CROWDSOURCING FOR DEFORESTATION DETECTION IN THE AMAZON

G. Bratic[1], M.A. Brovelli[2,*]

[1] Politecnico di Milano, Dept. of Civil and Environmental Engineering, Via Gaetano Previati 1/c, 23900 Lecco, Italy - gorica.bratic@polimi.it

[2] Politecnico di Milano, Dept. of Civil and Environmental Engineering, Piazza Leonardo da Vinci 32, 20133 Milan, Italy - maria.brovelli@polimi.it

Commission IV, WG IV/4

KEY WORDS: Crowdsourcing, High-Resolution Land Cover, Deforestation, Rainforest, Land Cover, OpenStreetMap

ABSTRACT:

Every year, deforestation results in the loss of wide stretches of forest which is worsening the state of air quality, biodiversity, indigenous cultures, climate, meteorological conditions, etc. According to the Monitoring of the Andean Amazon Project (MAAP), roughly 20 million hectares of land were lost due to deforestation in 2020. To address the issue of deforestation, this study proposes a derivation of the deforestation risk model to target the spread of deforestation, which is the first step towards its prevention. The region of interest - North West of Mato Grosso, Brazil - was selected based on two characteristics: it is a deforestation hotspot according to MAAP and it comprises 4 indigenous lands. The sequence for developing the risk model comprises reference information collection, information cleaning, classification, postprocessing, and change detection. In a crowdsourcing mapathon, reference data were gathered, and they were refined in an iterative process using existing land cover maps and photo interpretation. Google Earth Engine and the Random Forest algorithm were used to classify Sentinel-2 imagery for 2019 and 2020. The results obtained are land cover maps from 2019 and 2020 and land cover change, and the risk model. The results are not demonstrating intensive deforestation in the region of interest, however, the deforestation appears to be systematic in two subregions, indicating that it has the potential to spread. An additional concern in the case of these subregions is their proximity to the indigenous land.

1. INTRODUCTION

Deforestation is a worldwide growing problem that creates extensive consequences. Rainforests have a pivotal role in climate, biogeochemical and ecological processes, and as they are getting depleted, the well-being of humanity is becoming more and more uncertain (Wearn et al., 2012). Our largest rainforest, the Amazon, is not spared either. The increasing amount of forest fires in the Amazon rainforest in the last decade also arises the important question of deforestation. Forest fires naturally occur in the Amazon during the dry season which runs from July to October. But the forest fires that took place in 2019 are thought to have been started by human activities as it is the easiest way to clear the land for starting economical activities such as mining, timber exploitation, cattle ranching, agricultural production, etc (Geist and Lambin, 2002; Silveira et al., 2020; Sonter et al., 2017). Moreover, the consequences caused by the fire are severe and often irreversible.

A year-to-year surge in fires occurring in the Amazon rainforest shows the urgency for regulating climate and protecting biodiversity. This research is an answer to a research question: "What can be done in order to prevent/stop/reduce deforestation?". It takes care of identifying the areas in which deforestation happens by deriving a risk model that is nothing else but a land cover change model focused on the forest change. By locating areas at high risk of deforestation, we give a possibility to the local authorities to react and stop it. The research was done under the 2020 YouthMappers Research Fellowship which aims to promote the use of open, geospatial data in research on the resilience of vulnerable populations worldwide.

Deforestation happens everywhere, but since the Amazon is the largest rainforest, this was selected as the location of interest. To be more precise, the area of interest is an area of around 35000 km^2 around the Xingu river in Mato Grosso, Brazil. This area is interesting not only because it is close to one of the deforestation hotspots identified by the Monitoring of Andean Amazon Project (MAAP) (MAAP, 2021), but also because it includes the land of some of the indigenous peoples in Brazil.

The methodology to derive the deforestation risk model can be broken down into reference data collection, data cleaning, classification, postprocessing, and change detection. Reference data were collected in a crowdsourcing mapping event (mapathon) organized by PoliMappers, Youth Mappers chapter from Politecnico di Milano university. The mapping event was a part of the extra-curricular course of Politecnico di Milano university called Collaborative and Humanitarian Mapping (Gaspari et al., 2021). The mapping was based on Teach OSM projects (TeachOSM, 2021), thus all the contributions were stored in OpenStreetMap (OSM) database. Then data were validated, filtered, and adjusted when needed in order to get proper data for training and validation. Ultimately, 13000 sample points were extracted from collected reference data. The classification was carried out on Google Earth Engine (GEE). The Random Forest algorithm was trained by using about 9000 reference samples, and then the trained model was applied on Sentinel-2 imagery to derive land cover maps. The reference points that were not used for training are used for validation of the classification output. Thanks to the validation it was possible to recognize some major defects in the land cover map. The defects were mainly the consequence of the imperfection of reference data. To eliminate them, different solutions were tested during the postprocessing component of the process. The solution that proved to be the most efficient was the introduction of auxiliary land cover information from existing high-resolution land cover (HRLC), but in addition samples of some classes were photo-interpreted. After the reference samples were enhanced, they were used to classify Sentinel-2 im-

*Corresponding author

https://doi.org/10.5194/isprs-archives-XLIII-B4-2022-231-2022

agery for 2020 and 2019. Based on the land cover map for 2019 and for 2020, land cover change was estimated. Change from Forest to Grassland, and from Forest to Cropland are the most relevant changes when deforestation is concerned; Thus, these two types of change were extracted from the land cover change map as they represent the deforestation risk model.

