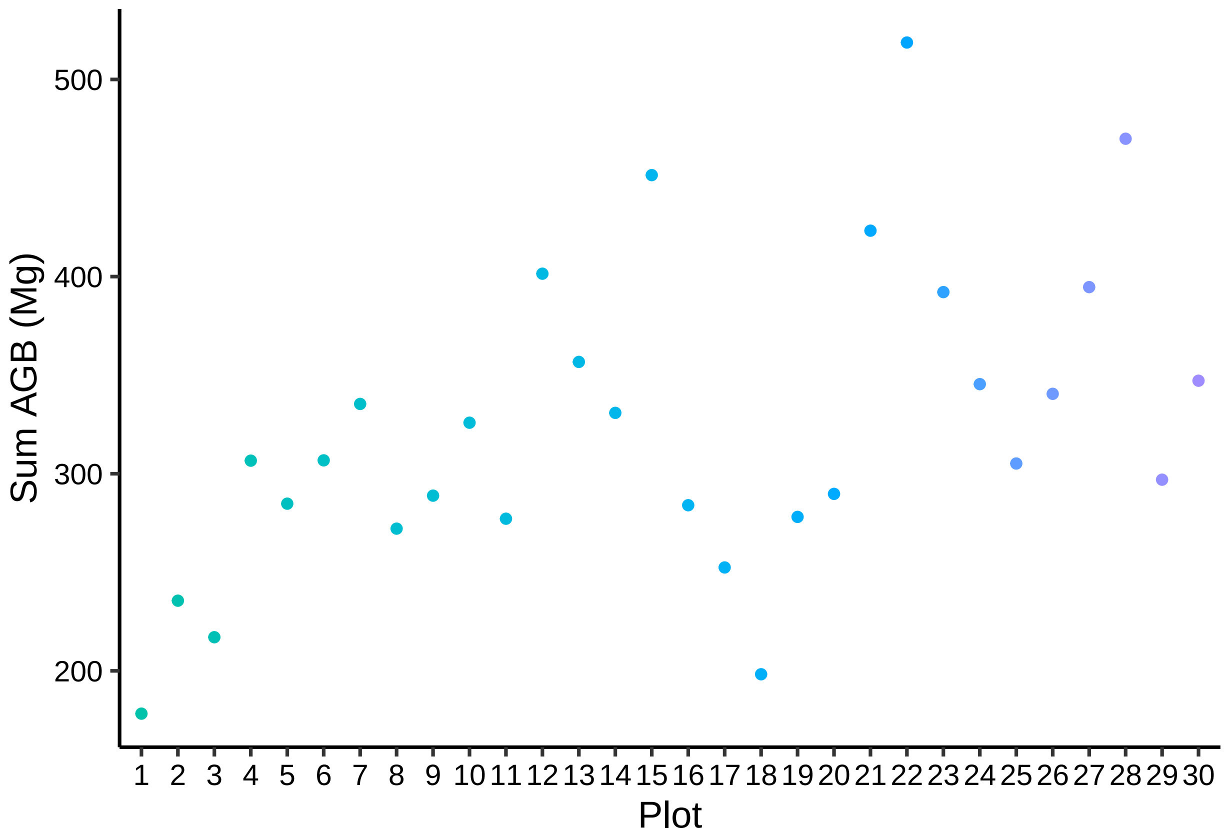

Figure 4: Relationship between total above ground biomass at each plot in 2013.

