# “The impact of road infrastructure on deforestation in Cameroon: A time-series analysis”

Master’s thesis by
Aleksandhros (Alex) Papa
2642873

Supervised by
Dr Erik Ansink

MSc Spatial, Transport, and Environmental Economics
School of Business and Economics
June 2022

# Abstract

Road infrastructure is strongly associated with environmental degradation in developing countries and specifically in tropical forests. However, such countries lack historical road data. Utilizing remote sensing techniques, this study creates a time series road network dataset for Cameroon from 2001 to 2020. At the same time, this body of work estimates the impact of distance to roads on deforestation using fixed-effects and propensity score matching models. The study concludes that the application of the historical data provides more accurate results of the effect, in comparison to the static roads which tend to overestimate the impact. Moreover, the year fixed-effect model indicates that the probability of deforestation increases by 0.98% when the distance to the closest road segment increases by 1km. The paper suggests that historical road data could be implemented in other studies and variables which use static observations such as road infrastructure, in order to generate robust results.


Keywords: road infrastructure, remote sensing, road detection, deforestation

## Table of Contents

## List of Figures

## List of Tables

## 1. Introduction

Public infrastructure is a necessary condition for economic growth. This is particularly true for developing economies that are striving to mobilise resources, allocate them efficiently and create conditions conducive to economic expansion. While this remains an undisputed path to sustainable growth and prosperity there are important trade-offs to consider: the case of roads is exemplary because constructing and expanding road networks leads to better access to resources, jobs, and markets creation; yet, roads generate negative externalities with environmental implications that policy experts and researchers need to analyse (Meijer et al., 2018). In other words, there is no such thing as a free lunch. Economic choices - no matter how neutral they might seem - inevitably bear significant costs for the environment. Namely, greenhouse emissions rise, and human proximity to the environment increases with unpredictable effects. Given the climate emergency that we face and the need for developing countries to upscale and upgrade their production by improving public infrastructure it is imperative to consider the environmental impact of road development.

According to the latest available data from the Global Forest Watch, Cameroon 'lost 1.53Mha of tree cover, equivalent to a 4.9% decrease in tree cover since 2001'. The main driver of this environmental crisis is a shift to agriculture in the national production of the country. Increased production of agricultural goods requires roads to facilitate access to land, reach untapped resources and facilitate mobility of goods. In the last decades, we have observed a steady increase in roads, especially in developing countries that are lagging behind. Nevertheless, the academic literature only recently started to analyse in a rigorous way the consequences of road development. What used to be a certainty in development economics (i.e., net positive effects of an investment in road infrastructure) is contested by researchers who reveal the tradeoff between road development and deforestation. For example, Kleinschroth et al. (2019) examined the impact of the road infrastructure on the ecosystem in Congo Basing and observed that road expansion deteriorates the environmental conditions. However, they only used two road datasets, one for the year 2003 and one for 2017. To the best of my knowledge, there is no other study that has analyzed historical road data to evaluate the impact on the environment and specifically on deforestation.

The present study focuses on Cameroon, a country in the Congo Basin that has experienced substantial forest loss in recent years. Moreover, it is important to mention that there is currently no road network panel data for Cameroon. Consequently, the current thesis aims to examine the impact of road proximity on deforestation across years by extracting a road dataset for the period of 2001 to 2020 employing remote sensing analysis. Using the derived time series road dataset, this thesis also compares the results to other relevant studies which use static roads in their analysis to examine whether the dynamic roads provide us with improved and accurate findings. In particular, the results are being compared to Panlasigui et al. (2018) who investigated the effect of the Forest Steward Council certificate on deforestation in Cameroon including proximity to roads. However, they analysed the effect by applying static roads, which did not vary over time, instead of dynamic roads.

The findings of the current research indicate that the inclusion of dynamic roads provides more precise estimations, while the static roads dataset leads to an overestimation of the environmental benefits or effects. This may motivate further studies to utilize historical road datasets in order to determine the effects that infrastructure has on the environment. Furthermore, the outcome may alert policymakers as well, since environmentally friendly certificates seem to be overestimated while the road panel data are not being incorporated. Additionally, the approach of including historical data can certainly be applied to other applications than deforestation, such as urban sprawl or economic growth. The same may apply to other kinds of manmade infrastructure than roads.

This study is divided into 7 sections. In section 2, relevant literature regarding the correlation between environmental degradation and road proximity is being discussed. Moreover, the section also reflects on the importance of utilizing remote sensing techniques and data to identify these effects. Section 3 elaborates on the road extraction methodology and on the data description. Additionally, section 4 discusses the empirical strategy which was performed in this research. Namely, the fixed effects, the pooled OLS and the Propensity Score Matching models. The results of the models are being presented in section 5. Furthermore, the empirical results are discussed in section 6. Finally, section 7 summarizes the study and discusses potential future research.

## 2. Literature Review

This section discusses the main findings of relevant studies regarding road infrastructure expansion and its effects. The approach of the following literature review is twofold. The first part elaborates on the impact of road network development on the ecosystem. Then, we take stock of the literature that weighs the benefits of using satellite and georeferenced data to examine the consequences of the expansion and battle the challenges that arise.

Barber et al. (2014) while studying the Amazon region, observed that the proximity to transportation networks and specifically to roads created by logging companies is a significant factor of environmental degradation in Amazonia. Moreover, they found that deforestation largely occurred in areas that were within a 5.5km radius of road segments. However, their research shows that the implementation of environmentally friendly policies would reduce the ecosystem's degradation. Kleinschroth et al. (2019), contend that in spite of the potential benefits of the road network on the development of isolated forest segments, such practice could also bear negative consequences on the ecosystem. The study focuses on the Congo Basin, and it records a positive correlation between proximity to roads and deforestation. In particular, they observe increased deforestation due to the unrestrained expansion of unpaved logging roads used by timber companies. Finally, the disuse of the newly created roads after the logging procedure, may benefit the ecosystem and rectify their negative impact (Kleinschroth et al., 2019).

Distance to roads is considered a stimulus for economic growth and poverty reduction by connecting remote areas (Jacoby 2000; Najman 2010). However, as mentioned before, it may also deteriorate the environmental conditions, by the expansion of the network. Another project that investigated road infrastructure development in the Brazilian Amazon is the one from Pfaff et al. (2007). The study observed that road investments in the Brazilian Amazon led to an increase in deforestation in a census tract within 100km. On the other hand, the examination of a distance between 100km and 300km produced contradicting results. Specifically, they argue that a road expansion or investment in such a distance may lead to a decrease in ecosystem deterioration. Yet, the latter part of the results is statistically insignificant and does not constitute a reliable source.

Moreover, the study by Laporte et al. (2007) focuses on Central Africa and finds that road infrastructure expanded significantly in the last decades. Indicatively, it rose from 156km a year$^{-1}$

during 1976–1990 to over 660 km a year$^{-1}$ after 2000. Laporte et al. (2007) used satellite data and observed that the logging roads represent one-third of the total roads in the Congo Basin. Additionally, they inspected the largest effect of deforestation in Cameroon.

From a different perspective, Lopez et al. (2017) examined the impact of road segments on biodiversity degradation. They detected that the majority of the fauna was affected in areas with road accessibility. Specifically, a considerable number of mammals became extinct in the area within a 7 km distance from road segments. However, the systematic review from Geist and Lambin (2002), even though they observed a negative association between road existence and environmental degradation in Latin America, concluded that the evidence from Africa and Asia was unreliable due to the lack of historical and accurate road infrastructure data.

