SAR DATA FOR LAND USE LAND COVER CLASSIFICATION IN A TROPICAL REGION WITH FREQUENT CLOUD COVER

V. H. R. Prudente[1,2], I. D. Sanches[1], M. Adami[3], S. Skakun[2,4], L. V. Oldoni[1], H. A. M. Xaud[5], M. R. Xaud[5], Y. Zhang[2]

[1]Remote Sensing Division, National Institute for Space Research, Sao Jose dos Campos SP 12227-010, Brazil
[2] Department of Geographical Sciences, University of Maryland, College Park MD 20742, USA
[3] Amazon Regional Center, National Institute for Space Research, Belem PA 66077-830, Brazil
[4]NASA Goddard Space Flight Center Code 619, Greenbelt MD 20771, USA
[5] Embrapa Roraima, Brazilian Agricultural Research Corporation, Boa Vista RR 69301-970, Brazil

ABSTRACT

This study aims at mapping Land Use and Land Cover (LULC) in the region of Roraima, Brazil, using time-series of Sentinel-1 Synthetic Aperture Radar (SAR) data. All available Sentinel-1 images covering the study area were used and classified using two machine learning algorithms, namely random forest and multilayer perceptron. LULC heterogeneity with the SAR process complexity makes the process challenging in distinguishing certain classes. Results show that SAR data could be used for LULC mapping, as rainforest, savannas, water, and sandbank/outcrop classes. But cannot provide accurate separation for all classes, mainly for those with similar geometrical structures, such as regeneration areas, perennial crops, and buritizais.

Index Terms—Roraima, machine learning, radar, Sentinel-1.

1. INTRODUCTION

Remote Sensing is a powerful tool for Land Use and Land Cover (LULC) monitoring and mapping [1]. But, it is a challenging task in the northern regions of Brazil, mainly because of the complex that involved the landscape [2] and frequent cloud cover [3].

These cloud limitations could be minimized by utilizing the microwave sensors data, such as the Synthetic Aperture Radar (SAR). SAR data collection can be performed under all-weather conditions, independently of lighting conditions, and are not affected by atmospheric and cloud conditions [4].

For this reason, it is possible to obtain a dense temporal series of SAR data, even in areas with large cloud cover, such as tropical regions. Thus, there is a significant potential for the SAR data to be utilized in mapping and monitoring LULC [5].

Radar backscatter is affected by factors related to crop biomass, structure, and ground conditions [1], and depends on the interaction among vegetation canopies and microwave energy. Those properties are influenced by the radar system itself such as polarization and frequency and/or by canopy conditions, such as dielectric constant, size, orientation, incidence angle, wetness, etc. [4], [5].

Different polarizations are more sensitive to certain canopies' characteristics. Cross-polarization is more sensitive to volume scattering, meanwhile, co-polarization is more directly related to direct backscattering [1]. Moreover, specific canopy properties cause different impacts on the SAR scattering intensity, scattering-type, and phase characteristics. These properties are specific for each vegetation type and vary with the phenology changes [5].

For the LULC classification process with SAR data, Random Forest (RF), Maximum likelihood (ML), and Neural Network (NN) are the more utilized [6]. RF is a robust classifier and can handle a high number of variables [7]. Moreover, RF is widely used in LULC approaches, mainly with SAR data [2]. Multilayer Perceptron (MLP) is a feed-forward artificial NN trained by the backpropagation method, designed to map a set of input vectors to a set of output vectors [3], [6].

In this context, this study has the objective to classify the LULC in a small area in Roraima for the 2019 year. For that, we used Sentinel-1 SAR images and tested different parameters of the RF and the MLP.

2. STUDY AREA & DATA DESCRIPTION

Roraima Brazilian state is located in the northern part of Brazil. It has an area of 224,300 km², with 15 municipalities and a population of 606,000 people. Roraima has three major natural formations: rainforest, campina-campinarana, and savannas (called the “lavrados”).

Altitude in Roraima ranges between 30 meters in the Amazon River to 2000 meters in the Roraima mountain [8]. This altitude range act as a natural barrier, blocking the moisture bought by the trade winds along the Intertropical Convergence Zone (ITCZ). These factors allowed a

precipitation gradient and a frequent cloud frequency in the state, which greatly limits the use of satellite optical images.