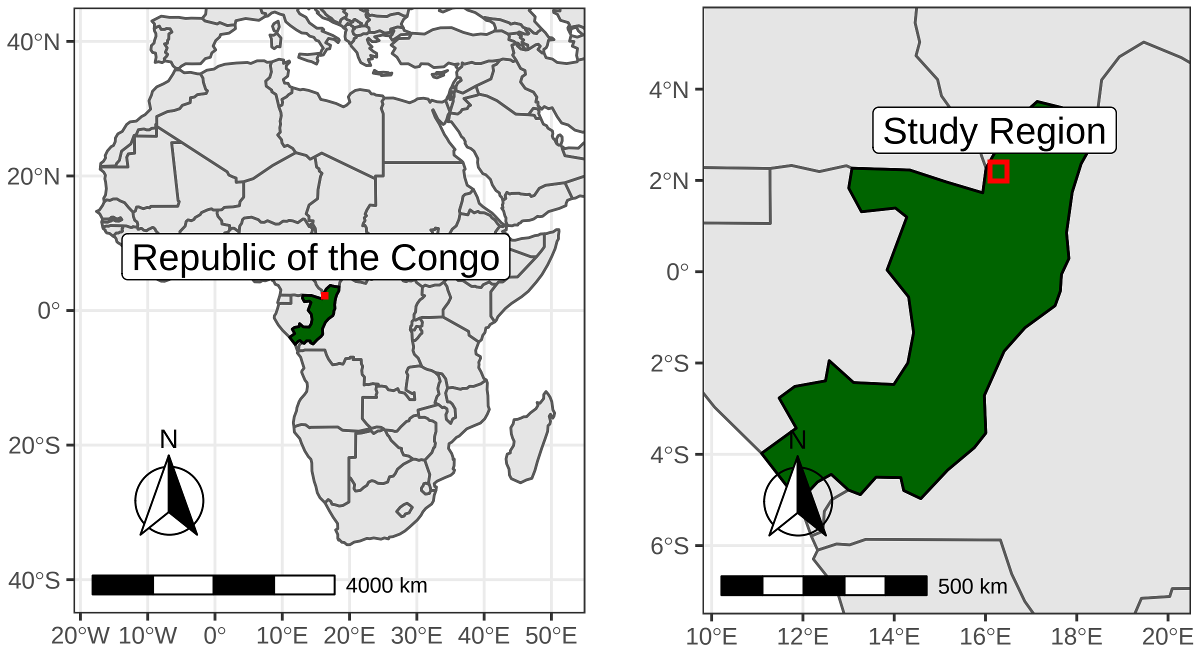

Figure 5: Map displaying the location of the forest plots within the Republic of Congo, Africa.

