Conference Proceedings Paper

Data Mining Using NDVI Time Series Applied to Change Detection

Andeise Cerqueira Dutra [1*], Yosio Edemir Shimabukuro [1] and Maria Isabel Sobral Escada [1]

[1] National Institute for Space Research (INPE), Av. dos Astronautas, 1758, São José dos Campos –SP, Brazil, CEP 12227-010

* Correspondence: andeise.dutra@inpe.br

Received: 2017.12.06; Accepted: 2017.12.08

Abstract: Quantifying and monitoring woody cover distribution in semiarid regions is challenging, due to their scattered distribution. Data mining has been widely used in remote sensing data for information extraction of spectral and temporal data in the analysis of change detection. The main objective of this study was to characterize the land cover and use over 2000-2010 time period for the brazilian Caatinga seasonal biome using a temporal NDVI series and Geographic Object-Based Image Analysis. For each of the target years was obtained NDVI images derived from MODIS (MOD13Q1, at 250 m spatial and 16-day temporal scale) sensor during the dry season to predict wood cover in the municipality of Buriti dos Montes, in the state of Piauí, Northeast region of Brazil (H13V09 *tile*). The images were automatically pre-processed and in the GEOBIA approach was performed image segmentation, spatial and spectral attribute extraction and labelled according to the following legend: Tree Cover (TC) and Cropland/Grass (CG), to obtain a classification using the decision tree supervised algorithm. Our results showed that approach using GEOBIA presented Kappa Index of 0.58 and Global Accuracy (GA) of 0.81% and showed better accuracy for the Tree Cover. Finally, we recommend new studies adding others parameters strongly related to vegetation of semiarid regions.

Keywords: Land cover change; deforestation; GeoDMA; semiarid; Caatinga.

1. Introduction

Semiarid regions present low and irregular precipitations, limited to a very short period of the year in large part of their extension. These regions are mainly characterized by a long period of rainfall reduction [1]. In Brazil, periods of drought are relatively frequent in the Northeast region as consequence of high interannual rainfall variability. Northeast Brazil has a type of vegetation adapted to semiarid conditions denominate caatinga (savanna), and the Caatinga biome cover an area of circa 844,453 km^2 or approximately 11% of the Brazilian territory and in which is populated by more than 27 million inhabitants [2].

However, the Caatinga biome is the third most degraded in Brazil and this region has suffered heavy losses of natural vegetation as a common practice for the preparation of land for agriculture, contributing to the loss of biodiversity. Moreover, partial or total removal of native vegetation has caused a reduction in the aerial biomass, a practice that has been carried out in a predatory manner due firewood be one of the main energy sources in these regions [3].

In this context, the monitoring and mapping is crucial to understand the vegetation and structure changes [4] and their variations over time. Thus, many studies have used remote sensing techniques to extract information in the analysis of temporal series through mining of spatial and spectral data to changes detection [5]. The Geographic Object-Based Image Analysis (GEOBIA) is based on topological information and geometric of the objects to the classification of images.

In this context, the aim of the present work is to classify the land cover and use of the brazilian Caatinga seasonal biome using Geographic Object-Based Image Analysis (GEOBIA) in a temporal NDVI series over 2000-2010 time period, in the municipality of Buriti dos Montes , Piauí.

2. Materials and Methods

2.1. Study Area

To accomplish our study goal, we investigated the municipality of Buriti dos Montes (Figure 1) located in the state of Piauí in the northwest Brazil. The municipality occupies an area of approximately 2653 km². Mean altitude is 500 meters and presents tropical climatic classification with dry season between July to October [6, 7].

The area is characterized by representative vegetation of semiarid regions, presenting tree and shrub savanna cover [8] and the main agricultural products are rice, beans and corn where native plants have been replaced [6].

