Mapping Grasslands Formations and Cultivated Pastures in the Brazilian Cerrado Using Data Mining

Wanderson Costa, Leila Fonseca, Thales Körting
Image Processing Division
National Institute For Spatial Research
São José dos Campos, Brazil
{wscosta, leila, thales}@dpi.inpe.br

***Abstract*—Cerrado is the second largest biome in Brazil. Among the land changes in the Cerrado, over 500,000 km² of the biome have been transformed into cultivated pastures in recent years. Distinguishing the native formations and identifying types of land use and cover in the Cerrado are important tasks for monitoring and protection policy of the biome. Within this context, this work aims at developing a methodology based on remote sensing techniques, to map pasture and native grassland areas in the Brazilian Cerrado. Data related to land use and cover, relief, spectral information from Landsat images and vegetation indices were used to perform the image classification. Decision trees and Support Vector Machines algorithms were used, and the results showed that the analysis and integration of different data sources can aid in the classification process. In order to discriminate areas of cultivated pastures and grassland formations, we obtained accuracies up to 82% in the study area, being able to identify attributes and data required to recognize these areas in the Brazilian Cerrado by remote sensing images.**

Keywords-Image Processing; Data Mining; Cerrado.

I. INTRODUCTION

Serious environmental risks can arise and pose a threat of global character when natural resources are not used properly [1]. Within this perspective, it can be mentioned the problems caused by changes in land use and land cover in the second largest biome in Brazil, the Cerrado [2]. More than half of Brazilian Cerrado's area has been transformed, mainly to make room for cattle and cash crops, losing more than 1,000,000 km² of its original vegetation [3].

Croplands cover more than 100,000 km² and pastures surpass 500,000 km², whereas protected areas comprise only about 33,000 km². The destruction of the Cerrado vegetation is three times larger than the amount of the deforested area in the Amazon region. Deforestation ranges from 22,000 to 30,000 km² per year, even higher than those observed in Amazon Forest. This scenario represents a high environmental cost and implies loss of biodiversity, soil erosion, vegetation degradation, water pollution, instability of the carbon cycle and probable regional climatic changes and variations in fire events, which are typical of the biome [4].

There is a large number of definitions for Cerrado and, as a result, several proposals for classification of native formations can be found in the literature. Among the physiognomic types presented in the biome, there are the grassland formations, which include the physiognomies of *Clean Field*, *Dirty Field* and *Rocky Field* [5]. We argue that with the identification and monitoring of these vegetation types using satellite imagery, policies can promote its physical, chemical and biological integrity, and estimate the productivity of the degraded regions [6].

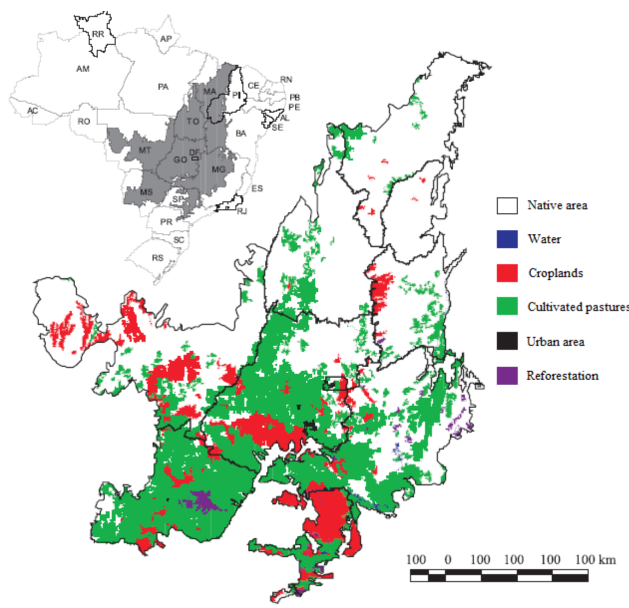

Figure 1. Spatial distribution of cultivated pastures in the Brazilian Cerrado biome. Source: [3].

