Analysis of Colombian deforestation, main agents and its relation to conflict associated variables.

[redacted]

December 12, 2019

Contents

1 Overview

The purpose of this analysis is to identify the effects of the post-conflict over deforestation status in Colombia. Around the world has been an increase in the deforestation level, however the biggest deforestation it's present in the countries located in the areas of tropical rainforest, and 40% of these countries present an internal conflict situation, this gives us a special research context for the relationship between deforestation and conflict variables.

We have selected several variables , some of them strongly related with the conflict analysis like coca crops density and civil attacks and control variables independent of the conflict context like weather, population growth and other crops density.

Our research variable it's the deforestation status, obtained as the difference in tree cover density between several satelital images, in our case and according to the general standard we classify an area as deforested if initially the tree cover density percentage was over 30% of his area and this percentage has reduced during the analysis period.

This analysis pursuits to identify if there's a connection between the country deforestation status and the peace agreement, recently declared. In other countries like Nepal, Sri Lanka, Ivory Coast and Peru, one of the unintended effects of the end of conflict was the loss of forest cover. Studies about those countries show an average 68.08% increase of annual forest loss in the five years following the end of conflict, while the worldwide mean is 7.20%.

Research shows that armed conflicts between government and rebel groups, allowed large regions under control of rebel groups to be inaccessible through the conflict. In turn, this inaccessibility allowed the region to grow as a biodiversity hotspot.

Once the conflict ends, a variety of factors accelerate the deforestation process, from the need of resources to reconstruction efforts to weak government and instability that allow uncontrolled exploitation.

This project seeks to find what happened in Colombia in the aftermath of the peace agreement between the government and the FARC-EP insurgent group, by modelling the forest coverage data from 2003 and 2018 against other variables related to the conflict. [Sin19]

1.1 Objective

Show what are the main factors (between the selected topics) related with the deforestation in Colombia, doing a comparison between the war and the post-conflict period. Once these factors are identified it's easier to focus on the alternative solutions.

2 Materials and Methods

2.1 Data

We use Colombian mainland departmental boundaries as the spatial unit of analysis. The panel dataset used consists of annual departmental observations from 2003 to 2018 (inclusive). Colombia has a total of 32 departments.

– **Tree Cover Loss:**

  The forest cover loss data shows the number of hectares in the department (year-by-year) which were classified as "tree loss" by [HPM$^+$13] dataset, which is based on a collaboration of the University of Maryland, Google, USGS, and NASA, and uses Landsat satellite images to map annual tree cover loss at a 30×30 meter resolution. The classification was done adjusting the minimum tree cover canopy (TCC) density for the visualization and analysis of the data to 30%.

  TCC density represents the estimated per cent of a pixel that was covered by tree canopy in the year 2000, as determined from the analysis of satellite imagery. For the tree cover loss data, TCC density therefore corresponds to the density of tree cover before loss occurred. For example,we selected 30% *as the minimum TCC density, so our information is the tree cover loss pixels for which the original tree cover density was greater than* 30%*.*

– **Conflict:**

  In order to examine the relationship between conflict and deforestation, we collected data about armed conflict carried out by the Revolutionary Armed Forces of Colombia (FARC) or the National Liberation Army (ELN) between 2003 and 2018 in each department. Armed actions were defined as the number of massacres, kidnappings, terrorist attacks and anti-personnel mine events. Also, we used data about the number of victims of this armed actions. Similarly, the variable coca area were used, and it's defined as the cultivated area (in hectares). Also, we collected data about the demobilized population. In general, conflict data were downloaded from: Open Data of Colombia, and Observatory of Memory and Conflict.

– **Crops:**

  We recorded crop statistics for 21 products from Food and Agriculture Organization of the United Nations (FAO); data is expressed in terms of harvested area. The objective is to comprehensively cover production of the crops for all the departments in Colombia. The selected crops for the analysis are: avocado, rice, pea, sugar, banana, coffee, coca, rubber, common bean, maize, mango, oil palm, potato, papaya, soy, tomato, tree tomato, cassava and carrot. We made a special emphasis on the coca crops density across the years.

