SENTINEL-1 TIME-SERIES ANALYSIS FOR DETECTION OF FOREST DEGRADATION BY SELECTIVE LOGGING

Henrique Luis Godinho Cassol[1], Yosio Edemir Shimabukuro[1], René Beuchle[3], and Luiz E. O. C. Aragão[1]

[1]National Institute for Space Research (INPE) – Remote Sensing Division, Av. dos Astronautas 1758, Jd Granja, São José dos Campos, São Paulo, Brazil; {henrique, yosio, laragao}@dsr.inpe.br; [2]Joint Research Centre – JRC, Via Enrico Fermi 2749, I - 21027 Ispra (VA), Italia, rene.beuchle@ec.europa.eu;

ABSTRACT

Forest degradation by selective logging is considered one of the main causes of biodiversity loss and CO_2 emissions in tropical regions. However, persistent cloud cover limits the detection of selective logging using optical satellite systems in the Brazilian Amazon. We develop a novel approach to detect selective logging using one-year time-series (TS) from Sentinel-1 RADAR data (C-band), based on state-of-art cloud computing using Google Earth Engine. The method consists of two temporal TS reductions. The first reduces the TS for the median monthly record while the second one computes annual statistics like mean, standard deviation, and amplitude. The result is a composite band used for classifying the annual TS through the application of a machine-learning algorithm (CART). Classification showed 69% overall accuracy within five classes; however, the misclassification of the degradation class was 54%. The classification accuracy has increased to 79% with the removal of the regrowth class, with 74% of the degradation correctly classified.

Key words — SAR system, machine-learning, cloud-computing, segmentation, regression trees.

1. INTRODUCTION

Forest degradation by selective logging is considered one of the main causes of biodiversity loss and CO_2 emissions in tropical regions [1, 2]. The estimates of greenhouse gas emissions (GHG) from forest degradation are around 30% of the total GHG emission from land use and land cover change (LULCC) across tropical countries [2]. Furthermore, forest under selective logging stores 40% less carbon and has, on average, 13% less species richness than undisturbed forests [3, 4]. However, the uncertainties related to the mapping and monitoring process are still unknown. The main limitation is the cloud cover persistency in tropical regions, which reduces the availability of cloud-free images from optical satellite systems and, consequently, decreases the detection accuracy of small spatiotemporal canopy disturbances [5].

In this sense, Synthetic Aperture Radar (SAR) systems such as a Sentinel-1, which operates on microwaves (C-band, 5.4 cm), are almost insensitive to clouds and capture images day and night. Moreover, cloud-free dataset in combination with the high revisit time (12 days on the Equator) allows for detection and monitoring of selective logging throughout the year, and thus not only in the dry season. Currently, illegal logging has been reported during the rainy season in order to avoid detection with optical sensors [6].

The main goal of the study is to assess Sentinel-1 time-series for detection of forest degradation by selective logging. The novel approach uses a state-of-art Google Earth Engine (GEE) cloud computing to handle the processing of large image data sets [7]. This is the first approximation of an almost fully automated algorithm to detect selective logging in the Brazilian Amazon, Mato Grosso (MT) State, using SAR data.

2. MATERIAL AND METHODS

The study area comprises an area of 2,662.45 km² and is located on both sides of BR-163 Highway, between the municipalities of Sinop and Itaúba. The area is located in the central-northern part of MT, within the Brazilian Amazon (**Figure 1**).

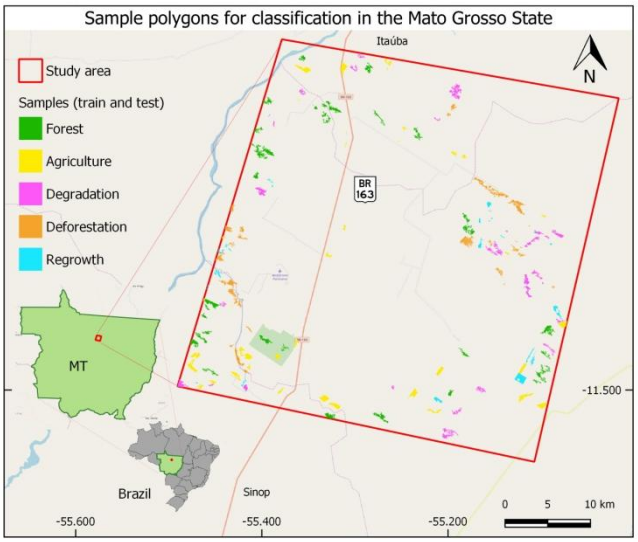

Figure 1. Location of the study area within Mato Grosso State.