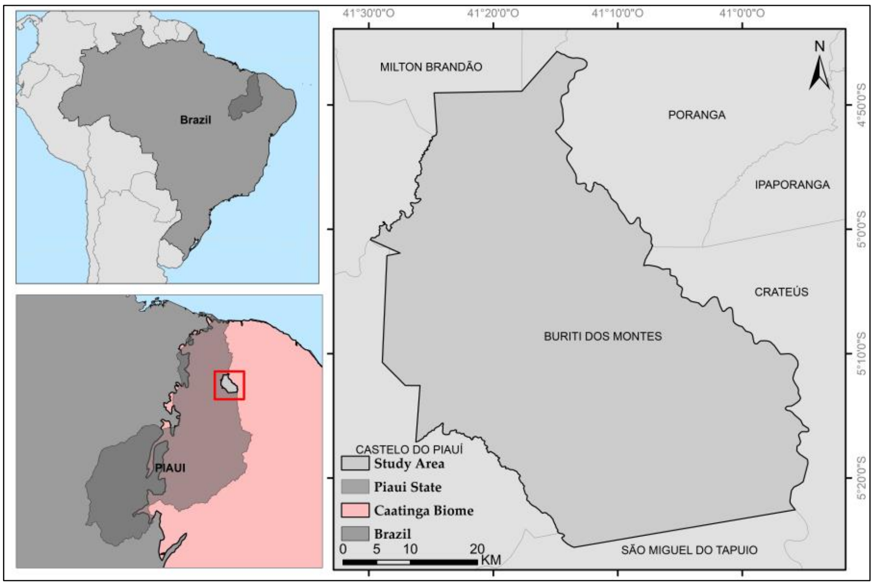

Figure 1 – Location of the study area in Piauí – Brazil.

2.2 Acquisition and pre-processing data

This study uses MODIS (Moderate Resolution Imaging Spectroradiometer) sensor imagery from MOD13Q1, a NDVI (Normalized Difference Vegetation Index) product available at a 250 meters spatial resolution [9], composed by a mosaic of 16 days of imaging. Based on the supposition that only trees and shrubs have active photosynthesis during dry season, we applied the methodology process only for the months of August, over the period 2000-2010 to the tile H13V09 available at EarthData – NASA (<https://ladsweb.modaps.eosdis.nasa.gov>) corrected atmospherically.

Two land use and cover maps were used for the years 2000 and 2010 as a reference for this study. The data were obtained from SAP (*Sistema de Alerta Precoce contra a Seca e Desertificação* - CCST / INPE) at a 30 meters spatial resolution derived from Landsat TM (Thematic Mapper) and Landsat ETM+ (Enhanced Thematic Mapper) sensors [8].

Very high spatial resolution sensor images from GeoEye 1 satellite were visualized in Google Earth Pro software. The images were used for visual interpretation of the targets and illustration of the land cover changes detected in MODIS time series.

In the pre-processing stage, the images of NDVI product were cut to the interest area in the limit of the municipality and then it were stacked to a single raster cube file in ENVI 5.1 environment and all the images were normalized to a range of 0 to 1 (Figure 2).

