

ASSESSING DIFFERENTIATION BETWEEN PASTURE AND CROPLANDS USING REMOTE SENSING IMAGE TIME SERIES METRICS

*Marcos Antônio de Almeida Rodrigues, Hugo do Nascimento Bendini, Anderson Reis Soares, Thales Sehn Körting, Leila Maria Garcia Fonseca*¹

¹ Image Processing Division, National Institute for Space Research (INPE, Brazil)

ABSTRACT

Pasture and croplands comprise two different types of land use, which are very common in Brazil. Mapping these areas using remote sensing techniques is a challenge when using a single date image due to their similarity in spectral response. Time series might aid in discrimination of these areas once it explores the temporal behavior of surface patterns. In this work we explore time series obtained from remote sensing images to separate pasturelands from croplands in Brazilian Cerrado biome, using metrics derived from a data cube. We used Landsat 8 imagery as data source to compose a time series of six bands from OLI sensor (2 to 7) for the year of 2018. Random Forest algorithm was elected to execute the classification obtaining global accuracy of 80% and 0.58 of Kappa.

Index Terms— Time Series, Data cube, Satellite imagery, Image metrics, Random Forest algorithm

1. INTRODUCTION

Pasture and croplands play an important role in the economic and political Brazil's scenarios. While pasturelands are considered in the country as land reserve, croplands figure in the frontline of the Amazon arc of deforestation [1]. Pasturelands are in the center of discussion concerning the intensification of the livestock production [2]. Surveys showed that in the latter years, pasturelands in Brazil faced a reduction due to this intensification process [1]. From 1985 to 2002, the country underwent a great increasing in pastureland extension, and from 2002 to 2017, this increasing was diminished by the growing of this intensification process, which went through in country territory. This transformation in the territory dynamics may have been driven by the increasing of this intensification process with the adoption of practices such as integrated crop-livestock systems [1, 2].

Deforestation actions are associated with the increasing of croplands [3]. Generally, deforestation areas are first converted into croplands [4]. This is what is seen in the north and mid-west regions of the country [1]. Soybeans, maize and rice lead the plantation chain, which conduct the land conversion in the states located in these regions [2]. Therefore, mapping land use and land change is mandatory for government strategy and administration.

Although remote sensing provides a bunch of tools that allow achieving the goals just mentioned, simply applying the tools do not guarantee success in the analyses, mainly using traditional techniques. It means that new approaches are required to deal with new challenging demands. A technique largely used in remote sensing is the image classification, which extracts from the satellite imagery the classes that compose land use and land cover. It is commonly applied in a single date satellite imagery. To obtain good results, targets to be classified must be spectrally distinct. Problems ascend, though, when dealing with targets on surface that presents similarity in the spectral response, although they hold different usages. Since classification algorithms use the feature space to find patterns, once we have different targets with similar spectral response, classifiers become not able to clearly differentiate them.

Given this contextualization, we are now able to say that this is the challenge behind the mapping of the pasture and croplands utilizing automatic techniques. In this work we tackle the problem of differentiating pasturelands from croplands in a subset of Brazilian Cerrado biome, utilizing new techniques of time series analyses, based on metric images created by aggregation of multitemporal satellite imagery.

2. MATERIALS AND METHODS

For the accomplishment of this work, it was used Landsat 8 imagery as data source [5]. The year of 2018 was chosen for the generation of time series once it is the most recent complete year before this work have been performed. Study area comprehends a subset dimensioned by 25 km x 25 km, approximately, completing an area of about 625 km of extension over the Brazilian Cerrado biome (Figure 1). This subset is located in the state of Goiás and takes part of four municipalities, which are Caçu, Cachoeira Alta, Itarumã, and Paranaiguara. This area was elected for holding great variability of both patterns that are being analyzed in this work, and thus playing as a good area to assess the methodology.

Considering that Landsat 8 has a temporal resolution of 16 days, and that the more observations, the better temporal characterization, study area was purposely disposed in the overlaying swath of two Landsat orbits, intending to increase the amount of observations, and thus attain more loyal time series to each pattern.