The study area encompasses not only municipalities with high legal timber production, but also areas with the high occurrence of illegal logging [8]. The vegetation is predominately Dense Ombrophylous Forest that includes commercial timber such as *Dinizia excelsea* Ducke, *Hymenaea courbaril* L., and *Bagassa guianensis* Aubl. [9]. Climate is humid tropical with very distinct dry and wet seasons that

extend from April through October, and from November through March, respectively. The average annual precipitation and temperature is 2000 mm and 26 °C, respectively [9].

2.1. Sentinel-1 time-series (TS) data

The collection of Sentinel-1 data on GEE Application Programming Interface (API) provides S-1 data from a dual-polarization C-band (VH and VV); it is processed in Sentinel-1 Toolbox to generate a TS composite of Ground Range Detected (GRD) images at 10 m nominal resolution. The Toolbox applies thermal noise removal, radiometric calibration, and terrain-correction to produce a calibrated orthorectified product [10]. In order to standardize the TS composite, only Interferometric Wide (IW) swath mode data from a descending orbit pass were collected. The revisit time for both S-1 satellites (A and B) is 6 days; the collection is updated weekly on GEE API. However, only S-1 A is available over the study area, reducing the revisiting time to 12 days. In consequence, the annual S-1 TS composite has about 30 images [7].

The backscattering coefficient (sigma naught σ^o) was converted to gamma naught γ° by applying the cosine of the average incidence angle θ at each resolution cell in both polarizations (VH and VV). This procedure performs the translation of a resolution cell perpendicular to a radar line-of-sight and smoothes topographic effects on the image [11]. Furthermore, we create an index called RGI (*Radar Gap Index*) which consists in the normalized difference between VV and VH, as RGI = (VV-VH)/(VV+VH). The RGI highlights the relative importance of canopy gaps in the total backscattering. The higher the value of VV-VH is the higher is the single-bounce contribution of bare soil from canopy gaps. The values are higher in open gaps and are sensitive to plant-soil moisture [12].

We chose an interval of one year, from June 2017 to July 2018, totaling an initial TS composite of 30 S-1 scenes. The TS was reduced to a median monthly record to avoid the influence of external factors such as rainfall, calibration and incidence angle at acquisition time. The resulting S-1 TS has 12 scenes with monthly smoothed lines. A boxcar filter (5x5 pixels size) was applied to the monthly S-1 TS for de-speckling.

A forest mask was applied to analyze only forest areas. This procedure is needed to avoid inserting unnecessary classes in the analysis, such as pastures, urban, and water, which inevitably add errors to the classification. For the forest mask, Hansen et al. [8] tree cover data greater than 90% was considered as forest in the year 2000; moreover, non-forest pixels were removed from the analysis.

2.2. Classification of Sentinel-1 TS data

The collection of training and testing data was done visually using an independent set of Sentinel-2 imagery data with 10 m spatial resolution (three visible and one near infra-red bands). A cloud quality filter was applied within GEE to select two cloud-free scenes: one on July 16, 2017 and other on July 26, 2018. The July 2017 image was segmented using the multi-resolution segmentation algorithm of eCognition® [13], with the defined parameters of: scale 100 m, shape 0.2, and compactness 0.3. As RADAR data has high local variability, the objective of image segmentation was to aggregate homogeneous regions to classify the TS by Geographic Object Based Image Analysis (GEOBIA).

Using both sets plus image difference, the segments were labeled in five classes through GEE: namely forest, agriculture, forest degradation, deforestation, and regrowth. A total of 198 collected samples were then labeled in the five classes previously defined. Finally, samples were divided in 60% for training and 40% for testing.

Twelve images from one year S1 TS were reduced using spatial and temporal statistics, resulting in a single image with 15 bands. The temporal statistics of the three original bands VH gamma, VV gamma, and RGI were mean (3 bands), standard deviation (3 bands), min and max (6 bands), and amplitude (3 bands). In addition, the spatial statistics were computed by changing the scale, i.e., the spatial resolution. The classification was performed by the Classification Algorithm and Regression Trees (CART) using the training data set [14]. The performance of classification algorithm was evaluated through the confusion matrix using the testing data set [15].