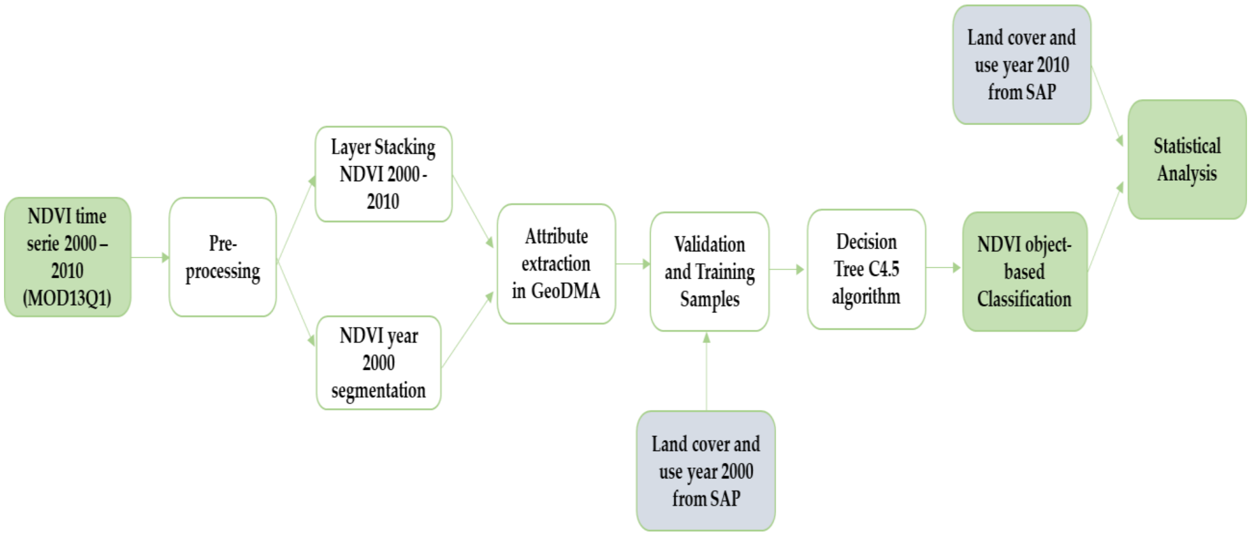

Figure 2 - Methodological work-flow including earth observation data for assessing changes in the vegetation.

The second part of the study was carried out in the TerraView 4.2.2 software to the processes involving data mining technique. The image segmentation, attribute extraction and sampling for training were the stages performed specifically through the GeoDMA (Geographical Data Mining Analyst) plug-in [10] which is used by GEOBIA for image classification [11].

For this procedure, the segmentation process was performed in the NDVI product referred to 2000 year. We used the segmentation algorithm based on the region growing [12], which Euclidean distance and minimum area parameters are used to divide the image in homogeneous spectrally regions. During this procedure, several segmentations were tested but the threshold that best fit the analyzed data was the values of 30 and 10 for Euclidean distance and minimum area respectively.

Subsequently, the spectral and spatial metrics were extracted using the segmentation results and the NDVI cube over the period. Thus, each object generated through segmentation has an attribute value calculated from the selected metrics.

For the classification process, it is necessary to select training and validation samples, which consists in the selection of pixels or homogeneous regions that best represent each one of the classes resulting in an object-based classification map. In this study we used the land cover and use mapping from SAP as a reference for year 2000. The selected samples were used for the classification based on decision tree by algorithm C4.5 contained in the GeoDMA plug-in. Objects were classified into two land cover classes: The Tree Cover (TC) class was defined in trees and shrubs savannas and Cropland/Grass (GC) class.

Due to the lack of appropriate field data required for assessing the quality of land cover map produced for our study, we opted to compare the land cover map from SAP of year 2010 as a reference mapping with the classification obtained by decision tree to evaluate the accuracy. The Kappa Index Kappa Index (Equation 1), obtained by the error matrix, Global Accuracy (Equation 2), Hypothesis Test by Z test (Equation 3), Producer Accuracy (Equation 4) and Consumer Accuracy (Equation 5).

$$\hat{k} = \frac{\theta_1 - \theta_2}{1 - \theta_2}, \quad where:\ \theta_1 = \frac{\sum_{k=1}^{c} x_{kk}}{n} \quad e\ \theta_2 = \frac{\sum_{k=1}^{c} x_{+k}}{n^2} \tag{1}$$

$$Global\ Accuracy:\ \frac{\sum_{k=1}^{c} X_{kk}}{n} \tag{2}$$

$$Z\ Test = \frac{\hat{k} - k}{\sqrt{Var(\hat{k})}} \sim N(0,1) \tag{3}$$

$$\textit{Producer Accuracy:}\ k = \frac{x_{kk}}{x_{+k}} \tag{4}$$

$$\textit{Consumer Accuracy:}\ k = \frac{x_{kk}}{x_{k+}} \tag{5}$$

Where kk is the sum of each element of the diagonal, + k is the total of the column of each class and k + is the total of the row of each class.

3. Results and Discussion

After performing the steps in GeoDMA, the C4.5 classifier generated the decision tree from the calculated spatial and spectral metrics, using those that best fit the data set.

Using the decision tree generated by the classification algorithm, we obtained the land cover and use classification map based on the NDVI time series. The generated map was compared in relation to the reference map of the year 2010 also of the SAP in order to generate the statistical analyzes (Figure 3).

