INACIO THOMAZ BUENO

CHANGE DETECTION IN THE BRAZILIAN SAVANNA BIOME

LAVRAS – MG

2018

INACIO THOMAZ BUENO

CHANGE DETECTION IN THE BRAZILIAN SAVANNA BIOME

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Engenharia Florestal, área de concentração em Ciências Florestais, para a obtenção do título de Mestre.

Prof. Dr. Fausto Weimar Acerbi Junior
Orientador

LAVRAS – MG
2018

Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).

Bueno, Inacio Thomaz.
Change detection in the Brazilian Savanna biome / Inacio Thomaz Bueno. - 2018.
104 p. : il.

Orientador(a): Fausto Weimar Acerbi Júnior.
.
Dissertação (mestrado acadêmico) - Universidade Federal de Lavras, 2018.
Bibliografia.

1. Land cover changes. 2. Cerrado. 3. Exploratory analysis. I. Acerbi Júnior, Fausto Weimar. . II. Título.

INACIO THOMAZ BUENO

CHANGE DETECTION IN THE BRAZILIAN SAVANNA BIOME
DETECÇÃO DE MUDANÇAS NO BIOMA DE SAVANA BRASILEIRO

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Engenharia Florestal, área de concentração em Ciências Florestais, para a obtenção do título de Mestre.

APROVADA em 08 de março de 2018.

Dr. Alan de Brito	FUNDEP
Dr. Luís Marcelo Tavares de Carvalho	UFLA

Prof. Dr. Fausto Weimar Acerbi Júnior
Orientador

LAVRAS – MG
2018

To my beloved family and dear friends

In memory of Demétrius Aguiar Barbosa Prado

AGRADECIMENTOS

A Deus, por ceder o dom da vida.

A minha família: pais, avós, tios, irmãos e primos, que com a simplicidade de um valor, nunca deixaram e vão deixar de influenciar na formação do meu caráter. Ana Claudia pelo companheirismo e motivação.

As minhas grandes amizades que conquistei por onde passei em especial aos amigos de Lavras e Juruaia pelos grandes momentos compartilhados. Minha República e meus irmãos pelos princípios aprendidos em função de uma perfeita convivência.

A Universidade Federal de Lavras, e todo seu corpo docente pelo conhecimento transmitido, em especial ao Laboratório de Estudos e Projetos em Manejo Florestal, LEMAF, pelo suporte durante esse tempo.

A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo fomento da bolsa de mestrado, e ao Programa de Pós Graduação em Engenharia Florestal.

Ao meu grande amigo e mestre, Professor Fausto, que sempre auxiliou na construção do profissional que sou hoje com toda sua sabedoria, paciência e orientação. Ao professor Luís Marcelo (Passarinho), e aos pesquisadores Alan de Brito e Eduarda Silveira pela colaboração e confiança no meu trabalho.

E por fim, a aqueles que indiretamente, mesmo que sendo com um simples bom dia no corredor, me passaram confiança para a realização deste trabalho, o meu muito obrigado.

ABSTRACT

Many remote sensing techniques have been developed for forest change detection but there is no optimal method without limitations that can be applied in all landscapes. In the Brazilian savanna biome is not different, the analysis and quantification of human induced deforestation in Cerrado areas proved to be a challenge regarding to the spectral information. This study was divided in two parts, the first one exploring the spectral and temporal information of land cover changes, and in the second we used meaningful information of these changes to discriminate human induced from seasonal changes by different machine learning algorithms. Chapter one evaluated the image data availability in the SF9 basin sampled areas based on cloud and shadows cover, and used filter-based feature selection methods and object-based image analysis to also evaluate Landsat 8 bands. These feature selection methods took red and short wave infrared bands as promisor bands to detect deforestation in savanna biome. In temporal context, free cloud cover presented good change detection accuracies even for distinct image frequencies. Chapter two used the promisor bands previous evaluated to compute spectral indices, which create an input dataset to three machine learning algorithms, Artificial Neural Network (ANN), Random Forest (RF) and Support Vector Machine (SVM), and also assessed individually spectral channels indices in all detections. Random Forest demonstrated the best results in test phase with overall accuracy of 92%. The short wave infrared spectral channel as well as the tasseled cap brightness and greenness transformations indices had positive influence in all machine learning algorithms. Thus, this study emerged new options to savanna change detection through a database exploratory analysis and different machine learning algorithms.

Keywords: Land cover changes. Cerrado. Exploratory analysis. Machine learning.

RESUMO

Técnicas em sensoriamento remoto vêm sendo desenvolvidas para se detectar mudanças em florestas, porém não há um método ótimo, ausente de limitações, que se aplique em qualquer tipo de paisagem. O bioma de savana brasileiro não é diferente, a quantificação de mudanças em áreas de Cerrado tem se tornado um desafio no âmbito espectral. Este estudo foi dividido em duas partes, a primeira através de uma análise exploratória de mudanças na cobertura do solo, e uma segunda que se utilizou de informações promissoras da primeira parte para discriminar mudanças decorrentes da ação humana de mudanças naturais através de algoritmos de aprendizado de máquina. O capítulo 1 avaliou a disponibilidade de imagens de satélite de áreas amostradas na Bacia SF9, e também aplicou métodos de seleção de atributos e segmentação multi-data para avaliar as bandas espectrais de imagens Landsat 8. Estes métodos selecionaram a banda do vermelho e a banda do infravermelho de ondas curtas como promissoras para detectar mudanças no Cerrado. Em relação à informação temporal, a total ausência de nuvens e sombras demonstrou boas acurácias mesmo em diferentes frequências de imagens. O capítulo 2 levou em consideração as bandas promissoras do capítulo 1 para o cálculo de índices espectrais, onde estes índices serviram como base de entrada para três algoritmos de aprendizado de máquina, Redes Neurais Artificias (RNA), *Random Forest* (RF) e *Support Vector Machine* (SVM). A importância individual de cada índice espectral também foi avaliada para todas as detecções. O algoritmo baseado em árvores de decisão *Random Forest*, gerou os melhores resultados na fase de teste, com acurácia global de 92%. O canal espectral do infravermelho de ondas curtas, assim como os índices transformação *tasseled cap brightness* e *greenness* se mostraram importantes no desempenho de todos os algoritmos avaliados. Assim, o estudo oferece novas opções para a detecção de mudanças no Cerrado através de uma análise exploratória das mudanças e avaliação de diferentes algoritmos na detecção.

Palavras-chave: Mudanças na cobertura do solo. Cerrado. Análise exploratória. Aprendizado de máquina.

LIST OF FIGURES

LIST OF TABLES

SUMMARY

1 GENERAL INTRODUCTION

The Earth surface is constantly changing when it comes to changes in land use and land cover. Global demand for food reflects the deforestation of preserved forests being replaced by croplands, pastures and urban areas (GIBBS et al., 2010). This scenario directly affects the biodiversity of biomes, climate regulation, carbon stock, and water resources (FOLEY et al., 2005; HANSEN et al., 2013). Therefore, land use and land cover monitoring programs have become essential for understanding the Earth's surface transformations, reporting and planning national needs, international treaties, and supporting scientific research (WULDER et al., 2008). This important information reflects the need of continuity and precision of such programs to support decision makers (COPPIN et al., 2004).

Remote sensing change detection provided many different methods over the past decades and have been reviewed by several authors (BANSKOTA et al., 2014; COPPIN et al., 2004; HUSSAIN et al., 2013; LU et al., 2004; TEWKESBURY et al., 2015; ZHU, 2017). The method accuracy is dependent on several aspects and there is no optimal algorithm that applies to all purposes (CHEN et al., 2012; COPPIN et al., 2004; HUSSAIN et al., 2013; TEWKESBURY et al., 2015). Among these circumstances, the seasonal differences caused by solar angle differences and phenological changes in vegetation are considered sources of noise for the majority of the change detection algorithms and must be avoided by selecting images from the same season or choosing methods unaffected by seasonality (ZHU, 2017). These phenological changes in vegetation are typical in the Brazilian tropical savanna, the Cerrado (HILL et al., 2017).

Cerrado represents a challenge to change detection methods based on remote sensing imagery. In addition to seasonal effect upon the biome, the major

conversion and fragmentation by land use change, and high vegetation heterogeneity with gradients of complex woody and herbaceous structure, makes it an appeal to change detection (HILL et al., 2017). According to Brasil (2015), almost 50% of the original 2 million km² Brazilian Cerrado were converted to agricultural land use until 2011, being the second Brazilian biome with more human-induce changes. Most of this large land conversion is due to agriculture expansion, wood predatory extraction, depletion of natural resources, and low rate of protected areas, where just 7.44% of the biome is protected, and only 2.91% are fully preserved (BRASIL, 2011).

Thus, Cerrado change detection background motivates the later chapter's assumptions where a better understanding of the change phenomenon at the biome is required to support change detection studies.

2 LITERATURE REVIEW

2.1 The Cerrado biome

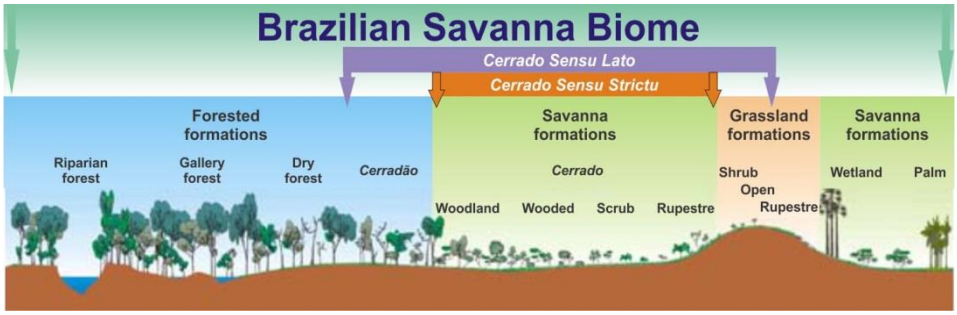
Savannas ecosystems comprises 20% of the global terrestrial surface spread at the Earth's tropics, mainly prevalent in Australia, Africa and the Americas (LEHMANN et al., 2011). In Brazil, the Cerrado tropical savanna, is the second largest biome in the country with 2 million of square kilometers approximately occupying 25% of the national territory, particularly in the States of *Goiás, Mato Grosso do Sul, Mato Grosso, Tocantins* and *Minas Gerais* (BRASIL, 2015).

The Brazilian savanna has the richest flora among the savannas in the world with a considerable number of endemic species, and it is considered a hotspot for conservation of biodiversity in Brazil (KLINK; MACHADO, 2005; MYERS et al., 2000).

Cerrado is unlike any other large savanna in the world (HILL et al., 2017; RATTER et al., 1997), since it combines an unique vertically structured mosaic of plant formations (FERREIRA et al., 2003). The Cerrado stricto sensu is the mainly formation, which better illustrate the savanna biome with sparse and short twisted trees. Woodland areas are represented by forest formations with dense canopy cover, known as *Cerradão*, riparian forests, and wetlands formations or *Veredas,* and grasslands completes the Cerrado mosaic (FIGURE 1) (RIBEIRO; WALTER, 2008). The combination of this very diverse woody physiognomy and highly biodiversity species composition make it a biodiversity hotspot with the richest flora among the savanna biomes (HILL et al., 2017; KLINK; MACHADO, 2005). The Cerrado territory is also included in the three biggest water basins in South America – *Amazônica/Tocantins, São Francisco*

and *Prata* – resulting in a significance influence on water resources (BRASIL, 2011).

Figure 1 – The Cerrado biome.



Source: Adapted from Ribeiro & Walter (2008).

Peel et al. (2007) describes the Cerrado climate as Tropical savanna (Aw) with dry winter and wet summer, with annual temperature around 22-23°C. Approximately 90% of the rains are concentrated from October to April with annual precipitation ranging from 1,200 to 1,800 mm. Dry season is quite distinct with monthly precipitation reaching zero millimeters. This drastic seasonal rainfall change results in a wide range of adaptive phenological strategies as leaf drop in Cerrado trees.

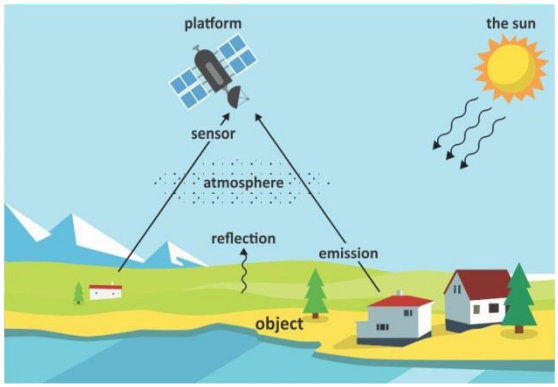
The Cerrado vegetation species developed strategies in order to overcome water scarcity. Most of the woody species are evergreen or deciduous formations, these last types of vegetation are represented by total or partial leaf fall during the dry season (FERREIRA; HUETE, 2004)

2.2 Remote sensing and vegetation

Among a collection of applications, the remote sensing science extracts information of a target on Earth's surface through a set of data provided by a sensor placed at an airplane or satellite (SCHOWENGERDT, 2007). The solar energy is the main source of energy in passive remote sensing. It can be absorbed, transmitted and/or reflected by a target where the ratio reflected energy to the total incident on a same target is quantified by the reflectance measure (LILLESAND; KIEFER; CHIPMAN, 2008).

Sensors are responsible to convert radiance received from a target into digital images or remote sensing images (SCHOWENGERDT, 2007). These sensors can be on board of airplanes or satellites and their main characteristics are spatial, spectral, temporal and radiometric resolutions (FIGURE 2).

Figure 2 – Principles of Remote Sensing.



Source: Author (2018).

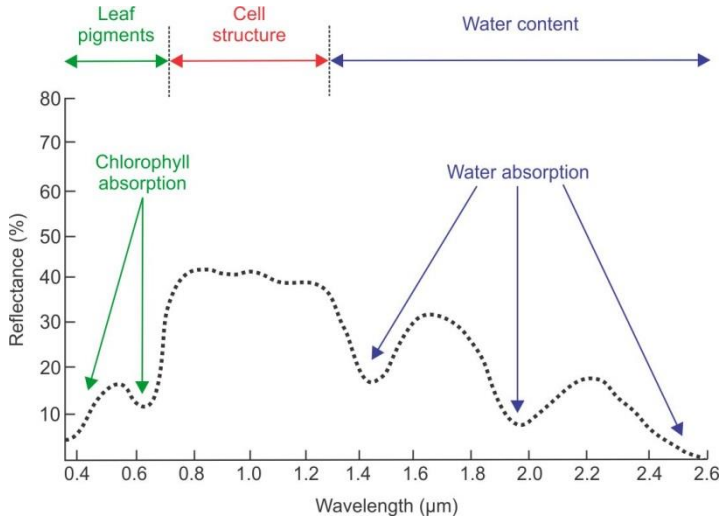
Remote sensing is very important and efficient for land cover mapping and monitoring purposes because it is generally faster and less costly than the information collected at the ground level. It also provides an aerial perspective

that allows better understanding of the spatial relationships and invisible data to human sense such as spectral information in the infrared region (CONGALTON; GREEN, 2009). Thus, remote sensing science has become an universal practice in government agencies, environmental organizations, industrial sector, and research institutions (KENNEDY; YANG; COHEN, 2010).

Among the large set of information provided by remote sensing on Earth's surface, the spectral information provides radiation reflectance of target in function of electromagnetic wavelengths. The reflectance extracted from vegetation dynamics can provide information based on stage of growth, vegetation health and moisture content (SCHOWENGERDT, 2007), since this information is determined by chemical and morphological characteristics of the surface of organs or leaves (PONZONI; SHIMABUKURO; KUPLICH, 2012).

According to Ponzoni, Shimabukuro and Kuplich (2012), the main applications for remote sensing of vegetation are based on the following light spectra: (a) the visible region, where the wavelength ranges from 0.4 to 0.72 μm; (b) the near infrared region 0.72 – 1,1 μm; and (c) the mid infrared 1,1 – 3,2 μm (FIGURE 3).

Figure 3 – Typical spectral response characteristics of a green leaf.



Source: Adapted from Ponzoni, Shimabukuro and Kuplich (2012).

In the visible spectra, cell pigments such as chlorophyll, carotene and xanthophyll, determine the shape of spectral signature making each curve unique by specie. In near infrared, there is a decrease of energy absorption by vegetation and an increase the internal scattering of incident radiation. And mid infrared contains information about the absorption of radiation by water (PONZONI; SHIMABUKURO; KUPLICH, 2012). According to Xue and Su (2017), the reflectance of the surface of leaves of a green plant without any biotic or abiotic stress is generally higher and restricted (0.96 – 0.99 µm) than of a dry plant (0.88 to 0.94 µm).

Another characteristic of remote sensing systems is the repetitive data acquisition of a same region on the Earth's surface (SCHOWENGERDT, 2007). The multi-temporal information relates the same area to periods of time and their changes during the period (LILLESAND; KIEFER; CHIPMAN, 2008).

2.3 Monitoring changes in Cerrado

The biome has been subject to the second most rapid land conversion in Brazil. Almost 50% were already transformed into others land uses due to agriculture expansion, wood predatory extraction, depletion of natural resources, and low rate of protected areas (BRASIL, 2011).

With the monitoring and conservation programs attention upon the amazon biome, the human intervention in Cerrado started to be seen by the scientific community only at the 90s (TRANCOSO; SANO; MENESES, 2015), making the Cerrado one of the least-studied biomes in the world (MENINO et al., 2012; RIBEIRO; WALTER, 2008).