In order to promote the recovery of degraded areas and the policies that protect the biome, it is essential to create maps to analyze the land use and land cover of the savanna. However, cultivated pastures, similarly to native grasslands, may vary from formations with predominance of grasses to areas that present dominance of pioneer trees and shrub species [7]. Therefore, the mapping of pasture areas and native formations in the Cerrado is a difficult task if only the spectral information obtained by satellite images is used [8].

To overcome this problem, this work proposes a methodology for mapping areas of Cultivated Pasture and Native Vegetation (Clean Field, Dirty Field and Rocky Grasslands), using data mining from integration of satellite imagery data in multiple resolutions.

The remainder of the paper is organized as follows. In Section 2, we present a brief description of the Brazilian

 ISBN: 978-1-61208-326-1

Cerrado. Section 3 describes the cultivated pastures in the biome. The methodological procedures and data used are depicted in Section 4. In Section 5, we discuss the results obtained in this work. Finally, we describe the conclusion and future work in Section 6.

II. THE BRAZILIAN CERRADO

Cerrado is the Portuguese word for Brazil's plateau of savannas, grasslands, woodlands, and gallery and dry forests [9]. One of the most biodiverse regions on the planet, the Brazilian Cerrado has an area of approximately 2 million km², comprising about 24% of the Brazilian territory [10]. Cerrado is the richest tropical savanna in the world [4] and the biome occupies the central region of Brazil, extending from the northeast coast of the Maranhão state (MA) to the north of Paraná (PR) state, as shown in Figure 1. Overall, Cerrado can be understood as a grass field coexisting with scattered trees and shrubs [11] and it is the second largest of Brazil's major biomes, after Amazonia [4].

The biome is characterized as one that presents three types of formations: forests, savannas and grasslands. Generally, the grassland formation refers to regions with predominance of herbaceous species and some shrubs, without the occurrence of trees in the landscape [5].

The vegetation included in the grassland formations comprises areas of Clean Field, Dirty Field and Rocky Field (Figure 2). Regions of Dirty Field and Rocky Field present a physiognomic type predominantly herbaceous and shrubby. However, Rocky Field areas include micro-relief landscapes with typical species, which usually occupy regions of rocky outcrops at elevations above 900 m. On the other hand, the phytophysiognomy of Clean Field has a predominance of grasses interspersed with underdeveloped woody plants, without the presence of trees [5].

Figure 2. Grassland formations in the Brazilian Cerrado. Sources: [12], [13].

More than half of Cerrado's original vegetation has been transformed into cultivated pasture areas, agriculture and other uses, as shown in Table I. In addition, studies indicate that changes in land use in the Cerrado occur with greater intensity than in the Amazon region [14] [15].

TABLE I. PRINCIPAL LAND USE IN THE CERRADO. SOURCE: [4]

Land use	*Area (ha)*	*Percent core area*
Native areas	70,581,182	44.53
Cultivated pastures	65,874,145	41.56
Agriculture	17,984,719	11.35
Urban areas/bare soil	3,006,830	1.90
Planted forests	116,760	0.07
Others	930,304	0.59
Total	158,493,921	

Hence, these changes in land use impose substantial threats to ecosystems and species of the biome. Only 2.2% of its area is legally protected and various species of animals and plants are endangered [4]. Furthermore, approximately 20% of endangered endemic species no longer exist in preserved areas. Besides the vegetation degradation and soil erosion, the introduction of non-native species and the use of fire to create pastures can have an adverse impact on ecosystems and cause significant loss of biodiversity in the Cerrado [4].

III. CULTIVATED PASTURES

Cultivated pastures represent about 500,000 km² of the biome. It is important to analyze the degradation of cultivated pastures, since nearly 50% of the planted areas in grasslands are severely degraded (Figure 3), causing increased erosion, loss of soil fertility and the predominance of invasive species [16]. Therefore, not only can the recovery of these areas increase the producers' income, but also it can reduce the environmental impact by decreasing erosion, emission of carbon dioxide and opening new areas for cattle [17].