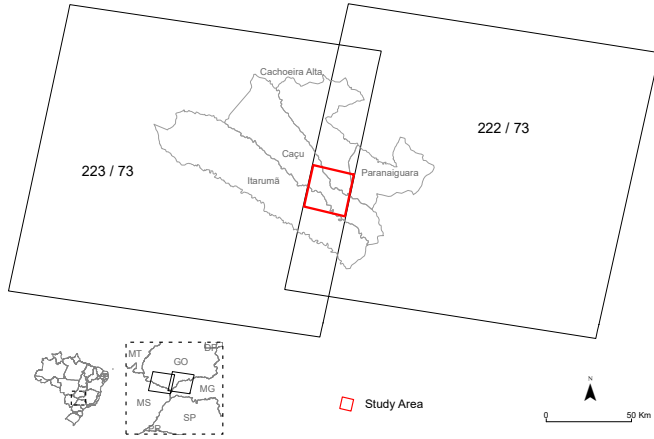


Fig. 1: Study area designed for methodology assessment.

With this strategy, we could enhance the quantity of observations, going from 16 days interval to 7-9 [6]. Two scenes were used to compose the time series, which are represented by path/row 223/73 and 222/73. Each Landsat scene disposes 22-23 images a year due to its revisit time. Working on the overlaying swath it was possible to obtain 45 images of the same area, once this area was seen for two Landsat orbits.

After downloading all images of both scenes from Earth Explorer repository, they were all involved in a first pre-processing step of reprojecting, once their reference system comes set to the Northern hemisphere. Afterward, they were clipped to the subset that was chosen as the study area. Regarding this work seek for assessing a methodology to well discriminate pasture from croplands, it would not be reasonable having other types of land cover in the study area, once this could even disturb the differentiation analyses between the main patterns herein explored. For this sake, all types of land use non pasturelands or crop areas were masked with the land cover mapping from Mapbiomas project, collection 4, for the same year of the images obtained in this work [7].

Six bands were used in this work, which comprise band 2 to band 7 of OLI sensor. Each band may be considered herein such as a dimension for the time series. Thus, for each band it was calculated seven temporal metrics [8], what consists in an aggregation of the pixel values along the time, creating new images with the time information embedded in [9, 10]. The metrics that were calculated are the following: mean, median, standard deviation, minimum, maximum, amplitude and sum. After these new images were created, they were all assembled in a single dataset containing 42 resulting metric images, which were obtained from the 7 metrics applied to the 6 spectral bands from the 45 Landsat images. This dataset was then used to feed the classifier algorithm in order to produce the classification image. Random Forest [11] were the elected algorithm to perform the classification, due to its ability to deal with great amount of data maintaining high level of quality. Flow chart depicted in Figure 2 presents the methodological draw.

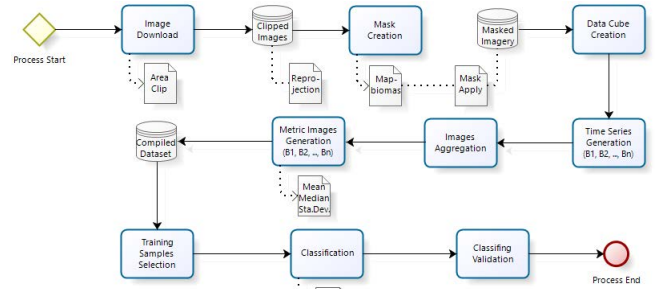


Fig. 2: Flow chart depicting the methodology.

Random Forest algorithm was trained utilizing 100 trees and Mapbiomas classification as ground truth reference. It was established 500 points as samples, which were widespread along the image (dataset) according to the size of each pattern previously observed, i.e. proportionally to its area [12]. Therefore, 210 samples were selected for croplands and 290 for pasturelands, once the latter holds a larger area than the first in study subset. Also, it is worth mentioning that samples were randomly and automated generated. As well as for training samples, accuracy assessment points were either automated, randomly and proportionally widespread [13], hereon with another set of samples comprised of 200 points, likewise using Mapbiomas classification as reference. Thus, 84 and 116 points compose the croplands and pasturelands accuracy set of samples, respectively.