– **Accessibility:**

  Finally, we collected variables related to the accessibility in each department. The first one is called: 'accessibility' and it's related to the average distance of each location in the department to a principal road. The second one is "distance to human settlements", which is the average distance of each location of the department to a principal city. The third one is defined and the average distance of each point to major rivers in the department.

3 Exploratory Analysis

Over the 2003-2018 period, on average Colombia has lost approximately 7181.54 hectares of tree cover. At the same period, the hectares of coca's harvested has been on average 3072 in the country. Most of the conflict intensity is attributed to anti-personnel mines, kidnapping and recruitment, which have left on average 68, 19 and 9 victims respectively per department and per year.

Correlations between the conflict variables and tree cover loss are not too strong (see Fig. 1), but there are some of them that influence the most to the response: hectares of cultivated coca (34%), the number of victims of anti-personnel landmines (51%), the number of victims of slaughter (16%) or murders, victims of recruitment (32%), and number of kidnapped people (16%) and kidnapping cases (26%).

– **Analysis by deparment:** San Andres has lost less tree cover over the years than the other departments (see Fig. 2); this might seem obvious because of their own features like island, like size and it's fertile land availability, however it's also the most distant department and the less directly affected by the conflict, and this might have impacted.

  Antioquia, Caqueta, Guaviare and Meta are the ones that have lost the most tree cover. Atlantico, Quindio and Risaralda didn't lose more than 10,000 hectares on average, over the years.

– **Analysis by year.** As our objective is to see how the tree cover loss has been over the years, we constructed a time series to visualize better possible patterns in the variable:

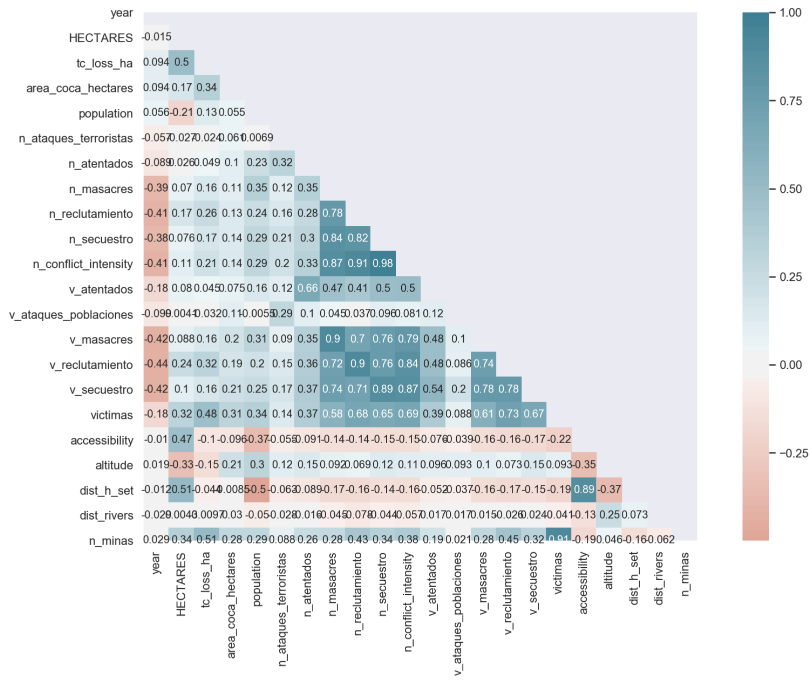

Figure 1: Correlation matrix

And from figure 3, it's possible to see that the loss of Colombian forest has been exponentially increasing from 2015, which is very interesting because in that year, although there were some violent actions in which military forces and FARC were involved, during this specific time period the government and FARC were in a negotiation stage in la Habana.

The final agreement on victims of the decades-long conflict was reached, offering mutually agreed-upon punishments for perpetrators of violence and remunerations for those affected.