Roraima rainy season is from May to August months, with 2000 mm accumulated rainfall. In the lavrados, the small lakes fill and connect in this period. The mean annual temperature is 28◦C [2].

Combining the climatic, altitude, and water availability factors, as well as affordable land prices and government subsidies, has encouraged agricultural exploitation (agriculture and livestock) in the state [9]. Moreover, the crop calendar, with harvest during the off-season for the other Brazilian states (April-September), favors better prices and facilitates production chain logistics.

Besides, agriculture in Roraima is in the process of expansion, mainly over the lavrados. The scenario is more intensify after 2010, due to the soybean crop expansion, causing environment threatening. This activity needs to be monitored and regulated; otherwise, it can affect the availability of natural resources (e.g. water, soil) [9].

Considering these factors and the big Roraima extensions, this study was used in a small area that represents a portion of all heterogenic LULC inside the state. The square is located between the coordinates 61.06° W, 2.38° N; 60.69° W, 2.92° N, with 2390.92 km². The location was chosen due to large agricultural expansion in this area.

Due to the frequent cloud cover, mainly during the crop seasons, it is almost impossible to obtain clear sky observations with optical data in this region [2]. Therefore, Sentinel-1 A and B SAR images data were utilized [10]. Sentinel-1 SAR has dense time-series data, can be used to improve and/or develop new methods for mapping and monitoring LULC [11]. It was utilized both polarizations, VH and VV, IW (Interferometric Wide swath) mode, GRD (*Ground Range Detected*) and 12 days for each satellite temporal resolution, 4 and 8 for both satellites.

We used field data for the classification process. These data were obtained in fieldwork (August-September – 2019), during the crop season.

3. METHODOLOGY

The process steps are showing in Fig 1. We preprocessed the SAR data and then we made the classification process.

Field data was collected roadside using the Locus Map Pro applications. After, with QGIS software and Sentinel 2 optical images true color (10 meters spatial resolutions), we drew the polygons (630 in total), avoiding mix pixels in the edges.

We used 13 LULC classes (rainforest, savanna, campinarana, water, regeneration areas, sandbank or outcrop, buritizais, annual crops, perennial crops, pasture, forestry, conversion, anthropic areas). The polygons were randomly separated in 75% for training and 25% for the validation process.

Sentinel-1 images were obtained through the Copernicus Open Access Hub (also known as the Sentinels Scientific Data Hub) (https://scihub.copernicus.eu/), with an open-source toolbox in a python routine, named SentinelSat (https://github.com/sentinelsat/sentinelsat) [12]. It was downloaded all 2019 Sentinel-1 images. The pre-processed were performed using the Sentinel Application Platform-SNAP applications and python routines. In the total, was used 59 images (118 bands), being 30 from Sentinel-1A and 29 from Sentinel-1B.

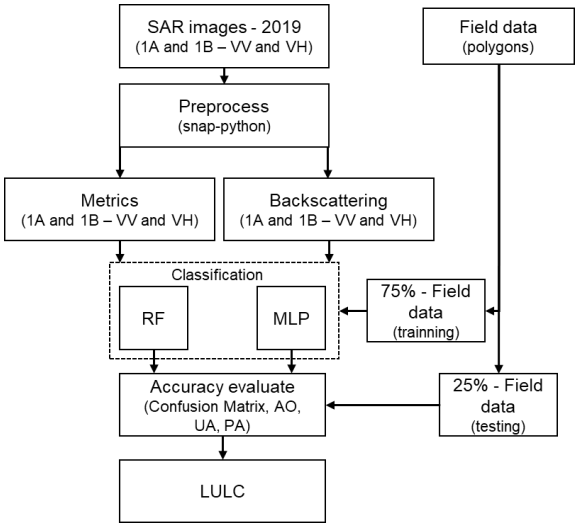

Fig. 1. Steps to process and evaluate the Land Use and Land Cover (LULC) classification with SAR images and Random Forest (RF) and Multilayer Perceptron (MLP) classifiers.

In the SAR data preprocessing, after several tests and bibliographic consults [13], we used the following procedure: apply orbit file; thermal noise removal without re-introduction; calibration: gamma enough; multilook with ground square pixel and number of ranges look equal to one, intensity output; speckle filter: refined lee; terrain correction: bilinear resample method, projection WGS84 22N UTM, 10 meters of spatial resolution and without mask out areas with elevations; and results were converted to decibels (dB). After the preprocessing, it was built virtual raster, utilizing the GDAL. This procedure is important to save disk space.