3. RESULTS AND DISCUSSION

Seeking for assessing time series in terms of cloud coverage, images were visually evaluated through their covering level. All images were then distributed among six classes, which expresses their cloud coverage level. These are the established classes, from the lowest to the highest level: (i) free at all; (ii) rather free; (iii) low medium; (iv) high medium; (v) almost total; and (vi) total. Regarding the amount of observations in the study area, images affected by clouds, according to the classes above described, follow these proportions: (i) 17.78%; (ii) 13.33%; (iii) 20.00%; (iv) 17.78%; (v) 15.56%; and (vi) 15.56%. Overall, it means that only 8 images, from the total of 45 observations in 2018, are completely free of clouds. Figure 3 depicts samples of images representative of cloud coverage classes.

Instead of aggregating all images into new metric images, we noticed that it would be necessary to remove from the time series those images affected by clouds. Even if few images affected by clouds were involved in aggregating process, they would still injury the resultant image with undesired artifacts, and ultimately harm the classification process. Next figure depicts the increase of image harming with the increasing of images with clouds. As shown in Figure 4, as the number of cloudy images increases, as harmer the resultant metric image becomes, even if only images with few clouds are involved in aggregating process.

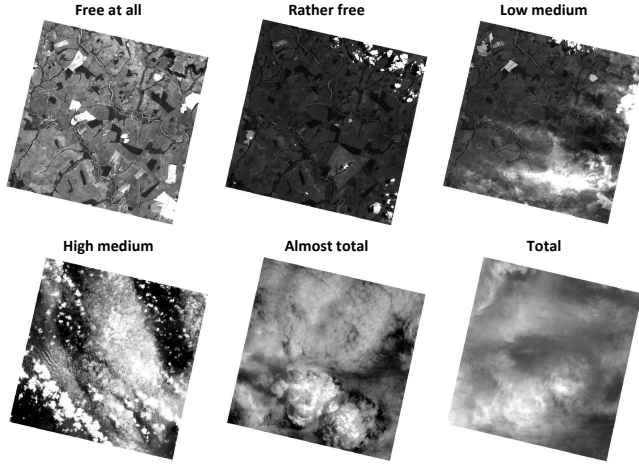


Fig. 3: Image samples depicting examples of the cloud coverage types utilized to evaluate every image.

Considering it, we saw the necessity of taking out of the time series all images affected by clouds, even those evaluated as rather free, for the reason above explained. This implies that only free at all (of clouds) images would work in our methodology. This way we had eight images composing the time series in each band/attribute, unlike the 45 initials. All metric images were then recreated, hereon using only images with no clouds, and the dataset, reconstructed. Although the number of images comprising the time series were diminished, the number of metric images in the final dataset remained the same, for quantity of calculated metrics and number of attributes keeps likewise the same. This change only affected the density of time series, making them less dense and irregularly spaced. Moreover, it may have mischaracterized the temporal behavior of targets, making it harder for the classifier to identify particularities in the time series that could cause each pattern to be unique and hence well classified by the algorithm.

Classification result with the Random Forest algorithm may be seen in Figure 5. Accuracy assessment is available at the confusion matrix presented in Table 1. As shown, global accuracy achieved a percentile of 80.00%, and a Kappa index of 0.58. User’s and producer’s accuracies of croplands reached the values of 80.56% and 69.05%, respectively. Regarding the pasturelands, these values are 79.69% and 87.93%, respectively.

Table 1: Confusion matrix.

		Ground Truth			User Accuracy	Comission Error
		Crop Areas	Pasture-lands	Totals		
Classification	Crop Areas	58	14	72	80.56%	19.44%
	Pasture-lands	26	102	128	79.69%	20.31%
	Totals	84	116	200	Global Accuracy	Kappa Index
	Producer Accuracy	69.05%	87.93%	-	80.00%	0.58
	Omission Error	30.95%	12.07%	-		
	Error					