But, why is this information important? because conflict may help preserve ecosystems, if the armed groups use forest coverage to hide by dissuading extractive economic activities through extortion, or don't letting people visit natural parks or forests because of the fear of being kidnapped or attacked [FL03]. So, it makes sense that after the agreements, people return to the rural zones of the country, and also some organizations started to make expeditions or working on the forest.

On the other hand, conflict can increase deforestation in function of their economic activities to support their activities, in the specific case

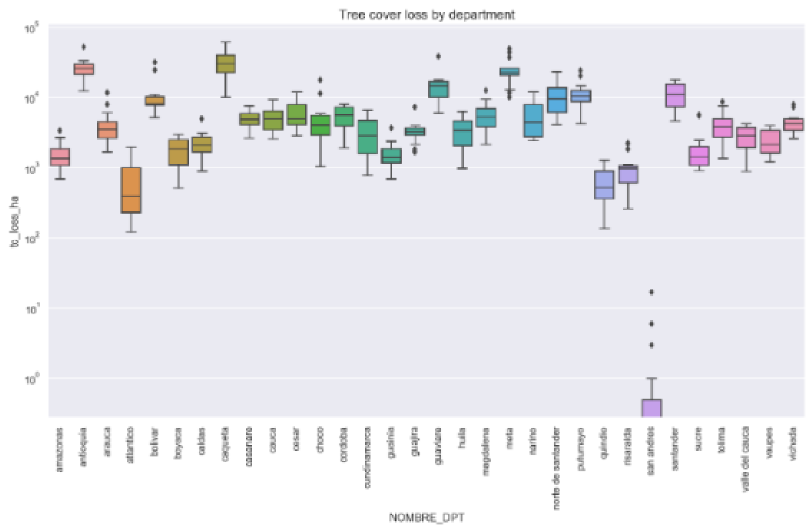

Figure 2: Boxplot of tree cover loss by department

of Colombia, the area for illegal crops harvesting, such as coca or the exploitation of natural resources, because these activities are very land-intensive.

– **Analysis by year and department:**

From the figure 4, we realized that in most of the departments, the tree cover loss increased from 2015, period in which the government and FARC were in a negotiation stage; it would means that the peace process made the tree cover decreased. But it's not directly related to the process itself, but it's related to the fact that people started to feel more secure and free to visit natural places, like forests, and also big international and national companies were allowed to explore and exploit natural resources without any obstacle.

In particular, Caqueta, Meta, Antioquia, Guaviare, Santander and Narino have had tree cover loss hectares greater than the national average of loss; this is interesting because one of these departments has a huge number of inhabitants (Antioquia), and the other ones have been affected by the coca crop cultivation.