Figure 3. Examples of managed (left) and degraded pastures (right). Source: [7].

Some types of data can be extracted and analyzed to support the detection of these regions, such as biophysical (soil type, percentage of green cover and biomass), radiometric (Enhanced Vegetation Index – EVI and Normalized Difference Vegetation Index – NDVI) and climatic (precipitation) data. Moreover, wet and dry seasons are well defined in the region and, therefore, this fact may assist to identify the radiometric and biophysical characteristics of planted pasture areas [16].

The classification of cultivated pastures is difficult because the degradation of these regions can, for example, influence the percentage of vegetation cover and the response of vegetation indices. Misclassification may occur when pastures are managed improperly, since those areas might be detected as invasive species or even the revival of species of native shrubs and trees in these regions [16].

Therefore, in order to improve the discrimination of such targets it is necessary to use temporal and field data and also to better understand the biophysical properties of these areas. [8]. Thus, the analysis using images with different spatial and temporal resolutions can allow the more accurate identification of spatial and temporal patterns of the targets in the Cerrado.

IV. DATA AND METHODS

This study encompasses a Cerrado area that comprises a region of Serra da Canastra National Park and neighboring regions (Figure 4). The region is located in the south-central state of Minas Gerais, southeastern of Brazil. The chosen scene contains the targets of interest, both native grassland areas and cultivated pasture.

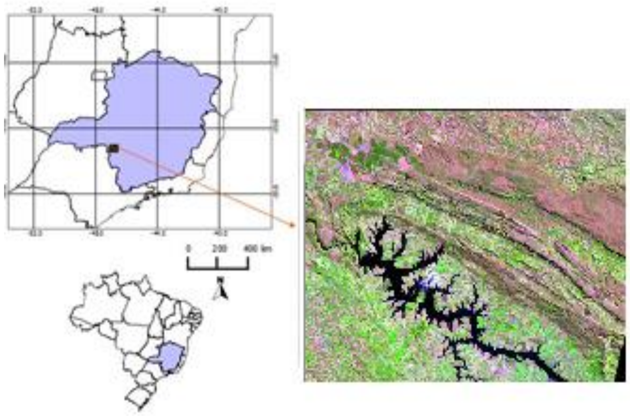

Figure 4. Location of the study area.

Information on cultivated pasture areas, for the year 2006, were provided by the Brazilian Ministry of the Environment. The regions of native areas, for the year 2009, were obtained from the Forest Inventory of Minas Gerais, created by the University of Lavras (UFLA), which ranks the grassland formations into two classes: Field (Clean and Dirty Fields) and Rocky Field. As a means to evaluate the classification results, these reference data were used as ground truth.

Therefore, the types of classes used in this work includes the following standards: Cultivated Pasture, Field and Rocky Field. The experiments were performed in the system GeoDMA [18]. This system performs image segmentation, classification, temporal analysis and data mining, and it is available on the internet.

Data related to relief, spectral information from Landsat images and vegetation indices EVI2 (Enhanced Vegetation Index 2) were used to perform the classification. For the classification of the areas, this work used the algorithms of decision tree C4.5 [19] and Support Vector Machines (SVM) [20], [21], using the system GeoDMA and the tools Weka [22] and LibSVM [23].

V. RESULTS

In order to implement an efficient methodology for distinguishing cultivated pastures and native vegetation, several experiments were performed. In the first experiment, it was used the Decision Tree (DT) classifier and only information about topography, obtained from altitude and declivity, called TOPODATA [24], with a spatial resolution of 30 meters, to map four classes: Cultivated Pastures, Field, Rocky Field and Others. The class Others covered all other native formations in the study that are neither Field nor Rocky Field. The results with 8,000 samples showed that regions below 871.47m were classified as Cultivated Pasture and areas above this value were classified as native vegetation. However, the accuracy of the results was only 60.0%.