After the opening of the Landsat archive in 2008 by the Landsat Global Archive Consolidation (LGAC) initiative, millions of images became available enabling time series studies throughout the world (ZHU, 2017). These studies have provided for change detection gains in resolution and accuracy, and also leveraged the development of programs based on satellite time series, as the Brazilian systems: *Programa de Monitoramento do Desmatamento da Amazônia* – PRODES, *Detecção do Desmatamento em Tempo Real* – DETER, *Projeto de Monitoramento do Desmatamento dos Biomas Brasileiros por Satélite* – PMDBBS (DEVRIES et al., 2016).

Government projects such as *Monitoramento do Desmatamento dos Biomas Brasileiros por Satélite* (PMDBBS) used Landsat TM and CBERS to monitoring and mapping deforestation events in the biome. Four deforestation reports was generated: until 2002 and from 2002-2008, 2008-2009, 2009-2010 and 2010-2011 (BRASIL, 2015).

Scientific studies also evaluated the biome in order to map and monitor. Machado et al. (2004) had the initiative of mapping the Cerrado in 2002 using MODIS satellite with better precision then global studies, registering an annual

loss of 2.2 million hectares in the biome. Sano et al. (2010) also mapped the entire biome in 2002 with better spatial resolution, using Landsat ETM+, and confirming the intensive land use pressure. In regional scale, Brannstrom et al. (2008) evaluated the land conversion in two Cerrado regions between 1986 and 2002, and founding different spatial patterns of Cerrado fragmentation with identical underlying drivers. Using an object based context, Grecchi et al. (2013), mapped Cerrado between 1985 and 2005, registering high rates of crop expansion in southeastern *Mato Grosso* State, and Grecchi et al. (2014) for the same area and period, linked the land use change with intrinsic environmental vulnerability, providing environmental indicators. Beuchle et al. (2015) also applied object-based approach to map Cerrado and Caatinga sample units between 1990 and 2010, and then estimate the loss in tree coverage and other wooded lands.

2.4 Change detection by object based image analysis

Change detection in remote sensing images is applied to two or more dates that do not represent the normal variation of a particular area (SHALABY; TATEISHI, 2007), and object-based change detection has the goal to identify these changes in geographic objects by object-based image analysis (CHEN et al., 2012).

Object based image analysis combines segmentation and remote sensing information along with analyst experience with image-objects in order to model geographic entities (BLASCHKE, 2010; CHEN et al., 2012). The segmentation is the most common approach for building objects in images (VIEIRA et al., 2012), which has the objective of creating groups of pixels spectrally similar and spatially adjacent from an image with the purpose of minimizing the within-

object variability compared to the between-object variability (DESCLÉE; BOGAERT; DEFOURNY, 2006).

According to Chen et al. (2012), the advantages of object-based based approach on pixel-based change detection methods are: (a) the segmentation procedure, which characterizes landscape elements by groups of homogeneous pixels, (b) reduction salt and pepper effect attributed by small spurious changes, and (c) extraction of sophisticated information of the image objects as geometry and texture.

Among the mainly object-based change detection (OBCD) methodologies described in scientific literature, Chen et al. (2012) classify them into four groups: (a) image-object change detection: similar to pixel-based, two or more image segmented are direct compared by the extraction of spectral and/or spatial information; (b) class-object change detection: comparison "from-to" of image-objects defined by landscape classes so additional classification information is required; (c) multi-temporal object change detection: objects are generated by two or more image with temporal information where a set of images create one segmentation; (d) hybrid change detection: involve the use of both object and pixel methodologies. Table 1 lists some of these OBCD methods.

Table 1 – Object-based change detection methods. (Continue)

Method	**Authors**
Image-object change detection	Hall and Hay (2003); Lefebvre, Corpetti and Huberty-Moy (2008); Miller, Pikaz and Averbuch (2005).
Class-object change detection	Desclée, Bogaert and Defourny (2006); Gamanya, De Maeyer and De Dapper (2009);

Table 1 – Object-based change detection methods. (Conclusion)

	Xian and Homer (2010).
Multi-temporal object change detection	Bontemps, Langner and Defourny (2012); Chen et al. (2013); Desclée, Bogaert and Defourny (2006); Lu et al. (2016).

Source: Adapted from Couto Júnior (2011).

Despite the of object-based change detections, pixel-by-pixel techniques are still widely used in many areas to measure changes using remote sensing data (HUSSAIN et al., 2013).

2.5 Data mining in remote sensing

A large amount of remote sensing data may represent a challenge to change detection algorithms. Sometimes, the data needs to be discriminated to reduce the number of original features that are highly correlated or might be redundant in order to improve the detection accuracy and decline the computational efforts in the classifier algorithm estimation parameters (PAL; FOODY, 2010).

Given the background, data mining is defined as the process of discovering valuable information in a large volume of data (WITTEN et al., 2016). It encompasses a set of techniques that allow to (a) search through large data sets, (b) search for spatio-temporal patterns, (c) extract knowledge and relationships, and (d) provide advanced clustering and classification algorithms (HUSSAIN et al., 2013).

Data mining techniques has been recently applied to remote sensing image interpretation using information processing and artificial intelligence to get information from a massive volume of satellite images (LARY, 2010). This large data set can be related to image time series in studies by Petitjean et al. (2010), which mined sequential patterns from pixels in time series, and Boulila et al. (2011) that predicted spatiotemporal changes in objects from satellite image databases. Silva et al. (2008) explored massive land change patterns in order to identify agents of change by a decision tree classifier.

Regarded to land cover classification, data mining had also been applied in decision tree classifiers by setting classification thresholds in pixel (OTUKEI; BLASCHKE, 2010) and object-based image classification (VIEIRA et al., 2012).

2.6 Machine learning algorithms

The science of learning is very involved in the fields of statistics, data mining and artificial intelligence, and has proven useful for a large number of applications in many fields of science (HASTIE; TIBSHIRANI; FRIEDMAN, 2009) such as geosciences and remote sensing (LARY, 2010).

Machine learning is based on experimental supervised methods for regression and/or classification of nonlinear systems. Such systems, as in all data mining fields, can be massively multivariate involving thousands of variables (LARY et al., 2016). The machine learning algorithms induces to information analyze, pattern recognitions, and prediction of accuracies through an automated and repeated learning from a set of training data (ROGAN et al., 2008).

According to Rogan et al. (2008), the increase of attention upon machine learning in remote sensing studies can be attributed by some follow advantages:

(a) non-parametric characteristic; (b) reduction of computational effort; (c) ability to investigate importance of variables; and (d) flexibility to accommodate categorical and continuous variables. Ali et al. (2015) demonstrated a relevant activity in the last few years of machine learning methods in biomass and soil moisture studies, also listing the most commonly used machine learning methods in recent scientific literature (TABLE 2).

Table 2 – Most commonly used machine learning algorithms.

Algorithms	**Examples**
Regression	Linear, power, logistic regression
Decision tree	Conditional decision trees, C5.0, decision stump
Bayesian	Bayesian network, naive, Gaussian naive and multinomial naive Bayes
Artificial neural network	Perceptron, back-propagation, radial basis function network
Deep learning	Deep belief networks, convolutional neural networks
Ensemble	Random forest, bagging, gradient bagging
Support vector	Support vector machines, support vector regression

Source: Ali et al. (2015).

A number of machine learning studies have focused on extract and analyze vegetation information, such as biomass estimation, land cover classification, change detection, and tree species identification

The aboveground biomass information has been computed in remote sensing imagery by machine learning. Algorithms as decision trees and support vectors were applied in Landsat images (AVITABILE et al., 2012; POWELL et

al., 2010), very high resolution imagery, such as Quickbird (CHEN; HAY; ST-ONGE, 2012) and WorldView-2 (MUTANGA; ADAM; CHO, 2012), Synthetic Aperture Radar data (CARREIRAS; VASCONCELOS; LUCAS, 2012; KARJALAINEN et al., 2012) and multi sources (GUO et al., 2012). Estimation of forest measures volume/ha, basal area/ha and stems/ha using field inventory and Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) data was demonstrated by Shataee et al. (2012) using three different types of machine learning algorithms.

Land cover classification carried out by remote sensing and machine learning was established by Pelletier et al. (2016) assessing the Random Forest to map land cover in large areas, Lardeux et al. (2009) classifying tropical vegetation by support vector machine using SAR data, and Juel et al. (2015) using object based analysis of aerial orthophoto and DEM data to classify coastal vegetation. In savanna regions, Mishra and Crews (2014) examined the contributions made by spectral and topographic variables calculated from objects in order to map the land cover.

The detection of changes were supervised detected by Rogan et al. (2008) comparing multiple machine learning algorithms in order to map land cover modifications, and Makkeasorn, Chang and Li (2009) detecting seasonal changes of riparian zones based on genetic programming. Bovolo, Bruzzone and Marconcini (2008) used unsupervised change detection in multispectral remote-sensing images using a selective Bayesian threshold.

In savanna regions, machine learning has been applied in tree species classification (ADELABU et al., 2013; COLGAN et al., 2012). Almeida et al. (2014) also applied using multiscale classifier to detect phenology patterns in Cerrado trees.

CHAPTER 1 EXPLORATORY ANALYSIS OF LAND COVER CHANGES USING THE SPECTRO-TEMPORAL CONTEXT

ABSTRACT

Human-induced land cover changes are commonly found in Cerrado formations, and its vegetation complexity has been proved to be a challenge regarding change detection techniques. The aim of this study was to carry out an exploratory analysis of Landsat OLI images based on data mining techniques and object-based image analysis. We used cloud cover and cloud mask to indicate image availability in the study area. In addition, OLI spectral channels were evaluated in order to discriminate human induced from natural changes. In the temporal context, image acquisition was linked to cloud possible noise and cloud mask edition, which can take expensive work and accuracy dependency into account based on cloud mask uncertainties. On the other hand, free cloud image compositions presented good change detection accuracies even for distinct image frequencies. In spectral background, four filter-based methods (a) correlation-based; (b) consistency-based; (c) gain ratio; and (d) Relief-F selected the Short-wave Infrared spectral channel of OLI sensor as the best band to detect deforestations in this vegetation type, followed by Red channel. Thus, this study emerged new options to savanna change detection through an exploratory analysis in change areas.

Keywords: Land cover changes. Cerrado. Exploratory analysis. Data mining.

RESUMO

Mudanças antrópicas são frequentemente descritas na paisagem do Cerrado, e a complexidade destas formações tem mostrado ser um desafio na detecção automática de mudanças. Neste estudo foi realizada uma análise exploratória através de técnicas de mineração de dados e análise orientada ao objeto em imagens Landsat OLI. A disponibilidade de imagens e cobertura de nuvem foi avaliada, e também os canais espectrais em relação à discriminação de mudanças antrópicas de mudanças naturais da paisagem. No contexto temporal, a seleção das imagens mostrou estar diretamente relacionado a cobertura de nuvem e a performance de uma máscara de nuvem, o que pode gerar um processamento oneroso além da dependência da acuracidade dos dados, porem, séries temporais compostas por imagens com total ausência de nuvens mostraram resultados satisfatórios mesmo com distinção no numero de imagens por área. Já no contexto espectral, quatro métodos de seleção baseados em filtro: (a) método de correlação; (b) método de consistência; (c) ganho de informação; e (d) algoritmo Relief-F, selecionaram a banda espectral do Infravermelho Médio do sensor OLI como a melhor na distinção de desmatamentos na vegetação do Cerrado, seguida da banda do Vermelho. Assim, este estudo pode gerar novas opções na detecção de mudanças da cobertura do solo neste tipo de vegetação através uma análise exploratória das áreas de mudança.

Palavras-chave: Mudanças na cobertura do solo. Cerrado. Análise Exploratória. Mineração de Dados.

1 INTRODUCTION

There are a myriad of techniques and algorithms in the scientific literature for monitoring, analyzing and detecting changes in native vegetation (HUSSAIN et al., 2013), but the current research is not enough to identify optimum approaches and to tackle persistent problems for change detection (TEWKESBURY et al., 2015). The complexity of Brazilian savanna biome generates a challenge for change detection studies (HILL et al., 2017) and a lack of information, which makes the Cerrado one of the least-studied biomes in the world (MENINO et al., 2012; RIBEIRO; WALTER, 2008).

The analysis and quantification of human induced deforestation in Cerrado areas have been proved to be a challenge regarding spectral information (TRANCOSO; SANO; MENESES, 2015). Although NDVI became the most frequently used vegetation index in the world, the exploration of new spectral metrics had been suggested in the biome (PEREIRA et al., 2016; TRANCOSO; SANO; MENESES, 2015). Recently, studies proposed post classification comparison in Cerrado change detection and monitoring (BEUCHLE et al., 2015; BRANNSTROM et al., 2008; GRECCHI et al., 2013, 2014) while in burned areas, spectral channels based on near infrared (DACAMARA et al., 2016; LIBONATI et al., 2011; PEREIRA et al., 2017) and middle infrared (DACAMARA et al., 2016; LIBONATI et al., 2011, 2012) have been used.

Additionally, large image archives does not avoid the difficulty to obtain dense time series due to cloud cover or revisiting cycle limitation (BONTEMPS et al., 2008; LU et al., 2016). For example, Kovalskyy and Roy (2013) estimated the probability, at global scale, of obtaining at least one free cloud image for the years 2000 and 2010 was 0.194 and 0.332 respectively. Also land cover monitoring methods for temperate forests generally are characterized by a higher

frequency of cloud-free observations than in tropical forests (SCHULTZ et al., 2016).

Chapter I makes a spectral analysis of the Operational Land Imager (OLI) sensor in Cerrado changed areas and its relation to cloud free image availability. This study was intended to answer two main questions: (a) which spectral band can better discriminate changed areas from seasonal areas (no change) in order to improve change detection approaches in zones with high phenological seasonal noise? (b) how cloud-free imagery affects the change detection on those seasonal areas?

2 OBJECTIVES

The main objectives of the chapter were: (a) to evaluate image data availability in change detection; and (b) to evaluate the best spectral bands to discriminate human-induced changes from seasonal changes in native vegetation.

3 METHODOLOGY

3.1 Study area

The study area is located in the *Bacia Hidrográfica dos Afluentes Mineiros do Médio São Francisco* (SF9), a basin in the northern State of Minas Gerais, Brazil. Placed between 43º00' to 46º00' West and 14º00' to 16º00' South, the study area occupies more than thirty one thousand square kilometers in 24 counties with total population of 284,475 people (IBGE, 2010).

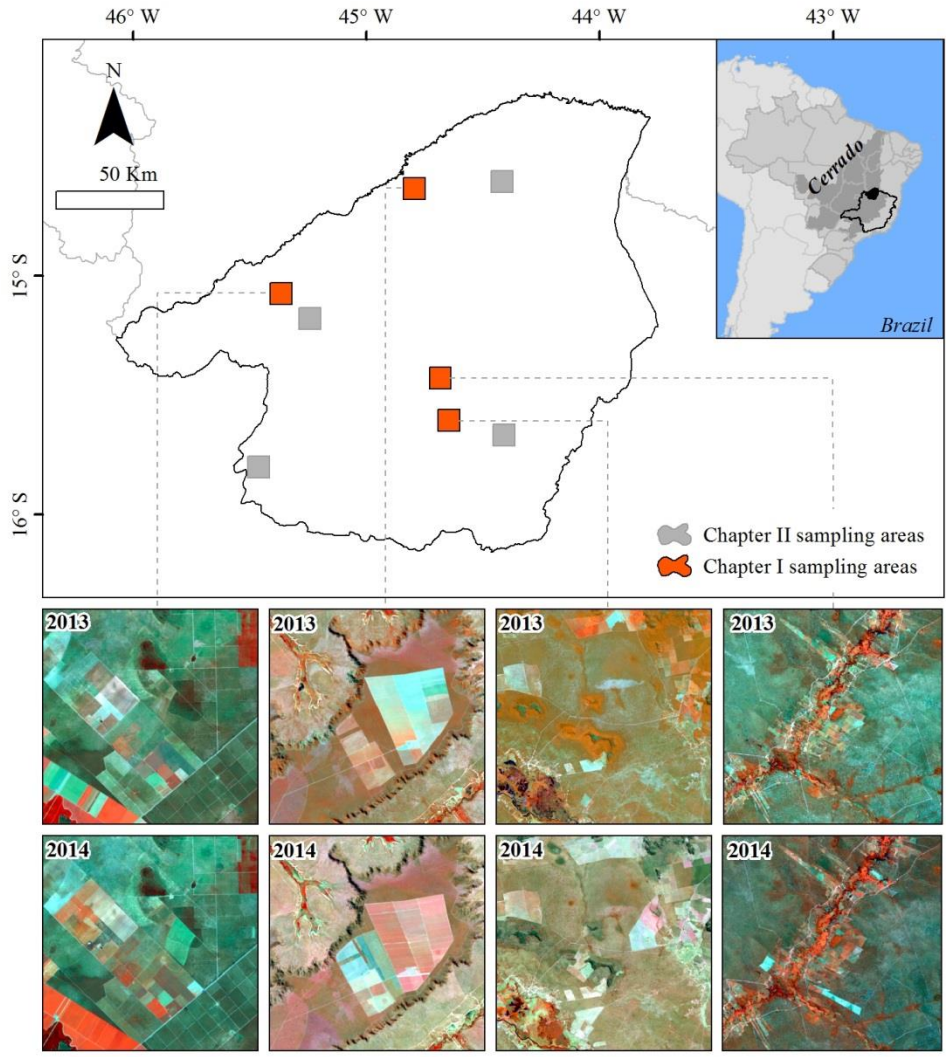
According to Carvalho and Scolforo (2008), about 59% of the area is represented by native vegetation with a wide range of types or physiognomies, where Cerrado is absolute in the area (about 37%) ranging from grasslands to heavily wooded areas and gallery forests.