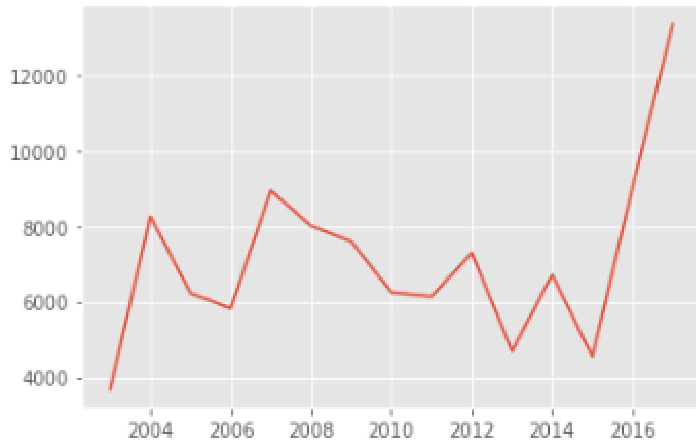

Figure 3: Tree cover loss over the 16 years

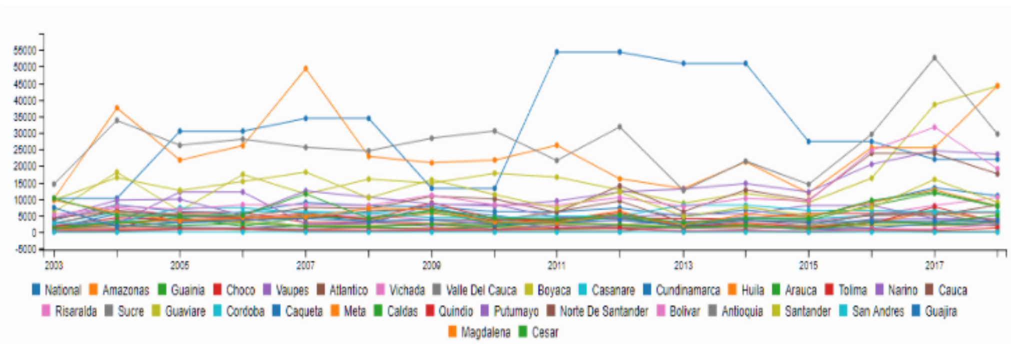

Figure 4: Tree cover loss over the time by department

4 The model

4.1 Using Conflict Data

As we have data about the number of hectares that have been classified as 'tree cover loss', we decided to model this response variable by a Poisson regression model because that is a count data. Then, because the model estimations are overdispersed, we fit a Negative Binomial regression, known as NB2, which is a generalization of the Poisson regression but it loosens the highly restrictive assumption that the variance is equal to the mean made by the Poisson model. The negative binomial regression model, is based on the Poisson-gamma mixture distribution, and is popular because it models the Poisson heterogeneity with a gamma distribution.

The Poisson probability distribution is given as:

$$f(y, \lambda) = \frac{e^{-\lambda}\lambda^{y}}{y!}.$$

However, almost all real count data are in fact over-dispersed. Therefore the negative binomial model has become a central model in the evaluation of count data. The negative binomial probability function is expressed in a variety of ways. The definition of the distribution is given as:

$$f(y; \lambda, \alpha) = \binom{y + \frac{1}{\alpha} - 1}{\frac{1}{\alpha} - 1}\left(\frac{1}{1 + \alpha\lambda}\right)^{\frac{1}{\alpha}}\left(\frac{\alpha\lambda}{1 + \alpha\lambda}\right)^{y}.$$

Thus, we computed a Poisson regression model and then an NB2 for its estimations. we excluded the number of recruitment cases and victims. We also assessed the accuracy of the model by looking at its Log-Likelihood, deviance and the Pearson-Chi square.

Additionally, we fit the models excluding the year by looking at the average of tree cover loss for each department through the time period. For a new model, we excluded the departments by looking at the average tree cover loss for each year in Colombia.

We computed a linear regression model to compare its performance with the other ones, and constructed the time series of the average tree cover loss in Colombia.

In order to assess the risk of each department of having a tree cover loss above the median loss for each one, we fit a Random Forest classifier which predicted the probability that a department was in one of those two categories: 1 means equal or above the median loss of the department, 0 means below the median loss, in 2019 and 2020. For doing that, we simulated the values of the variables for each department in those years as:

1. We extracted a 'with replacement' sample of the target variable 1000 times.
2. We calculated the mean in each of the 1000 samples.
3. We constructed a 95% confidence interval of the mean.
4. We generated numbers between the upper and lower value of the interval
5. We extracted a sample of the numbers and calculated the mean

Finally, with the value's variables we classified the departments in the categories defined above and then predicted the risk according to the results of the trained Random Forest model.

4.2 Using conflict and crop data:

Initially, to grant the condition of independent variables, we identified the variables presenting multicollinearity by 2 methods: PCA(Principal Component Analysis) and VIF(Variance Inflation Factor). In the analysis some of the variables that were removed because of the multicollinearity were: The mean and max temperature (keeping the min temperature), and some crops of relatively low incidence like some fruits and tubers that probably grow in similar areas explaining the relationship.