In the second experiment, two Landsat-5 TM images (spatial resolution of 30m) were used for the orbit/point 220/74, acquired on 06/01/2006 and 07/27/2009. These images were georeferenced from NASA database (Global Land Cover Facility – GLCF). Various spectral attributes, such as mean, amplitude, homogeneity and entropy were extracted for bands 1-5 and 7. These were combined with the relief data previously used to classify the four classes. The classification accuracy based on decision tree slightly improved to 63.50%.

Because of the confusion between Field and Rocky Field classes, these were combined into a single class. For 12,000 samples, the results using the DT algorithm increased the accuracy to 71.65%. Finally, the SVM was also tested on the same set of samples, and the classification accuracy was 73.65%.

In the last experiment, it was added to the previously used set of attributes the time series of images EVI2 MODIS sensor (250m spatial resolution), with an interval of 16 days for the years 2006 and 2009. In addition to measuring accuracy, Kappa Coefficient [25] showed significant values in this experiment. Using the typology of four classes, the decision tree resulted in 60.15% of accuracy and Kappa equal to 0.468, while an accuracy of 67.20% and Kappa of 0.563 for SVM. With the combination of classes Field and Rocky Field, the accuracy increased to 73.27% (DT) and 82.12% (SVM), as well as Kappa Coefficient, that corresponded to 0.599 (DT) and 0.7319 (SVM). Table II summarizes the experiments performed so far.

TABLE II. EXPERIMENTS

Data used	*N. of classes*	*Accuracy (%)*	*Algorithm*
TOPODATA	4	60.00	DT
TOPODATA + Landsat	4	63.50	DT
TOPODATA + Landsat	4	71.65	DT
TOPODATA + Landsat	3	73.65	SVM
TOPODATA + Landsat + EVI2	4	60.15	DT
TOPODATA + Landsat + EVI2	4	67.20	SVM
TOPODATA + Landsat + EVI2	3	73.27	DT
TOPODATA + Landsat + EVI2	3	82.12	SVM

As shown in Table II, by the end of this stage of work, the experiment that combined different spatial resolutions (TOPODATA, Landsat and EVI2) and the SVM algorithm obtained the best classification, with an accuracy of 82.12% and Kappa of 0.7319.

VI. CONCLUSION AND FUTURE WORK

In order to develop a methodology to combine image processing and analysis to identify areas of pasture and grassland formations of the Brazilian Cerrado, some experiments were performed. The best classification result was the one that combined different spatial resolutions and

 ISBN: 978-1-61208-326-1

the SVM algorithm, with an accuracy of 82.12% and Kappa of 0.7319, distinguishing areas of cultivated pasture and native grassland.

However, further experiments will be carried out, by integrating images from different sources to aid in the characterization of targets. Images of high spatial resolution, as well as techniques of Linear Spectral Mixture Model and principal components will be tested to evaluate the most proper features for each of the targets.

REFERENCES

[1] J. Beddington, “Food, energy, water and the climate: A perfect history of global events?,” *Lecture to Sustainable development UK 09,* 2009, pp. 1-9.

[2] J. A. Ratter, J. F. Ribeiro, and S. Bridgewater, “The Brazilian Cerrado and threats to its biodiversity,” *Annals of Botany,* vol. 80, 1997, pp. 223-230.

[3] R. B. Machado et al., “Estimated loss of the Brazilian Cerrado area” Brasília, DF, 2004, pp. 1-26.

[4] C. Klink and R. Machado, “Conservation of the Brazilian Cerrado,” *Conservation Biology,* vol. 19, no. 3, Jun. 2005, pp. 707-713.

[5] J. F. Ribeiro and B. M. T. Walter, “The main phytophysiognomies in Cerrado,” in *Cerrado: Ecologia e Flora*, vol. 1, Brasília, EMBRAPA, 2008, pp. 152-212.

[6] R. R. Rodrigues and S. Gandolfi, “Recovery of Native Forests: General Principles and Allowances for a Methodology Definition,” *Revista Brasileira de Horticultura Ornamental,* vol. 2, no. 1, 2001, pp. 4-15.