3. RESULTS

In general, the cross-polarized (VH) channel had the highest accuracy for separating the main classes (**Figure 2)**, followed by the vertical co-polarized (VV) channel and the radar index (RGI).

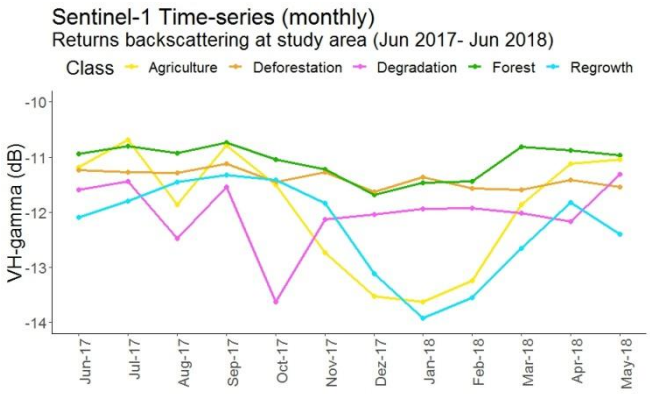

Figure 2. Monthly composite of the Sentinel-1 time-series depicting the annual behavior of the five classes at the VH channel TS.

The forest class, defined as low-disturbance forest, has a consistently high backscattering coefficient and a low standard deviation (γ° = -11.1±0.3 dB) throughout the year, due to signal saturation of the C-band on canopies with high biomass. The variation of γ° during the rainy season (Nov-Feb) occurs as a consequence of the increased plant moisture. The radar signal is sensitive to water content, reducing the power signal by

decreasing the dielectric constant during the dry season [16]. The degradation class, corresponding to selectively logged forests, showed a non-uniform annual behavior with a slight low backscattering and an intermediate standard deviation (γ° = -12±0.6 dB) compared to low-disturbed forest. The extraction of timber reduces the average backscatter somewhat, allowing the penetration of the wavelength in the canopy gaps, which, in turn, decreases the portion of volumetric scattering from VH channel.

The agriculture class shows a typical seasonal crop cycle, which extends from the end of the dry season (Aug-Oct) to the middle of the rainy season (Apr-Jun). The annual cropping phenology is characterized by low backscattering during the sowing period (Oct-Nov) due to the low roughness of soil and specular scattering backward to the RADAR line of sight, and to high backscattering before harvesting (Apr-May). This behavior is similarly observed on optical indices and is characterized by a high annual standard deviation γ°_{sd} = 1.1 dB. Deforestation is characterized by the strong reduction of the backscattering value related to an abrupt increase of the soil contribution and the decrease of the volumetric scattering. Because deforestation events often occur at the end of the dry season, the reduction of γ° values is most common at this period (Sep-Oct). Finally, the annual behavior of regrowth class varies in parallel to the agriculture class, as young secondary forests have high seasonality due to the variations in soil-plant moisture. However, seasonality in regrowth is less pronounced than in the agriculture class, in consequence it has the lowest annual standard deviation γ°_{sd} = 0.86 dB.

3.2. Classification results

Due to fast cloud-computing on GEE, the user can change some parameters of classification and quickly evaluate the performance of the algorithm. In this context, we have tested the forest degradation analysis on six spatial scales: 10 m (original data), 30 m, 100 m, 150 m, 200 m, 300 m and 500 m. The overall accuracy of classification was between 58% at 10 m scale (pixel-by-pixel) and 48% at 500 m scale, with the best result achieved with the 200 m scale (69%). The spatial scale of 200 m was selected as a trade-off between the speckle reducing and spatial smoothing and the scale that each class event occurs in nature.

We also tested the classification using separate bands (VH, VV, or RGI) in order to evaluate the strength of a single band regarding the algorithm performance. The best result was received using only VH-gamma (67%), followed by VV-gamma (59%), and RGI (56%). Changing band numbers from the temporal statistics, the overall accuracy using only the annual mean of VH-gamma was with 69%, in the same magnitude as if using all 15 bands (mean, standard deviation, min-max, and amplitude). According to our analysis, it seems there is a covariance among of prediction variables, leading to the conclusion that the number of temporal reductions bands should be decreased as well as the polarizations. However, the misclassification of the degradation was lower when all 15-bands were used together although the overall accuracy was the same; so we kept all polarization bands and temporal reductions.