Meijer et al. (2018) showcase the need for accurate spatial road datasets to underpin strategic spatial planning in order to reduce the impacts of roads in the remaining unaffected ecosystems. In particular, the study inspected the socio-economic effect of road expansion and they suggest that georeferenced information on road infrastructure is a significant tool for spatial planning and environmental impact assessments. As a response, Burke et al. (2021) argue that modern satellite-based methods can produce precise data. In many cases, satellite data provide the same degree of accuracy as their ground-based counterparts. The study notes that these techniques are frequently used by researchers to estimate land-use activity, economic development and policy assessment. Namely, the study from Yeh et al. (2020) uses publicly available satellite imagery and deep learning for the evaluation of economic development in Africa. Accordingly, Laporte et al. (2007) indicate that regular monitoring with satellite remote sensing allows for a consistent examination of the effects on land-use changes. However, the research from Burke et al. (2021) highlights that despite the strong potential for spatial analysis of satellite-based approaches, they cannot substitute the ground data but only augment their validity. Meijer et al. (2018) support that the need for historical and accurate data is prevalent since large increases in road length were projected for developing nations. Additionally, Burke et al. (2021) indicated that at least half of African nations lack consistent data, as they conduct ground-based surveys once every 6.5 years.

Last, Kleinschorth et al. (2019) using satellite-based data, created a road network dataset on the Congo Basin for the year 2017, including logging concession roads, which could not be found in any other road dataset so far. The present paper was significantly motivated by the approach from Kleinschorth et al. (2019). Namely using remote sensing data as well, this work aims to contribute to the existing literature by the creation of a time series road network dataset on Cameroon from 2001 to 2020. Due to the fact that the data in developing countries remain scarce, it would be compelling to identify patterns by analyzing historical evidence. Furthermore, unlike previous research, this study utilizes the derived dataset in order to evaluate the impact of the road network on Cameroon's forests over time.

## 3. Methodology and Data Collection

### 3.1 Classification approach

This section discusses the main methods that have been used in order to create a dynamic road dataset for Cameroon. The first part elaborates on the remote sensing approach, the classification model, and its accuracy. While the second part describes the time-series road selection process.

Our goal was to create a panel road dataset for Cameroon as there is currently no adequate data for the period before 2015. Taking into account the substantial deforestation that has occurred in the Congo Basin over the last two decades, we extracted road segments for the time period between 2001 and 2020. This was based on the Hansen et al. (2013) updated dataset which provides georeferenced deforestation data for the same study period. By analyzing these datasets, we were able to examine the causal effects of road infrastructure on deforestation in Cameroon.

In order to do so, we analyzed public access satellite images using the Google Earth Engine (GEE) platform; specifically, we performed supervised pixel-based image classification on satellite images derived from Landsat 7 and Landsat 8. The resolution of each pixel accounts for a 30m X 30m surface. The classification modelling was a complicated process since Cameroon is located in the Congo Basin rainforest. The majority of the area is facing storms and rains throughout the year. Therefore, cloud-free satellite images are scarce, hence we had to exclude some regions which would bias our sample. Specifically, the area close to Douala - even though it is the biggest city in Cameroon - had to be excluded since it was covered with clouds during the whole-time frame of our study. At the same time, the constant cloud existence affected road detection as well.

To address this challenge, we applied five-year intervals in the period of interest so we could extract accurate results. In particular, we utilized a mosaic of satellite images for the following years: 2005, 2010, 2015, and 2020. The mosaic application selects the less cloudy images over the years. As presented in Figure 3.1. the majority of the forest cover area is located in the Central, East, and South parts of Cameroon. On the other hand, the northern part consists of mountains and almost no forest or vegetation. Since the main objective of this study is to identify the impact of the road network on deforestation in Cameroon, the northern part of the country was not included in the analysis.

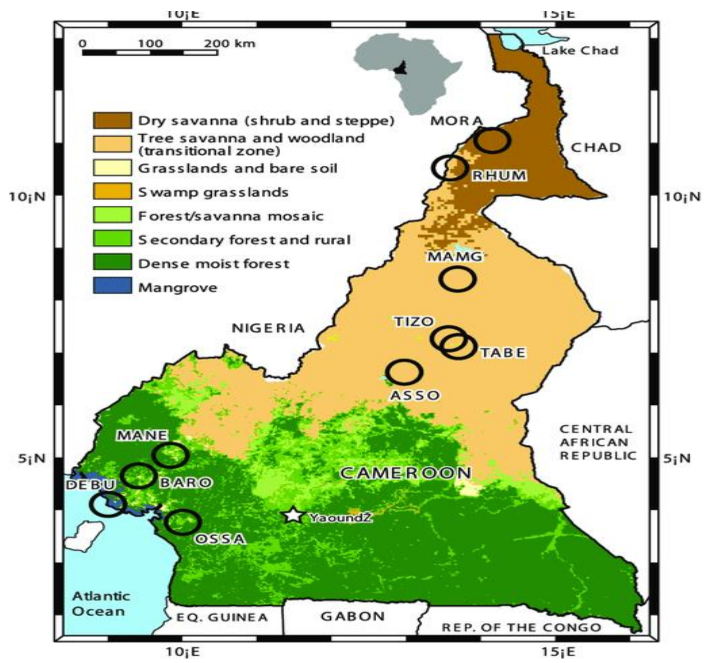



Figure 3.1: Morphological Map of Cameroon. Source: Schwab et al. 2015

Overall, for the classification, we used 8 satellite images and managed to cover 182.535 square km each year. In order to provide a consistent structure for the classification, we based our classes on the Land Cover Classification System (LCCS) provided by FAO. For our analysis, we had to modify and add extra values to tackle visual challenges, such as cloud existence and the Landsat 7 sensor’s defect. Thus, we created the following main classes: waterbodies (sea, rivers, lakes), vegetation (forests, land crops, trees), soil (agriculture and grasslands), urban (cities, infrastructure, buildings), roads (paved and unpaved roads, mud or sand roads) and clouds (clouds and satellite images combination defects). Finally, we introduced the supervised learning algorithm provided by GEE based on Breiman (2001) and (2017). We trained the algorithm by adding land-use values that have been manually identified from satellite images. After the training process, the algorithm was able to classify each pixel of the satellite image (Figure 3.2). As a result, as illustrated in Figure 3.3, a classified map indicating the land cover characteristics was generated.

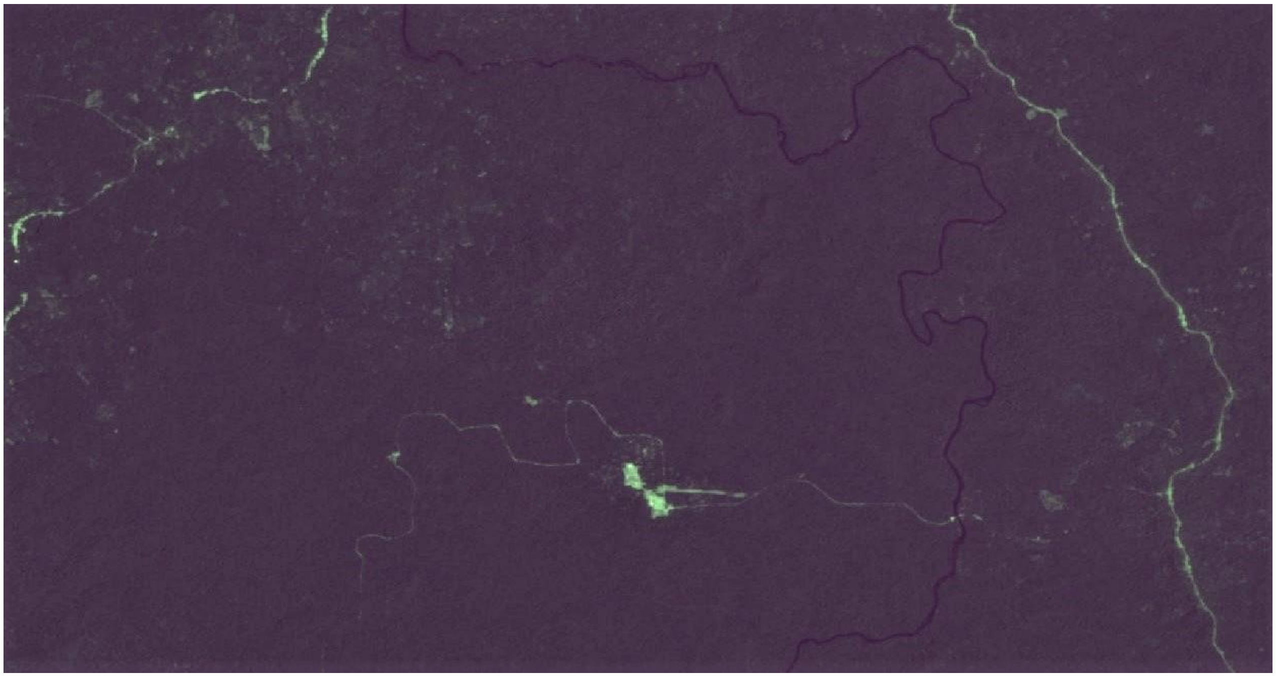


Figure 3.2: Satellite image of a region in Cameroon before the supervised classification intervention.