Peel et al. (2007) described the climate as Tropical savanna (Aw) with distinct dry winter and wet summer, which influence the vegetation's spatial and spectral dynamics.

The deforestation rates are still high in the study area. According to the *"Monitoramento do Desmatamento dos Biomas Brasileiros por Satélite"* project, about 6,631 km² of Cerrado was deforested until 2002 and 1,244 km² between 2002 and 2008 (BRASIL, 2009), also 584 km² between 2008 and 2009 (BRASIL, 2011), 175 km² between 2009 and 2010 (BRASIL, 2011), and 118 km² between 2010 and 2011 (BRASIL, 2015).

Eight sampling areas of 100 km² mostly covered by native vegetation with change events were particular selected at the study site, being four samples to use in this chapter and the others four in the next one (FIGURE 4).

Figure 4 – Study area.



Legend: Landsat OLI image false color composition R = NIR, G = SWIR, B = Red.
Source: Author (2018).

3.2 Data

All available Landsat Operational Land Imager (OLI) imagery acquired from 2013 to 2014 with no cloud cover restriction per scene and processing level L1C Higher Level from United States Geological Survey for Earth Observation and Science (USGS/EROS) were downloaded. The data quality of L1C Higher Level product supports time series analyses and data stacking with high precision (RMSE < 12m) (ZHU, 2017) and bottom of atmosphere reflectance calculated by Vermote et al. (2016).

All OLI bands with 30 meters of resolution were used (TABLE 3) except Band 1 (Ultra Blue) and Band 9 (Cirrus) which are useful for aerosol studies and cirrus cloud detection.

Table 3 – Landsat OLI bands used in the study.

Bands	Abbreviation	Wavelength (μm)
Band 2 – Blue	B	0.452 – 0.512
Band 3 – Green	G	0.533 – 0.590
Band 4 – Red	R	0.636 – 0.673
Band 5 – Near Infrared	NIR	0.851 – 0.879
Band 6 – Shortwave Infrared	SWIR 1	1.566 – 1.651
Band 7 – Shortwave Infrared	SWIR 2	2.107 – 2.294

Source: Adapted from Roy et al. (2014).

3.3 Cloud cover

The presence of clouds and shadows in remote sensing imagery is considered noise by change detection approaches. The high brightness values of

clouds and the darkening effect of cloud shadows can be confused with land cover change if they are not evaluated (ZHU; WOODCOCK, 2012).

The cloud detection algorithm Fmask, developed by Zhu and Woodcock (2012) and available at Landsat L1C product from USGS/EROS, was applied to analyze the cloud scene frequency in each sampling area. Parallel to frequency analysis, cloud and shadow free images per sample were selected to the next steps.

3.4 Ancillary data

A land cover classification was produced in the first image, year 2013. The classification method adopted was the object-based image classification performed in *eCognition Developer 8.0* (DEFINIENS, 2009). The multi-resolution segmentation algorithm described by Baatz and Schäpe (2000) was applied and its criteria adjusted to a scale parameter (h_{sc}) equal to 50 and the compactness (w_{cp}) and shape (w_{sp}) equal to 0.3 for both parameters (a clearly description about segmentation is given in section 3.5). The fuzzy logic classified Cerrado objects based on spectral parameters selection. Classification post processing, such as manual edition, corrected class errors improving the final map accuracy.

High resolution imagery provided by GeoEye and Rapideye satellites supported an accuracy analysis based on a confusion matrix and validation Cerrado samples.

The classification map created a Cerrado mask in order to support the study where change areas only limited to this vegetation type were taken into analysis.

3.5 Multidate segmentation

Image segmentation has the objective of creating groups of pixels spectrally similar and spatially adjacent from an image with the purpose of minimizing the within-object variability compared to the between-object variability (DESCLÉE; BOGAERT; DEFOURNY, 2006).

According to Tewkesbury et al. (2015), the definition of the unit of analysis in object-based change detection can be classified in (a) image-object change overlay, where the segmentation is applied in one of the images and a comparison against other images is then made by simple overlay; (b) image-object comparison, where objects are created in each image in the time series individually; and (c) multi-temporal image-object, the segmentation is applied in the entire times series together.

In this study, multi-temporal image-objects were created by segmenting difference images of sequential periods, and defined in a single operation from the whole set of spectral bands using all sequential difference images together. Adapted from Desclée, Bogaert and Defourny (2006), this method assumes spectral, spatial and temporal information of difference images which creates objects based on the feature dynamics in time, as example, change events in a vegetation background.

The algorithm is the multi-resolution segmentation as performed in the land cover classification (EQUATION 1).

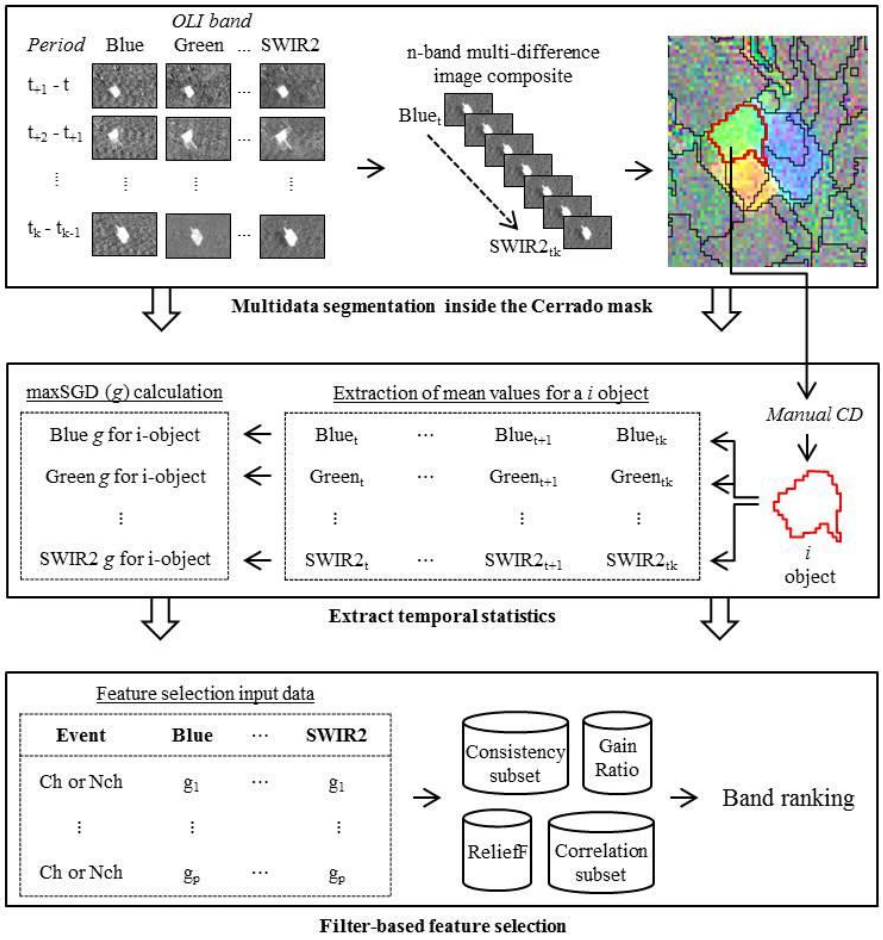
$$w_{sp}\sum_{nb} w_b\sigma_b + \left(1 - w_{sp}\right)\left[w_{cp}\frac{l}{\sqrt{np}} + \left(1 - w_{cp}\right)\frac{l}{lr}\right] \leq h_{sc} \qquad (1)$$

Where there are parameters set by user's visual analysis as w_{sp} is the spectral parameter ranging from 0 to 1 (zero is the maximum weight to spectral

homogeneity and 1 is the opposite for object shape); w_{cp} is the same meaning of w_{sp} but to compactness and smoothness parameters, adjusting the object shape between compact objects and smooth boundaries; and the scale parameter h_{sc} sets the object size. The others parameters are number of spectral bands *nb*; within-object variance for a spectral band σ_b; object border length *l*; number of pixels *np*; and the shortest possible length *lr* given the rectangle bounding the pixels.

The methodology for the multidate segmentation and the next three steps in sections 3.6 to 3.8 are illustrated by a framework in Figure 5.

Figure 5 – Framework of the OLI band analysis for land cover change.



Legend: *i* is an object created by segmentation, *t* is an image, *tk* is the last image of a sampling area, *k* is the dimension of an image difference composite, *g* is the maximum spectral difference gradient; and *p* is the last observation in feature selection.
Source: Author (2018).

3.6 Extract temporal statistics

In order to extract information from the objects, the mean pixel value inside an object was calculated for all OLI spectral bands.

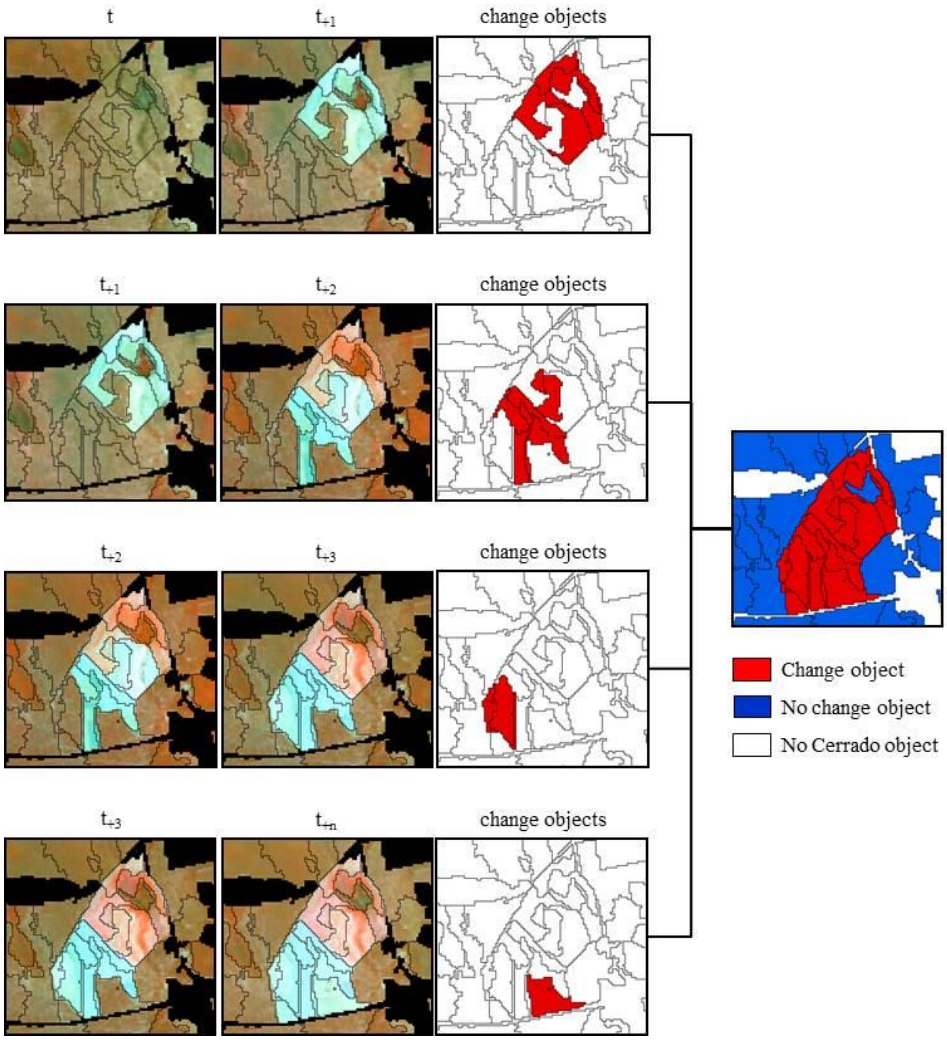
The maximum spectral gradient difference (SGD), adapted from Chen et al. (2013) and Lu et al. (2016), was calculated based on the objects means *i*, where a greater positive absolute value of *g(i, t, t -1)* in a band *b* indicates a larger change from time index *t-1* to *t* (EQUATION 2).

$$g(i,b)_{max} = \left[B_{(b,i,t)} - B_{(b,i,t-1)}\right]_{max} \quad (2)$$

3.7 Change objects validation

The validation of change objects, basically deforested areas in Cerrado, was manually performed by visual image interpretation in the whole data set (FIGURE 6). High resolution imagery as provided by Google Earth or Bing platforms is not always available for validation in a given period, so the changed objects were confirmed by analyst experience. Burn events were not representative in the sampling areas, so not considered.

Figure 6 – Validation of change and no change objects.



Legend: Landsat OLI image false color composition R = NIR, G = SWIR, B = Red.
Source: Author (2018).

3.8 Feature selection

Feature selection methodologies were basically divided in three categories: Filter, Wrapper and Embedded. The first division was executed in this study due to some advantages as: (a) independent assessment based on general characteristics of the data; (b) not dependent on classifiers; (c) usually faster than wrapper based methods; and (d) low computational complexity.

The goal of this analysis was to make feature selection as exploratory analysis with the purpose of not taking into account data reduction for future procedures, so filter-based facilities support the chapter aim. However, wrapper and embedded methods are also appropriate.

Beside the large amount of feature selection methods in scientific literature, four of them were applied in the database, two multivariate: Correlation-based and Consistency-based feature selection, and two univariate: Gain ratio and Relief-F. The feature selection input is represented by the instances (objects) and attributes (bands SGD_{max} values).

In the multivariate methods, the Best First Search heuristic was applied in order to identify a good subset by evaluating the five last subsamples that not increased the maximum Merit, and then set the subsample as stop criteria. A 10-fold cross validation was fitted in the selection methods and the number of folds containing a particular attribute ranked the band performance. In order to group the results, a nominal arrangement was created based on the number of folds in cross validation that selected a particular band. Thus, a band selected in nine or ten folds was registered with high significance; selection in seven or eight folds has medium significance; five or six low significance; and less than five was not relevant.

For univariate, a Ranker procedure sorted attributes by their individual evaluations. Ranking methods were used due their simplicity and good results for practical applications (CHANDRASHEKAR; SAHIN, 2014).

The four selection methods results were combined and re-established in an ordinal ranking in order to analyze the result by an intuitive view.

The fitted models, search methods and cross validation were performed at Weka 3.8 software (FRANK et al., 2016).

3.8.1 Correlation-based

The correlation-based feature selection (CFS) selects a subset of attributes based on a subsample heuristic evaluation considering the attribute-attribute and attribute-class correlation (HALL, 1999). The CFS heuristic evaluates the worth or merit *s* of a subset of features (EQUATION 3)

$$Merit(s) = \frac{f \times C_{ci}}{\sqrt{f + f \times (f-1) \times C_{ii}}} \quad (3)$$

Where *f* is the number of attributes; C_{ci} is the attribute-class correlation; and C_{ii} is the attribute-attribute correlation.

3.8.2 Consistency measure

Developed by Liu and Setiono (1996), consistency feature selection (CSY) evaluates a subset of attributes by the degree of consistency in class values when the training instances are projected onto the set (WITTEN et al., 2016). The degree of consistency is defined by an inconsistency rate where two instances are considered inconsistent if they have the same feature values but

different attributes. So, the goal is to select attributes that better allow to define consistent logical hypothesis about the training data set (ARAUZO-AZOFRA; BENITEZ; CASTRO, 2008).

3.8.3 Information Gain Ratio

Information Gain Ratio (IGR) elects attributes by ranking the entropy reduction. The information gain or entropy reduction is defined as the difference between the prior uncertainty and expected posterior uncertainty, defined in this study by the Shannon entropy. The ratio purpose solves the bias of huge databases increasing sensitivity to the analysis by an amount of split information (EQUATION 4).

$$\mathrm{IGR} = \frac{-\sum_{i=1}^{d} P_D(C_i) \times \log_2 P_D(C_i) - \sum_{j=1}^{p} \frac{D_i}{D} \sum_{i=1}^{d} P_{D_j}(C_i) \times \log_2 P_{D_j}(C_i)}{-\sum_{j=1}^{p} \frac{D_i}{D} \times \log_2 \frac{D_i}{D}} \quad (4)$$

Where d is the number of dimensions of the variable; $P_{D(Ci)}$ is the probability of the variable value = C_i; p is the number of dimensions of attribute A; D_j is the number of cases for A's each dimension j, and D is the total number of cases for A.

3.8.4 Relief-F

The Relief-F algorithm (RLF), developed by Kononenko (1994) is an instance based that selects instances randomly and compute a weight values based on nearest neighbors of the same and different classes. Near instances with different values for an attribute makes it irrelevant and computes a decrease of weight for the specific attributes, and vice-versa (JIA et al., 2013) (EQUATION 5).

$$W_f^i = W_f^{i-1} + \sum_{c \neq class(x)} \frac{\frac{p(x)}{1-p(class(x))} \sum_{j=1}^{k} diff_f(x, M(x))}{m \times k} - \sum_{j=1}^{k} \frac{diff_f\big(x, H(x)\big)}{m \times k} \quad (5)$$

Where f is the feature; i is the randomly selected instance; m is the sample size; *diff()* is the distance between samples; *p()* is the probability; k is an instance selected in every class; c is a class different from *class()*; *M(x)* is nearest-neighbor sample with the same class; and *H(x)* is nearest-neighbor sample with different class.