We performed the average, median, mean, standard deviation, variance, range and percentiles (25%, 50%, and 75%) metrics for each satellite and each polarization data set, total 36 metrics bands. For this step, we used the panda library in python routine.

For the LULC classification process, we utilized two machine learning classifiers, RF, and MLP. We performed in a python routine with the scikit-learn library [14]. In the RF, we tested different tree numbers, and in the MLP different layers size, alpha, and learning rate values. In total, were building 12 scenarios for each classifier. These scenarios were applied for the backscattering and the metrics Sentinel-1 SAR images. Accuracies were evaluated using a confusion matrix, performed the overall accuracy (OA), user's accuracy (UA), and producer's accuracy (PA) [15].

4. RESULTS

Fig. 2 shows the Overall Accuracy (OA) and training time results for each image data set and each classifier.

In general, MLP applied on backscattering images classification had better performance overall (scenario 10, layers:50, alpha: 10e-2, learning rate: 0.005, OA: 80.37%, Fig 3). RF shows better results for the metrics images, using 300 trees (scenario 12, OA: 75.05%).

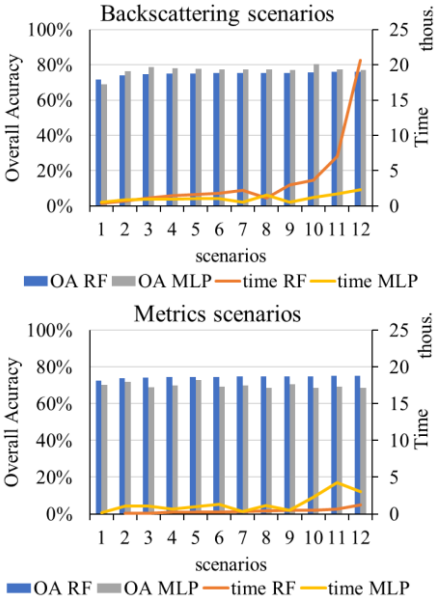

Fig. 2. Overall Accuracies (OA) and training time (in seconds) for each scenario on backscattering and metrics images.

Fig. 3 shows the better classification map achieved (Fig.2), with MLP classifier on the backscattering images. The accuracy was assessed for this better map, using the Users (UA) and Producers (PA) accuracies (Fig. 4). Fig. 5 shows Sentinel 1A and 1B backscattering temporal profile to help understand the results with the class seasonality.

Even the best accuracy does not provide higher accuracies (UA, PA) for all LULC classes. Some classes were better classified, like rainforest and savannas. However, for other classes, the classification is less possible or almost impossible only with SAR data use.

Water and sandbank/outcrop classes had similar accuracies with some confusion between them. The sandbanks are seasonal, occurring inside the river and depends on the rain season, which difficult the separation.

Annual crops and pasture classes showed confusion between them. Annual crops have PA: 46%, due to most of his reference data, were mapped as pasture. In this way, pasture class has a large commission error, UA: 39%, due to the confusion with annual crops. These two classes have similar temporal backscattering profiles between themselves and with savannas (Fig. 5), that difficulty the accurate separation.