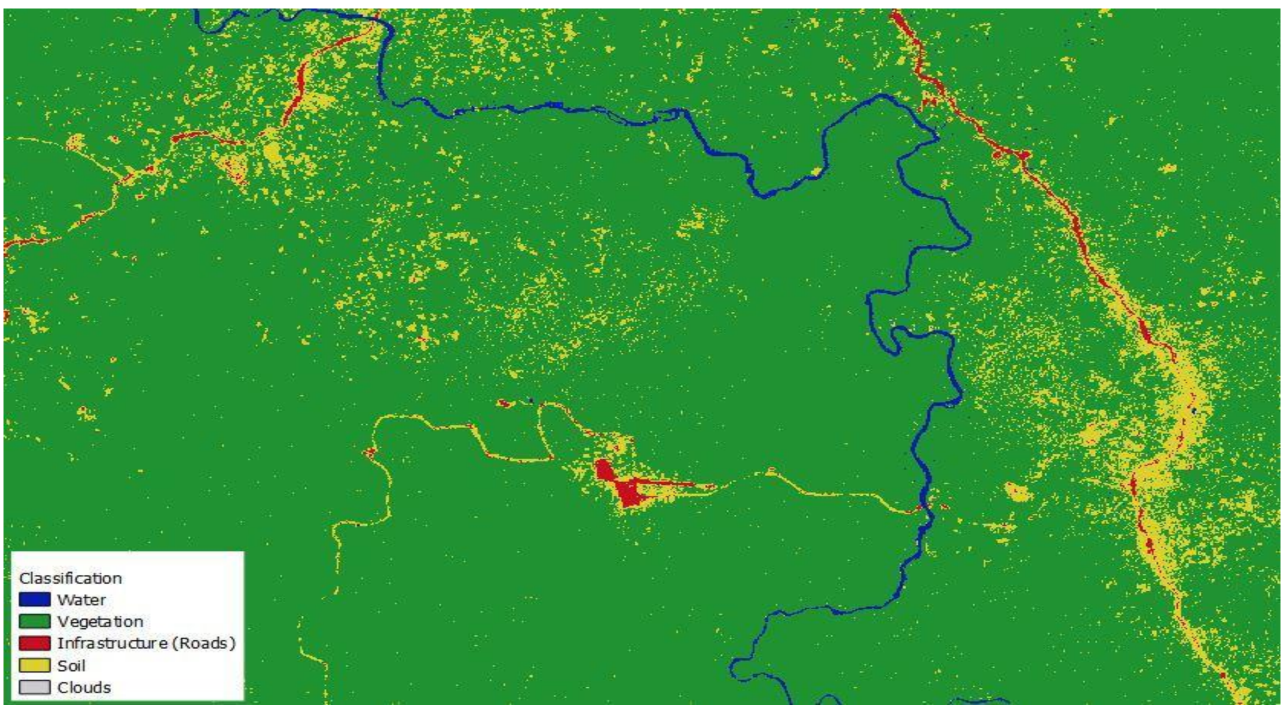




Figure 3.3: A classified image of the same region illustrated in Figure 3.2

## 3.2 Accuracy assessment of the classification model

In this section, we examine the accuracy of the classification. The class accuracies are determined by comparing test pixels with the corresponding location in the classified image. In a perfect world, we would be able to use field verified ground reference locations for the test pixels. Of course, this is not always possible, and, in this case, we may also select references that have been visually identified from the imagery. The pixels should be evenly distributed across the image. To evaluate our supervised learning algorithm, we perform an accuracy assessment on Landsat 7 and Landsat 8 satellite images. In particular, we selected 840 (ground truth) reference pixels indicating the corresponding class of the land cover. So, we chose the pixels that we were certain that were roads, and pixels that were not. Given that we know that our satellite images include a significant amount of forests and vegetation, we adjusted the number of pixels on the percentage of the actual land cover.

In order, to assess the classification accuracy, we used again the Google Earth Engine platform. We generated two samples from our data: one for training and one for testing. The training sample is used to train the classifier, while the testing one is being used as a test to get a confusion matrix representing the (expected) test set accuracy. Confusion matrixes are a widely accepted method of determining the accuracy of the classification (Breimann 2001; Foody 2002). However, it is important to remember that the biases that are present in the test pixels will also bias the accuracy of the confusion matrix.

We used 70% training and 30% testing. In particular, the 30% of the testing accounted for 252 pixel points. As the main focus of this study is to correctly identify road segments across years, we clustered into two classes, so we can examine the road pixel classification accuracy. The first part includes the potential road pixels (infrastructure, roads, and soil parts), while the second part consists of the non-road pixels (water, vegetation, and clouds). In Table 3.1, we can observe that the classification scored high overall accuracy. The producer’s accuracy achieved 96.3 % while the user’s accuracy notched a 95.5%. In particular, the classification model counted 5 out of 111 road pixels as a vegetation land cover. This was quite expected since the infrastructure and soil land cover reflection was similar to vegetation in many regions across South Cameroon.

Table 3.1: Accuracy assessment results

| | Roads | Other | Ground Truth | Users Accuracy |
|---|---|---|---|---|
| Roads | 137 | 4 | 141 | 97.16% |
| Other | 5 | 106 | 111 | 95.49% |
| Total | 142 | 110 | 252 | |
| Producer's Accuracy | 96.47% | 96.36% | | Total Accuracy = 96.5% |
| | | | | |
| Kappa Coefficient | 0.927 | | | |

Overall accuracy essentially tells us out of all of the reference sites what proportions were mapped correctly. In our case, we scored an overall accuracy of 96.5% which indicates that the majority of the ground truth pixels were classified correctly. The Kappa Coefficient is generated from a statistical test to evaluate the accuracy of classification. In our model, the coefficient stands for 0.927. A value close to 1 indicates that the classification is significantly better than random. Furthermore, the ground truth pixels have been as near to evenly distributed as possible in order to avoid potential bias in the accuracy. So, in order to ensure that the training samples are uncorrelated with the evaluation sample, we removed samples that are within 10km of any other sample using spatial join. In that way, we avoided spatial autocorrelation.

## 3.3 Road Selection Methodology

It was essential to have accurate classified images in order to implement our methodology for the time series road dataset production. By introducing these images in QGIS, we were able to reclassify the images and extract only the pixels that could be accounted for as part of roads (road pixels). We utilized the 2021 Open Street Maps (OSM) road dataset as a benchmark and compared each road segment to the road pixels. Our approach indicates that if the potential road pixels overlap with the 2021 OSM roads, then the whole road segment overlapping with the road pixels is selected for that year of the road dataset. However, the potential road pixels had a small deviation of five meters from the actual road segments, as the satellite images represented 30m X 30m pixels, while the OSM road segments had ground measurements.