4 RESULTS AND DISCUSSION

4.1 Cloud free data

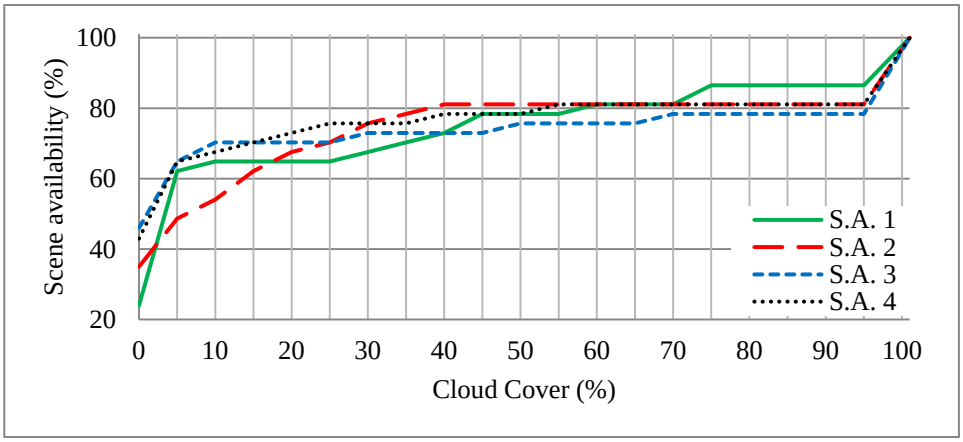
The cloud free image analysis returns scene availability from 24 to 46% per sampling area by a total of 37 acquisitions in 2013-2014 period. However, an increase of 19 to 38% in scene availability is registered when cloud cover was allowed until 5% (TABLE 4).

Table 4 – Percentage of Landsat 8 OLI images free of clouds.

S.A.	Free cloud cover scene availability	5% cloud cover scene availability
1	9 – (24%)	23 – (62%)
2	13 – (35%)	18 – (49%)
3	17 – (46%)	24 – (65%)
4	16 – (43%)	24 – (65%)

Source: Author (2018).

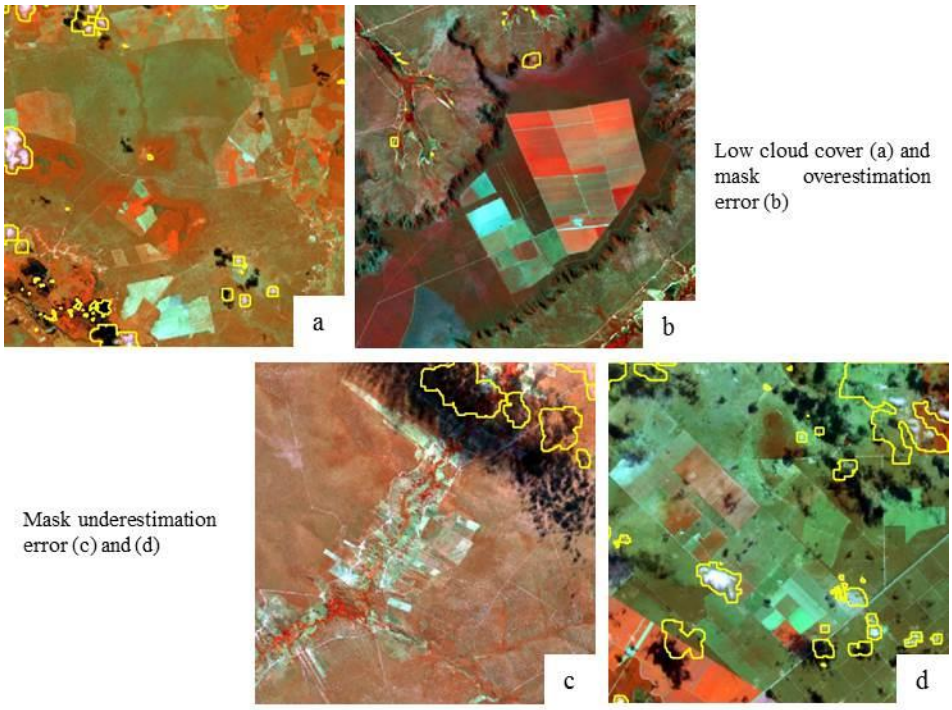
Figure 7 – Data availability based on cloud cover.



Source: Author (2018).

Figure 7 registered an abrupt increase of scene availability between 0 and 5% cloud cover, a smoothed increase until 95% of cloud cover, and again an abrupt to 100%. The first gain, which has the biggest attention to change detection image selection, might be justified by two situations: a) a true low cloud cover (FIGURE 8a); and b) mask errors that overestimates the cloud cover (FIGURE 8b).

Figure 8 – Low cloud cover situation and mask errors.



Legend: Landsat OLI image false color composition R = NIR, G = SWIR, B = Red.
Source: Author (2018).

Mask errors can be represented by the confusion between clouds and land cover with high brightness as bare soils (FIGURE 8b) or also cloud shadows with low brightness pixels as deep water bodies.

When a maximum cloud cover is allowed, underestimated errors can also be detected (FIGURE 8c and 8d). This sort of error needs to be checked and edited in the cloud mask because it may exceed the maximum cloud cover at the area, making the study unfeasible.

Optimal image selection is difficult to access as requirements area are often application specific (BANSKOTA et al., 2014). The database creation is also linked to cloud possible noise and cloud mask edition, which takes

expensive work and accuracy dependency into account. However, the cloud mask algorithms are necessary in the database creation, once the imagery amount needs to be controlled based on the maximum cloud cover allowance.

4.2 LULC classification

The classification map presented accuracies higher than 90%. These high accuracies are required in this mask process because it indicates an accurate data input for change detection and a low error rate for the next procedures.

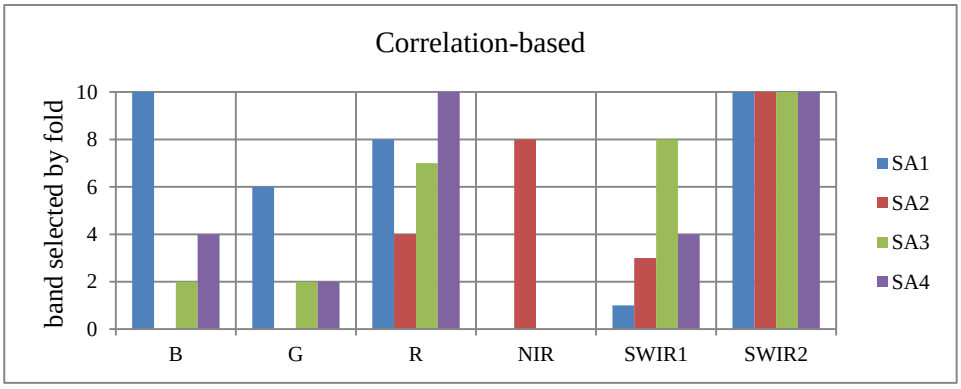
Cerrado physiognomies were not distinguished since these vegetation gradients have high structural and spectral similarity (SCHWIEDER et al., 2016), where this classification can reduce land cover mapping accuracy and has no meaningful impact on change analysis.

4.3 Feature selection

The correlation method (FIGURE 9) checked Blue, Red and SWIR2 bands with high significance, where SWIR2 was selected in all folds and all sampling areas, resulting in a perfect performance for this method. Red, NIR and SWIR1 were also checked with medium significance for change detection in some SA. Green band and other SA were weak or irrelevant in the analysis.

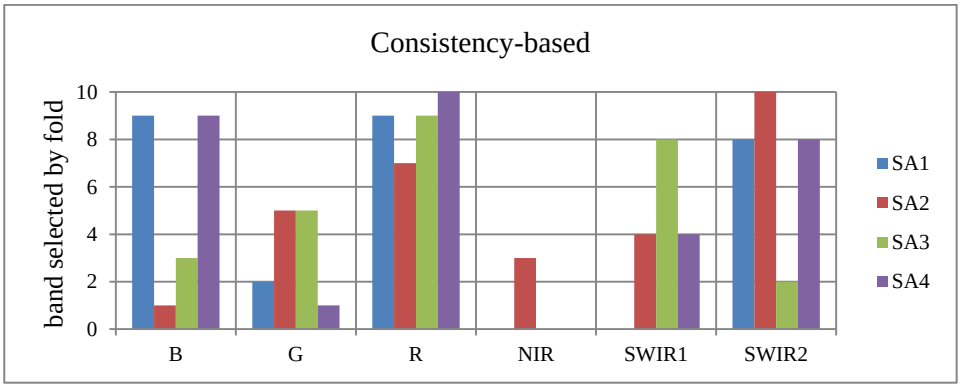
Figure 9 drafts Red and SWIR2 as the most significance bands in the subsets created by correlation based selection.

Figure 9 – Band subset selection by correlation based method.



Source: Author (2018).

Figure 10 – Band subset selection by consistency based method.

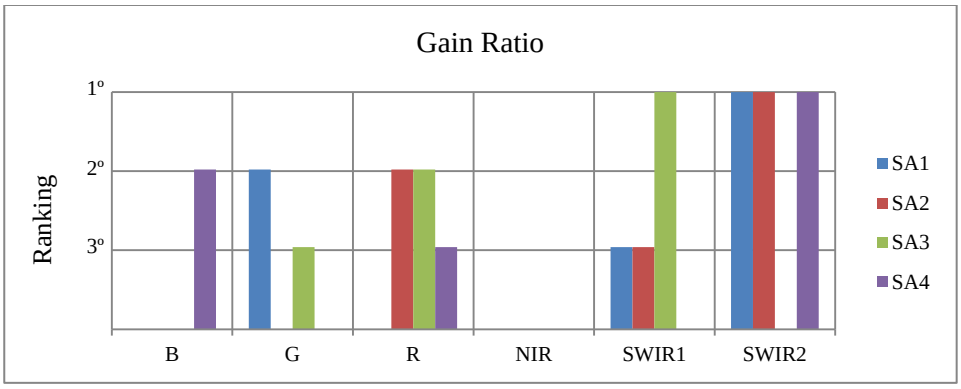


Source: Author (2018).

Figure 10 presents de band selection computed by the consistency method. As the CFS, Red and SWIR2 presented higher significance results in this selection, red band reach better results in the sampling areas one, three and four, and a better general classification. Blue and SWIR1 presented high and medium significance for particulars SA as well.

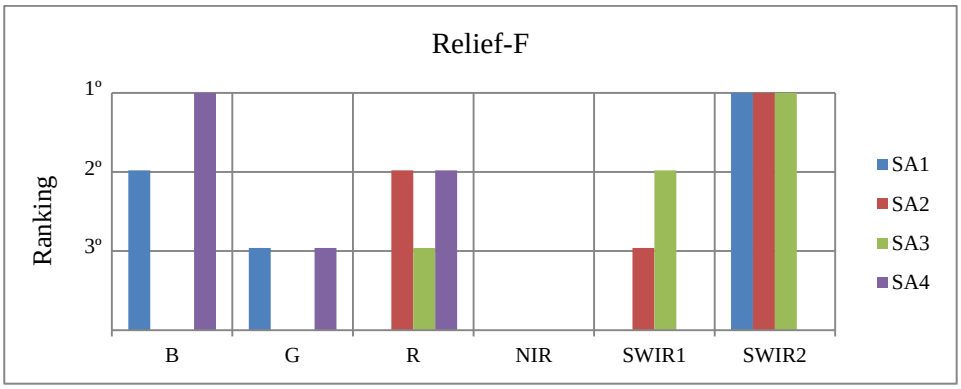
The NIR band was not significant in the multivariate procedures as a similar result founded by Trancoso, Sano and Meneses (2015). For univariate procedures, bands were ranked and Figures 11 and 12 registered the first three best bands selected for each sampling area.

Figure 11 – Band selection by information gain ratio method.



Source: Author (2018).

Figure 12 – Band selection by relief-F method.



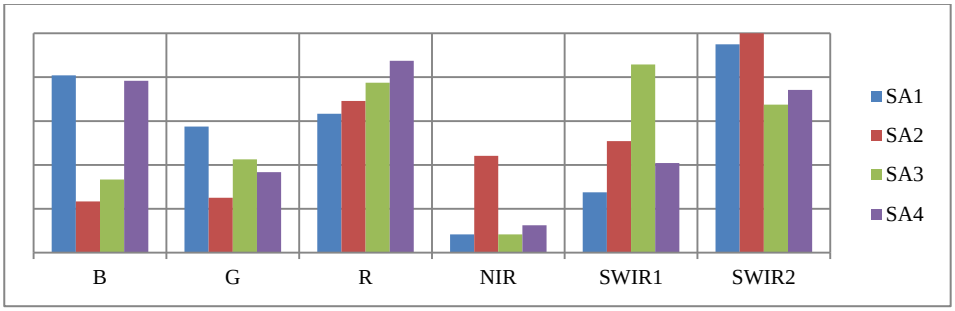
Source: Author (2018).

The second shortwave infrared was the most significant band for almost all sampling areas. It reached the best results in IGR method except for SA number three and for RLF method except for SA number four. SWIR1 presented the best discrimination for SA3 in IGR, and the blue band for SA4 in RLF.

Univariate results were not so far different from multivariate since the red band scoring the second best discrimination for the areas two and four, and the third position for area three. Near infrared was not present in the first three positions for any univariate method.

The four selection methods ranked are seen in Figure 13.

Figure 13 – General selection ranking by each sampling area.



Source: Author (2018).

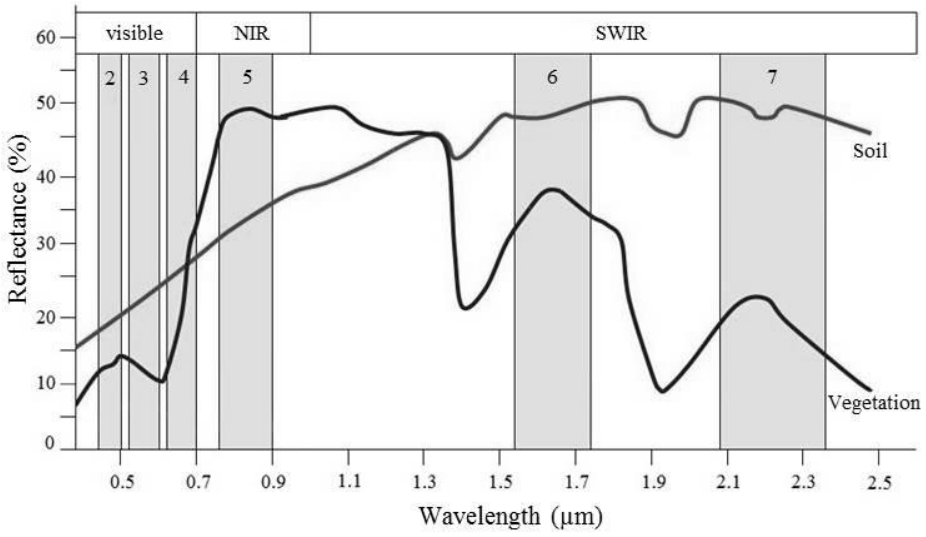
Based on the selection methods used in this study, SWIR2 was the best band to detect human-induced changes for two sampling areas with high seasonality. Red and SWIR1 achieved the first position for the others SA. Blue, green and NIR bands complete the order of significance.

The results presented a scene dependency that can be justified by aerosol effects or different seasonal influence since the large size of the study area where the sampling areas may reach 120 km from each other.

4.4 Emerging options for Cerrado change detection

In remote sensing science, the NDVI is the most frequently used index. Based on the spectral signature of vegetation and soil (FIGURE 14), reflectance values in the SWIR region are divergent where soil reflectance is greater than vegetation reflectance. However to red and NIR bands (NDVI composition), the reflectance values between soil and vegetation are less divergent. NDVI may be much applied on vegetation studies but not quite relevant on soil applications compared to SWIR bands.

Figure 14 – The spectral signatures of soil and vegetation, and spectral bands of Landsat OLI.



Source: Jensen (2009).

Some authors have indicated the use of new metrics in order to describe and detect changes in the Cerrado biome. A spectral context based on SWIR channel is reported by Guerschman et al. (2009) and Hill et al. (2017) to

characterize Cerrado vegetation, and suggested by Trancoso, Sano and Meneses (2015) to detect deforestations in those areas. Acerbi Júnior et al. (2015) and Silveira et al. (2017) used spatial context in order to detect deforestation in Cerrado areas with seasonal noise.

A new option to detect Cerrado changes is to look forward the change event. Land cover class post change (bare soil, initial forest regeneration, initial crops fiends, etc.) must be comprehended as the pre change (native vegetation) and it needs to be part of the change analysis.

5 CONCLUSION

Clouds and cloud shadows will always be a noise for most change detection studies over the world and the use of cloud mask algorithms for large areas and monitoring programs is very important in order to mitigate change detection errors. However, cloud mask algorithms present miscalculations and a thorough cloud and mask analysis can support study planning and decision making. The free cloud cover method applied in the study presented good change detection accuracies even for distinct image frequencies.