The risk model suggests that cattle ranching is the main driver of the rainforest in the region of interest, but that the deforestation rate is not intensive. About 80km^2 of the forest was replaced by Grassland and 27 km^2 by Cropland. Yet, two subregions in the North-East of the region of interest contain a deforestation pattern that is characteristic of systematic deforestation that is likely to be continued. The two subregions are in proximity to Xingu indigenous land, which increases the gravity of the situation.

The rest of the paper is structured as follows. Section 2 contains the description of the region of interest. In Section 3 the different segments of methodology (reference data collection, data cleaning, classification, postprocessing, and change detection) are described. Results that consist of land cover maps, and land cover change are reported and discussed in Section 4, while conclusions are made in Section 5.

2. REGION OF INTEREST

Region of interest was selected by taking into account the risk of deforestation and the presence of indigenous land. In 2020 MAAP identified 8 main deforestation hotspots, which are mostly located in the southern part of the Amazon rainforest (Figure 1).

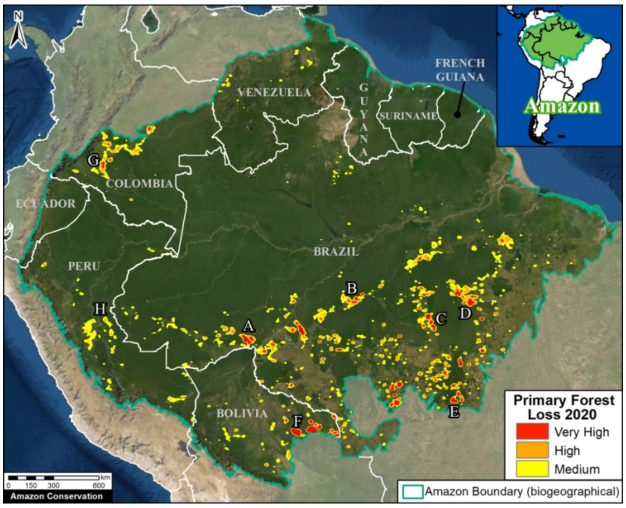

Figure 1: Hotspots of primary forest loss in Amazon in 2020 (Source: MAAP)

The region of interest (Figure 2) is North-East of Mato Grosso state in Brazil around Xingu river valley. It extends along hotspot E displayed in Figure 1. The area was limited to 35000 km^2 in order to collect a sufficient amount of reference data for the training and validation of the risk model. The region includes the indigenous land of Batovi and Pequizal do Naruvôtu, and southern halves of the indigenous territories of Xingu and Wawi.

3. METHODOLOGY

The work done for this project had multiple steps that can be split into reference data collection, data cleaning, classification, postprocessing, and change detection (Figure 3). Each of these steps will be further explained in the following subsections.