We performed here a linear regression model and a decision tree to assess the importance of the variables

5 Results

5.1 Model with conflict data

5.1.1 Department and Year

- **Poisson Model:**

  In the Poisson model, all the variables were significant and the performance was good. Now, checking the dispersion of the estimated parameters, it looks very over-dispersed over time:

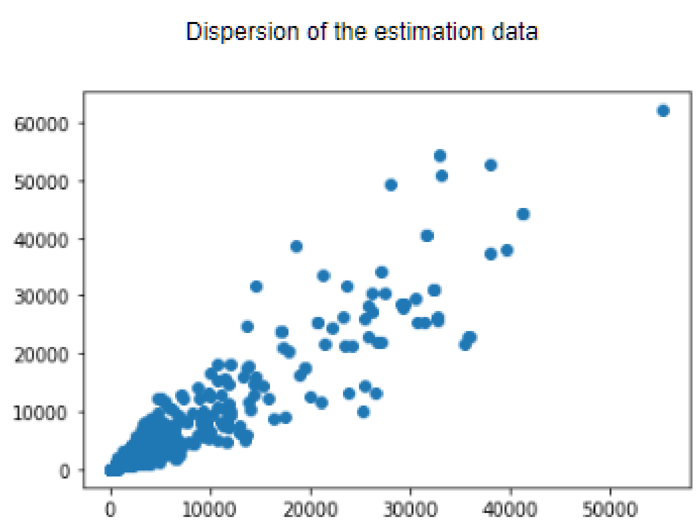

Figure 5: Scatter plot of the Poisson parameter's estimation

– **Negative binomial model:**

So now, fitting the negative binomial model we got better performance, however, the area of the department in hectares, number of attack's cases and victims, massacre's cases and victims, and kidnapping cases and victims, are not significative.

Predicting the testing data, we got a good accuracy:

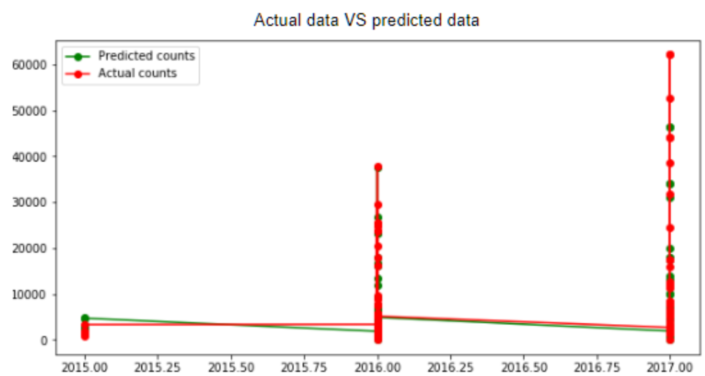

Figure 6: Accuracy of the Negative binomial model

– **Linear Regression Model:**

The model reached an R-square of 80.2% and an AIC of 9631, which is the minimum with the variables we have. Maybe, for future work, it would be better to group the departments in political regions of the country and repeat the models. It's possible to see that most of the conflict variables increase the average tree cover loss when they increase too, and the variables that affect the most this response are the number of attack's cases and victims, where particularly the average tree cover loss increases 664.32 hectares for each new attack, and 656.42 for each attack to the population, and it seems like as long as there are more victims the tree cover loss decreases, it means that if there are less people in the department, the forests are better conserved. Additionally, it's interesting that for each new hectare of cultivated coca the average tree cover loss increases 0.25 hectares. About the year, we noticed that every new year the response variable increase 254.21 hectares, it means that the tree cover in every Colombia's department has been increasing and will continue doing it. Finally, for each new anti-personnel landmine victim, the average tree cover loss decrease 22.21 hectares, and that's interesting because mines are supposed to hide in the woods, so if one of them explode, the trees should affect, but this is not the case, so our guest is that these mines are mostly improvised or handcrafted and don't cause major damage to the tree cover in the zone.

Checking at the accessibility variables, we noticed that as long as the forests are near roads, the average tree cover loss increases 31.23 hectares, and that's because is easier to access them. Also, if the forests are near major rivers and cities or human settlements, the average tree cover loss decreases a lot, and we suppose that's because the cities that are near are not too crowded, because even the population size is a factor that decreases, not too much, the tree cover loss.