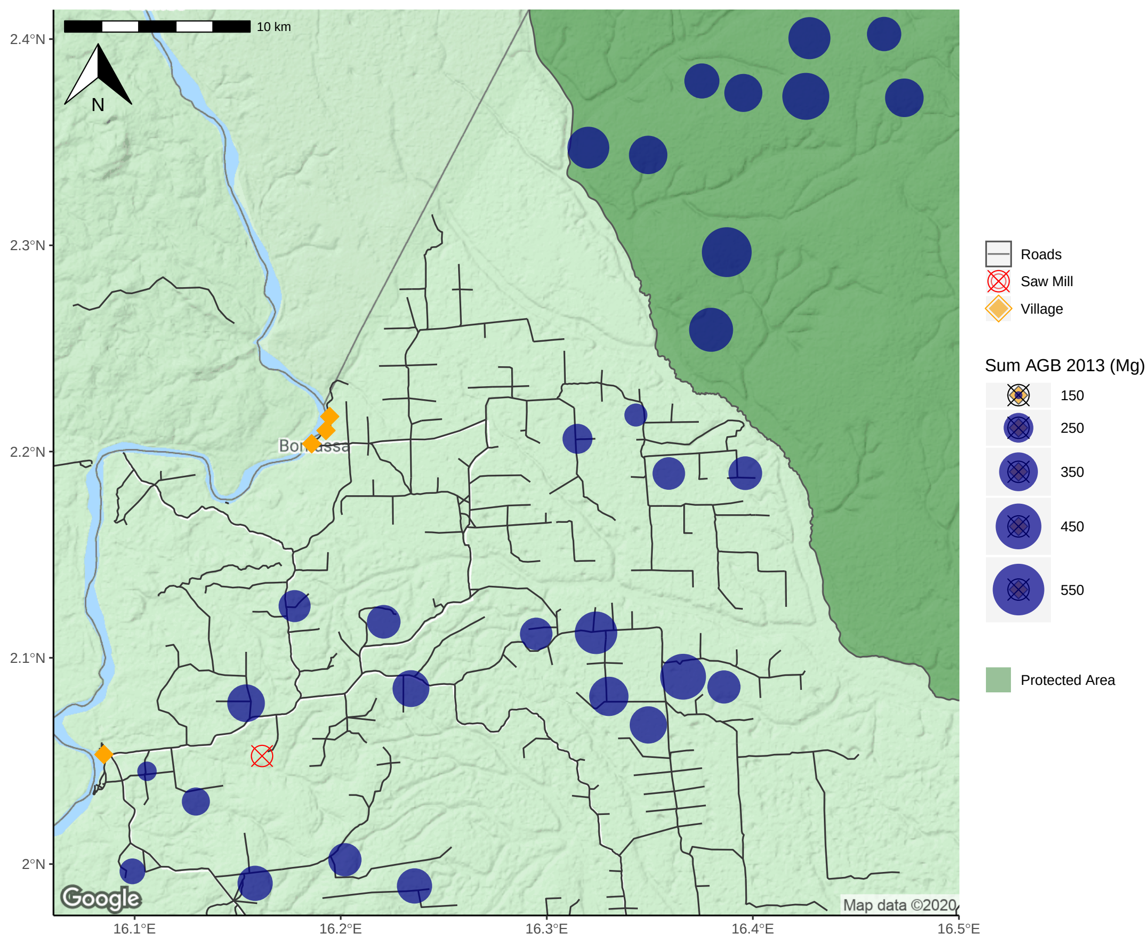

Figure 6: Sum of above ground biomass (Mg) at each plot location in 2013. Nearby road, village, saw mill, and protected area locations are plotted to visualize the relationship between these variables and the amount of above ground biomass.

4.2 Question 1: Is there significant difference in AGB between each of the years of data collection (2005, 2009, and 2013)?

In order to determine if there is significant increases in AGB between the years of data collection, a one-way anova was utilized to assess the effect that year has on the amount of AGB. A post-hoc Tukey test was implemented on the anova model to determine if there were significant differences between AGB for each year compared to the others.

The results show that there was a significant increase in AGB from 2005 to 2013, however not from 2005 to 2009 (ANOVA, $p<0.001$, $F_{2,32750} = 42.6$). The Tukey test reveals that there is no significant difference between the AGB values in 2005 compared to 2009 ($p=0.287$), whereas there is a significant difference between total AGB in 2005 and 2013 ($p < 0.001$) and between 2009 and 2013 ($p < 0.001$).

4.3 Question 2: Which environmental factors are significant predictors of AGB?

In order to determine which variables significantly impact amount of AGB at a plot, a generalized linear model was implemented. For simplicity, only data from 2013 were analyzed. In this model, HFI, GlobCover, soil type, precipitation sum, distance to road, distance to village, distance to protected areas, and distance to saw mills were all considered as the predictor variables, and sum of AGB was the response variable. The step function was implemented to determine the best model by testing the different combination of predictor variables with the response variable, and designating the model with the lowest Akaike's Information Criterion (AIC) value. The model with the best fit included precipitation, distance to road, distance to village, distance to protected area, and distance to saw mills as predictor variables. Of these variables, all of them apart from distance to roads represented statistically significant impacts on change in AGB.

The results of this model show that AGB increased with increasing precipitation, distances further from roads and villages, and distances closer to saw mills and protected areas (linear regression, $R^2 = 0.31$, df = 5,24, $p = 0.01$) (Figure 7)

Figure 7 shows that the model fits the data well. The QQplot displays that the data are not normal. The Residuals vs Fitted and the Scale-Location graph show that the data are not gathered strongly to either side, and are evenly distributed around the line. Finally, the Residual's vs Leverage plot shows that the points are all within Cook's Distance, thus there are no drastic outliers.

