
Quantify the deforestation in Amazon forest by using Machine Learning Algorithm

Baptiste Poffet ¹

Abstract

Deforestation has been an important topic for many years, especially in the Amazon forest, the biggest in the world. Amazon forest shelters hundreds of species of animals and trees. This report will focus on the Bolivian part of the Amazon forest and the analysis will go from 2016 and 2021. We will use satellite data and classify the land cover based on supervised classification. The results show that deforestation is still an actual problem. Our results show the area of forest lost between 2016 and 2021.

1. Introduction

As the world is trying to deal with climate change, overpopulation, and many other worldwide problems, trees and forests are key to trying to reduce these impacts. Purely, deforestation means the purposeful clearing of forested land [2]. Deforestation is accentuated by many factors including politics and economics. As explained by Meyerfeld [3], Jair Bolsonaro (President of Brazil since the 1st January 2019) maintains the high level of deforestation started a long time ago. To give an example, the forest loss in the Brazilian Amazon forest from August 2019 to July 2020 is around 11'000 km² [3]. As a comparison, 11'000 km² is more than a quarter of the area of Switzerland.

Deforestation is a big issue for population, animals and of course... flora. As said by Nunez [4], 80% of land animals and plants are living in forests. Deforestation threatens many species of animals, birds, and plants. Loss of trees causes loss of canopy which preserves stable temperature during nights and days which is important for plants and animals.

Amazon forest countries helped by governments made a lot of efforts in the last 15 years to reduce deforestation. As we can see in Figure 1, the level of deforestation in the Amazon forest decreased from 2004 to 2011, then remained stable from 2011 to 2015 and now the tendency is to increase again

¹Master in Spatial Analysis and Complex Systems, University of Lausanne, Switzerland.

and accelerate since early 2019.

In this work, we are using satellite images from two dif-

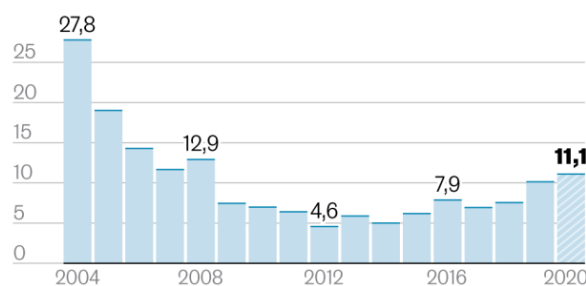


Figure 1. Evolution of deforestation in Amazon forest, in thousands of km². (Source :[3])

ferent satellites (Landsat 8 and Sentinel 2) to perform a supervised classification of the land cover in two classes and be able to quantify the land cover transformation and especially the area of forest loss. To do it we first downloaded the images from the satellite databases. We trained our algorithm on a supervised data set, and as the results were satisfying we used this algorithm to classify the pixels of our images into two classes: Forest or Non-forest. After the classification was done for our two years of study (2016 and 2021), we calculated the area of forest loss and the forest gain, then we created some maps to see it better than only with numbers. All these steps will be explained longer in the next sections.

2. Methodology

2.1. Study Area

To see deforestation, the area of interest has been defined in Bolivia close to the city of Santa Rosa in the department of Santa Cruz. Santa Cruz is known as a state impacted by deforestation. Our area of interest is around 17'000 km² and as you can see in Figure 2 it's mainly covered by forest but close to a small village. This area looked particularly interesting as we might observe (or not) that the village grows and the forest surface is accordingly changed. This would be less visible with an area of interest located in the middle of the forest.

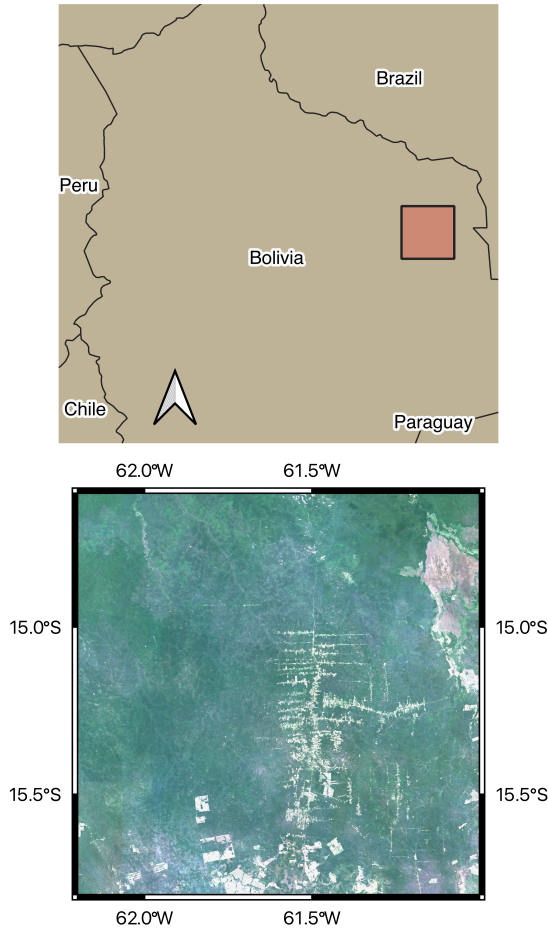


Figure 2. Area of study, around 17'000 km².