[7] Embrapa and Inpe, “Survey information of use and land cover in the Amazon,” Brasília, 2011, pp. 1-20.

[8] L. G. Ferreira, E. E. Sano, L. E. Fernandez and F. M. Araújo, “Biophysical characteristics and fire occurrence of cultivated pastures in the Brazilian savanna observed by moderate resolution satellite data,” *International Journal of Remote Sensing*, vol. 34, no. 1, 2013, pp. 154-167.

[9] G. Eiten, "Delimitation of the concept of Cerrado," vol. 21, Rio de Janeiro: Arquivos do Jardim Botânico, 1977, pp. 125-134.

[10] M. Brossard and A. O. Barcellos, “Cerrado conversion into cultivated pastures and operation of latosols,” *Cadernos de Ciência & Tecnologia,* vol. 22, no. 1, 2005, pp. 153-168.

[11] B. M. T. Walter, “Phytophysiognomies of Cerrado biome: terminological synthesis and floristic relationships,” Brasília, 2006, pp. 1-373.

[12] IBGE, "Technical Manual of Brazilian Vegetation," 2 ed., vol. 1, Rio de Janeiro: Manuais Técnicos em Geociências, 2012.

[13] S. M. d. C. Coura, "Vegetation Mapping of Minas Gerais State Using MODIS Data," São José dos Campos, SP, 2006, pp. 1-150.

[14] E. E. Sano, A. O. Barcellos, and H. S. Bezerra, “Assessing the spatial distribution of cultivated pastures in the Brazilian savanna,” *Pasturas Tropicales,* vol. 22, 2001, no. 3, 2001, pp. 2-15.

[15] D. L. Skole, C. W. H. W. A. Salas, and C. A. Nobre, “Physical and human dimensions of deforestation in Amazonia,” *Biosciences,* vol. 44, no. 5, 2012, pp. 314-322.

[16] L. G. Ferreira, et al., “Biophysical Properties of Cultivated Pastures in the Brazilian Savanna Biome: An Analysis in the Spatial-Temporal Domains Based on Ground and Satellite Data,” *Remote Sensing,* vol. 5, no. 1, 2013, pp. 307-326.

[17] J. M. Chaves, L. Moreira, E. E. Sano, H. S. Bezerra, and L. Feitoza, “Using the segmentation technique to identify the main types of cultivated pastures in the Cerrado,” em *Simpósio Brasileiro de Sensoriamento Remoto, 10 (SBSR)*, Foz do Iguaçu, 2001, pp. 31-33.

[18] T. Korting, L. M. G. Fonseca, and G. Câmara, “GeoDMA - Geographic Data Mining Analyst: a framework for GIScience,” *Computers & Geosciences,* 2013, pp. 133-145.

[19] I. H. Witten and E. Frank, Data mining: practical machine learning tools and techniques with Java implementations, San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2000.

[20] C. Cortes and V. Vapnik, “Support-Vector Networks,” *Maching Learning,* vol. 20, no. 3, Set. 1995, pp. 273-297.

[21] F. Bovolo, G. Camps-Valls, and L. Bruzzone, “A support vector domain method for change detection in multitemporal images,” *Pattern Recognition Letters,* vol. 31, no. 10, 2010, pp. 1148-1154.

[22] M. Hall et al., “The WEKA Data Mining Software: An Update,” *SIGKDD Explorations,* vol. 1, no. 1, 2009.

[23] C.-C. Chang and C.-J. Lin, “Libsvm: a library for support vector machines,” *ACM Transaction on Intelligent Systems and Technology,* vol. 2, no. 3, 2011, pp. 1-27.

[24] M. d. M. Valeriano, “Digital model of morphometric variables with SRTM data nationwide: the project TOPODATA,” *Anais do XII Simpósio Brasileiro de Sensoriamento Remoto,* 2005, pp. 3595-3602.

[25] J. A. Cohen, “Coefficiente of agreement for nominal scales,” Educational and Psychological Measurement, no. 20, pp. 37-46, 1960.

 ISBN: 978-1-61208-326-1