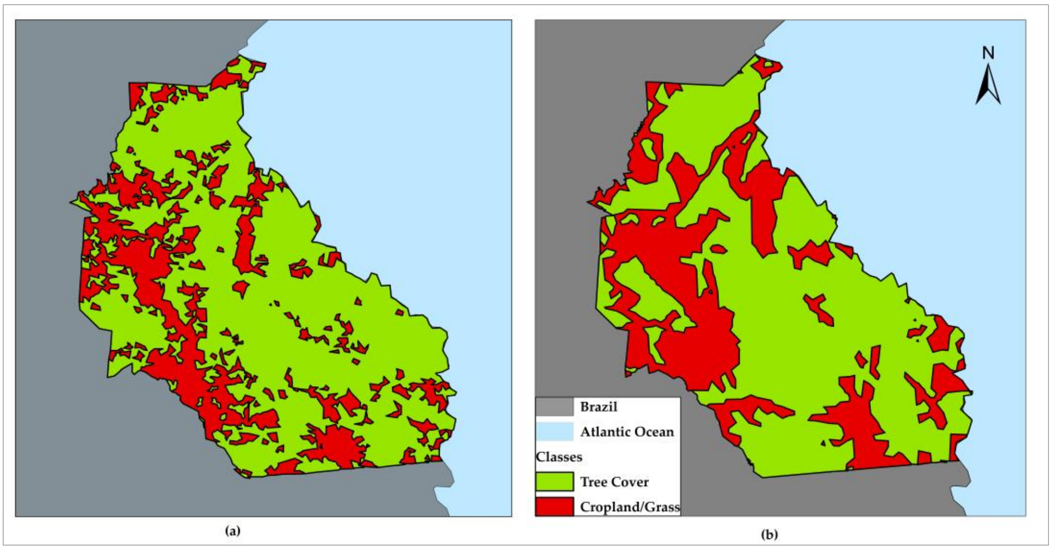

Figure 3 – (a) NDVI time series object-based Classification Map and (b) Land cover and land use reference data for the 2010 year by SAP adapted by author.

Based on the map generated and the reference map, it was calculated the extension for each class obtaining a result of 1927 km² for Tree Cover and 725 km² for Cropland/Grass using the classification based on the NDVI time series. On the other hand, for the reference map, the TC area covers approximately 1746 km² of the municipality of Buriti dos Montes and 907 km² for Cropland/Grass (Table 1).

Table 1. Area of Tree Cover and Cropland/Grass classes for object-based Classification and SAP Mapping for the year 2010 as a reference.

	Area 2010 (Km ²)	
	NVDI classification	SAP Reference
Tree Cover	1927	1746
Cropland/Grass	725	907
Total	2653	2653

According to [8] the municipality of Buriti dos Montes had about 81.5% of it is entire territory covered by Caatinga (tree and shrub savanna) in year 2000, however, there was a reduction to 65.82% in 2010, while there was a considerable increase of 15, 64% in agriculture.

In addition, the authors report twice the average of the outbreaks of fires for the year 2010, ranging from an average of 50 to 100 fires detected. They also point to the increase of the environmentally susceptible area index (IAS) reaching attributes of moderate and high susceptibility to the municipality.

The error matrix (Table 2) was generated after crossing the 265 sampling points with the mapping obtained with object-based classification. The Kappa Index found for this classification process was 0.58 and Global Accuracy of 81%. The main source of error occurred in the classification of 31 TC as a CG class. This pattern found in this study suggest that there are large variations between classes due to the patchiness of tree cover, also to scattered distribution of wood plants and distribution of cultivated fields. As well as values of producer and consumer accuracies are observed (Table 3).

Table 2. Error matrix for object-based classification.

Classes	Tree Cover	Cropland/Grass	Total
Tree Cover	151	31	182
Cropland/Grass	19	64	83
Total	170	95	265
Kappa Index			*0,58*
Global Accuracy			*0,81*
Z Test			*10,88*

Considering the application of the Z test, at 5% significance, in order to verify if there is agreement between the NDVI time series in a objet-based classification and the reference map fot the year 2010 used, we observed that a Z value of 10,88 means that there is agreement between both maps.

Table 3. Producer and Consumer Accuracy for NDVI Classification (%).

Classes	Producer Accuracy	Omission Error	Consumer Accuracy	Inclusion Error
Tree Cover	88,82	11,17	82,96	17,03
Cropland/Grass	67,36	32,63	77,10	22,89

It can be attributed that the result of the classification obtained from the time series of the NDVI did not obtain excellent results possibly due to differences in spatial resolution of the data used and the reference data.