2.2. Data

To be able to perform a good analysis we needed good images. To get good data we used Google Earth Engine and used the images of Landsat 8 and Sentinel 2 collections. The data used are images from 2016 (Landsat 8) and 2021 (Sentinel 2). We decided to use both satellite collections because Sentinel 2 did not provide images before 2017 for our area of interest and it was more relevant to have more than 2 years between the two points of comparison. So we decided to use the collection of Landsat 8 and Sentinel 2, and from these collections, we used images of our area of interest and filtered them by date and cloud percentage. We used images from the 1st May to the 30th September and only images with less than 5% of clouds to be sure to have good images. The last step of the preparation of our data was to make a median of all images from the same year to have only one image by year based on all the images corresponding to our criteria.

2.3. Classification and analysis Method

In this analysis, we are using a supervised classification algorithm using Random Forest as explained by Brovelli et al. [8]. We need to classify the pixels into two classes: Forest or Non-Forest. We are training the Random Forest algorithm on manually supervised data of forest and non-forest in the Amazon forest. These data have been downloaded separately from another region of the amazon forest.

All the process of downloading images has been done with Google Earth Engine. The training and use of the classification algorithm are also done with Google Earth Engine. First, we train the algorithm on the training data set, and we evaluate the performance on the validation data set. Our training and validation data sets were randomly divided into 80% for the training data set and 20% for the validation data set. After verifying the score of the algorithm, we used it on the images downloaded from the satellite collections of Google Earth Engine to classify all the pixels into two classes: Forest or Non-Forest. After the classification is completed we are able to create two two-tone maps (one for 2016 and the other one for 2021) with forest pixels in one color and Non-Forest pixels in another. We will also create a single map to see the changes between 2016 and 2021, it will be a three-tone map with a different color for pixels that were forest in 2016 and Non-Forest in 2021, pixels that were Non-Forest in 2016 and Forest in 2021 and pixels that did not change of classes. Finally, we will quantify the area of forest loss or gain between 2016 and 2021 in our area of study by calculating the number of Forest pixels in 2016 and 2021. We will show the results in the next section.

3. Results & Discussion

3.1. Classification results

The classification results based on our Random Forest algorithm are visible here after. We perform the same algorithm for the two analyzed periods and we can see the binary classification for each period in Figure 3. We can see in Figure 3 the result of our classification. We can notify changes already on these maps but as it is not easy to know the area of forest loss, we will calculate the difference between 2016 and 2021 and show all the pixels that changed categories from Forest to Non-Forest in another figure. Regarding the performance of our classification with the validation data, we have the following error matrix:

3121	25
24	2518

Table 1. Error matrix

The validation accuracy is greater than 99%.