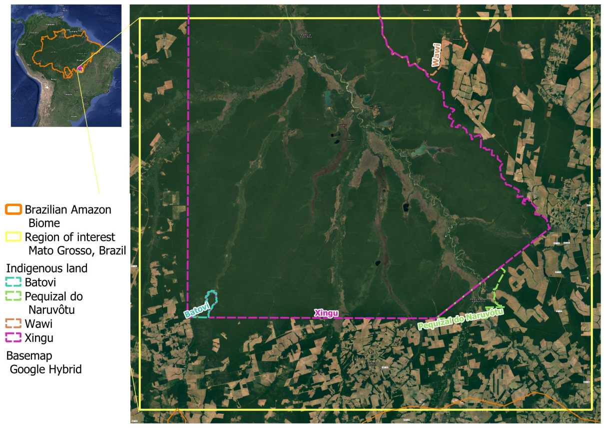

Figure 2: Selected region of interest

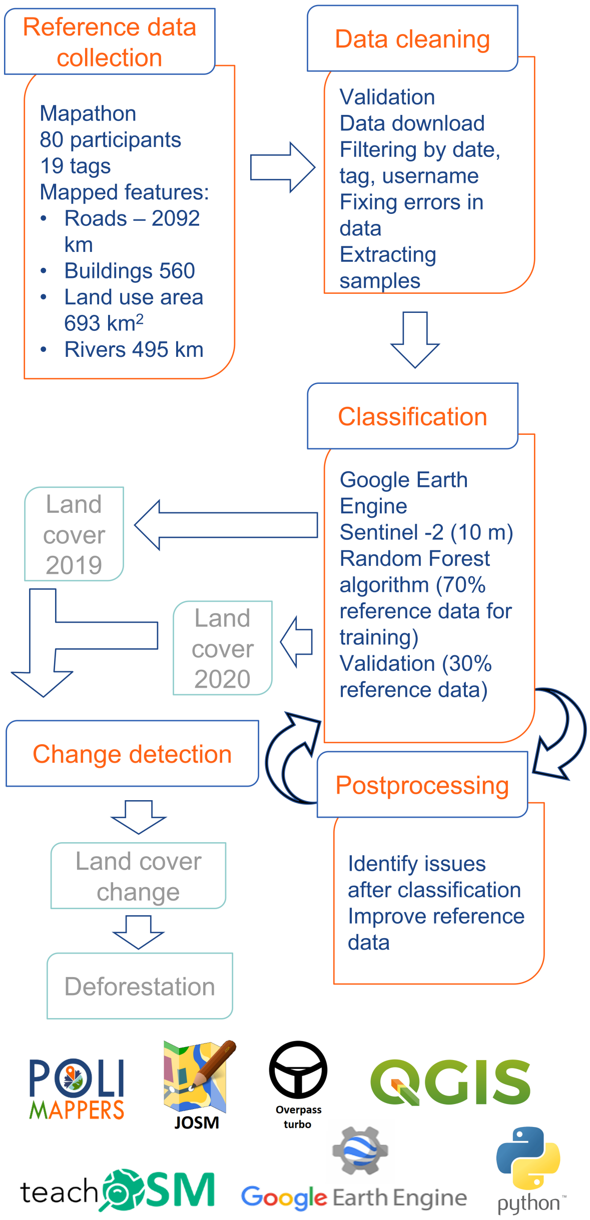

Figure 3: Workflow

3.1 Reference Data Collection

The final goal of reference data collection was to collect information related to 8 generic land cover classes present in the region of interest to be used for training and validation of a machine learning algorithm. These classes are Cropland, Forest, Grassland, Shrubland, Wetland, Water, Built-up, and Bareland. Training and validation data were collected during the Collaborative and Humanitarian Mapping course (Gaspari et al., 2021). This course is an extra-curricular course of Politecnico di Milano university organized and supervised by Prof. Maria Antonia Brovelli and Prof. Ludovico Giorgio Aldo Biagi. It is an innovative course focused on promoting technical skills related to open-source geospatial tools, but also soft skills. The total duration of the course was 20h that were distributed in the 7 online sessions. The session dedicated to the 2020 YouthMappers research fellowship took place on 12. March 2021. It was named ”Mapping deforestation in the Amazon rainforest with JOSM” and it had a form of mapathon. During this session, a tutorial was given on how to use JOSM (Java OpenStreetMap editor), but also how to contribute to TeachOSM projects, and to recognize, trace and tag various features on satellite imagery. To keep mapping in the target area, seven TeachOSM projects were created. Since the mapping was done on Teach OSM, all the contributions were stored in OpenStreetMap (OSM) database. A total of 19 OSM tags were proposed for mapping. OSM tags are typically more detailed than target land cover classes, therefore for some of the target land cover classes, there was more than one corresponding OSM tag (Table 1). To make it easier for participants to keep track of the numerous tags, a preset of tags was prepared and it was shown to participants how to integrate and use it with JOSM. Moreover, participants had access to the very detailed written support materials, but also every TeachOSM project was accompanied by a list of tags to be used, and a description of how to recognize corresponding features in the basemaps. One week after the session dedicated to deforestation, an additional session was organized to address the doubts or curiosities that students have had.

The amount of collected data categorized by different features is reported in Table 2.

There were around 80 participants in the mapathon. They are from different continents: South America, Africa, Asia, and Europe which makes the project cross-continental.

Data collected in the mapathon were validated by PoliMappers officers. During the validation, the minor errors were fixed, while for the major errors we gave instructions to the OSM user who introduced the error how to fix it. Prevalently, the errors were related to the buildings that are not orthogonalized, landuse features connected to the roads, highways mistakenly named road, and landuse overlapping. These errors are not related only to deforestation mapping but are typical errors in mapping on OSM.

3.2 Data Cleaning

We used Overpass turbo to download OSM data in the region of interest, then we filtered data by date, tags, username, and hashtags. Data filtering was done by using a custom Python script. It was necessary to filter data to eliminate all features that are not collected during the mapathon and that therefore could be inadequate or unnecessary for the scope of this work. After filtering, we made a visual inspection of data in QGIS. The visual inspection resulted in the identification of some errors such as missing tags, presence of two mutually exclusive tags (i.e. natural=wetland and natural=grassland), confusion of farmland with grassland and vice versa, confusion of wood with scrub and vice versa, etc. The identified errors were corrected both locally using QGIS and on OSM using JOSM. Tags were translated into land cover classes (Table 1), and features that belong to the same land cover class were dissolved. At the end of the data cleaning procedure, 2000 samples per class were extracted from the available data polygons to be used in the classification.

OSM tags / combination of tags	LC class	Value
landuse=farmland	Cropland	1
natural=wood	Forest	2
natural=grassland	Grassland	3
natural=scrub	Shrubland	4
natural=wetland	Wetland	5
waterway=stream	Water	6
natural=water		
natural=water, water=lagoon		
natural=water, water=lake		
natural=water, water=oxbow		
natural=water, water=pond		
natural=water, water=reservoir		
natural=water, water=river		
natural=water, water=stream		
natural=water, waterway=riverbank		
waterway=river		
building=house	Built-up	8
building=yes		
highway=path		
highway=residential		
highway=secondary		
highway=tertiary		
highway=track		
highway=unclassified		
natural=sand	Bareland	9
natural=scree		

Table 1: OSM tags and their correspondence to the land cover classes

Feature	Amount
Roads	2092.3 km
Buildings	560
Land use area	693.22 km^2
Rivers	495.3 km

Table 2: Mapping results. Source: resultmaps.neis-one.org/, https://mapathon.cartong.org/

3.3 Classification

Classification of Sentinel-2 imagery was carried out on GEE by using the Random Forest training algorithm.