Landsat OLI spectral channels based on shortwave medium infrared (2.107 – 2.294 μm) and red (0.636 – 0.673 μm) presented the best results that discriminated Cerrado changes for a group of feature selection algorithms. These results may emerge more opportunities to detect changes in vegetation areas affected by seasonal noise. Spectral indices based on SWIR and Red wavelengths must be evaluated in order to improve deforestation detection in the Cerrado.

CHAPTER 2 EVALUATING MACHINE LEARNING ALGORITHMS IN CERRADO CHANGE DETECTION

ABSTRACT

There is no optimal method without limitation that can be applied in Cerrado formations, so that a wide range of change detection methods are still being proposed and assessed. This study used the promisor bands previously evaluated in chapter one to calculate spectral indices, in order to create an input dataset for three machine learning algorithms, Artificial Neural Network (ANN), Random Forest (RF) and Support Vector Machine (SVM). A parameter simulation was carried out for each algorithm and the best architecture was selected to generate change maps and an attribute importance analysis. The three algorithms presented satisfactory performances, but Random Forest demonstrated the best results in test phase with overall accuracy of 92% and low error variation in the parameter simulation phase. Spectral channels as short wave infrared, tasseled cap brightness and greenness transformations had positive influence in the three algorithms while NDVI showed poor influence in neural network and Random Forest algorithms.

Keywords: Land cover changes. Cerrado. Object-based image analysis. Machine learning.

RESUMO

A presença de um método de detecção de mudanças que se aplique no Cerrado sem limitações é algo ilusório na ciência atual. Várias metodologias de detecção de mudanças nessas formações vêm sendo desenvolvidas e avaliadas na tentativa de mitigar erros de detecção. Este estudo utilizou bandas promissoras do capítulo 1 para o cálculo de índices espectrais, onde estes serviram como base de entrada para três algoritmos de aprendizado de máquina, Redes Neurais Artificias (RNA), *Random Forest* (RF) e *Support Vector Machine* (SVM). Uma simulação de parâmetros dos três algoritmos foi realizada onde a melhor configuração produziu mapas de mudança das áreas e uma análise da importância de cada atributo espectral. Os três algoritmos apresentaram resultados satisfatórios, porem o algoritmo baseado em árvores de decisão *Random Forest*, gerou os melhores resultados na fase de teste, com acurácia global de 92%, além de baixa variação do erro de predição na fase de simulação. Canais espectrais como o infravermelho de ondas curtas, transformação *tasseled cap brightness* e *greenness* se mostraram importantes no desempenho dos três algoritmos enquanto o índice NDVI se mostrou inviável para os algoritmos de redes neurais e o *Random Forest*.

Palavras-chave: Mudanças na cobertura do solo. Cerrado. Análise orientada ao objeto. Aprendizado de máquina.

1 INTRODUCTION

Change detection in forest environments is not a trivial task. There is no optimal method without limitations that can be applied in all landscapes (HUSSAIN et al., 2013). The Brazilian savannas, known as Cerrado, is one of the most complex global environments (HILL et al., 2017), its high seasonal noise, vegetation heterogeneity, and high anthropic pressure on the biome make good change detections accuracies a challenge (TRANCOSO; SANO; MENESES, 2015).

Remote sensing studies have recently focused on assessing land cover changes in Cerrado (BEUCHLE et al., 2015; BRANNSTROM et al., 2008; COELHO et al., 2014; GRECCHI et al., 2013, 2014) and searching for options in order to mitigate change detection misclassifications on deforestation (ACERBI JÚNIOR et al., 2015; TRANCOSO; SANO; MENESES, 2015; SILVEIRA et al., 2017), and burn events (DACAMARA et al., 2016; LIBONATI et al., 2011, 2012; PEREIRA et al., 2017).

In addition, a wide range of change detection methods in these particular landscapes continues to be proposed and assessed, once simple classifiers may reach their limits in some applications (BELGIU; DRĂGUŢ, 2016). Machine learning algorithms have proven to be useful tools for a large number of applications, including remote sensing and geosciences (LARY et al., 2016). Specifically on vegetation change detection they have been recently applied to regional (BRANDT et al., 2012; DEVANEY et al., 2015; DEVRIES et al., 2016; PEREIRA et al., 2017), national (GUINDON et al., 2014; HEALEY et al., 2018), and global (HUANG et al., 2008) scales.

Therefore, Chapter II proposed to evaluate the issue of detecting deforestation in seasonal areas using spectral information based on the results of Chapter I through machine learning algorithms. The study was motivated by the

following research question: Is it possible to detect human-induces land cover changes in areas affected by seasonality with high accuracies?

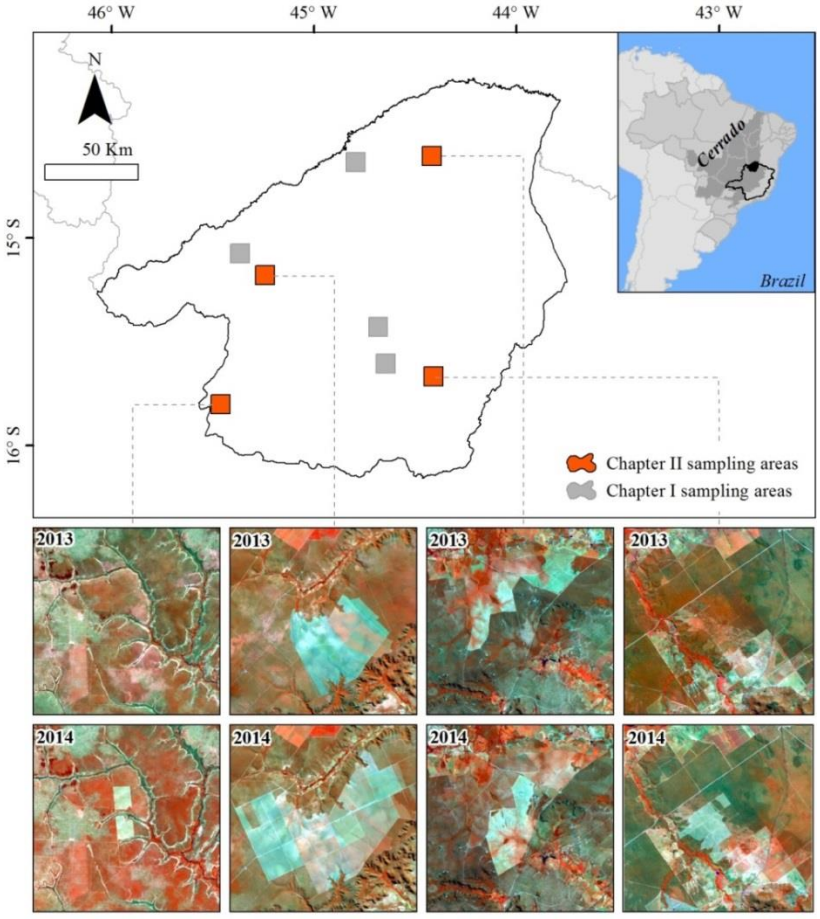
2 OBJECTIVES

The main objective of this chapter was to evaluate the use of machine learning algorithms for Cerrado change detection based on the spectral indices selected in Chapter One. The analysis was divided into: (a) evaluation of different algorithms for Cerrado change detection, and (b) assessment of individual spectral channels and indices in all algorithms.

3 METHODOLOGY

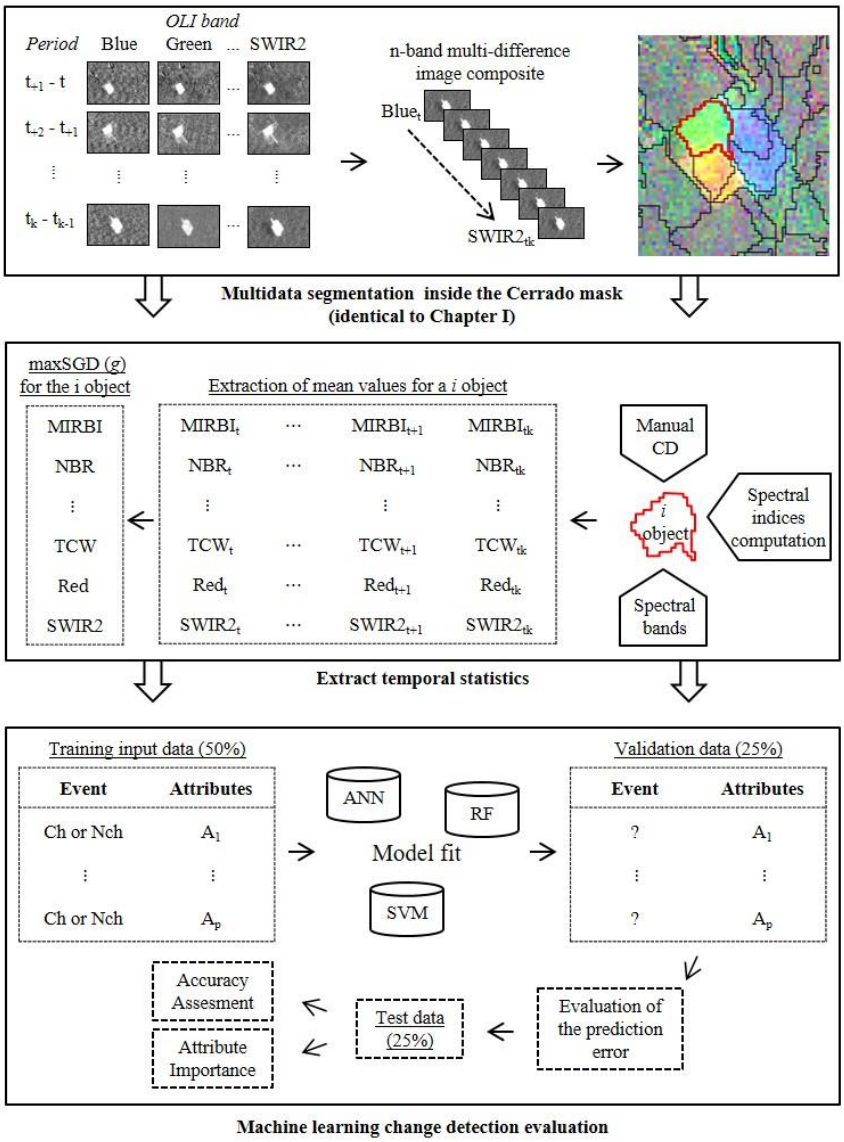
In this chapter, some initial methodologies sections are in common with Chapter 1. Except for the study area section with difference in the location of the sampling areas (FIGURE 15), data preparation and segmentation procedures are equal and no longer described in this chapter but shown in Figure 16.

Figure 15 – Study area.



Legend: Landsat OLI image false color composition R = NIR, G = SWIR, B = Red.
Source: Author (2018).

Figure 16 – Framework of land cover change methodology.



Legend: *i* is an object created by segmentation, *t* is an image, *tk* is the last image of a sampling area, *k* is the dimension of an image difference composite, *g* is the maximum spectral difference gradient; and *p* is the last observation in change detection.
Source: Author (2018).

3.1 Spectral indices

A set of eight spectral indices, including the Normalized Difference Vegetation Index (NDVI), were selected (TABLE 5). The NDVI was chosen because it is the most frequently used spectral index in remote sensing science (SCHULTZ et al., 2016). In addition, Red and Shortwave Infrared channels of OLI sensor, which were best ranked in chapter one, completed the input dataset (TABLE 6).

Despite a large list of indices that handles shortwave channels can be found in scientific literature, the selection took into account band computations that are sensitive to vegetation characteristics, states or changes (DEVRIES et al., 2016; SCHULTZ et al., 2016).

Table 5 – Spectral indices used in this study. (Continue)

Name	Abv.	Equation	Reference
Mid Infrared Burned Index	MIRBI	$10 * SWIR2 + 9.8 * SWIR1 + 2$	(TRIGG; FLASSE, 2001)
Normalized Burn Ratio	NBR	$\frac{NIR - SWIR2}{NIR + SWIR2}$	(KEY; BENSON, 2006)
Normalized Burn Ratio 2	NBR2	$\frac{SWIR1 - SWIR2}{SWIR1 + SWIR2}$	(DEVRIES et al., 2016)
Normalized Difference Moisture Index	NDMI	$\frac{NIR - SWIR1}{NIR + SWIR1}$	(JIN; SADER, 2005)
Normalized Difference	NDVI	$\frac{NIR - R}{NIR + R}$	(ROUSE et al., 1973)

Table 5 – Spectral indices used in this study. (Conclusion)

Vegetation Index			
Tasseled Cap Brightness	TCB	$b_1 * B + b_2 * G + b_3 * R + b_4 * NIR + b_5 * SWIR1 + b_6 * SWIR2$	(BAIG et al., 2014)
Tasseled Cap Greenness	TCG	$g_1 * B + g_2 * G + g_3 * R + g_4 * NIR + g_5 * SWIR1 + g_6 * SWIR2$	
Tasseled Cap Wetness	TCW	$w_1 * B + w_2 * G + w_3 * R + w_4 * NIR + w_5 * SWIR1 + w_6 * SWIR2$	

Source: Adapted from DeVries et al. (2016).

Table 6 – Landsat OLI bands used in the study.

Bands	**Abbreviation**	**Wavelength (µm)**
Band 4 – Red	R	0.636 – 0.673
Band 7 – Shortwave Infrared	SWIR 2	2.107 – 2.294

Source: Adapted from Roy et al. (2014).

3.2 Data input

Objects of the four sampling areas were combined and a sample of 100 objects per class (change or no change) was set. The sample was set according to Belgiu and Dragut (2016) requirements: (a) the databases must be statistically independent; (b) training samples must contain almost the same number of

observations per class, and be representative in the target class; (c) training sampling must be large enough to avoid Hughes phenomenon. The sampled data were divided into three parts, which is suggested by Hastie et al. (2009) when the study area is encompassed by a "data-rich situation", as described below.

1. Training set (50% of the data): used to fit the models.
2. Validation set (25% of the data): used to compare the performances of the prediction algorithms that were created based on the training set. The performances were evaluated by the prediction error for the models with data outside the training set.

   This data set contains the pre-classified results but they were not used for fitting the model.
3. Test set (25% of the data): used to assess the generalization error of the final chosen model by each algorithm. The test data set ran against the fitted model and the results compared to the unused pre-classified data and evaluated by a confusion matrix.

3.3 Change detection

Three machine learning algorithms, Artificial Neural Network, Random Forest, and Support Vector Machine, were chosen to detect the vegetation changes (BELGIU; DRĂGUȚ, 2016; MAS; FLORES, 2008; MOUNTRAKIS; IM; OGOLE, 2011). These non-parametric supervised classifiers did not make assumptions regarding frequency distribution, a good characteristic to represent remotely sensed data (LARY, 2010).

3.3.1 Artificial Neural Network

The first machine learning method evaluated was the artificial neural network (ANN) which have become a usual technique in the analysis of remotely sensed data since the evolution of artificial intelligence and the computational effort (MAS; FLORES, 2008).

In this study, the multilayer perceptron architecture was chosen because it is the most used class of ANN in applied fields. It is divided in a set of inputs, which were the spectral attributes of each object observed, a set of computation nodes or hidden layers, and two sets of output nodes that were *change* or *no change* object situation.

Backpropagation algorithm was applied together with the multilayer perceptron architecture. It is divided in a initialization phase assigning random numbers to the synaptic weights; forward computation that propagate the node values in the hidden layers (EQUATION 6); and back propagation computes the error signals corresponding to each neuron (EQUATION 7) (MAS; FLORES, 2008).

$$v_k = \sum_{j=0}^{m} w_{kj} * x_j \tag{6}$$

$$e_k = d_k - \varphi_k(v_k) \tag{7}$$

Where a neuron v_k is the defined by a synaptic weight w_{kj} and a signal x_j or a connection j. The error computed by the desired output of the neuron d_k and the activation function φ_k, which is defined in this study by the logistic activation function (EQUATION 8)

$$\varphi_k = \frac{1}{1+ e^{-a*v}} \tag{8}$$

In this study, the optimization of ANN includes the choice of hidden layers, learning rate and momentum for dataset validation. A set of configurations were tested in order to find the most successful and robust architecture. The parameters learning rate (LR) and momentum (Mom) were set in 0.1, 0.5 and 0.9 for five, ten and fifteen hidden layers (HL) (TABLE 7).

Table 7 – Parameter simulation in the multilayer perceptron architecture.

Parameter	**Simulations**								
LR	0.1	0.1	0.3	0.3	0.5	0.7	0.7	0.9	0.9
Mom.	0.7	0.9	0.7	0.9	0.5	0.1	0.3	0.1	0.3
HL	5	5	5	5	5	5	5	5	5
LR	0.1	0.1	0.3	0.3	0.5	0.7	0.7	0.9	0.9
Mom.	0.7	0.9	0.7	0.9	0.5	0.1	0.3	0.1	0.3
HL	10	10	10	10	10	10	10	10	10
LR	0.1	0.1	0.3	0.3	0.5	0.7	0.7	0.9	0.9
Mom.	0.7	0.9	0.7	0.9	0.5	0.1	0.3	0.1	0.3
HL	15	15	15	15	15	15	15	15	15

Legend: LR = Learning Rate; Mom. = Momentum; and HL = Hidden Layers.
Source: Author (2018).