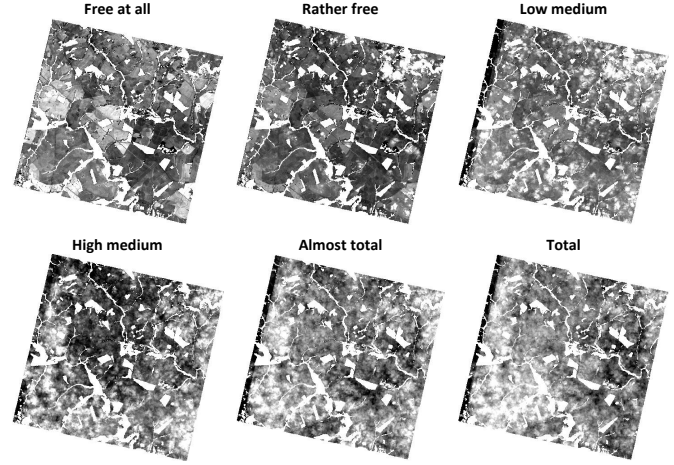


Fig. 4: Mean images of band 2 depicting the harm increasing with the increase of cloudy images in the aggregating process.

Regardless user and producer percentages, overall results indicate that the classification was slightly well performed, and our proposed methodology was able to catch the temporal behavior of the analyzed patterns, it is, their specificities along the time series, and thus discriminate one land use from another. In spite of global accuracy, user and producer accuracies of either croplands and pasturelands revealed some misclassification, mainly when considering the commission and omission errors of both land uses, which were a little higher than expected, but not enough to disqualify the classification mapping. The main aspect concerning the single date classification, and which causes it to be voided to separate croplands from pasturelands is the different appearance experienced by croplands throughout the year (i.e. fallow and growing seasons), whereas pasturelands remain regular the year along. And when we look at the classification result and compare it to the original images, we do see these different characteristics of the croplands assembled in the same class in the map. This reinforces the thesis that our methodology was capable to deal with patterns variations along the time and perform a good classification based on it.

Despite this approach, our results may have been limited by issues encountered during the process, such as cloud coverage in most images of the analyzed period. This caused our time series to be quite reduced and not equally spaced, and thus, certainly much important specificities concerning the patterns, which could boost the classification and the differentiation between them, were lost.

4. CONCLUSIONS AND FINAL REMARKS

The use of time series and metric images generated by aggregation of multitemporal satellite imagery presented a great potential to classify our dataset and separate targets spectrally similar, such as pasturelands and croplands.

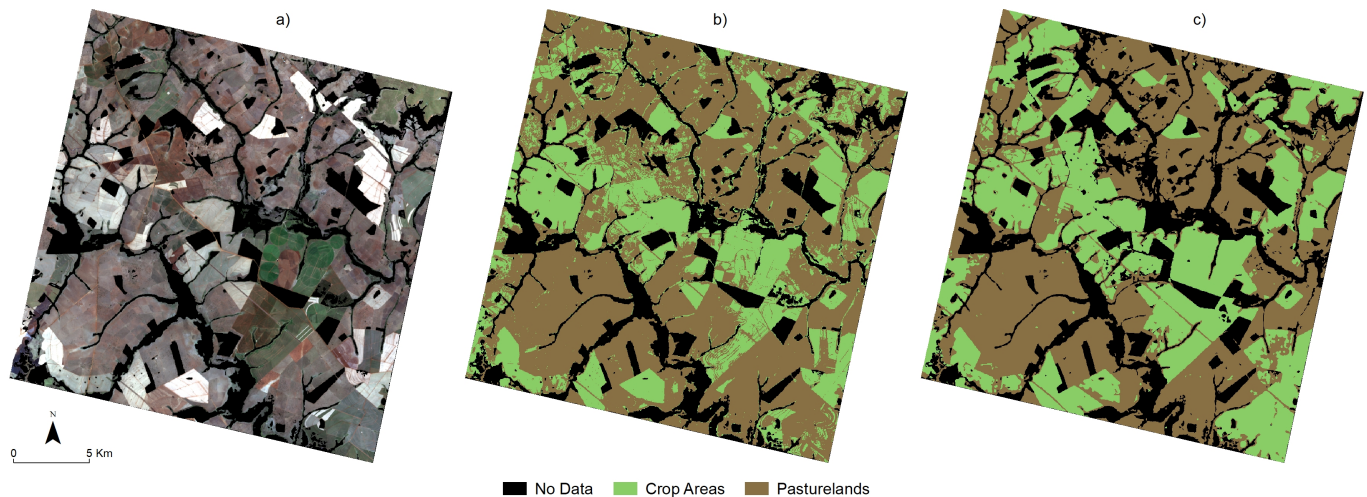


Fig. 5: Comparison between: a) True Color Image (09/Set. /2018); b) Classification result; and c) Mapbiomas reference.