Sentinel-2 was selected for classification because it is imagery with the best resolution available publicly. In addition, it is accessible through GEE Catalog which is one more advantage given that it could be directly processed on GEE. Sentinel-2 is the imagery of the Copernicus Sentinel-2 mission. An image consists of 4 bands at 10m resolution, 6 bands at 20m resolution, and 3 bands at 60m resolution. In this work, only bands at 10m and 20m resolution and scenes with less than 10% cloud coverage were used. The two separate classifications were executed - one of imagery of 2019 and one of 2020. In both cases, Sentinel-2 imagery at the L2A processing level was considered. At this level, imagery values refer to *Bottom of atmosphere* (BOA) reflectance which means that the image reflectance is corrected for absorption and scattering caused by the presence of gases and aerosols in the atmosphere.

The classification procedure started by loading all images for 2020 with cloud coverage lower than 10% from Earth Engine Data Catalog. At this point, a cloud mask was applied to remove clouds and cirrus. The cloud mask is provided for Sentinel L2A products as the band QA60. In the following step, 10 bands to be used in the classification are specified. The bands selected are bands 2, 3, 4, 5, 6, 7, 8, 8A, 11, and 12 that correspond to Blue, Green, Red, Red Edge 1, Red Edge 2, Red Edge 3, NIR, Red Edge 4, SWIR 1, and SWIR 2 electromagnetic spectrum. Then, samples extracted after reference Data Cleaning were imported into GEE assets. 70% of reference samples were used for the training of the Random Forest algorithm with 100 trees. The output raster was exported from GEE to Google Drive and then to local storage. The intermediate step of exporting to Google Drive was needed because there is no direct way to export data from GEE to local storage. Finally, the classification output was validated with 30% of the remaining reference samples. The output of the validation was error matrix and accuracy indexes - Overall Accuracy (OA), User's Accuracy (UA), and Producer's Accuracy (PA).

3.4 Postprocessing

The validation results of the first classification outputs were showing relatively low accuracy - OA = 63%. By visually checking the resulting raster, it was evident that the Built-up and Wetland classes were overestimated. This was an indicator that the reference data might not be sufficiently accurate and that they need further refinement. At this point, we decided to use additional information about land cover in this area to increase the reliability of reference data collected in the crowdsourcing approach. The additional sources of the information were existing HRLC maps at a resolution of 30m or better and available for the year 2016 or later. Existing HRLC maps and data collected in OSM were combined in a particular way in order to try to extract only those information with a higher probability to be correct in each individual map/data source.

Existing HRLC maps that were used for reinforcing reference data collected on TeachOSM are listed in Table 3 together with their main characteristics - name, resolution, and baseline year.

Dataset name	Year	Resol.	Reference
FROM GLC (Finer Resolution Observation and Monitoring of Global Land Cover)	2017	10 m	Gong et al. (2019)
MapBiomas	2018	30 m	MapBiomas (2019)
GHS BU LDS (Global Human Settlement Built-Up Grid – Sentinel-1)	2016	20 m	Corbane et al. (2018)
WSF (World Settlement Footprint)	2019	10m	Marconcini et al. (2020)
FNF (Forest / Non-Forest)	2016	25 m	Shimada et al. (2014)
GSW seasonality (Global Surface Water seasonality)	2019	30 m	Pekel et al. (2016)
GL30 (GlobeLand30)	2017	30m	Chen et al. (2015)

Table 3: Existing high-resolution land cover dataset

Every land cover map aims at representing material on the Earth's surface as accurately as possible. Nevertheless, every map contains certain errors depending on imagery type, preprocessing, training data, classification algorithm, etc. When existing land datasets are compared among themselves, the area in which they all show coherent information is the area with the highest probability that they are correct. Therefore, if we intersect multiple land cover maps, the areas in which they share information can be used to extract training samples for deriving a new land cover map. Figure 4 illustrates how coherent land cover information can be extracted from existing HRLC and reference data collected on OSM.