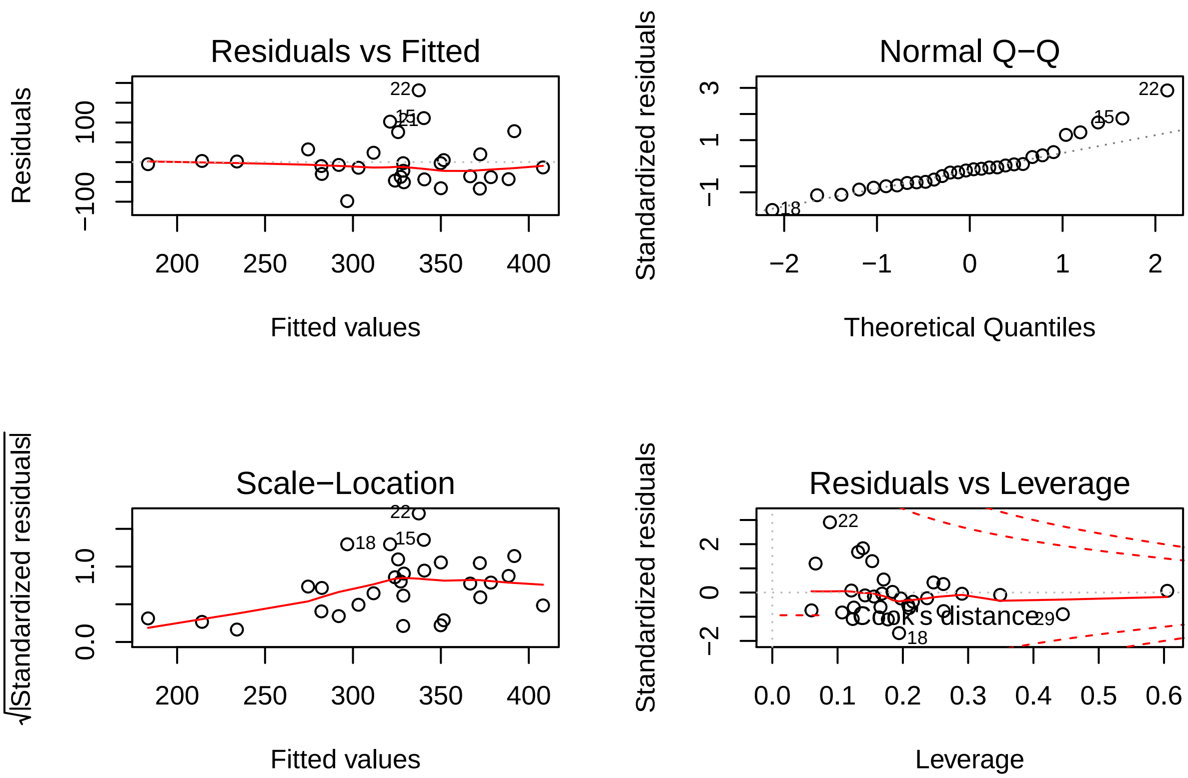

Figure 7: Fit of the model that represents the effect of total precipitation, distance to the nearest road, distance to the nearest village, distance to the nearest protected area, and distance to the nearest saw mill on total AGB at the plots in 2013

5 Summary and Conclusions

Overall, this study found that there was no significant increase in total AGB across all plots between 2005 and 2009. However, there was a significant increase from 2005 to 2013, and from 2009 to 2013. This is not what one would expect to find in an undisturbed area, given that trees continue to grow with time. However, it is possible that the rate of logging exceeded the rate of tree growth between the years 2005 and 2009. Further research would have to be done to determine the cause of this lack of AGB increase. This study also found that precipitation, distance to villages, distance to protected areas, and distance to saw mills all had a significant impact on the amount of AGB within a plot. AGB increased with increasing precipitation and increasing distance from the villages. AGB also increased in plots that were closer, or were located within the protected area. This all logically makes sense, given that rainfall increases tree growth, and the unprotected plots that are closer to villages are at an increased risk of logging. However, we did not expect to find that AGB would increase the closer the distance to the saw mill. It is clear from the lack of AGB increase from 2005 to 2009 and from the inverse relationship between AGB and distance to saw mills that further analysis is needed to explain some of these unexpected results. One possible source of error is the blank tree measurements in the data mentioned early in this report, or it is possible that illegal or legal logging is occurring at unprecedented rates. If the latter is true, studies like these have the potential to bring awareness to practices that cause hard to the environment, and could impact the

6 References

Chave, J., Réjou-Méchain, M., Búrquez, A., Chidumayo, E., Colgan, M. S., Delitti, W. B., ... Vieilledent, G. (2014). Improved allometric models to estimate the aboveground biomass of tropical trees. Global Change Biology, 20(10), 3177–3190. doi: 10.1111/gcb.12629

Duncanson, L., Armston, J., Disney, M. et al. The Importance of Consistent Global Forest Aboveground Biomass Product Validation. Surv Geophys 40, 979–999 (2019). https://doi.org/10.1007/s10712-019-09538-8