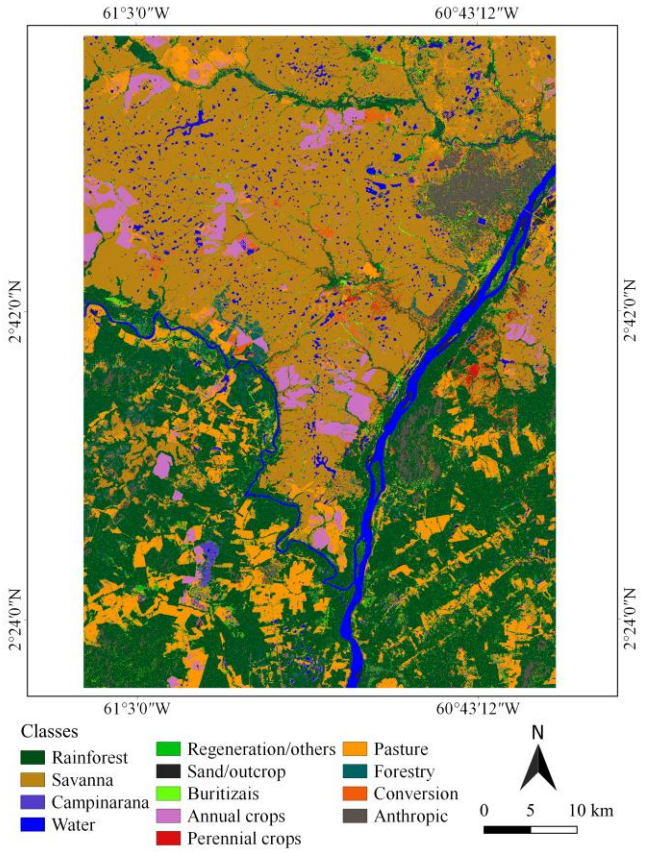

Fig. 3. Best classification scenario map with backscattering images and MLP algorithm.

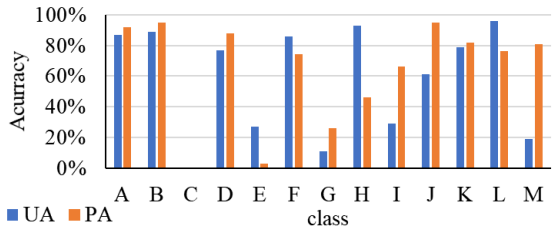

Fig. 4. Confusion matrix and User (UA) and Producer Accuracies (PA) for the best classification scenario with backscattering images and MLP algorithm. A: rainforest; B: savannas; C: campinarana; D: water; E: regeneration and others natural; F: sandbank and outcrop; G: buritizais; H: annual crops; I: perennial crops; J: pasture; K: forestry; L: conversion areas; M: anthropic areas

Regeneration areas, perennial crops, and burizitais classes have a similar geometric structure with rainforest and forestry. For this reason, these LULC classes have similar backscattering temporal profiles (Fig. 5). Consequently, SAR data do not provide accurate separation among them, occurring misclassification, mainly for the small classes (Fig. 3 and 4).

Nevertheless, with SAR data was possible to identify some land uses, as annual crops and pasture, and some land covers, like water, savannas, and rainforest. In this way, it is

possible to identify humans' impacts in environments, like the agriculture expansion over the small lakes in the lavrados.

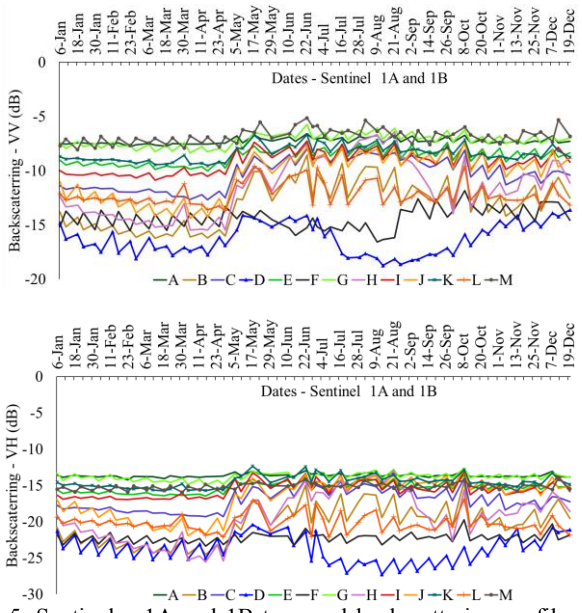

Fig. 5. Sentinel – 1A and 1B temporal backscattering profiles for VV and VH polarization in 2019. A: rainforest; B: savannas; C: campinarana; D: water; E: regeneration and others natural; F: sandbank and outcrop; G: buritizais; H: annual crops; I: perennial crops; J: pasture; K: forestry; L: conversion areas; M: anthropic areas.

5. CONCLUSIONS

This study showed the potential of SAR satellite data to provide the LULC mapping in a tropical region with high cloud cover. With the Sentinel-1 SAR data, it was possible to discriminate the more heterogeneous LULC classes, such as rainforest, savannas, sandbank/outcrop, and water. However, only the SAR data could be not enough to discriminate similar classes, such as buritizais, perennial crops, regenerations areas. Future work will include efficient use of optical data for the classification process and it will quantify the multi-year LULC changes.