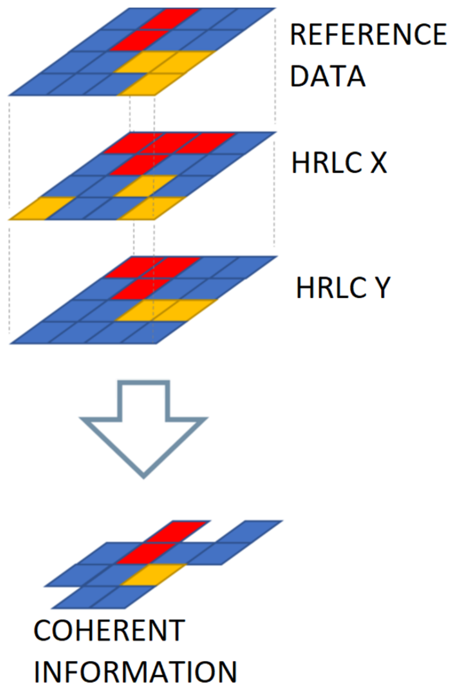

Figure 4: Extracting coherent information from multiple datasets

Before any type of data comparison/intersection, it was necessary to harmonize data with respect to their coordinate reference system, legend, and resolution. Furthermore, OSM data were rasterized to comply with the existing HRLC data type. The coordinate reference system selected was WGS84 (EPSG:4326) and all datasets were reprojected to 10m resolution. The legends (pixel values and labels) of the existing HRLC classes were adapted to correspond to 1 - Cropland, 2 - Forest, 3 - Grassland, 4 - Shrubland, 5 - Wetland, 6 - Water, 8 - Built-up, and 9 - Bareland. The harmonized data were intersected and coherent information was extracted and stored in a raster format. Hereafter, the raster will be referred to as a map of agreement. It was computed by using GRASS GIS and Python.

The map of agreement had one shortcoming - classes Bareland and Built-up were missing. This means that intersected datasets were not showing consistency with respect to these two classes. Nonetheless, it was possible to extract samples for other classes i.e., Cropland, Forest, Grassland, Shrubland, Wetland, and Water. Also, in this case, 2000 samples per class were extracted. To overcome the issue with missing information for Built-up and Bareland classes multiple potential solutions were tested:

- Using originally collected reference samples for these two classes.
- Using originally collected reference samples for these two classes, but removing Built-up class samples coinciding

with roads. The reason for doing so is the fact that the roads are relatively narrow features that might be subject to positional error. In case of positional errors, it could happen that the samples that should represent road actually represent surrounding features but are labeled as Built-up class samples.

- Using originally collected reference samples for Bareland class, but extracting only buildings centroids as samples for Built-up class.

Each test was one iteration of modification of the reference data, classification, validation, and export of the classification results for visual inspection. None of the previously mentioned solutions was giving satisfactory results, therefore the initially collected reference samples of Bareland and Built-up classes were revised and corrected by comparing them visually with Sentinel-2 basemap for 2020. In the case of the Built-up class, revision showed that only buildings with an area bigger than 100 m^2 were detectable on Sentinel-2 imagery. Thus, all the buildings with sizes smaller than 100m were discarded from the reference data. When Bareland class is concerned, the only areas where we noticed this class was present are sand accumulations in the Xingu river. The reference samples collected for this class did not resemble the situation displayed Sentinel-2 basemap for 2020. Most probably the reason behind is that the basemap used to collect reference data (Maxar Standard Imagery (Beta)) was not from the year 2020, and sand accumulation can move or change its form over time. As a solution, the collected Bareland samples were compared visually against Sentinel-2 basemap for 2020, and those samples that were not aligned with sand accumulation in the basemap were deleted. The revision yielded 52 samples of Built-up and 919 samples of Bareland.

Once the samples were refined, there were utilized to derive two land cover maps - one for 2019 and one for 2020. The intersection of reference data with existing land cover produced in the period 2016-2019 assured that the land cover did not change by 2020 therefore it was reliable to use the same samples for classifying imagery in 2019.

Modification of reference data and visual inspection of the results were done in QGIS, while classification and validation were done on GEE.

3.5 Change detection

Change detection was the last data processing exercise. Firstly, the land cover change was computed by using r.cross of GRASS GIS. This tool produces output in the form of raster whose values are all unique combinations of category values found in the raster input layers (land cover maps). Furthermore, starting from the land cover change map, changes related to deforestation were extracted - change of forest to grassland or cropland. Finally, the changes of forest to grassland or forest to cropland were examined visually to understand the presence of a pattern that indicates systematic deforestation.

4. RESULTS AND DISCUSSION

The results of the previously presented procedure include both maps and accuracy figures.