3.3.2 Random Forest

Random Forest (RF) is a well-known ensemble learning method that combines classification or regression decision trees (BREIMAN, 2001). These decision trees are built by performing an individual learning algorithm that randomly splits the input dataset at each node into subsets based on an attribute

value test. In the output, all value tests at each tree are assessed and the data is classified by a majority of votes (PELLETIER et al., 2016)

Two parameters need to be set in order to build the classification trees: the number of attributes to be tested for the best split in the growing trees phase (Mtry), and the number of decision trees to be generated (Ntree). In this study, the parameters simulation were tested based on recent literature where the majority of the studies reported Ntree value to 500 and Mtry equals to the square root of the number of input variables (BELGIU; DRĂGUȚ, 2016). For the simulation procedure, Mtry was established into 2, 4 and 6 attributes per subset, and Ntree equals to 100, 500 and 1,000 (TABLE 8).

Table 8 – Parameter simulation in the Random Forest classifier.

Parameter	**Simulations**								
Mtry	2	4	6	2	4	6	2	4	6
Ntrees	100	100	100	500	500	500	1000	1000	1000

Legend: Mtry = Number of attributes to be tested; Ntree = Number of decision trees to be generated.
Source: Author (2018).

3.3.3 Support Vector Machine

Upgraded by Cortes and Vapnik (1995), the support vector machine (SVM) basically trains a group of objects to find a hyperplane that separates the dataset based on a predefined number of classes. The optimal hyperplane is set according to the decision boundary or support vectors that minimizes misclassifications (MOUNTRAKIS; IM; OGOLE, 2011).

The nu-SVC classifier, implemented in LIBSVM (CHANG; LIN, 2011), was used in the study. This classifier fit a support vector machine based on a *nu*

parameter which has the same function of *C* cost of general models but ranging from zero to one. *C* is the parameter related to soft margin cost function which controls the influence of each individual support vector and is related to the ratio of support vectors and the ratio of the training error.

In a complex space of features, a kernel function is applied to SVM in order to transform the attributes and then covert the non-linear space into a linear space. According to Mountrakis, Im and Ogole (2011), the kernel choice is the major challenge concerning the applicability of SVM, so in this study three kernel functions were computed in the dataset: polynomial (EQUATION 9), radial based (EQUATION 10) and sigmoidal (EQUATION 11).

$$K_{polynomial}(x_i,x_j) = (-\gamma * x_i * x_j + coef(\theta))^d \quad (9)$$

$$K_{radial}(x_i,x_j) = e^{-\gamma * \|x_i - x_i\|^2} \quad (10)$$

$$K_{sigmoidal}(x_i,x_j) = \tanh(-\gamma * x_i * x_j + coef(\theta)) \quad (11)$$

Where the parameters γ, *d*, and *nu* must be set by the user. In this study, a short parameter simulation was taken in order to evaluate and choose the best model (TABLE 9). The designation of an advanced selection was not considered because the three parameters can be quickly evaluated by a short number of tests.

Table 9 – Parameter simulation by different kernel functions.

Parameter	**Simulations**								
γ	0.1	0.5	1.0	0.1	0.5	1.0	0.1	0.5	1.0
nu	0.01	0.01	0.01	0.1	0.1	0.1	0.5	0.5	0.5
d*	1	1	1	1	1	1	1	1	1
γ	0.1	0.5	1.0	0.1	0.5	1.0	0.1	0.5	1.0
nu	0.01	0.01	0.01	0.1	0.1	0.1	0.5	0.5	0.5
d*	2	2	2	2	2	2	2	2	2
γ	0.1	0.5	1.0	0.1	0.5	1.0	0.1	0.5	1.0
nu	0.01	0.01	0.01	0.1	0.1	0.1	0.5	0.5	0.5
d*	3	3	3	3	3	3	3	3	3

Legend: γ = Gamma parameter of kernel functions; nu = parameter which has the same function of *C* cost of general models but ranging from zero to one; d = *degree* parameter (only simulated to polynomial kernel function).
Source: Author (2018).

3.4 Accuracy analysis

Whatever the procedure that generates a map is (change map, land cover map, etc.), its accuracy must be known. Among a set of reasons to assess the accuracy of a map, the need to improve the map quality and the possibility to compare techniques or algorithms are very important (CONGALTON; GREEN, 2009).

The accuracy analysis was divided into two parts: (1) evaluation of the validation datasets based on the multiple fitted models, and (2) accuracy of the change maps based the final chosen model by each algorithm.

The validation dataset was evaluated by the square root of the residual difference between the observed data and the predicted data, the root mean

squared error (EQUATION 12). For each machine learning algorithm, the error measure was computed for all simulations and the best model (lowest RMSE) was selected to test data input. Boxplot charts displayed the dataset variability in order to support the analysis.

$$RMSE = \sqrt{\frac{\sum_{j=1}^{N}\left(y_j - \hat{y}_j\right)^2}{N}} \quad (12)$$

The second part of accuracy analysis consisted to compare the changed maps provided by the best model fit. The classified objects as change or no change in the test dataset were evaluated by high resolution imagery as Rapideye imagery and Google Earth/Bing platforms (when available). As in chapter one, burn events were unrepresentative in the sampling areas, so not included in the analysis.

A confusion matrix was build and overall accuracy (EQUATION 13), user's accuracy (inversely related to commission error) (EQUATION 14), producer's accuracy (inversely related to omission error) (EQUATION 15), and Kappa index (EQUATION 16) were calculated.

$$\text{Overall accuracy} = \frac{\sum_{i=1}^{k} n_{ii}}{n} \quad (13)$$

$$\text{User's accuracy} = \frac{n_{ii}}{n_{i+}} \quad (14)$$

$$\text{Producer's accuracy} = \frac{n_{ii}}{n_{+j}} \quad (15)$$

$$K = \frac{\frac{\sum_{i=1}^{k} n_{ii}}{n} - \frac{\sum_{i=1}^{k}\left(n_{i+} * n_{+j}\right)}{n^2}}{1 - \frac{\sum_{i=1}^{k}\left(n_{i+} * n_{+j}\right)}{n^2}} \quad (16)$$

Where, n_{ii} is the correct classified objects in the major diagonal by a k number of classes; n is the total of observations; n_{i+} is the total in line; and n_{+j} is the total in column.

The validation and test datasets were joined in order to create and visualize the change maps in the study area.

3.5 Attribute evaluation

For the three changed maps produced (ANN, RF and SVM), the attribute importance was computed by the analysis of RMSE variation when a particular spectral channel is removed from the training dataset, where an error increase leads to an importance increase for an attribute and vice versa. The difference between the general error and the attribute evaluation error ranked the importance of the fourteen attributes for a particular machine learning algorithm.

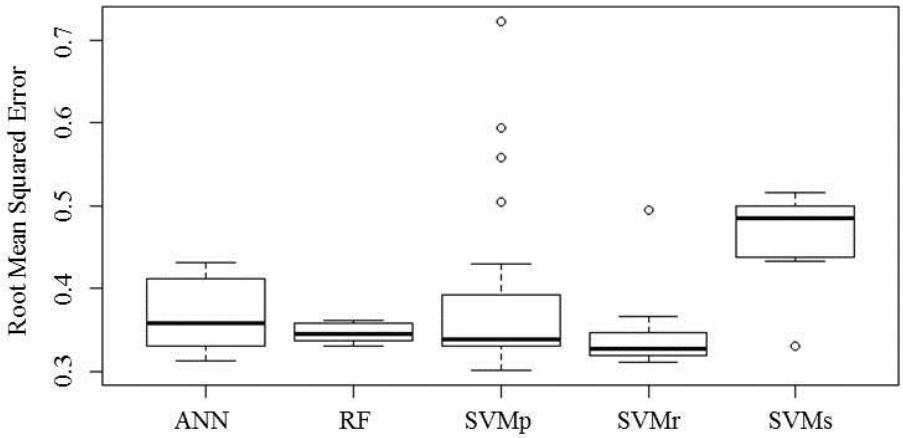
4 RESULTS AND DISCUSSION

4.1 Model accuracies

The error measures for the three machine learning algorithms simulations are registered in Table 10. The best parameter simulation for ANN classifier, which generates a RMSE of 0.3125 was the leaning rate set in 0.1, momentum in 0.9 and 10 hidden layers; for Random Forest, 2 attributes subset for 500 trees generated an error of 0.3296; and for SVM, a polynomial kernel function with gamma parameters set in 1.0, *nu* equal to 0.1, and degree equal to 4 generated an error of 0.3005.

Besides SVM had generated the lowest error among the MLA, Random Forest performed the lowest error variation among simulations followed by ANN, and SVM with outliers (FIGURE 17). Based on these parameters tests, RF was less sensitive and still returned good results. The support vector machines presented high error variability with outliers for the three kernel functions.

Figure 17 – Model simulation boxplots.



Source: Author (2018).

In order to reach low errors measures in the fit step, model parameters need to be comprehended. In RF and SVM radial kernel, the soft parameter simulation presented subtle error variations and provided good information to model choice, but advanced simulations involving more parameters and combinations are also appreciated with regards to improve change detection accuracies.

Table 10 – Model simulations and the respectively root mean squared errors. (Continue)

Parameters								Root Mean Squared Error				
LR	**Mom**	**HL**	**Mtry**	**Ntrees**	**γ**	**nu**	**d**	**ANN**	**RF**	**SVMp**	**SVMr**	**SVMs**
0.1	0.7	5	2	100	0.1	0.01	1	0.3481	0.3373	0.4294	0.4952	0.4999
0.1	0.9	5	4	100	0.5	0.01	1	0.3160	0.3413	0.4050	0.3665	0.4416
0.3	0.7	5	6	100	1.0	0.01	1	0.3858	0.3609	0.3394	0.3180	0.4853
0.3	0.9	5	2	500	0.1	0.1	1	0.3221	0.3296	0.3332	0.3465	0.4379
0.5	0.5	5	4	500	0.5	0.1	1	0.4262	0.3454	0.3585	0.3318	0.4327
0.7	0.1	5	6	500	1.0	0.1	1	0.3847	0.3573	0.7235	0.3143	0.5165
0.7	0.3	5	2	1,000	0.1	0.5	1	0.4194	0.3335	0.3295	0.3266	0.3294
0.9	0.1	5	4	1,000	0.5	0.5	1	0.4081	0.3468	0.3294	0.3188	0.5156
0.9	0.3	5	6	1,000	1.0	0.5	1	0.3186	0.3573	0.3294	0.3103	0.5004
0.1	0.7	10	-	-	0.1	0.01	2	0.3357	-	0.4201	-	-
0.1	0.9	10	-	-	0.5	0.01	2	0.3125	-	0.5942	-	-
0.3	0.7	10	-	-	1.0	0.01	2	0.3203	-	0.3278	-	-
0.3	0.9	10	-	-	0.1	0.1	2	0.3281	-	0.5589	-	-
0.5	0.5	10	-	-	0.5	0.1	2	0.3313	-	0.3438	-	-

Table 10 – Model simulations and the respectively root mean squared errors. (Conclusion)

0.7	0.1	10	-	-	1.0	0.1	2	0.3253	-	0.3065	-	-
0.7	0.3	10	-	-	0.1	0.5	2	0.3361	-	0.3308	-	-
0.9	0.1	10	-	-	0.5	0.5	2	0.3407	-	0.3312	-	-
0.9	0.3	10	-	-	1.0	0.5	2	0.3763	-	0.3312	-	-
0.1	0.7	15	-	-	0.1	0.01	3	0.3582	-	0.3807	-	-
0.1	0.9	15	-	-	0.5	0.01	3	0.3391	-	0.3370	-	-
0.3	0.7	15	-	-	1.0	0.01	3	0.4146	-	0.3080	-	-
0.3	0.9	15	-	-	0.1	0.1	3	0.3633	-	0.5039	-	-
0.5	0.5	15	-	-	0.5	0.1	3	0.4233	-	0.3059	-	-
0.7	0.1	15	-	-	1.0	0.1	3	0.4141	-	0.3005	-	-
0.7	0.3	15	-	-	0.1	0.5	3	0.4315	-	0.3379	-	-
0.9	0.1	15	-	-	0.5	0.5	3	0.4162	-	0.3381	-	-
0.9	0.3	15	-	-	1.0	0.5	3	0.3594	-	0.3381	-	-

Legend: LR = Learning Rate parameter in ANN; Mom. = Momentum parameter in ANN; HL = Hidden Layers parameter in ANN; Mtry = Number of attributes to be tested in RF; Ntree = Number of decision trees to be generated in RF; γ = Gamma parameter of kernel functions in SVM; nu = parameter which has the same function of *C* cost of general models but ranging from zero to one in SVM; d = *degree* parameter in SVM polynomial. Source: Author (2018).

4.2 Change maps accuracies

Based on the results of the validation phase, the performance of Artificial Neural Network (LR = 0.1; Mom = 0.9; HL = 10), Random Forest (Mtry = 2; Ntree = 500), and Support Vector Machine (γ = 1.0; nu = 0.1; d = 3) on the test data resulted in the confusion matrixes in Figure 18. In general, ANN and RF methods showed high accuracies; follow by a moderate accuracy in SVM.

Although Support Vector Machine had presented the lowest RMSE in model choice, the overall accuracy and kappa are the poorest among the algorithms, 80.0% and 0.60 respectively. Omission and commission errors were the most inferior as well, 28.0% and 20.0% respectively.

Random Forest presented better general results with overall accuracy of 92.0% and Kappa index of 0.84. It also showed a satisfactory commission error of 4.3%, best among the MLA, and an omission error of 12.0%. This low commission error is a good point in order to detect changes in the study area, which means the algorithm was almost not affected by the seasonal noise.

As in Random Forest, the neural network showed satisfactory overall accuracy of 86% and kappa of 0.72. The omission error of 4.0% was the lowest among the others methods, but the highest commission error of 20.0%.

The root mean squared error reached in the validation phase did not follow a rank pattern in the test phase. This divergence might be related to dataset dependence, once the test and validation sets came from different regions of the study area.

The change maps drawing the sampled objects used in the validation and test phases are presented in the Figure 19 to Figure 22, according to the sampling areas.

Figure 18 – Confusion matrixes and accuracy results for selected models.

Artificial Neural Network

Classified Data	Reference Data: Ch	Reference Data: Nch	Row Total
Ch	24	6	30
Nch	1	19	20
Column Total	25	25	50

Random Forest

Classified Data	Reference Data: Ch	Reference Data: Nch	Row Total
Ch	22	1	23
Nch	3	24	27
Column Total	25	25	50

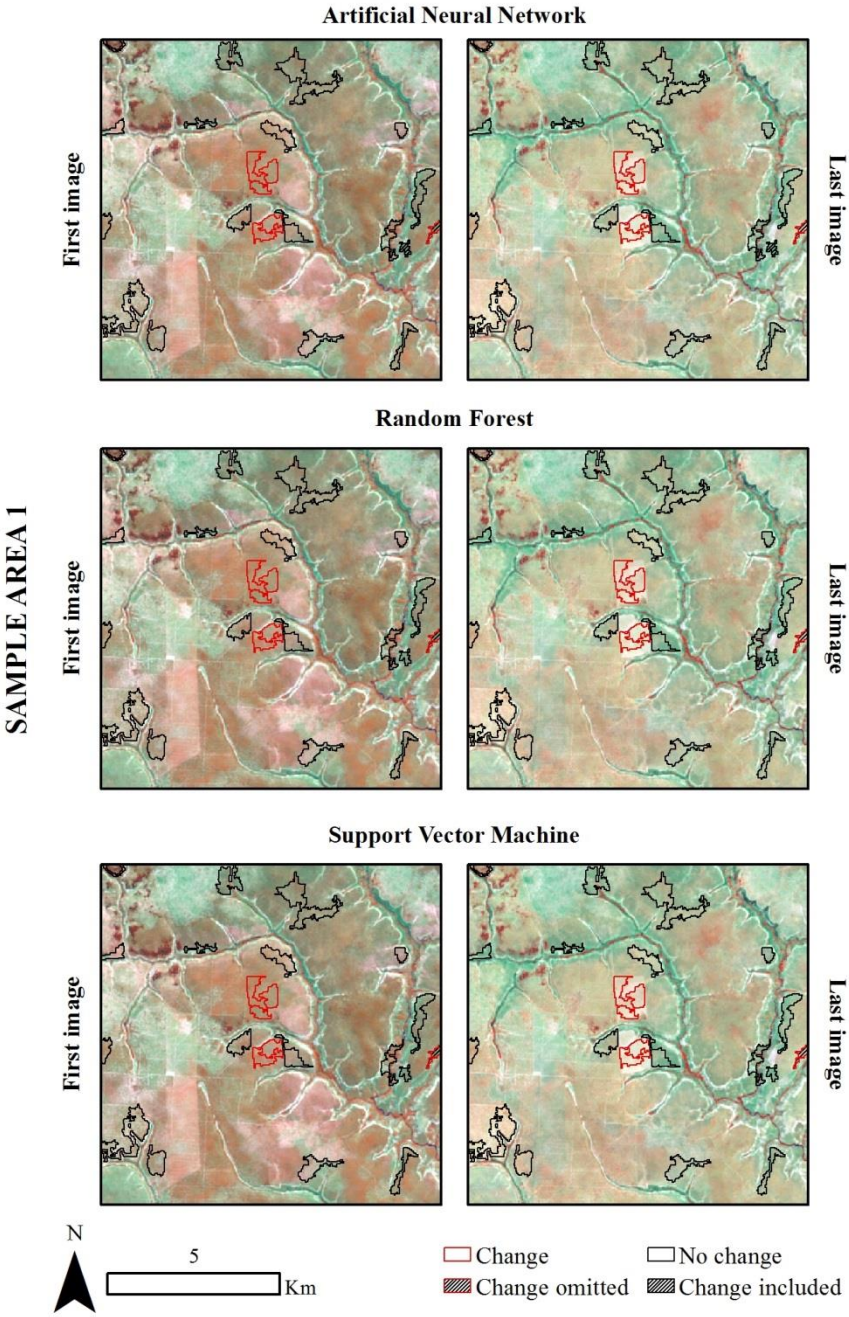
Support Vector Machine

Classified Data	Reference Data: Ch	Reference Data: Nch	Row Total
Ch	18	3	21
Nch	7	22	29
Column Total	25	25	50