Variable	Coefficient
Intercept	- 390626.28
Antioquia	34700.28
Arauca	13618.36
Atlantico	- 86051.82
Bolivar	- 25773.66
Boyaca	21398.07
Caldas	- 29870.36
Caqueta	16583.3
Casanare	- 24972.30
Cauca	- 19091.52
Cesar	- 25553.85
Choco	- 62297.97
Cordoba	- 12357.06
Cundinamarca	- 10768.82
Guainia	- 28280.25
Guajira	- 25227.69
Guaviare	- 37133.76
Huila	- 15338.77
Magdalena	- 101460.07
Meta	- 31037.62
Narino	33190.45
Norte De Santander	47461.05
Putumayo	- 21966.83
Quindio	- 18184.06

Variable	Coefficient
Risaralda	- 18252.50
San Andres	- 26881.23
Santander	41420.87
Sucre	- 37169.79
Tolima	- 38635.53
Valle Del Cauca	- 12922.03
Vaupes	13692.17
Vichada	15645.32
Year	254.21
Hectares	- 0.00
Area_Coca_Hectares	0.25
Population	- 0.0087
N_Ataques_Terroristas	656.42
N_Atentados	664.32
N_Masacres	- 44.75
N_Secuestro	- 6.73
V_Atentados	- 594.04
V_Ataques_Poblaciones	110.2
V_Masacres	- 16.80
V_Secuestro	11.14
Accessibility	31.23
Altitude	- 27.77
Dist_H_Set	- 36691.69
Dist_Rivers	- 14742.25
N_Minas	- 22.21

– **Time Series**

We realized that the tree cover loss in Colombia across those years follows the same trend that the departments which has lost the most tree cover

over the period. It's interesting to see that in 2012, when the negotiations started and a ceasefire was agreed, the average tree cover of Colombia was the greatest since the last 2 and half years.

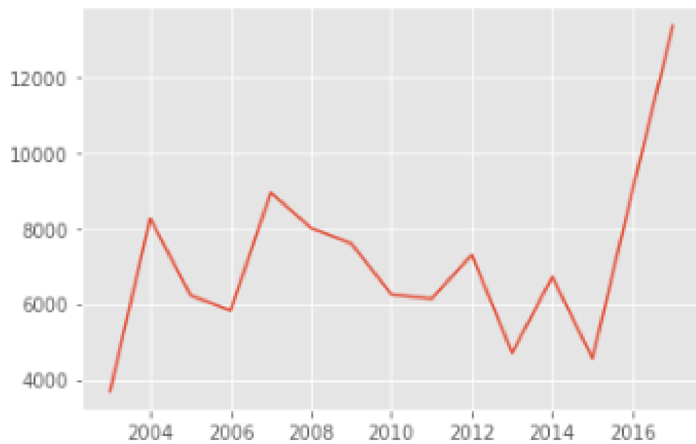

Figure 7: Tree cover loss over time

- **Risk predictions**

  The next, are the Random Forest predictions if tree cover loss in 2019 and 2020 and the probability of each department to be in the category **1** (above the tree cover loss median) or **0** (below):

Department	**2019**	**2020**
Amazonas	77%	63%
Antioquia	76%	56%
Arauca	54%	40%
Atlantico	60%	39%
Bolivar	58%	41%
Boyaca	56%	39%
Caldas	64%	57%
Caqueta	59%	57%
Casanare	51%	54%
Cauca	50%	43%
Cesar	42%	50%
Choco	51%	49%
Cordoba	38%	52%
Department	**2019**	**2020**
Cundinamarca	42%	48%
Guainia	62%	64%
Guajira	62%	50%

Guaviare	56%	50%
Huila	46%	51%
Magdalena	61%	89%
Meta	53%	97%
Narino	47%	39%
Norte De Santander	49%	40%
Putumayo	46%	49%
Quindio	38%	43%
Risaralda	58%	51%
San Andres	52%	52%
Santander	43%	50%
Sucre	40%	49%
Tolima	59%	48%
Valle Del Cauca	65%	52%
Vaupes	58%	63%
Vichada	67%	57%

It's possible to see that the probability will decrease in 2020, and that's because the average tree cover loss will increase at that time 508.42 hectares approximately (according to the linear model), compared with 2018 and the tree cover loss won't be significantly different, so the number of departments with a tree cover loss over the median will be less.