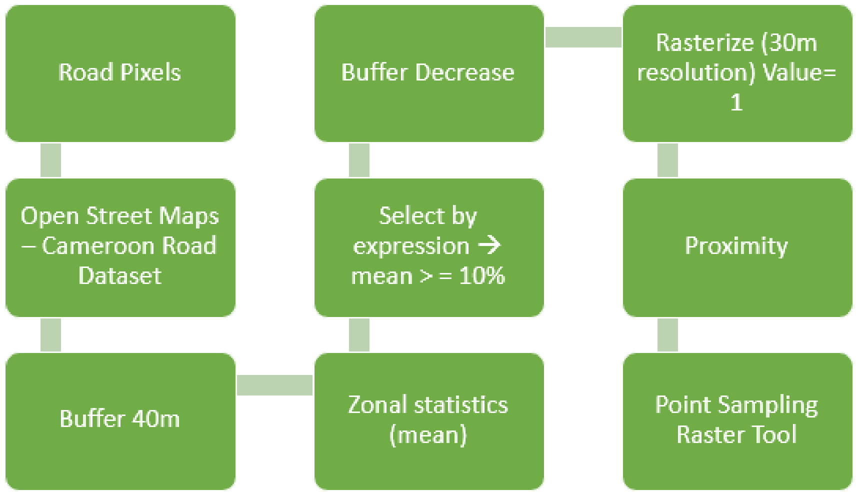



Figure 3.4: The methodology process for road segment selection

To address this problem, we buffered the OSM segment by 40m in total, so it would include the road pixels inside the segment surface area. However, this also led to having a bigger surface of the actual road segments. Consequently, a significant decrease in the percentage of road pixels in each road segment was observed. Our approach compares the surface of the potential road pixels to the surface OSM road segments and selects the ones that cover at least 10% of the road segments. A manual accuracy assessment was performed by randomly testing 100 road segments. We concluded that by defining a threshold of 10% surface inclusion of the classified pixels inside the area of the OSM road segments the selection became accurate and precise. In particular, while setting the threshold higher than 10% in our selection process, the potential road pixels were underestimated, and a significant amount of road segments were excluded. At the same time, while the threshold was set lower than 10% the roads were overestimated. We also excluded in our formula single pixels in order to avoid potential bias. Figures 3.5 and 3.6, show the way that a road is being selected. Specifically, in Figure 3.5, which is a reclassified image, we can observe in black the parts (pixels) that are considered as potential road segments and in white all the rest. The red lines represent the 2021 OSM roads. In Figure 3.6, our model selects only the segment -the yellow line- that exists in Figure 3.5 too. The rest segments (blue lines) are being dropped from our dataset since they do not exist in the examined period. In that way, we derived four road datasets for the years 2005, 2010, 2015 and 2020. Finally, we removed the buffer intervention from the selected roads in order to obtain actual measurements.


Figure 3.5: A reclassified image of a region in Cameroon in 2005 obtained from Landsat 7

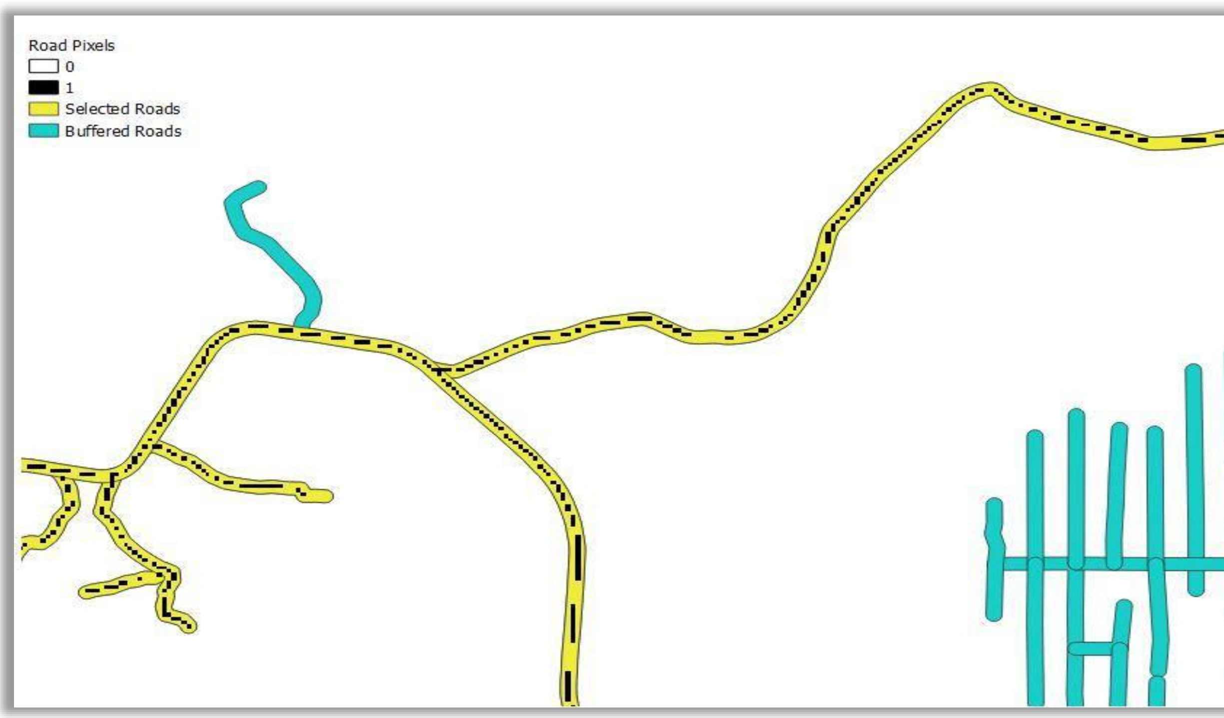



Figure 3.6: Illustration of the road selection for the same region as Figure 3.5

## 3.4 Data

This section presents and discusses the data that have been utilized for this study. The data was acquired from several sources. Specifically, we derived our satellite images from Google Earth Engine and explicitly from Landsat 7 and Landsat 8 provided by the United States Geological Survey (USGS) and NASA. The 2021 road dataset that was utilized as a benchmark in our methodology was derived by Open Street Maps. The Euclidean distance to roads was estimated through QGIS and calculated in km. Moreover, the deforestation and the forest cover data were obtained from the Hansen et al. (2013) database for global deforestation from 2001 to 2020. Forest loss is defined as a stand-replacement disturbance (a change from a forest to a non-forest). While forest cover is defined as a standard percentage through all the years of interest. In particular, as a canopy closure for all greenery greater than 5m in height. Regarding the biophysical features, such as elevation and slope, were obtained by Jarvis et al. (2008) topographic mission. The Hijmans et al. (2005) climate dataset was utilized in order to extract the annual average precipitation values. The remaining geographical variables were calculated in GEE. Last, the FSC data stands for the Forest Steward Council certification which was made available by the FSC organization website and the World Resource Institute.

Table 3.2: Data Sources

| **Data** | **Source** | **Details** |
|---|---|---|
| Satellite Images | USGS - NASA (2001-2020) | Landsat7 and Landsat8 - Processed on Google Earth Engine |
| 2021 Road Dataset | Open Street Maps (OSM) | Processed on QGIS |
| Deforestation | Hansen et al. (2013) | 2021 Updated version of deforestation data (2001 - 2020) |
| Forest Cover | Hansen et al. (2013) | Standard percentage through all the years (2001 - 2020) |
| Annual Average Precipitation | Hijmans et al. (2005) | Processed on QGIS |
| Elevation and Slope | Jarvis et al. (2008) | Processed on QGIS |
| Distances | USGS - NASA | Processed on QGIS |
| FSC certificate | FSC.org (2020) and World Resource Institute (2021) | Processed on QGIS |

The region of interest consists of 182,535 square km which account for more than 300-million-pixel points with a spatial resolution of 30m x30m (see Figure 3.7). We analyzed one million random pixel observations for each time period so we could eliminate potential spatial biases. While implementing the pixel panel data, the total observations reached 4 million, one million per time period. As we wanted to examine the causal effects of road proximity on deforestation, we divided the sample into two parts. The first 500k pixels were located in areas that were deforested between 2001 and 2020. While the rest 500k pixels were selected in areas where no deforestation occurred. In the second part, we included only pixels in areas with forest cover higher than 30% based on Panlasigui et al. (2018). We excluded the pixels that had forest cover lower than 30% from our data set since no deforestation can arise in areas where no forest exists.