Despite the reduce time series underwent due to cloud coverage in most images on the analyzed period, classification attained satisfactory results, achieving 80% global accuracy and kappa index of 0.58. Misclassification may be addressed to the lack of temporal information concerning the time series.

Increasing the time series by either amplifying the period of analyzes or observations amount, might improve the classifier differentiation ability, since it will be using a more accurate dataset in terms of temporal behavior. This strategy may be useful to reduce the impact magnitude of cloud coverage over the time series.

5. ACKNOWLEDGMENTS

This work was developed as part of the project “Development of systems to prevent forest fires and monitor vegetation cover in the Brazilian Cerrado”, with financial support of the Forest Investment Program (World Bank Project #P143185).

6. REFERENCES

- [1] L. Parente, V. Mesquita, F. Miziara, L. Braumann, and L. Ferreira, “Assessing the pasturelands and livestock dynamics in brazil, from 1985 to 2017: a novel approach based on high spatial resolution imagery and google earth engine cloud computing,” *Remote Sensing of Environment*, vol. 232, pp. 111301, October 2019.
- [2] L. Parente, L. Ferreira, A. Faria, S Noqueira, F. Araújo, L. Teixeira, and S. Hagen, “Monitoring the brazilian pasturelands: a new mapping approach based on the landsat 8 spectral and temporal domains,” *International Journal of Applied Earth Observation and Geoinformation*, vol. 62, pp. 135–143, January 2017.
- [3] L. Parente and L. Ferreira, “Assessing the spatial and occupation dynamics of the brazilian pasturelands based on the automated classification of modis images from 2000 to 2016,” *Remote Sensing*, vol. 10, pp. 606, April 2018.
- [4] L.L. Parente and L. Ferreira, “Desafios e oportunidades para o monitoramento da cobertura terrestre brasileira utilizando séries temporais landsat,” in *Proceedings... Simpósio Brasileiro de Sensoriamento Remoto - SBSR*. INPE, 2017, vol. XVIII, p. 2146–2152.
- [5] D.P. Roy, M.A. Wulder, and T.R. Loveland et al., “Landsat-8: Science and product vision for terrestrial global change research,” *Remote Sensing of Environment*, vol. 145, pp. 154–172, April 2014.
- [6] A.J.B. Luiz, I.D. Sanches, and K. et al. Trabaquini, “Dinâmica agrícola em área de sobreposição de órbitas adjacentes dos satélites landsat,” in *Proceedings... Simpósio Brasileiro de Sensoriamento Remoto - SBSR*. INPE, 2015, vol. XVII, p. 6381–6388.
- [7] A. Alencar, J.Z. Shimbo, and F. Lenti et al., “Mapping three decades of changes in the brazilian savanna native vegetation using landsat data processed in the google earth engine platform,” *Remote Sensing*, vol. 12, pp. 924, March 2020.
- [8] T.S. Korting, G. Câmara, and L.M.G. Fonseca, “Land cover detection using temporal features based on polar representation,” in *Proceedings... Conference on Graphics, Patterns, and Images*. SIBGRAPI, 2013, vol. XXVI.
- [9] J. Gregory, “The cf metadata standard,” *CLIVAR Exchanges*, vol. 8, pp. 4, 2003.
- [10] M. Appel and E. Pebesma, “On-demand processing of data cubes from satellite image collections with the gdalcubes library,” *Data*, vol. 4, June 2019.
- [11] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, pp. 5–32, 2001.
- [12] R.R. Colditz, “An evaluation of different training sample allocation schemes for discrete and continuous land cover classification using decision tree-based algorithms,” *Remote Sensing*, vol. 7, pp. 9655–9681, July 2015.
- [13] P. Olofsson, G. 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 Sensing of Environment*, vol. 148, pp. 42–57, May 2014.