Finally, computing the variable importance for doing the classification in the model, we identified that the top 6 most important variables were: year, population, number of anti-personnel landmine cases, victims and hectares of cultivated coca. We must clarify that victims is the sum of kidnapping, massacres and attacks victims in each department and each year. The less important is the department, so it suggests that the tree cover loss.

5.2 Model with conflict and crop data

Linear Regression:

Fitting a multiple linear regression, in order to check the relations between the variables and the tree cover loss, we found that each 1-degree increase in the temperature multiplies the expected value of the tree cover loss by 2.8 hectares approximately, and if the increase were in the tons of mine extraction, the expected value would increase 2.7 hectares.

```
Log(Tree Cover Loss + 1) =  0.02841 Min_temp + 0.0000685 Mine_extraction (Ton)
+ 0.0002 Area_Coca_Hectares + 0.0000026 Population + 0.8905 N_terrorist_attacks
- 0.1483 N_attacks_population - 0.1727 N_Masacres + 0.1366 N_recruitment
+ 0.00002 Avocado + 0.000017 Rice + 0.0000587 Cocoa - 0.000006855 Sugar_cane
- 0.00001416 Forage_corn + 0.00001681 Oil_palm - 0.00006 Papaya
+ 0.000004566 Pineapple - 0.00001 Banana
```

We applied a logaritmic transformation to the target variable, it was suggested by the Box-Cox coefficient. The model explained 73% of the variability of the logritm of tree cover loss

6 Conclusions

With this study, we figured out some characteristics of the relation between deforestation and armed conflict in Colombia.

Our findings suggest that Colombia's armed conflict affect department's tree cover mostly in a positive way because some of the most affected areas by conflict presented a lower level of deforestation against the national average, and one of the reasons why it happens is that when the armed groups inhabited the forests and jungles of the country, nobody dared to go inside, so they meant a kind of "protection" for the biodiversity of each department. Particularly, as long as there are more murders or kidnappings, the forests lose less tree cover, which would mean that population is one of the principal causes of deforestation; however, Valle del Cauca has had one of the highest population averages through the years and it's not the department with the highest tree cover loss, so we think that the tree cover loss decrease when there are more people afraid of being killed or victims of any other class of violence by the armed groups.

Now, we know that during the conflict period, the groups like FARC-EP, ELN and other illegal armed groups had great extensions of cultivated coca in the forests, it affected the tree cover and caused the cover loss increasing per cultivated hectare. One interesting thing is that during the 2003-2018 period, there were no reports of coca crops in Casanare despite the fact the close departments has a strong presence, we haven't figured out why was it happen. In the departments with the highest cultivation of coca, the tree cover loss increased at the same time.

There are more factors that increased the average tree cover loss in each department, for example, the easy access to the forest by a principal road, it means that the further the forest is from the road, the fewer people can access it; however, given that the human settlements closest to the forests are not too crowded, the tree cover is not significatively affected by this fact because probably people in the towns are mostly farmers.

For 2019 and 2020 is highly probable that Amazonas suffers a great tree cover loss even though the biodiversity had been starting to grow up in the last 2 years. Another department with a high probability of losing great hectares of tree cover is Antioquia, and most of the departments have a risk above 50% of great loss. Comparing these years, we noticed that the risk could decrease for 2020, however, in departments like Magdalena, Meta, Putumayo, Quindío, Sucre, Vaupes, Huila, Cundinamarca, Cordoba, Cesar y Casanare, the risk of having a loss above the median is very high.

One of the main conclusions we have is that since the peace agreements were signed, the tree cover in different Colombia's departments has increased in alarming quantities.