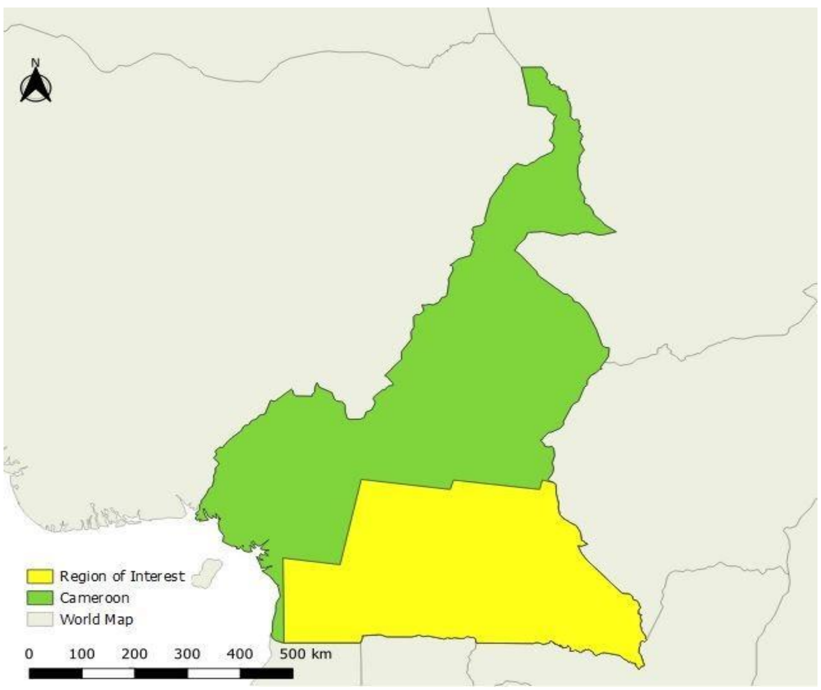



Figure 3.7: Map of Cameroon. The green area highlights the Cameroon boundaries, while the yellow area indicates the region where the study was performed.

Furthermore, we divided the Hansen et al. (2013) deforestation dataset into four time periods, so it can correspond with the derived dynamic roads. Thus, the first period consists of the accumulated forest loss for the years 2001 up to 2005. The second period is from 2006 to 2010, the third from 2011 to 2015, and the last one from 2016 to 2020. Regarding the dynamic road data, the variable for the Euclidean distance changes every five years. In that way, we managed to correlate the five-year road evolution to the concentrated forest loss for the same period. We had to use five-year intervals as it was not possible to extract road segments for each year due to poor resolution and permanent cloud existence.

We observed a substantial increase in deforestation after 2010 (Appendix A, Figure A1). Explicitly, in our sample, the deforestation points in 2005 accounted for 36,382 while in 2020 for 269,869. Regarding the road network, the 2021 OSM road dataset accounted for 47.949 segments in the region of study. In 2005, we observed 37.508 parts. In 2010 there was an increase to 43.938 roads, while in 2015 the derived roads were 44.296. Finally, in 2020 the number of roads was close enough to the benchmark as we managed to observe 46.413 segments (Appendix A, Figure A2).

## 4. Empirical Strategy

This section elaborates on the econometric models that have been used in this study. Specifically, to identify the effect of the road infrastructure proximity on deforestation across time, we implemented a fixed-effect model, and two pooled Ordinary Least Square (OLS) models, one for static roads and one for dynamic roads. Moreover, we used two matched fixed-effects models derived from the propensity score matching approach, so we could replicate the methodology from Panlasigui et al. (2018) and compare the variation of the FSC certificate impact on deforestation between static and dynamic roads. It is also important to mention that we dropped the observations in the years after deforestation. In the table below, we introduce a brief description of the variables within the panel dataset and in the following models.

Table 4.1: Description of the variables and summary statistics

| **Variable** | **Description** | **Obs** | **Mean** | **St.Dev** | **Min** | **Max** |
|---|---|---|---|---|---|---|
| Deforestation | Dummy for deforestation. Value 1 if deforestation occurred | 3648804 | 0.13703 | 0.34388 | 0 | 1 |
| Dynamic Roads | Euclidean distance to roads in km across years 2005-2020 | 3648804 | 3.41379 | 5.06344 | 0 | 45.868 |
| Static Roads | Euclidean distance to roads in km on year 2021 | 3648804 | 2.54215 | 3.82198 | 0 | 36.544 |
| FSC | Dummy for FSC certification. Value 1 if certified. In unbalanced certifications across years, the median was used as the value. | 3648804 | 0.27933 | 0.16477 | 0 | 1 |
| Forest Cover | Percentage of forest cover | 3648804 | 80.2546 | 16.2546 | 0 | 100 |
| Water Proximity | Euclidean distance to water bodies in km | 3648804 | 43.8332 | 26.3244 | 0 | 112.46 |
| Cities Proximity | Euclidean distance to cities in km | 3648804 | 15.6771 | 13.1961 | 0 | 109.95 |
| Elevation | Elevation in meters | 3648804 | 606.062 | 152.994 | 0 | 1347 |
| Slope | Slope in degrees | 3648804 | 3.90884 | 3.14247 | 0 | 48.179 |
| Precipitation | Annual average precipitation | 3648804 | 1692.21 | 247.116 | 1297 | 2790 |

## 4.1 Fixed Effects

To identify the effect of the proximity to roads on deforestation, it is quite significant to control for observed and unobserved factors that affect deforestation. In order to calculate these effects, we utilized the pixel panel dataset and estimated model (1) using both five-year fixed effects and pixel fixed effects. The fixed effect model is presented below.

$$DY_{it} = R_{it}\,\beta_1 + \gamma_i + \delta_t + \varepsilon_{it} \quad (1)$$

The DY is a dummy variable that indicates if the forest pixel was deforested or not. Specifically, $DY_{it} = 1$ if the pixel was deforested in the year *t* and $DY_{it} = 0$ if it was not deforested in the year *t*. R stands for the Euclidean distance to road segments over time. $\gamma$ and $\delta$ represent the fixed effects. Specifically, the $\gamma$ accounts for the pixel fixed effects, while $\delta$ for the five-year fixed effects. The subscript *i* stands for pixel observation, while the subscript *t* represents the year. Last, the error term is defined by the $\varepsilon$.

## 4.2 Pooled OLS

We performed pooled OLS regressions for two main reasons. First, the time-invariant control variables drop in the fixed-effect model, so it was impossible to control for them. Second, we wanted to compare the difference between the dynamic and static roads. That said, we performed two pooled OLS regressions (2) and (3). The first one includes the Euclidean distance to roads which varies over time (dynamic roads) as the independent variable. However, pooled OLS ignores time and treats the dynamic roads as a continuous variable. On the second one, we applied as an independent variable the Euclidean distance to the static roads of 2021. Both models included a dummy of deforestation as a dependent variable and extra control variables. We included geographic variables like forest cover and elevation. Distance variables such as Euclidean distance to cities and Euclidean distance to water were added as well. Again, we dropped the observations in the years after deforestation.

$$DY = \beta_0 + DynRoads\,\beta_1 + X\,\beta_2 + \varepsilon \quad (2)$$

$$DY = \beta_0 + StatRoads\,\beta_1 + X\,\beta_2 + \varepsilon \quad (3)$$

## 4.3 Propensity Score Matching – Matched Fixed Effects

The analysis from Panlasigui et al. (2018) performs a propensity score matching approach to examine the effect of the FSC certificate. In particular, the method matches the certified pixels with uncertified control pixels based on specific characteristics such as distances and physical conditions. Regarding the distance to roads, they use static roads. In this study, we replicate the Panlasigui et al. (2018) propensity score matching, and we create two models. The first one includes static roads and the second one dynamic roads (see Appendix A, Table A1 and Table A2). The main idea is to identify if the effect of the FSC differs when we use the dynamic road dataset instead of the static one.