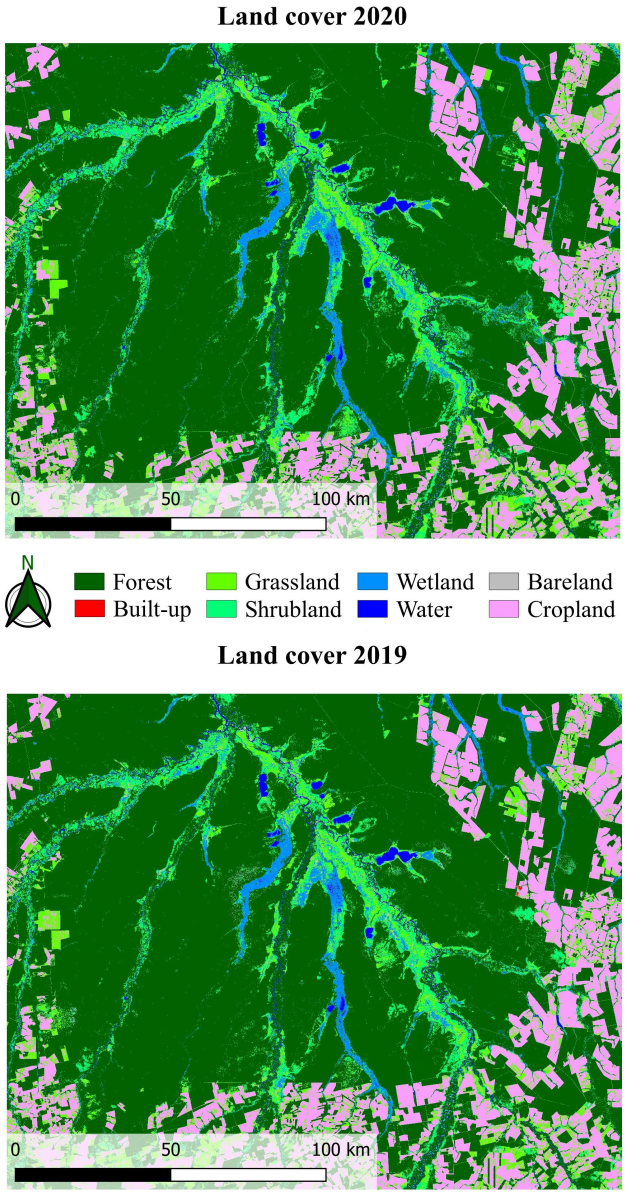

Figure 5: Land cover for 2020 (up) and 2019 (down)

4.1 Land cover

When the land cover is concerned there were two maps derived, one for 2019 and another one for 2020 (Figure 5). The two maps look alike, but there are small differences that cannot be appreciated at this scale.

The accuracy of two land cover maps was computed based on approximately 4000 samples (30% of reference samples). Due to the refinement of the reference data, the area of the reference data was reduced, and thus training and validation data might be biased due to proximity. However, the validation was performed anyway in absence of another reference data source. The error matrix and accuracy indexes for the map of 2019 are reported in Table 4, and for the map of 2020 in Table 5. In both cases, accuracy figures indicate a very high overall accuracy of 93%. PA of class Built-up is low for both years which indicates that there is omission error in this class i.e., many Built-up areas are not classified as such, but rather as Bareland or Cropland. However, this is acceptable in this study given that this class is very small in the region of interest and in general this class does not affect deforestation significantly. Somewhat lower accuracy is evident

https://doi.org/10.5194/isprs-archives-XLIII-B4-2022-231-2022 | © Author(s) 2022. CC BY 4.0 License.

also in the case of the Grassland class. PA of about 85% is indicating that some pixels of this class are not classified as such, but also UA of about 85% is indicating that other classes were classified as Grassland by mistake. This class was mostly confused with Cropland due to their similarity. Also, this confusion should not affect deforestation detection because both classes are contributing to deforestation.

		Reference							
		Cropland	**Forest**	**Grassland**	**Shrubland**	**Wetland**	**Water**	**Built-up**	**Bareland**
Classification	**Cropland**	668		63	1	1		3	19
	Forest		563	3	16	3			
	Grassland	41	5	517	12	4	2		1
	Shrubland		7	22	535	22			
	Wetland		2	1	9	542	9		
	Water					4	601		
	Built-up							7	
	Bareland	2		1				10	262
	PA [%]	**94**	**98**	**85**	**93**	**94**	**98**	**35**	**93**
	UA [%]	**88**	**96**	**89**	**91**	**96**	**99**	**100**	**95**
	OA [%]	**93**							

Table 4: Error matrix and accuracy indexes for land cover for 2020

		Reference							
		Cropland	**Forest**	**Grassland**	**Shrubland**	**Wetland**	**Water**	**Built-up**	**Bareland**
Classification	**Cropland**	652	4	56				3	5
	Forest	4	537	1	13	7			
	Grassland	58	1	514	22	7			1
	Shrubland	1	6	42	575	15			
	Wetland			2	10	602	3		
	Water					7	585		2
	Built-up							7	
	Bareland	1						3	281
	PA [%]	**91**	**98**	**84**	**93**	**94**	**99**	**54**	**97**
	UA [%]	**91**	**96**	**85**	**90**	**98**	**98**	**100**	**99**
	OA [%]	**93**							

Table 5: Error matrix and accuracy indexes for land cover for 2019

4.2 Land cover change

The final result - the deforestation risk model - represents the map of changes of Forest to Grassland or Forest to Cropland. The risk model is displayed in Figure 6. The figure includes also the borders of several indigenous lands in the region of interest. Overall, deforestation does not seem to be intensive in this region in the period from 2019 to 2020. Land cover maps are never completely correct, so the change includes also some errors. However, there are two regions emphasized in red boxes and denoted with numbers 1 and 2 in the figure in which deforestation seems to be systematic. The deforestation in the red boxes was compared visually against Sentinel-2 basemaps for 2019 and 2020, it seems that changes in these regions are real, and not a consequence of classification errors.