MLA	Overall Accuracy (%)	Kappa	Omission Error (%)	Commission Error (%)
ANN	86.0	0.72	4.0	20.0
RF	92.0	0.84	12.0	4.3
SVM	80.0	0.60	28.0	14.3

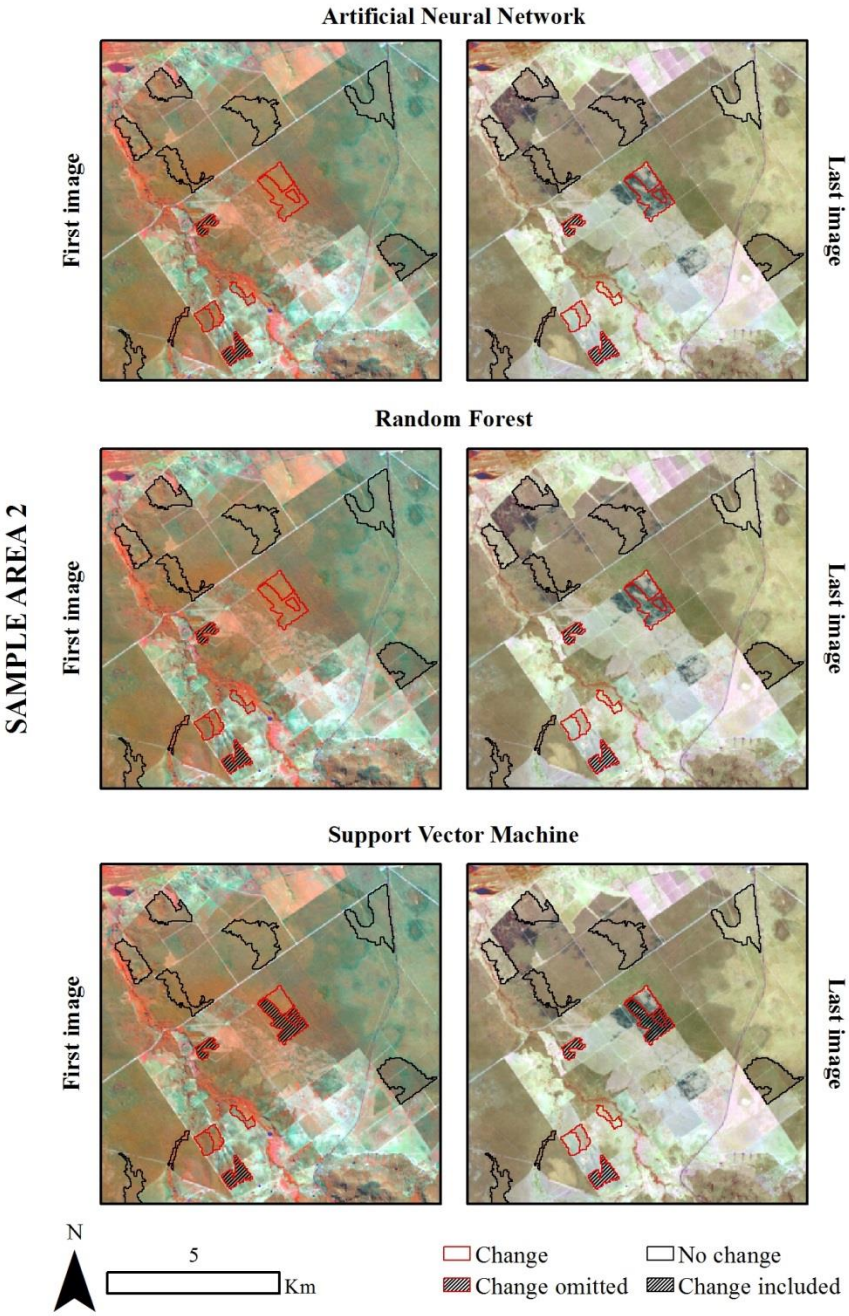
Source: Author (2018).

Figure 19 – Change map in the sampling area number one.



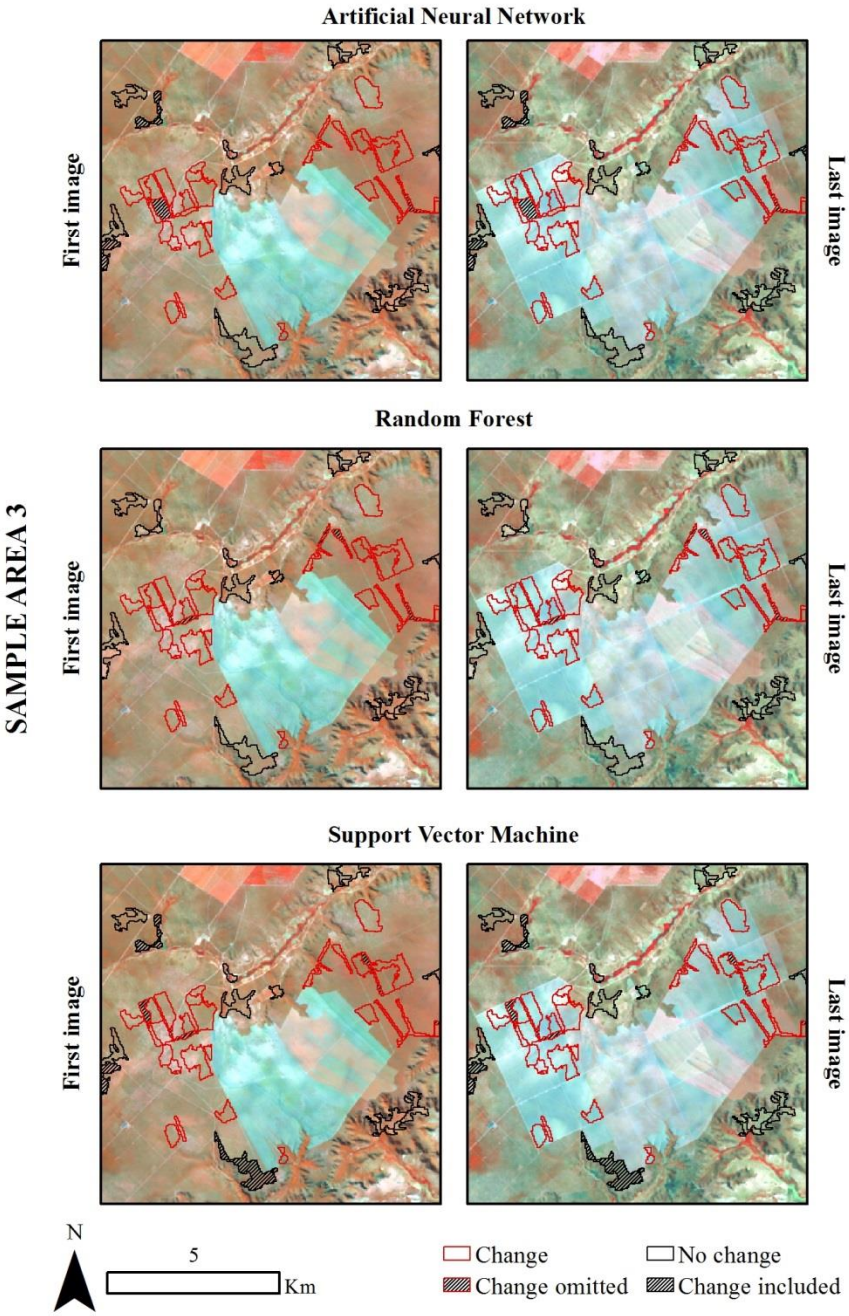
Source: Author (2018).

Figure 20 – Change map in the sampling area number two.



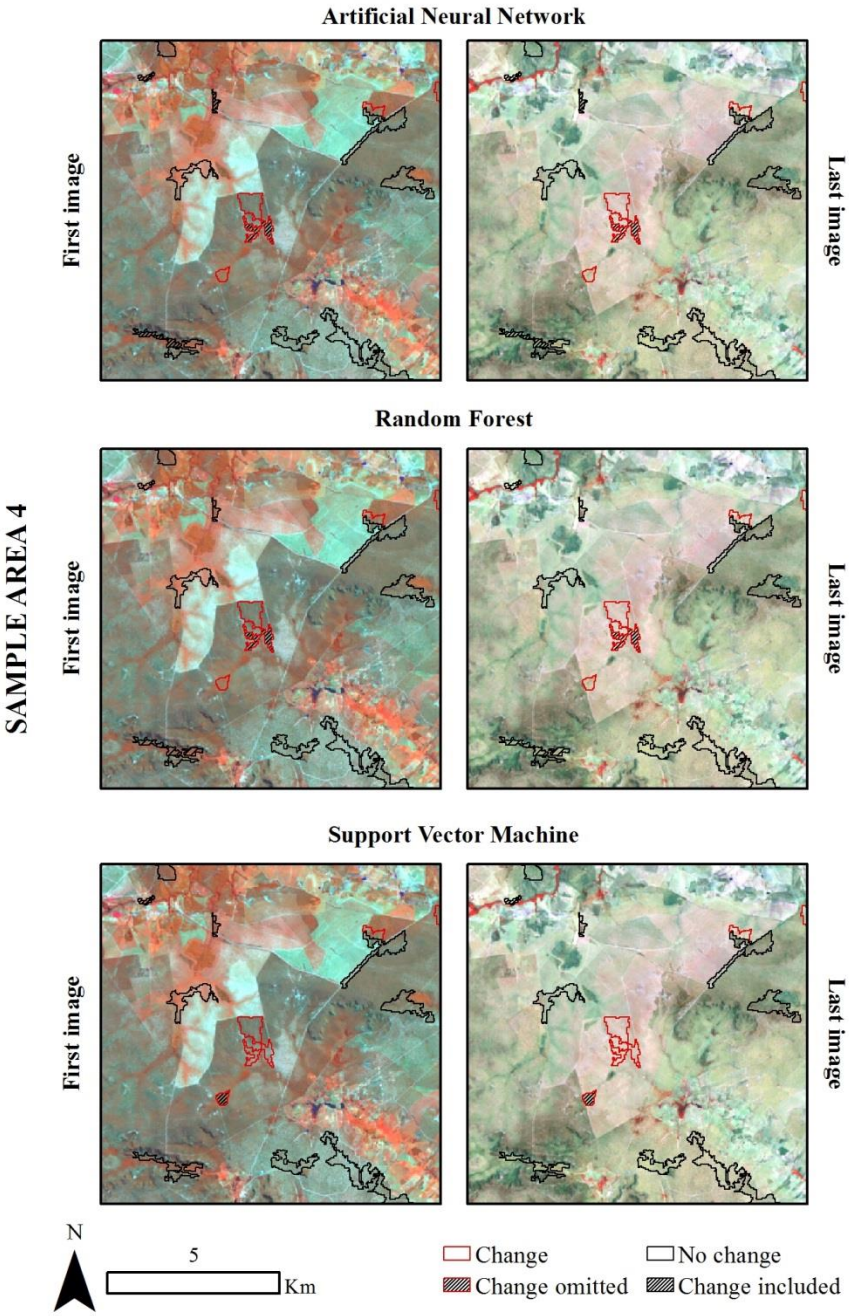
Source: Author (2018).

Figure 21 – Change map in the sampling area number three.



Source: Author (2018).

Figure 22 – Change map in the sampling area number four.



Source: Author (2018).

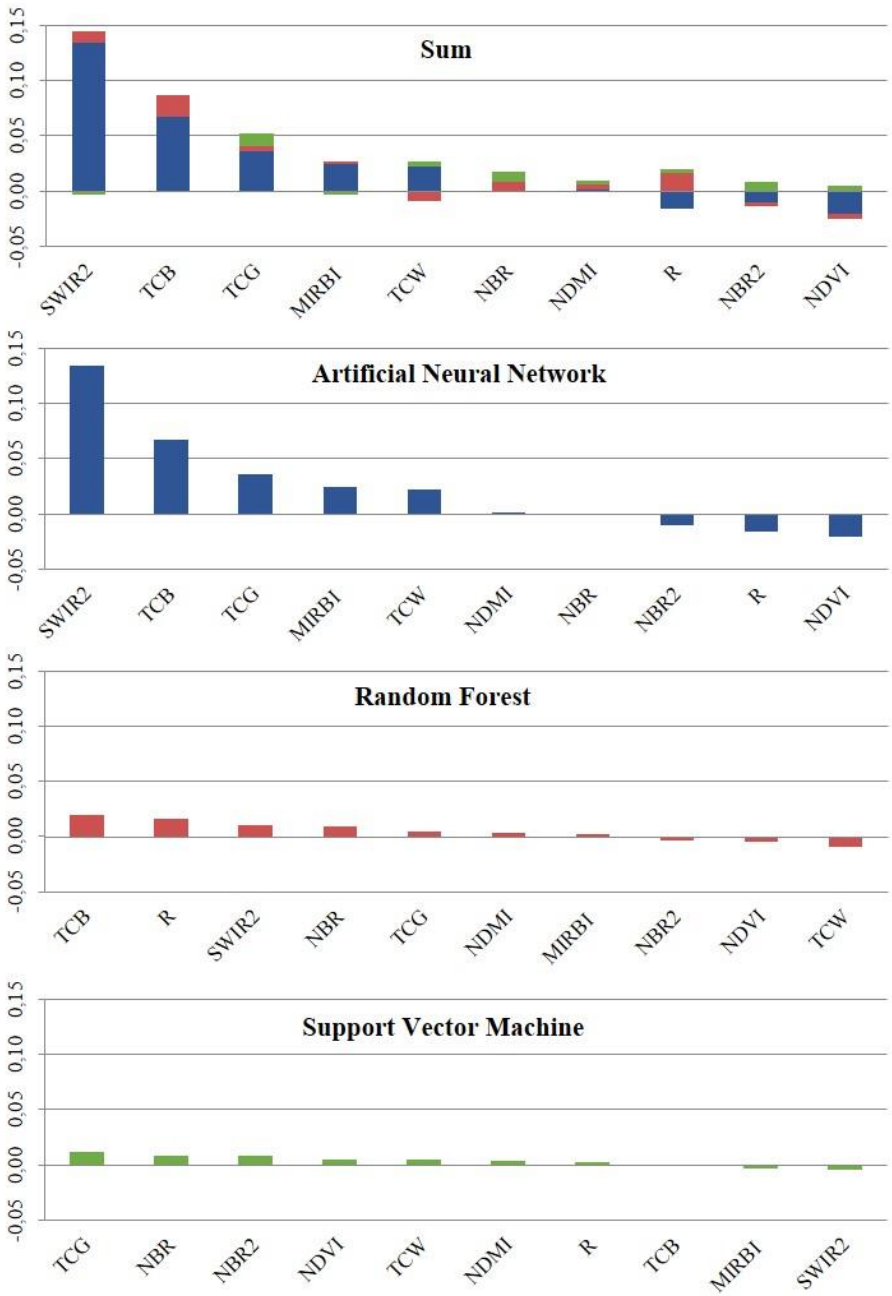
4.3 Importance of spectral channels

The importance of each spectral channel related to RMSE are shown in Figure 23. SWIR2 and TCB emerged as two of the most important attributes when ANN and RF were considered. TCG was the most important attribute in Support Vector Machine and was also present in the top five for the others merhods.

Artificial neural network demonstrated to be very sensitive to attribute permutation, which might be relate to the algorithm architecture. While SWIR2 had an increase of 43% and NDVI a decrease of 6% of the RMSE in neural network, Random Forest and Support Vector Machine had a RMSE increase of 6% and 5% and a decrease of 3% and 0.3% respectively. This less sensitivity by these algorithms may assurance a good point in order to reach high accuracy in change detection.

Some attributes presented divergences between methods as SWIR2 being the most important attribute for ANN and the worse one for SVM; also TCB important in ANN and RF, and almost irrelevant in SVM. These differences bring attention to the use of certain attributes and algorithms in univariate change detection studies, which can generate poor accuracies.

Figure 23 – Spectral channels importance for individual MLA.



Source: Author (2018).

4.4 Method limitations

Some limitations to the method used in this chapter are discussed in this section as sampling considerations, timing of changes, appliance in large datasets, and extend knowledge in machine learning algorithms. DeVries et al. (2016) also reported some of these limitations but focused on uncertainties in local expert data.

The data used to evaluate the machine learning was sample at the area in order to reduce computational and operational efforts. Heterogeneous landscapes in terms of vegetation gradients must be stratified in order to reduce mixed information once large areas can provide change uncertainties.

The timing of change is an important feature of forest monitoring. Even though the multidate segmentation procedure has the temporal information and detects the changes with high accuracy in time, it did not track the change timing for a particular image-object. This limitation is a crucial point in order to give continuity and consistency to the method in detection and monitoring studies.