For the Propensity Score Matching (PSM) approach, we randomly selected a sample of almost 1 million pixels. The certified pixels accounted for 30,543. We used the nearest neighbour covariate matching without replacement in both models, as it was the most robust and significant. In that way, we managed to match the certified pixels with the same amount of control pixels and extract two datasets. Both datasets consist of the same treated (FSC) pixels. Regarding the control pixels, the first dataset is being matched by static roads, while the second one is by dynamic roads. Finally, we perform pixel and 5-year fixed effects on both datasets in order to examine if the effect of the FSC diverges significantly while the dynamic roads are being included. Last, since Panlasigui et al. (2018) used a greater area of Cameroon's surface, but also implemented their analysis between 2000 and 2013, we expect our results to differ.

# 5. Results

This section discusses the results of the models that have been used for the analysis. In particular, the fixed-effect model is being discussed in the first place, the pooled - OLS model follows, and finally the matched fixed effects outcome.

## 5.1 Fixed Effects

The results from the pixel fixed effects indicate that there is a negative correlation between distance to roads and deforestation. Specifically, Column (1) in table 5.1, points out that an increase of 1km in distance to a road generates a 0.79% decrease in the probability of deforestation. All coefficients in this model are statistically significant. However, these estimations do not account for trends over time and might therefore be biased. As soon as we include 5-year fixed effects, we observe opposite results for the Euclidean distance to roads. According to column (2), for every 1 km increase in distance from roads, the likelihood of deforestation increases by 0.98%.

## 5.2 Pooled OLS

Table 5.1 also illustrates the results from the Pooled OLS regressions. In particular, column (3) represents a pooled OLS containing all the control variables and the distance to the roads from 2001 to 2020 which varies through time, while column (4) includes the distance to 2021 static roads instead. Both models produced similar results and almost the same coefficients. All the coefficients were reported statistically significant and along with the expected results. Specifically, the distance to the roads resulted in almost the same coefficient in comparison to the pixel fixed-effects model.

The dynamic roads model indicated a 0.80 % decrease in the likelihood of deforestation as the distance increases by 1km. While the static roads of 2021 implied a 0.89%. Regarding the rest control variables, we observed the following results (Appendix A, Table A3). As the forest cover percentage increases, the probability of logging activity decreases, which can be explained, as the denser the forest becomes, the more difficult is to access and harvest. The same stands for the elevation, as the forest is located at a higher altitude, the accessibility options diminish. Concerning the distance to cities, again the coefficient is negative indicating that it is less possible to encounter deforestation close to cities, as they lack forest cover. Last, the distance to the water suggests a positive association with deforestation, however, the percentage is quite small.

Table 5.1: Results from fixed effects and ordinary least square (OLS) estimations on deforestation

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | FE_Dynamic | FE_Dynamic_Year | OLS_Dynamic_Roads | OLS_Static_Roads |
| VARIABLES | deforestation | deforestation | deforestation | deforestation |
| Dynamic Roads | -0.0079*** | 0.0098*** | -0.00802*** | |
| | (6.05e-05) | (6.59e-05) | (2.77e-05) | |
| Static Roads | | | | -0.00892*** |
| | | | | (3.36e-05) |
| Constant | 0.164*** | -0.052*** | 0.317*** | 0.314*** |
| | (0.0002066) | (0.000354) | (0.00129) | (0.00129) |
| Observations | 3,648,804 | 3,648,804 | 3,648,804 | 3,648,804 |
| R-squared | 0.0027 | 0.274 | 0.039 | 0.036 |
| Number of ids | 1,000,000 | 1,000,000 | | |
| Control Variables | No | No | Yes | Yes |
| Pixel Fixed Effects | Yes | Yes | - | - |
| 5-Year Fixed Effects | No | Yes | - | - |

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

## 5.3 Matched Fixed Effects

The results from the pixel and 5-year fixed effects are illustrated in Table 5.2. In particular, the first two columns (1) and (2) account for the matched dataset using dynamic roads, while columns (3) and (4) include the 2021 static roads. In terms of the dynamic roads, again the results from pixel fixed effects and 5-year fixed effects differ. On the pixel fixed effects (1) the possession of the FSC certificate seems to increase the probability of deforestation by 5%. However, when we account for time trends (2), the certificate reduces the likelihood of logging by 1%. Both coefficients are statistically significant. We observe the same outcome for the static roads as well. Specifically, in the pixel fixed effects model (3), the deforestation risk is being increased by 4% if an area is certified. On the other hand, the 5-year fixed effects (4) indicate a negative correlation between FSC certification and logging. The ownership of the certificate leads to a decrease of 3% in the likelihood of deforestation. In conclusion, the most important remark of the matched fixed-effects approach is that we identify different outcomes from the dynamic and the static roads' models. Explicitly, when the static roads are being utilized in the analysis, the FSC certificate impact seems to be overestimated. Specifically, the effect is three times larger in comparison to the dynamic road dataset while we control for time trends.

Table 5.2: Results from fixed-effects estimations on FSC after Propensity Score Matching

| | (1)<br>FE_Dynamic | (2)<br>FE_Year_Dynamic | (3)<br>FE_Static | (4)<br>FE_Year_Static |
|---|---|---|---|---|
| VARIABLES | Deforestation | Deforestation | Deforestation | Deforestation |
| FSC | 0.0513*** | -0.0100* | 0.0394*** | -0.0308*** |
| | (0.00277) | (0.00518) | (0.00175) | (0.00451) |
| Constant | 0.0226*** | 0.000995 | 0.0218*** | 0.00358* |
| | (0.00139) | (0.00311) | (0.000874) | (0.00189) |
| Observations | 61,086 | 61,086 | 61,086 | 61,086 |
| Number of ids | 48,665 | 48,665 | 42,346 | 42,346 |
| R-squared | 0.025 | 0.050 | 0.020 | 0.047 |
| Pixel Fixed Effects | Yes | Yes | Yes | Yes |
| 5-Year Fixed Effects | No | Yes | No | Yes |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

## 6. Discussion

The present research used a fixed-effect model to calculate the impact of road proximity on deforestation and two different econometric models to identify the differences between dynamic and static roads. Our findings from the PSM and the fixed effects on the matched sample confirmed our assumption. In particular, following the methodology of Panlasigui et al. (2018), our study suggests that static roads overestimate the impact of the environmental policies, specifically the FSC certificate, in contrast to dynamic roads which implied a smaller effect. While utilizing static road data we cannot control for unobserved time-varying effects, so the FSC impact may be biased. Since the FSC certificate changes over time, it is important to include variables, such as road infrastructure data, that are time-variant. Additionally, our outcome highlights the need for accurate historical road panel data while we examine the effects on the ecosystem and want to tackle the challenges that arise.

Furthermore, on the pooled OLS method, all the results agree with the existing literature. In this approach, we were able to compare the differences between the dynamic and static road effects on deforestation. Both datasets concluded into almost similar results and the outcome is in line with Pfaff et al. (2007), who argue that in a census tract of 100km the closest is a road segment to a forest area the higher the chance of deforestation.

In the fixed-effects model, we encounter inconsistent results. For instance, in the pixel fixed effects model, we observe a negative association between Euclidean distance to roads and the likelihood of deforestation. Whereas, once we control for time trends the results differ, and indicate a positive correlation. The latter might have happened for several reasons: one reason may be that this study did not take into consideration road segments that have been abandoned during the examined period; another reason may be the limitation of the methodology to identify yearly road evolution. Thus, the selection of 5-year fixed effects instead may have affected the outcome. Additionally, the time lag between the road creation and the moment of deforestation may have also affected our results.