About 80 km^2 of Forest was replaced by Grassland and 27 km^2 by Cropland, which indicates that cattle ranching is probably the

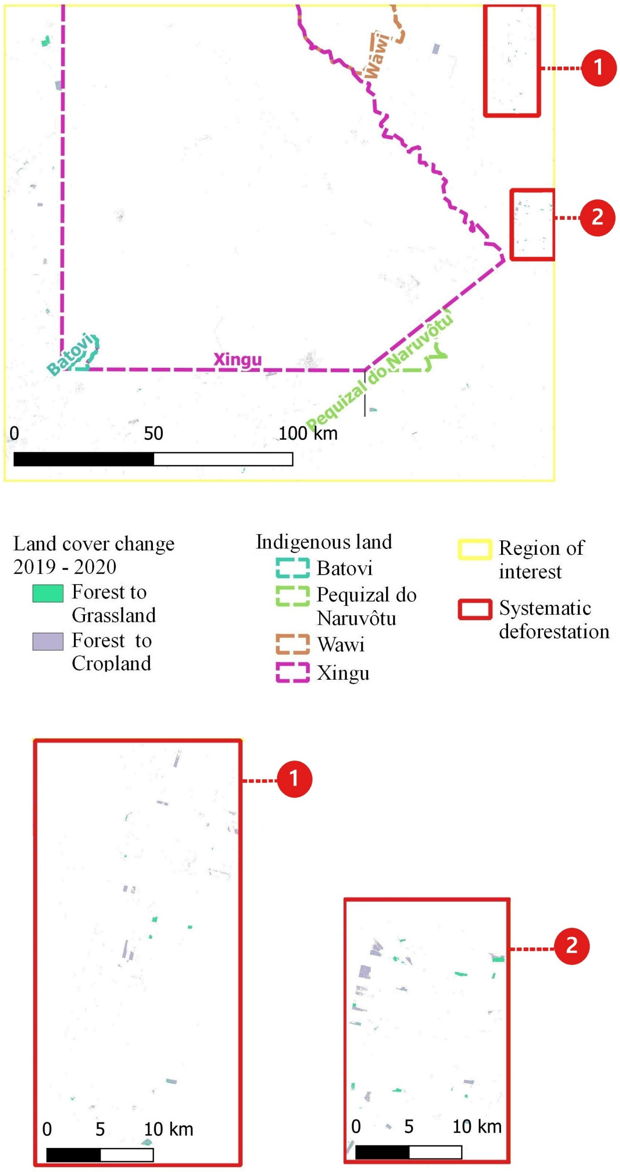

Figure 6: Land cover change – Forest to Cropland or Grassland

main driver for deforestation. From the figure, it can be also observed that the lower red box is very close to the boundary of Xingu indigenous land. The distance is approximately 3km, therefore if deforestation in this area continues, it might invade the indigenous land.

5. CONCLUSION

The research presented in this report was focused on the development of the deforestation risk model for the Amazon rainforest. The research was driven by the motivation to contribute to deforestation prevention. Deforestation seriously suppresses the ecological services that we owe to the rainforests, such as oxygen production, air quality regulation, and regulation of weather on a local and regional scale. Moreover, rainforests are home to a wide variety of plant and animal species, as well as of indigenous people.

The research tackles the deforestation problem by providing a deforestation risk model. The risk model represents a land cover change map focused only on changes from Forest to Cropland or Forest to Grassland between 2019 and 2020. The risk model concerns a region of approximately 200 km x 180 km in the North-East of Mato Grosso state in Brazil. The region comprises 4 indigenous land - Batovi, Pequizal do Naruvôtu, Xingu, and Wawi. The process of the deforestation risk model derivation has 5 main components: reference data collection, data cleaning, classification, postprocessing, and change detection. Reference data collection was based on a mapathon during which participants were trained how to map features that are useful input for classification. After the training, the participants mapped many different features. The mapathon was a part of the Collaborative and Humanitarian extracurricular course of Politecnico di Milano, sustained by the local Italian YouthMappers chapter - PoliMappers. The features collected in the mapathon were validated by PoliMappers officers, and they were adapted to serve as the reference samples for classification. Then, classification and processing were running repetitively until satisfactory results of classification were obtained. The iterations were needed because the first classification results were worse than expected, and therefore it was necessary to remove the source of problems to improve the results. The source of the problem was in the reference data. They were refined by using auxiliary land cover information (existing HRLC) as well as by photo-interpretation of samples for those classes for which information was not available in the existing HRLC maps. 70% of the refined reference samples were fed to Random Forest in order to produce land cover maps for 2019 and 2020, based on Sentinel-2 imagery. Land cover change raster was computed for the period 2019–2020, and thereupon the changes that are the most relevant for deforestation were extracted - Forest to Cropland or Forest to Grassland.

Derived land cover maps were validated against 30% of reference data, and they showed an overall accuracy of 93%. Somewhat lower accuracy was present in the case of the Built-up class, which was an indicator of underestimation of this class. However, due to the small area of this class in the region of interest, the underestimation is not significant for the risk model. Slightly lower accuracy was noticeable also in the case of the Grassland class. With UA and PA of 85%, the accuracy figures are suggesting that this class was committed and omitted on behalf of other classes, mainly Cropland. Due to the particular refinement of the reference samples, the area from which training and validation samples were extracted was rather narrow, and therefore validation results might not be completely unbiased from the training data.

Regarding the deforestation risk model, it does not show extensive deforestation in the analyzed region. The change from Forest to Grassland was estimated to be approximately 80 km^2, and in case of change from Forest to Cropland 20 km^2. Even though the change does not seem significant in terms of its extent, the analyses of the risk model show some patterns that imply systematic deforestation with evolving potential. Moreover, the areas in which systematic pattern is detected are at the 3 km distance from Xingu indigenous land boundaries, hence the potential expansion could lead to deforestation of indigenous land.