When we reviewed the analysis using variables like other crops and mining extraction, we found a strong impact on the deforestation, following the next order:

1. Mining extraction
2. Petroleum production
3. Normal crops:
   - Avocado
   - Rice
   - Cocoa
   - Sugar cane
   - Corn
   - Oil palm
   - Banana crop

The effects resulting from the mining exploitation probably does not seem unexpected, however, the increase in other crops, with exportation purposes, some of those extremely media discussed like the avocado are also impacting negatively in the deforestation in our country.

7 Dashboard and backend implementation

The backend is a set of web services (API) that allow the front-end to access the data to build and present charts.

The API allow those queries:

- URL: **api/departments** returns a JSON file that contains data and the geometry of each department.
- URL: **api/metrics** returns the values of all metrics in the database.

– URL: **api/metrics/<m>** returns the values for a specific metric <m>, this metric can be: $aguacate_{c}rop$, $arroz_{c}rop$, $arveja_{c}rop$, $attempt$, $cacao_{c}rop$, $cafe_{c}rop$, $caucho_{c}rop$, $caa_{c}rop$, $coca_{c}rop$, $frijol_{c}rop$, $maiz_{c}rop$, $maiz_{f}orrajero_{c}rop$, $malanga_{c}rop$, $mango_{c}rop$, $palma_{d}e_{a}ceite_{c}rop$, $papa_{c}rop$, $papaya_{c}rop$, $pia_{c}rop$, $platano_{c}rop$, $soya_{c}rop$, $terrorist_{a}ttack$, $tomate_{c}rop$, $tomate_{d}e_{a}rbol_{c}rop$, $yuca_{c}rop$, $zanahoria_{c}rop$, $conflict_{i}ntensity$, $displacement$, $kidnapping$, $mining$, $petroleum$, $population$, $recruitment$, $slaughter$, $tree_{c}over$ and $weather$.

– URL: **api/metrics/<m>/<y>** returns the values for a specific metric <m>, for that year <y>. This API is built using Ruby on Rails and deploy on AWS EC2 using nginx linked to the RoR App through a socket with Unicorn (a ruby gem)

References

[CNA17] Sossa C. Castro-Nunez A., Mertz O., *Geographic overlaps between priority areas for forest carbon-storage efforts and those for delivering peacebuilding programs: implications for policy design.*, Environ. Res. (2017).

[FL03] James D. Fearon and David D. Laitin, *Ethnicity, insurgency, and civil war*, The American Political Science Review **97** (2003), no. 1, 75–90.

[Goe08] Juanita Goebertus, *Oil palm and forced displacement in zona bananera: "pathways" between natural resources and conflict*, Colombia Internacional **67** (2008), 152–175.

[HPM+13] M. C. Hansen, P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, S. V. Stehman, S. J. Goetz, T. R. Loveland, A. Kommareddy, A. Egorov, L. Chini, C. O. Justice, and J. R. G. Townshend, *High-resolution global maps of 21st-century forest cover change*, Science **342** (2013), no. 6160, 850–853.

[JdW18] Matt Piotrowski Joeri de Wilde, Tim Steinweg, *Deforestation risk in colombia: Beef and dairy sectors may expose investors*, Chain reaction research (2018).

[MHK16] Lamin Mansaray, Jing-feng Huang, and Alimamy Kamara, *Mapping deforestation and urban expansion in freetown, sierra leone, from pre- to post-war economic recovery*, Environmental Monitoring and Assessment **188** (2016).

[PSV18] M Prem, S Saavedra, and J.F Vargas, *End-Of-Conflict Deforestation: Evidence From Colombian's Peace Agreement*, Documentos de Trabajo 017068, Universidad del Rosario, December 2018.

[Sin19] Nelson Grima; Simron Jit Singh, *How the end of armed conflicts influence forest cover and subsequently ecosystem services provision? an analysis of four case studies in biodiversity hotspots*, Land use policy. The International Journal Covering All Aspects of Land Use (2019), no. 81, 267–275.