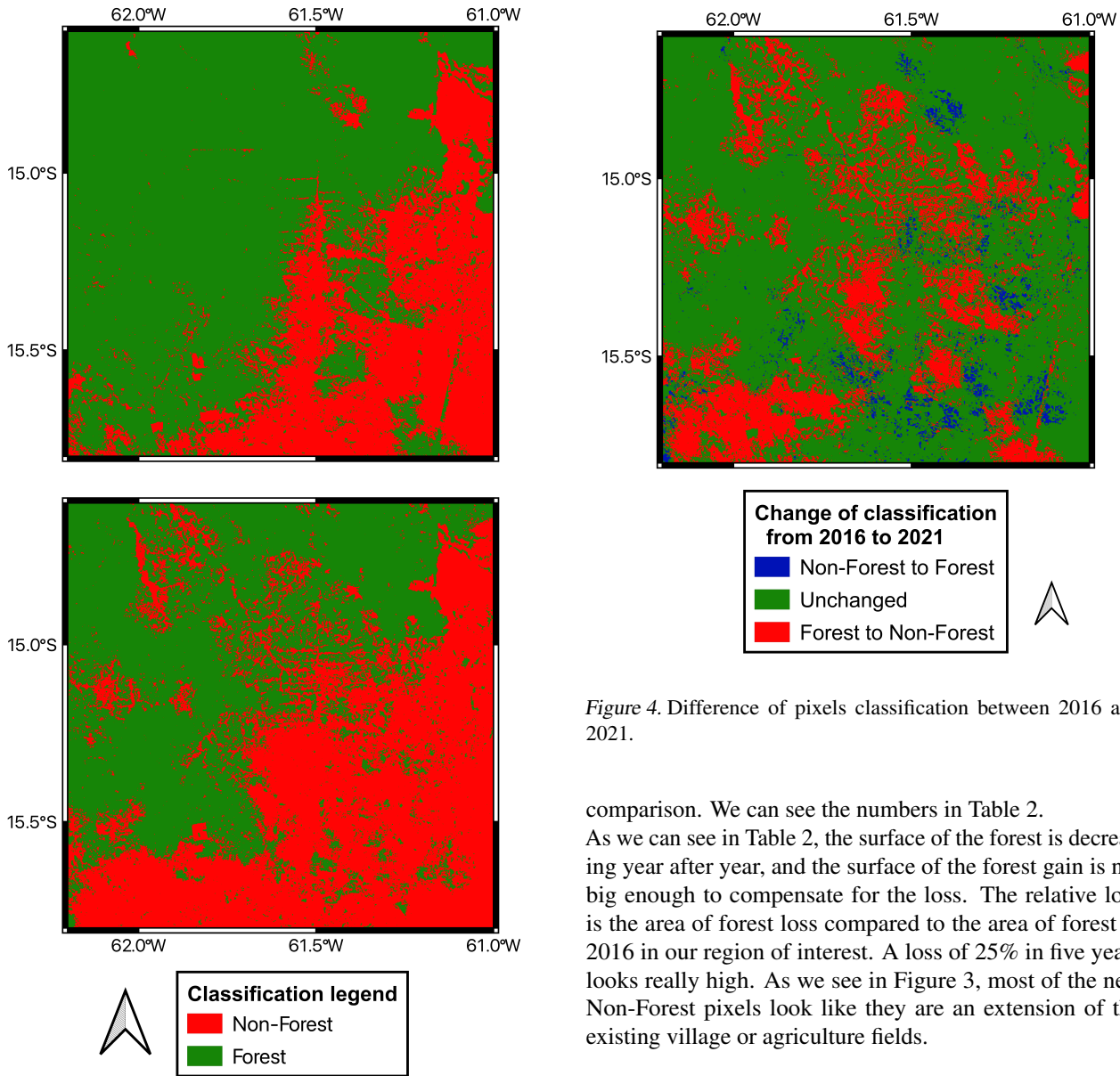


Figure 3. Classification in two classes

Figure 4. Difference of pixels classification between 2016 and 2021.

comparison. We can see the numbers in Table 2. As we can see in Table 2, the surface of the forest is decreasing year after year, and the surface of the forest gain is not big enough to compensate for the loss. The relative loss is the area of forest loss compared to the area of forest in 2016 in our region of interest. A loss of 25% in five years looks really high. As we see in Figure 3, most of the new Non-Forest pixels look like they are an extension of the existing village or agriculture fields.

Years	2016-2021
Forest loss	3543 km ²
Forest gain	485 km ²
Difference	-3058 km ²
Relative loss	25.9%

Table 2. Area of forest loss and gain between 2016 and 2021.

3.2. Forest changes

In this subsection, and in Figure 4 we can see the forest changes between 2016 and 2021. All the green pixels have the same land cover in 2016 and 2021, all the red pixels were forest in 2016 and changed to non-forest in 2021 and all the blue pixels were non-forest in 2016 and are forest in 2021.

We can clearly see that the forest decrease a lot from 2016 to 2021. To have a better idea of what the new forest pixels and the new Non-Forest pixels represent, we calculate the areas of forest lost and gained between our two years of

4. Conclusion

In conclusion of this work, we can see that the methodology gave us good and usable results. It's not easy to see the differences on the maps, it would be great to have maps with a smaller scale to be able to see these changes in land cover. But with the numbers in Table 2, we can have a better

idea of the surface of forest loss over five years. A total of 3058 km² of forest disappeared, to give a comparison scale, the area of Paris is about 100 km². Unfortunately, we did not have time to use another classification method to compare our results with. We could have done a support vector machine classification to see if we have better or worse results. Another improvement to this paper would be to classify our pixels into more than 2 categories. It could be interesting to have categories like "Urban area", "Water", "Roads" or even just different kinds of vegetation to have a more precise classification. These are the next steps of our study.

5. Code & Data

Code and data are available following this link: [ML project data and code](#)

6. Bibliography

References

- [1] Maria Antonia Brovelli, Yaru Sun, and Vasil Yordanov. "Monitoring Forest Change in the Amazon Using Multi-Temporal Remote Sensing Data and Machine Learning Classification on Google Earth Engine". In: *ISPRS International Journal of Geo-Information* 9.10 (2020). ISSN: 2220-9964. DOI: 10.3390/ijgi9100580. URL: <https://www.mdpi.com/2220-9964/9/10/580>.
- [2] *Deforestation* — National Geographic Society. <https://education.nationalgeographic.org/resource/deforestation>.
- [3] Bruno Meyerfeld. "Au Brésil, la déforestation de l'Amazonie au plus haut depuis 2008". In: *Le Monde.fr* (Dec. 2020).
- [4] Christina Nunez. *Deforestation and Its Effect on the Planet*. <https://www.nationalgeographic.com/environment/article/deforestation>. July 2022.