Besides the immediate results relative to the region of interest, the research path is taken to derive the deforestation risk model can serve as an example of how OSM data can be combined with existing HRLC in order to derive land cover maps, land cover change maps, and/or deforestation risk models in general. In this case, the mapathon was organized to map specific features in the region of interest which were later used for classification. In the future, it would be interesting to examine the potential to directly use data already available in the OSM in combination with existing HRLC to produce a deforestation risk model.

ACKNOWLEDGEMENT

This project received funding from the YouthMappers Research Fellowship Program, supported by USAID Award AID-OAA-G-15-00007 and Cooperative Agreement Number: 7200AA18CA00015, anonymous donor, and in-kind support from Texas Tech University, The George Washington University, West Virginia University, and Arizona State University. We are very grateful for the funds. We would also like to thank the YouthMappers chapter PoliMappers for the assistance in organizing the mapathon for reference data collection, as well as for their help for the validation of the mapping tasks.

REFERENCES

Chen, J., Chen, J., Liao, A., Cao, X., Chen, L., Chen, X., He, C., Han, G., Peng, S., Lu, M., Zhang, W., Tong, X. and Mills, J., 2015. Global land cover mapping at 30m resolution: A POK-based operational approach. ISPRS Journal of Photogrammetry and Remote Sensing 103, pp. 7–27.

Corbane, C., Florczyk, A., Pesaresi, M., Politis, P. and Syrris, V., 2018. GHS built-up grid, derived from Landsat, multitemporal (1975-1990-2000-2014), R2018A. type: dataset.

Gaspari, F., Stucchi, L., Bratic, G., Jovanovic, D., Ponti, C., Biagi, L. G. A. and Brovelli, M. A., 2021. Innovation in Teaching: The Polimappers Collaborative and Humanitarian Mapping Course at Politecnico di Milano. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLVI-4/W2-2021, pp. 63–69.

Geist, H. J. and Lambin, E. F., 2002. Proximate Causes and Underlying Driving Forces of Tropical Deforestation. BioScience 52(2), pp. 143.

Gong, P., Liu, H., Zhang, M., Li, C., Wang, J., Huang, H., Clinton, N., Ji, L., Li, W., Bai, Y., Chen, B., Xu, B., Zhu, Z., Yuan, C., Ping Suen, H., Guo, J., Xu, N., Li, W., Zhao, Y., Yang, J., Yu, C., Wang, X., Fu, H., Yu, L., Dronova, I., Hui, F., Cheng, X., Shi, X., Xiao, F., Liu, Q. and Song, L., 2019. Stable classification with limited sample: transferring a 30-m resolution sample set collected in 2015 to mapping 10-m resolution global land cover in 2017. Science Bulletin 64(6), pp. 370–373. Number: 6.

MAAP, 2021. Monitoring of the Andean Amazon Project. https://maaproject.org/. [Online; accessed 25-November-2021].

MapBiomas, 2019. Mapbiomas Brasil | Estatística de acurácia. https://mapbiomas.org/estatistica-de-acuracia. [Online; accessed 06-March-2021].

Marconcini, M., Metz-Marconcini, A., Üreyen, S., Palacios-Lopez, D., Hanke, W., Bachofer, F., Zeidler, J., Esch, T., Gorelick, N., Kakarla, A., Paganini, M. and Strano, E., 2020. Outlining where humans live, the World Settlement Footprint 2015. Scientific Data 7(1), pp. 242. Number: 1 Publisher: Nature Publishing Group.

Pekel, J.-F., Cottam, A., Gorelick, N. and Belward, A. S., 2016. High-resolution mapping of global surface water and its long-term changes. Nature 540(7633), pp. 418–422. Number: 7633 Publisher: Nature Publishing Group.

Shimada, M., Itoh, T., Motooka, T., Watanabe, M., Shiraishi, T., Thapa, R. and Lucas, R., 2014. New global forest/non-forest maps from ALOS PALSAR data (2007–2010). Remote Sensing of Environment 155, pp. 13–31.

Silveira, M. V. F., Petri, C. A., Broggio, I. S., Chagas, G. O., Macul, M. S., Leite, C. C. S. S., Ferrari, E. M. M., Amim, C. G. V., Freitas, A. L. R., Motta, A. Z. V., Carvalho, L. M. E., Silva Junior, C. H. L., Anderson, L. O. and Aragão, L. E. O. C., 2020. Drivers of Fire Anomalies in the Brazilian Amazon: Lessons Learned from the 2019 Fire Crisis. Land 9(12), pp. 516.

Sonter, L. J., Herrera, D., Barrett, D. J., Galford, G. L., Moran, C. J. and Soares-Filho, B. S., 2017. Mining drives extensive deforestation in the Brazilian Amazon. Nature Communications 8(1), pp. 1013.

TeachOSM, 2021. Amazon Deforestation - Projects #1054-#1058; #1060-#1061. https://tasks.teachosm.org/. [Online; accessed 03-December-2021].

Wearn, O. R., Reuman, D. C. and Ewers, R. M., 2012. Extinction Debt and Windows of Conservation Opportunity in the Brazilian Amazon. Science 337(6091), pp. 228–232.