The performance of the classification is shown in the confusion matrix (**Table 1**). The overall accuracy of classification in 2017/2018 was 69%, using the test data set for validation. The forest degradation class showed a higher producer accuracy (67%), then user accuracy (43%). The low user accuracy was due to misclassification amongst all other classes. The omission error varied from 3% to 14% for agriculture and deforestation classes, respectively. The commission error was at 11% for the forest class and at 18% for the agriculture class.

Table 1. Confusion Matrix for CART.

Classes	For.	Agr.	Deg.	Def.	Reg.	User Accuracy
For.	135	0	3	1	4	0.94
Agr	22	54	2	2	4	0.64
Deg.	11	18	43	14	14	0.43
Def.	4	0	9	64	0	0.83
Reg.	9	2	7	10	4	0.13
Producer Accuracy	0.75	0.73	0.67	0.70	0.15	**0.688**

As expected, the forest class showed the best classification result of producers and user accuracy, 75% and 94%, respectively. The regrowth class showed the worst results of 15% and 13%, respectively.

Considering the low accuracy of regrowth and the small number of training and test data sets, we exclude this class from the analysis. The overall accuracy of the classification with four classes has augmented by 15%, with a substantial improvement of the producer and user accuracy of the forest degradation class with 76% and 71%, respectively. The classification map is shown in **Figure 3**.

4. DISCUSSION

S-1 TS has a huge potential for forest degradation monitoring due to selective logging in such persistent cloud-cover environments as the Brazilian Tropical forests [16]. However, some issues have arisen from our analysis: 1) Reducing the number of bands in the one-year time-series analysis, such as mean and standard deviation, leads to a loss of information in the near-real time detection, inasmuch as it is feasible through this approach to identify forest degradation, but not the date of occurrence [16]; 2) Inherent speckle noise of RADAR precludes an application of the technique pixel by pixel, which makes an object-based (segmentation) approach necessary. A simple spatial aggregation is not sufficient for circumventing this problem, since it tends to homogenize distinct classes. In consequence, the algorithm cannot be fully automated yet.

In the future, other classifiers could be tested, as Neural Network, Support Vector Machine, or Random Forest, which are reported to produce high accuracy for TS classification [17]. In addition, we expect to implement the TS analysis over a longer time-series in the future, similar to LandTrendr [16],

which could detect better the monthly rupture of the TS, and would including synergistic data analysis with Sentinel-2 imagery.

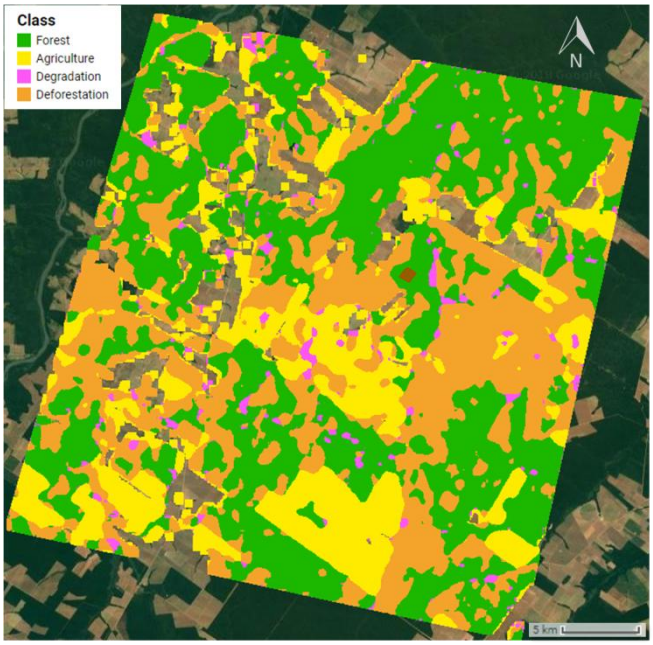

Figure 3. Classification result with 4 classes using CART.