6. ACKNOWLEDGEMENT

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001 and the National Council for Scientific and Technological Development (CNPq). We appreciate the Embrapa Roraima fieldwork support by Dr. Haron Xaud and Dr. Maristela Xaud.

7. REFERENCES

[1] A. Veloso et al., "Understanding the temporal behavior of crops using Sentinel-1 and Sentinel-2-like data for agricultural applications," Remote Sens. Environ., vol. 199, pp. 415–426, 2017.

[2] J. A. P. Pavanelli, J. R. dos Santos, L. S. Galvão, M. R. Xaud, and H. A. M. Xaud, "PALSAR-2/ALOS-2 and OLI/Landsat-8 data integration for land use and land cover mapping In Northern Brazilian Amazon," Bol. Ciências Geodésicas, vol. 24, no. 2, pp. 250–269, 2018.

[3] F. F. Camargo, E. E. Sano, C. M. Almeida, J. C. Mura, and T. Almeida, "A Comparative Assessment of Machine-Learning Techniques for Land Use and Land Cover Classification of the Brazilian Tropical Savanna Using ALOS-2/PALSAR-2 Polarimetric Images," Remote Sens., vol. 11, no. 13, p. 1600, Jul. 2019.

[4] C. Liu, J. Shang, P. W. Vachon, and H. McNairn, "Multiyear Crop Monitoring Using Polarimetric RADARSAT-2 Data," IEEE Trans. Geosci. Remote Sens., vol. 51, no. 4, pp. 2227–2240, Apr. 2013.

[5] S. C. Steele-Dunne, H. McNairn, A. Monsivais-Huertero, J. Judge, P.-W. Liu, and K. Papathanassiou, "Radar Remote Sensing of Agricultural Canopies: A Review," IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., vol. 10, no. 5, pp. 2249–2273, May 2017.

[6] S. Skakun, N. Kussul, A. Y. Shelestov, M. Lavreniuk, and O. Kussul, "Efficiency Assessment of Multitemporal C-Band Radarsat-2 Intensity and Landsat-8 Surface Reflectance Satellite Imagery for Crop Classification in Ukraine," IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., vol. 9, no. 8, pp. 3712–3719, 2016.

[7] R. Jhonnerie, V. P. Siregar, B. Nababan, L. B. Prasetyo, and S. Wouthuyzen, "Random Forest Classification for Mangrove Land Cover Mapping Using Landsat 5 TM and Alos Palsar Imageries," Procedia Environ. Sci., vol. 24, pp. 215–221, 2015.

[8] R. I. Barbosa and C. G. Bacelar-Lima, "Notas sobre a diversidade de plantas e fitofisionomias em Roraima através do banco de dados do herbário INPA," Amaz. Ciência Desenvolv., vol. 4, no. 7, pp. 131–154, 2008.

[9] W. D. de Carvalho and K. Mustin, "The highly threatened and little-known Amazonian savannahs," Nat. Ecol. Evol., vol. 1, no. 4, p. 0100, Mar. 2017.

[10] R. Torres et al., "GMES Sentinel-1 mission," Remote Sens. Environ., vol. 120, pp. 9–24, May 2012.

[11] T. Tamm, K. Zalite, K. Voormansik, and L. Talgre, "Relating Sentinel-1 Interferometric Coherence to Mowing Events on Grasslands," Remote Sens., vol. 8, no. 10, p. 802, Sep. 2016.

[12] J. Hu, P. Ghamisi, and X. Zhu, "Feature Extraction and Selection of Sentinel-1 Dual-Pol Data for Global-Scale Local Climate Zone Classification," ISPRS Int. J. Geo-Information, vol. 7, no. 9, p. 379, Sep. 2018.

[13] D. Small, "Flattening Gamma: Radiometric Terrain Correction for SAR Imagery," IEEE Trans. Geosci. Remote Sens., vol. 49, no. 8, pp. 3081–3093, Aug. 2011.

[14] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," J. Mach. Learn. Res., vol. 12, pp. 2825–2830, 2011.

[15] P. Olofsson, G. M. Foody, M. Herold, S. V. Stehman, C. E. Woodcock, and M. A. Wulder, "Good practices for estimating area and assessing accuracy of land change," Remote Sens. Environ., vol. 148, pp. 42–57, May 2014.