Last, significant regions, such as the biggest city, Douala, were excluded from the research. As stated earlier, Cameroon is located in the Congo Basin region, an area that flourishes in rainforests. This means that persistent clouds were observed in the satellite imagery. Thus, it was extremely difficult to extract clear patterns of road segments as the clouds made it almost impossible to detect them. This limitation may have led to potential spatial and selection bias, as a considerable number of roads have not been evaluated. Yet, our methodological approach complements other studies and underlines the relevance of research in the association between infrastructure expansion and environmental degradation.

## 7. Conclusion

This study applied remote sensing and machine learning techniques to produce a panel data set that did not exist until then. Utilizing public access satellite imagery, a road database for Cameroon was derived for the years between 2001 and 2020. The detection accuracy seemed to be quite successful. Despite the challenges, the findings agree with Meijer et al. (2018), who observe a substantial increase in road expansion in developing countries. In Cameroon, 10,441 new road segments were created from 2001 to 2021.

Our main findings suggest that the inclusion of static roads in analyses tends to overestimate the environmental impact, while the dynamic road dataset provides more reliable conclusions. Regarding the impact of road infrastructure on the environment, the results from the 5-year fixed effects estimation indicate that there is a 0.98 % increased probability of deforestation as the distance between the forest and the road increases by 1km. Yet, while we do not control for time trends the effect becomes the opposite.

Analyzing historical observations can help researchers and policymakers to estimate causal effects on the environment, prevent policy mistakes and market failures. As technology evolves, and access to innovative techniques, such as high-resolution satellite images, becomes easier we can improve our tools and techniques in order to make a more accurate analysis of the impact and evaluation of development trends. Despite the fact that this dissertation was limited in geographical scope and had methodological caveats, our results provide evidence on alarming trends in Cameroon and potentially in other developing economies that try to strike the right balance between investment in infrastructure and environmental sustainability. From the safe distance of academic institutions in Europe, our empirical analysis might seem far from the adverse economic reality of Cameroon. Yet, we are a step closer to establishing causal patterns between complex and multidimensional trends. New methods and creative use of data sources can ameliorate policy and practice to protect the environment and achieve sustainable growth.

## References

Barber, C. P., Cochrane, M. A., Souza, C. M., & Laurance, W. F. (2014). Roads, deforestation, and the mitigating effect of protected areas in the Amazon. *Biological Conservation*, *177*, 203–209. https://doi.org/10.1016/j.biocon.2014.07.004

Benítez-López, A., Alkemade, R., Schipper, A. M., Ingram, D. J., Verweij, P. A., Eikelboom, J. A., & Huijbregts, M. A. (2017). The impact of hunting on tropical mammal and bird populations. *Science*, *356*(6334), 180–183. https://doi.org/10.1126/science.aaj1891

Breiman, L. (2001). *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/a:1010933404324

Breiman, L. (2017). *Classification and regression trees*. CRC Press.

Burke, M., Driscoll, A., Lobell, D. B., & Ermon, S. (2021). Using satellite imagery to understand and promote sustainable development. *Science*, *371*(6535). https://doi.org/10.1126/science.abe8628

FAO. (2000). On definitions of forest and forest change. FRA Working Paper No. 33. Rome.

FAO. (2010). Global forest resources assessment 2010: main report. Rome.

FAO. (2018). Forests pathways to sustainable development. Rome: FAO.

Foody, G. M. (2002). Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, *80*(1), 185–201. https://doi.org/10.1016/s0034-4257(01)00295-4

Forest Stewardship Council. (2020). FSC Facts & Figure. FSC. Bonn, Germany. http://info.fsc.org/. 12pp.

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S. J., Loveland, T. R., Kommareddy, A., Egorov, A., Chini, L., Justice, C. O., & Townshend, J. R. (2013). High-resolution global maps of 21st-century forest cover change. *Science*, *342*(6160), 850–853. https://doi.org/10.1126/science.1244693

Hijmans, R.J., S.E. Cameron, J.L. Parra, P.G. Jones & A. Jarvis. (2005). Very High Resolution Interpolated Climate Surfaces for Global Land Areas. *International Journal of Climatology 25: 1965-1978.* doi:10.1002/joc.1276.

Jacoby, H. G. (2000). Access to markets and the benefits of rural roads. *The Economic Journal*, *110*(465), 713–737. https://doi.org/10.1111/1468-0297.00562

Jarvis, A., H.I. Reuter, A. Nelson, E. Guevara. (2008). Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m Database: https://srtm.csi.cgiar.org.

Kleinschroth, F., Laporte, N., Laurance, W. F., Goetz, S. J., & Ghazoul, J. (2019). Road expansion and persistence in forests of the Congo Basin. *Nature Sustainability*, *2*(7), 628–634. https://doi.org/10.1038/s41893-019-0310-6

Laporte, N. T., Stabach, J. A., Grosch, R., Lin, T. S., & Goetz, S. J. (2007). Expansion of industrial logging in Central Africa. *Science*, *316*(5830), 1451–1451. https://doi.org/10.1126/science.1141057

Leblois, A., Damette, O., & Wolfersberger, J. (2017). What has driven deforestation in developing countries since the 2000s? evidence from New Remote-sensing data. *World Development*, *92*, 82–102. https://doi.org/10.1016/j.worlddev.2016.11.012

Meijer, J. R., Huijbregts, M. A., Schotten, K. C., & Schipper, A. M. (2018). Global patterns of current and Future Road Infrastructure. *Environmental Research Letters*, *13*(6), 064006. https://doi.org/10.1088/1748-9326/aabd42

Najman, B., Gachassin, M., & Raballand, G. (2010). The impact of roads on poverty reduction : A case study of cameroon. *Policy Research Working Papers. The World Bank Africa Region Transport Unit*. https://doi.org/10.1596/1813-9450-5209

Panlasigui, S., Rico-Straffon, J., Pfaff, A., Swenson, J., & Loucks, C. (2018). Impacts of certification, uncertified concessions, and protected areas on forest loss in Cameroon, 2000 to 2013. *Biological Conservation*, *227*, 160–166. https://doi.org/10.1016/j.biocon.2018.09.013

Pesaresi, M., Florczyk, A., Schiavina, M., Melchiorri, M., & Maffenini, L. (2019, June 28). *GHS-SMOD R2019A - ghs settlement layers, updated and Refined Regio Model 2014 in application to GHS-built r2018a and GHS-pop R2019A, multitemporal (1975-1990-2000-2015)*. Joint Research Centre Data Catalogue - GHS-SMOD R2019A - GHS settlement layers, updated a... - European Commission. Retrieved September 7, 2021, from https://data.europa.eu/89h/42e8be89-54ff-464e-be7b-bf9e64da5218

Pfaff, A., Robalino, J., Walker, R., Aldrich, S., Caldas, M., Reis, E., Perz, S., Bohrer, C., Arima, E., Laurance, W., & Kirby, K. (2007). Road Investments, spatial spillovers, and deforestation in the Brazilian Amazon. *Journal of Regional Science*, *47*(1), 109–123. https://doi.org/10.1111/j.1467-9787.2007.00502.x

Schwab, V. F., Garcin, Y., Sachse, D., Todou, G., Séné, O., Onana, J.-M., Achoundong, G., & Gleixner, G. (2015). Effect of aridity on δ 13 C and ΔD values of c 3 plant- and C 4 graminoid-derived leaf wax lipids from soils along an environmental gradient in Cameroon (Western Central Africa). *Organic Geochemistry*, *78*, 99–109. https://doi.org/10.1016/j.orggeochem.2014.09.007

Yeh, C., Perez, A., Driscoll, A., Azzari, G., Tang, Z., Lobell, D., Ermon, S., & Burke, M. (2020). Using publicly available satellite imagery and deep learning to understand economic well-being in Africa. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-16185-w