5. CONCLUSIONS

The overall classification with 5 land cover classes was at 69%, but misclassification of forest degradation with other classes was at 67%. The VH-gamma was the best predictor variable, followed by VV-gamma and Radar Gap Index (RGI). The overall accuracy has improved from 69% to 79% with the removal of the vegetation regrowth class from the analysis. The producer and user accuracy of the forest degradation class with 4 classes was at 71% and 86%, respectively. Currently, this novel approach method only detects forest degradation or deforestation a few months after the event. A refined method will enable the monitoring of forest degradation in near-real time once in the future. The GEE code script is: https://code.earthengine.google.com/cd0443453e6156b152847b94d16880b4.

6. REFERENCES

[1] Cazzolla Gatti, R. et al., "The impact of selective logging and clearcutting on forest structure, tree diversity and above-ground biomass of African tropical forests," *Ecol. Res.*, vol. (30), no. (1), pp. (119–132), 2014.

[2] Pearson, T.R.H. et al., "Greenhouse gas emissions from tropical forest degradation: An underestimated source," *Carbon Balance Manag.*, vol. (12), no. (1), 2017.

[3] Berenguer, E. et al., "A large-scale field assessment of carbon stocks in human-modified tropical forests," *Glob. Chang. Biol.*, vol. (20), no. (12), pp. (3713–3726), 2014.

[4] Whitworth, A. et al., "Long lasting impressions: After decades of regeneration rainforest biodiversity remains differentially affected following selective logging and clearance for agriculture," *Glob. Ecol. Conserv.*, vol. (13), p. (e00375), 2018.

[5] Asner, G.P., "Selective Logging in the Brazilian Amazon," *Science*, vol. (310), no. (5747), pp. (480–482), 2005.

[6] Revista EXAME. Available on: https://exame.abril.com.br/revista-exame/brasil-destroi-128-campos-de-futebol-de-floresta-por-hora/. Accessed on: 29 Sep. 2018.

[7] Gorelick, N. et al., "Google Earth Engine: Planetary-scale geospatial analysis for everyone," *Remote Sens. Environ.*, vol. (202), pp. (18–27), 2017.

[8] Grecchi, R.C. et al., "An integrated remote sensing and GIS approach for monitoring areas affected by selective logging: A case study in northern Mato Grosso, Brazilian Amazon," *Int. J. Appl. Earth Obs. Geoinf.*, vol. (61), no. (4), pp. (70–80), 2017.

[9] Feldpausch, T.R. et al., "When big trees fall: Damage and carbon export by reduced impact logging in southern Amazonia," *For. Ecol. Manage.*, vol. (219), no. (2–3), pp. (199–215), 2005.

[10] Snoeij, P. et al., "Sentinel-1 In-Orbit Calibration Plan," pp. 334–336, 2012.

[11] Small, D., "Flattening gamma: Radiometric terrain correction for SAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. (49), no. (8), pp. (3081–3093), 2011.

[12] Hird, J.N. et al., "Google earth engine, open-access satellite data, and machine learning in support of large-area probabilisticwetland mapping," *Remote Sens.*, vol. (9), no. (12), 2017.

[13] Baatz, M. and Schape, A., "Multiresolution Segmentation – an optimization approach for high quality multi-scale image segmentation," In: Strobl, J. et al. Angew. Geogr. Inform-Verarb. XII. Wichmann Verlag, Karlsruhe, Germany, pp. (12–23), 2000.

[14] Breiman, L. et al., "Classification Algorithms and Regression Trees," Wadsworth Stat., pp. (246–280), 1984.

[15] Witten, I.H. et al., Data Mining: Practical Machine Learning Tools and Techniques (Google eBook). 2011.

[16] Reiche, J. et al., "Improving near-real time deforestation monitoring in tropical dry forests by combining dense Sentinel-1 time series with Landsat and ALOS-2 PALSAR-2," *Remote Sens. Environ.*, vol. (204), no. (4), pp. (147–161), 2018.

[17] Shao, Y. and Lunetta, R. S., "Comparison of support vector machine, neural network, and CART algorithms for the land-cover classification using limited training data points," *ISPRS J. Photogramm. Remote Sens.*, vol. (70), pp. (78–87), 2012.

7. ACKNOWLEDGMENTS

This work is part of the institutional capacity building program from INPE. H..L.G.C has been funded by CNPq (grant 300807/2018-0). L.E.O.C.A. acknowledges the support of CNPq (grant 305054/2016-3).