4. Conclusions

The use of the GeoDMA computational application for extracting spatial and spectral metrics through data mining has proved to be an efficient and accessible tool for classifying orbital images of temporal NDVI series using the C4.5 algorithm.

The results indicate that the process of classification by data mining method allows to detect changes in land cover through from the NDVI product in a long period, especially in what concerns the expansion of agriculture in the municipality of Buriti dos Montes.

Finally, we recommended new approaches using earth observation data in a higher spatial resolution for better comparison with the reference data used in this work, as well as the addition of other parameters strongly related to vegetation of semiarid regions.

Acknowledgments: The authors thank the National Institute for Space Research (INPE/Brazil) and CAPES.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Jacques, D. C., Kergoat, L., Hiernaux, P., Mougin, E.; Defourny, P. Monitoring Dry Vegetation Masses In Semi-Arid Areas With Modis Swir Bands. *Remote Sensing of Environment*, **2014**, *153*, 40-49.
2. Ministério do Meio Ambiente (Mma). Biomas - Caatinga. <www.mma.gov.br/biomas/caatinga>. (accessed on 08 August. 2017).
3. Silva, A. C. D. C., Prata, A. P. D. N., Souto, L. S.; Mello, A. A. D. Aspectos de ecologia de paisagem e ameaças à biodiversidade em uma unidade de conservação na Caatinga, em Sergipe. *Rev. Árvore*, **2013**, Viçosa, V. 37, N. 3, P. 479-490.
4. Suemi Saito, N., Paiva Arguello, F.V., Alves Moreira, M., Rosa Dos Santos, A., Coelho Eugenio, F., Costa Figueiredo, A. Uso da geotecnologia para análise temporal da cobertura florestal. *Cerne*, **2016,** *22*(1).
5. Korting, T.S., GeoDMA: A Toolbox Integrating Data Mining with Object-Based and Multi-Temporal Analysis of Satellite Remotely Sensed Imagery. (Doctoral Thesis in Remote Sensing). Inpe-Instituto Nacional De Pesquisas Ambientais. São José Dos Campos), 2012.
6. Ibge. Instituto Brasileiro De Geografa E Estatística. Banco De Dados Agregados. <www.ibge.gov.br>. (accessed on 08 August. 2017).
7. Kottek, M., Grieser, J., Beck, C., Rudolf, B. And Rubel, F. World Map of The Köppen-Geiger Climate Classification Updated. *Meteorologische Zeitschrift*, **2006**, *15*(3), Pp.259-263.
8. Vieira, R. M. S. P.; Tomasella, J.; Alvalá, R. C. S.; Sestini, M. F.; Affonso, A. G.; Rodriguez, D. A.; Barbosa, A. A.; Cunha, A. P. M. A.; Valles, G. F.; Crepani, E.; Oliveira, S. B. P.; Souza, M. S. B.; Calil, P. M.; Carvalho, M. A.; Valeriano, D. M.; Campello, F. C. B.; Santana, M. O. Identifying Areas Susceptible To Desertification In The Brazilian Northeast. *Solid Earth*, **2015**, 6, 347-360, Doi:10.5194/Se-6-347-2015.
9. Kamel, D. Mod13q1 Modis/Terra Vegetation Indices 16-Day L3 Global 250m Sin Grid. Nasa Lp Daac. University Of Arizona, Alfredo Huete - University Of Technology Sydney And Modaps Sips – Nasa, 2015.
10. Korting Ts, Garcia Fonseca Lm, Câmara G. Geodma—Geographic Data Mining Analyst. *Computers & Geosciences*, **2013**. Doi: 57:133-145
11. Korting, T. S.; Fonseca, L. M.; Escada, M. I. S.; Silva, F. C.; Silva, M. P. S. Geodma: A Novel System For Spatial Data Mining. Ieee International Conference On Data Mining Workshops, 2008.
12. Bins, L.; Fonseca, L. M.; Erthal, G. J.; Mitsuo Ii, F. "Satellite Imagery Segmentation: A Region Growing Approach". In: Simpósio Brasileiro De Sensoriamento Remoto, 8., 14-19, Salvador, Brasil Abr. 1996, Anais... São José Dos Campos: Inpe, P. 677-680.