WRI. (2021, June). Cameroon | Dashboards | Global Forest Watch. Retrieved June 20, 2021, from Global Forest Warch website: https://www.globalforestwatch.org/dashboards/country/CMR?widget=treeLossTsc&treeLossTsc=eyJoaWdobGlnaHRlZCI6ZmFsc2V9

## Appendix A

Table A1. Average characteristics for matched controls using Propensity Score Matching with and without replacement. Moreover, comparison of land average characteristics of treatment and control groups for FSC pixels using dynamic roads.

| Characteristics | Treated (FSC) | All Control | PSM with replacement | PSM without replacement |
|---|---|---|---|---|
| Distance to dynamic roads | 4.40 | 4.07*** | 4.92*** | 4.88*** |
| Distance to cities | 23.85 | 18.05*** | 24.07** | 23.98 |
| Elevation | 624.21 | 601.4*** | 634.83*** | 635.7*** |
| Precipitation | 1684.1 | 1693.6*** | 1662.8*** | 1662.4*** |
| Slope | 4.082 | 3.976*** | 4.04** | 4.05 |
| East | 0.63 | 0.59*** | 0.67*** | 0.67*** |
| South | 0.36 | 0.40*** | 0.32*** | 0.32*** |
| Mean Bias | | 13.1 | 6.8 | 6.5 |

** Note:* *p<0.1, **p<0.05, ***p<0.01

Table A2. Average characteristics for matched controls using Propensity Score Matching with and without replacement. Moreover, comparison of land average characteristics of treatment and control groups for FSC pixels using 2021 static roads.

| Characteristics | Treated (FSC) | All Control | PSM with replacement | PSM without replacement |
|---|---|---|---|---|
| Distance to static roads | 3.18 | 3.09*** | 3.24** | 3.27*** |
| Distance to cities | 23.85 | 18.05*** | 23.85 | 23.86 |
| Elevation | 624.21 | 601.4*** | 631.52*** | 632.79*** |
| Precipitation | 1684.1 | 1693.6*** | 1669.8*** | 1667.5*** |
| Slope | 4.082 | 3.97*** | 4.03** | 4.03** |
| East | 0.63 | 0.59*** | 0.63 | 0.64** |
| South | 0.37 | 0.40*** | 0.36 | 0.36** |
| Mean Bias | | 12.4 | 2.4 | 3 |

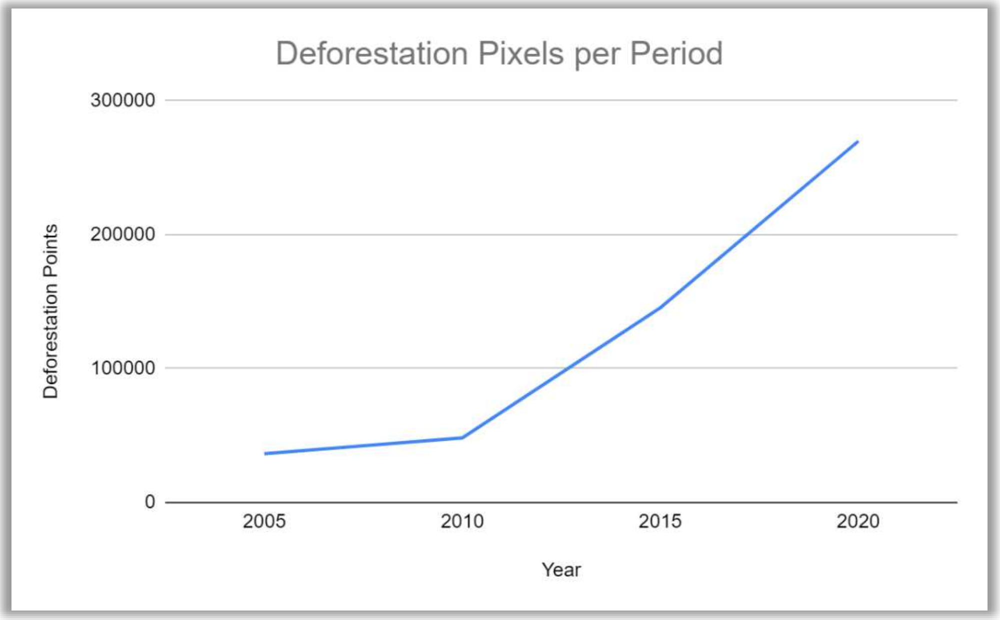
** Note:* *p<0.1, **p<0.05, ***p<0.01

Table A3. Pooled OLS Regressions with control variables

| | (1) | (2) |
|---|---|---|
| | OLS_Dynamic_Roads | OLS_Static_Roads |
| VARIABLES | Deforestation | deforestation |
| | | |
| Dynamic Roads | -0.00803*** | |
| | (2.77e-05) | |
| Static Roads | | -0.00892*** |
| | | (3.36e-05) |
| FSC | -0.0673*** | -0.0670*** |
| | (0.000616) | (0.000618) |
| Forest Cover | -0.00115*** | -0.00115*** |
| | (1.25e-05) | (1.26e-05) |
| Water Proximity | 0.000468*** | 0.000451*** |
| | (7.15e-06) | (7.19e-06) |
| Cities Proximity | -0.00186*** | -0.00219*** |
| | (1.21e-05) | (1.18e-05) |
| Elevation | -8.26e-05*** | -7.54e-05*** |
| | (1.37e-06) | (1.38e-06) |
| | | |
| Constant | 0.317*** | 0.314*** |
| | (0.00129) | (0.00129) |
| Observations | 3,648,804 | 3,648,804 |
| R-squared | 0.039 | 0.036 |

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1



,

Figure A1: Deforestation evolution in Cameroon over time

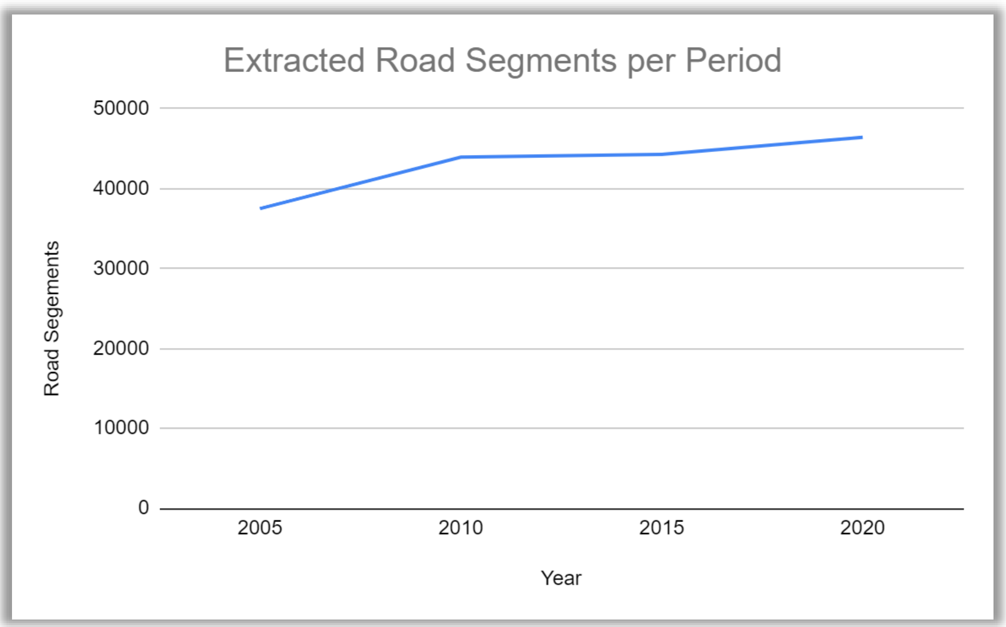



Figure A2: The number of detected roads on each time period