The use of a large set of images in order to create time series can be a storage problem in monitoring studies but as discussed in this study, large sets can be reduced once small variations of the number of images may not affect the detection of changed areas. Nevertheless, as discussed in Chapter 1 section 4.3, the expansion of the temporal series increases temporal and spectral information in the segmentation process and might make the multidate segmentation difficult to be explored by creating small objects. Also, a large amount of temporal (long time series), spectral (multiple bands) and spatial information (large areas) can take an extensive computational effort when these data is processed are the segmentation procedure.

In this study, machine learning algorithms were run in Weka 3.8 software with a few simulations and all others parameters set by the software

default. Extended knowledge on these algorithms can fit alternative architectures and provide better results for change detection.

These limitations must be more explored in savanna regions, since the challenge of detect change is recognized in these areas.

5 CONCLUSION

This chapter compared the performance of Landsat OLI spectral channels and spectral indices in Cerrado change detection using machine learning algorithms, and evaluated the importance of these channels.

The three algorithms presented satisfactory performances to discriminate human-induced change events from seasonal noise but Random Forest demonstrated the best results in the test phase with high overall accuracy. It also presented low error variation in the parameter simulation.

Spectral channels as shortwave infrared and tasseled cap greenness indicated positive influence in the change detection. The attribute ranking does not bring to univariate studies since the complexity of the Cerrado biome, but it may support options to create spectral subsets to eventual studies.

Future researches can also follow-up this approach by exploring different databases such as Sentinel imagery, and looking forward some limitation issues as the sampling factor, timing of changes, extended knowledge in machine learning, and the manipulation of massive sets of spatial (extensive areas) or spectral data (long time series).

BIBLIOGRAPHY

ACERBI JÚNIOR, F. W. et al. Change detection in brazilian savannas using semivariograms derived from NDVI images. **Ciência e Agrotecnologia**, Lavras, v. 39, n. 2, p. 103-109, 2015.

ADELABU, S. et al. Exploiting machine learning algorithms for tree species classification in a semiarid woodland using RapidEye image. **Journal of Applied Remote Sensing**, Bellingham, v. 7, n. 1, p. 73480, 2013.

ALI, I. el al. Review of machine learning approaches for biomass and soil moisture retrievals from remote sensing data. **Remote Sensing**, Basel, v.7, n. 12, p. 16398-16421, 2015.

ALMEIDA, J. et al. Applying machine learning based on multiscale classifiers to detect remote phenology patterns in Cerrado savanna trees. **Ecological Informatics**, Amsterdam, v. 23, p. 49-61, 2014

ARAUZO-AZOFRA, A.; BENITEZ, J. M.; CASTRO, J. L. Consistency measures for feature selection. **Journal of Intelligent Information Systems**, Dordrecht, v. 30, n. 3, p. 273-292, 2008.

AVITABILE, V. et al. Capabilities and limitations of Landsat and land cover data for aboveground woody biomass estimation of Uganda. **Remote Sensing of Environment**, New York, v. 117, p. 366-380, 2012.

BAATZ, M.; SCHÄPE, A. Multiresolution segmentation: an optimization approach for high quality multi-scale image segmentation. In: STROBL, J; BLASCHKE, T; GRIESBNER, G. **Angewandte Geographische Informations-Verarbeitung, XII.** Karlsruhe: Wichmann, 2000. v. 58, n. 3-4, p. 12-23.

BAIG, M. H. A. et al. Derivation of a tasselled cap transformation based on Landsat 8 at-satellite reflectance. **Remote Sensing Letters**, Abingdon, England, v. 5, n. 5, p. 423-431, 2014.

BANSKOTA, A. et al. Forest Monitoring Using Landsat Time Series Data: A Review. **Canadian Journal of Remote Sensing**, Kanata, v. 40, n. 5, p. 362-384, 2014.

BELGIU, M.; DRĂGUŢ, L. Random forest in remote sensing: A review of applications and future directions. **ISPRS Journal of Photogrammetry and**

Remote Sensing, Amsterdam, v. 114, p. 24-31, 2016.

BEUCHLE, R. et al. Land cover changes in the Brazilian Cerrado and Caatinga biomes from 1990 to 2010 based on a systematic remote sensing sampling approach. **Applied Geography**, Kidlington, v. 58, p. 116-127, 2015.

BLASCHKE, T. Object based image analysis for remote sensing. **ISPRS Journal of Photogrammetry and Remote Sensing**, Amsterdam, v. 65, n. 1, p. 2-16, 2010.

BONTEMPS, S. et al. An object-based change detection method accounting for temporal dependences in time series with medium to coarse spatial resolution. **Remote Sensing of Environment**, New York, v. 112, n. 6, p. 3181-3191, 2008.

BONTEMPS, S.; LANGNER, A.; DEFOURNY, P. Monitoring forest changes in Borneo on a yearly basis by an object-based change detection algorithm using SPOT-VEGETATION time series. **International Journal of Remote Sensing**, Abingdon, England, v. 33, n. 15, p. 4673-4699, 2012.

BOULILA, W. et al. A data mining based approach to predict spatiotemporal changes in satellite images. **International Journal of Applied Earth Observation and Geoinformation**, Amsterdam, v. 13, n. 3, p. 386-395, 2011.

BOVOLO, F.; BRUZZONE, L.; MARCONCINI, M. A novel approach to unsupervised change detection based on a semisupervised SVM and a similarity measure. **IEEE Transactions on Geoscience and Remote Sensing**, Piscataway, v. 46, n. 7, p. 2070-2082, 2008.

BRANDT, J. S. et al. Using Landsat imagery to map forest change in southwest China in response to the national logging ban and ecotourism development. **Remote Sensing of Environment**, New York, v. 121, p. 358-369, 2012.

BRANNSTROM, C. et al. Land change in the Brazilian Savanna (Cerrado), 1986-2002: Comparative analysis and implications for land-use policy. **Land Use Policy**, Kidlington, v. 25, n. 4, p. 579-595, 2008.

BRASIL. Ministério do Meio Ambiente. **Mapeamento do uso e cobertura do Cerrado:** projeto TerraClass Cerrado 2013. Brasília, DF, 2015. 67 p.

BRASIL. Ministério do Meio Ambiente. **Monitoramento do Desmatamento nos Biomas Brasileiros Por Satélite: Cerrado 2002-2008**. Brasília, DF, 2009.

BRASIL. Ministério do Meio Ambiente. **Monitoramento do Desmatamento**

nos Biomas Brasileiros Por Satélite: Cerrado 2009-2010. Brasília, DF, 2011.

BRASIL. Ministério do Meio Ambiente. **Monitoramento do Desmatamento nos Biomas Brasileiros Por Satélite: Cerrado 2010-2011**. Brasília, DF, 2015.

BREIMAN, L. Random forests. **Machine Learning**, Dordrecht, v. 45, n. 1, p. 5-32, 2001.

CARREIRAS, J. M. B.; VASCONCELOS, M. J.; LUCAS, R. M. Understanding the relationship between aboveground biomass and ALOS PALSAR data in the forests of Guinea-Bissau (West Africa). **Remote Sensing of Environment**, New York, v. 121, p. 426-442, 2012.

CARVALHO, L. M. T.; SCOLFORO, J. R. S. **Inventário florestal de Minas Gerais: Monitoramento da Flora Nativa 2005-2007**. Lavras: Editora UFLA, 2008.

CHANDRASHEKAR, G.; SAHIN, F. A survey on feature selection methods. **Computers and Electrical Engineering**, Kidlington, v. 40, n. 1, p. 16-28, 2014.

CHANG, C. C.; LIN, C. J. A Library for Support Vector Machines. **ACM Transactions on Intelligent Systems and Technology**, New York, v. 2, n. 3, p. 27, 2011.

CHEN, G. et al. Object-based change detection. **International Journal of Remote Sensing**, Abingdon, England, v. 33, n. 14, p. 4434-4457, 2012.

CHEN, G.; HAY, G. J.; ST-ONGE, B. A GEOBIA framework to estimate forest parameters from lidar transects, Quickbird imagery and machine learning: A case study in Quebec, Canada. **International Journal of Applied Earth Observation and Geoinformation**, Amsterdam, v. 15, n. 1, p. 28-37, 2012.

CHEN, J. et al. A spectral gradient difference based approach for land cover change detection. **ISPRS Journal of Photogrammetry and Remote Sensing**, Amsterdam, v. 85, p. 1-12, 2013.

COELHO, V. H. R. et al. Dinâmica do uso e ocupação do solo em uma bacia hidrográfica do semiárido brasileiro. **Revista Brasileira de Engenharia Agrícola e Ambiental**, Campina Grande, v. 18, n. 1, p. 64-72, 2014.

COLGAN, M. S. et al. Mapping savanna tree species at ecosystem scales using support vector machine classification and BRDF correction on airborne hyperspectral and LiDAR data. **Remote Sensing**, Basel, v. 4, n. 11, p. 3462-

3480, 2012.

CONGALTON, R. G.; GREEN, K. **Assessing the accuracy of remotely sensed data: principles and practices**. 2nd ed. Boca Raton: CRC Press, 2009.

COPPIN, P. et al. Digital change detection methods in ecosystem monitoring: a review. **International Journal of Remote Sensing**, Abingdon, v. 25, n. 9, p. 1565-1596, 2004.

CORTES, C.; VAPNIK, V. Support Vector Networks. **Machine Learning**, Dordrecht, v. 20, n. 3, p. 273-297, 1995.

COUTO JÚNIOR, A. C. de. S. **Monitoramento do Cerrado em Minas Gerais usando análises estatísticas baseadas em objetos: uma abordagem em diferentes escalas**. Lavras: Editora UFLA, 2011.

DACAMARA, C. C. et al. A User-Oriented Simplification of the (V,W) Burn-Sensitive Vegetation Index System. **IEEE Geoscience and Remote Sensing Letters**, Piscataway, v. 13, n. 12, p. 1822-1826, 2016.

DEFINIENS, A. Definiens eCognition Developer 8 User Guide. Munich: Definiens AG. 2009.

DESCLÉE, B.; BOGAERT, P.; DEFOURNY, P. Forest change detection by statistical object-based method. **Remote Sensing of Environment**, New York, v. 102, n. 1-2, p. 1-11, 2006.

DEVANEY, J. et al. Forest cover estimation in Ireland using radar remote sensing: A comparative analysis of forest cover assessment methodologies. **PLoS ONE**, San Francisco, v. 10, n. 8, p. 1-27, 2015.

DEVRIES, B. et al. Characterizing Forest Change Using Community-Based Monitoring Data and Landsat Time Series. **PLOS ONE**, San Franscico, v. 11, n. 3, p. e0147121, 2016.

FERREIRA, L. G.; HUETE, A. R. Assessing the seasonal dynamics of the Brazilian Cerrado vegetation through the use of spectral vegetation indices. **International Journal of Remote Sensing**, Abingdon, England, v. 25, n. 10, p. 1837-1860, 2004.

FERREIRA, L. G. et al. Seasonal landscape and spectral vegetation index dynamics in the Brazilian Cerrado: An analysis within the Large-Scale Biosphere-Atmosphere Experiment in Amazonia (LBA). **Remote Sensing of**

Environment, New York, v. 87, n. 4, p. 534-550, 2003.

FOLEY, J. A. et al. Global consequences of land use. **Science**, Washington, D.C., v. 309, n. 5734, p. 570-574, 2005.

FRANK, E.; HALL, M. A.; WITTEN, I. H. **The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques"**. 4th ed. Burlington: Morgan Kaufmann, 2016.

GAMANYA, R.; MAEYER, P. de; DAPPER, M. de. Object-oriented change detection for the city of Harare, Zimbabwe. **Expert Systems with Applications**, Kidlington, v. 36, n. 1, p. 571-588, 2009.

GIBBS, H. K. et al. Tropical forests were the primary sources of new agricultural land in the 1980s and 1990s. **Proceedings of the National Academy of Sciences**, Washington, D.C., v. 107, n. 38, p. 16732-16737, 2010.

GRECCHI, R. C. et al. Assessing the spatio-temporal rates and patterns of land-use and land-cover changes in the Cerrados of southeastern Mato Grosso, Brazil. **International Journal of Remote Sensing**, Abingdon, England, v. 34, n. 15, p. 5369-5392, 2013.

GRECCHI, R. C. et al. Land use and land cover changes in the Brazilian Cerrado: A multidisciplinary approach to assess the impacts of agricultural expansion. **Applied Geography**, Kidlington, v. 55, p. 300-312, 2014.

GUERSCHMAN, J. P. et al. Estimating fractional cover of photosynthetic vegetation, non-photosynthetic vegetation and bare soil in the Australian tropical savanna region upscaling the EO-1 Hyperion and MODIS sensors. **Remote Sensing of Environment**, New York, v. 113, n. 5, p. 928-945, 2009.

GUINDON, L. et al. Annual mapping of large forest disturbances across Canada's forests using 250 m MODIS imagery from 2000 to 2011. **Canadian Journal of Forest Research**, Ottawa, v. 44, n. 12, p. 1545-1554, 2014.

GUO, Y. et al. Optimal Support Vector Machines for forest above-ground biomass estimation from multisource remote sensing data. In: **Geoscience and Remote Sensing Symposium (IGARSS)**. Munich: IEEE International, 2012. p. 6388-6391.

HALL, M. A. **Correlation-based Feature Selection for Machine Learning**. 1999. 178 p. Thesis-University of Waikato, Hamilton, 1999.

HALL, O.; HAY, G. J. A Multiscale Object-Specific Approach to Digital Change Detection. **International Journal of Applied Earth Observation and Geoinformation**, Amsterdam, v. 4, n. 4, p. 311-327, 2003.

HANSEN, M. C. et al. High-resolution global maps of 21st-century forest cover change. **Science**, Washington, D.C., v. 342, n. 6160, p. 850-853, 2013.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical Learning**. 2nd ed. New York: Springer, 2009.

HEALEY, S. P. et al. Mapping forest change using stacked generalization: An ensemble approach. **Remote Sensing of Environment**, New York, v. 204, p. 717-728, 2018.

HILL, M. J. et al. Relationships between vegetation indices, fractional cover retrievals and the structure and composition of Brazilian Cerrado natural vegetation. **International Journal of Remote Sensing**, Abingdon, England, v. 38, n. 3, p. 874-905, 2017.

HUANG, C. et al. Use of a dark object concept and support vector machines to automate forest cover change analysis. **Remote Sensing of Environment**, New York, v. 112, n. 3, p. 970-985, 2008.

HUSSAIN, M. et al. Change detection from remotely sensed images: From pixel-based to object-based approaches. **ISPRS Journal of Photogrammetry and Remote Sensing**, Amsterdam, v. 80, p. 91-106, 2013.

JENSEN, J. R. **Remote Sensing of the Environment: An Earth Resource Perspective**. 2nd ed. Upper Saddle River: Pearson Prentice Hall, 2009.

JIA, J. et al. Object-oriented feature selection of high spatial resolution images using an improved Relief algorithm. **Mathematical and Computer Modelling**, Kidlington, v. 58, n. 3-4, p. 619-626, 2013.

JIN, S.; SADER, S. A. Comparison of time series tasseled cap wetness and the normalized difference moisture index in detecting forest disturbances. **Remote Sensing of Environment**, New York, v. 94, n. 3, p. 364-372, 2005.

JUEL, A. et al. Spatial application of Random Forest models for fine-scale coastal vegetation classification using object based analysis of aerial orthophoto and DEM data. **International Journal of Applied Earth Observation and Geoinformation**, Amsterdam, v. 42, p. 106-114, 2015.

KARJALAINEN, M. et al. Prediction of plot-level forest variables using TerraSAR-X stereo SAR data. **Remote Sensing of Environment**, New York, v. 117, p. 338-347, 2012.

KENNEDY, R. E.; YANG, Z.; COHEN, W. B. Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr — Temporal segmentation algorithms. **Remote Sensing of Environment**, New York, v. 114, n. 12, p. 2897-2910, 2010.

KEY, C. H.; BENSON, N. C. Landscape assessment: ground measure of severity, the composite burn index; and remote sensing of severity, the normalized burn ratio. In: LUTES, D. C. et al. **FIREMON: Fire effects monitoring and inventory system.** Ogden: USDA Forest Service, Rocky Mountain Research Station, 2006. p. LA1-LA51.

KLINK, C. A.; MACHADO, R. B. Conservation of the Brazilian Cerrado. **Conservation Biology**, Hoboken, v. 19, n. 3, p. 707-713, 2005.

KONONENKO, I. Estimating attributes: Analysis and extensions of RELIEF. In: BERGADANO, F; RAEDT, L. de. **Machine Learning: ECML-94**. Berlin: Springer, 1994. p.171-182.

KOVALSKYY, V.; ROY, D. P. The global availability of Landsat 5 TM and Landsat 7 ETM+ land surface observations and implications for global 30m Landsat data product generation. **Remote Sensing of Environment**, New York, v. 130, p. 280-293, 2013.

LARDEUX, C. et al. Support vector machine for multifrequency SAR polarimetric data classification. **IEEE Transactions on Geoscience and Remote Sensing**, Piscataway, v. 47, n. 12, p. 4143-4152, 2009.

LARY, D. J. Artificial Intelligence in Geoscience and Remote Sensing. In: IMPERATORE, P; RICCIO, D. **Geoscience and Remote Sensing New Achievements**. Vukovar: InTech, 2010. p.1-24, 2010.

LARY, D. J. et al. Machine learning in geosciences and remote sensing. **Geoscience Frontiers**, Beijing, v. 7, n. 1, p. 3-10, 2016.

LEFEBVRE, A.; CORPETTI, T.; HUBERT-MOY, L. Object-Oriented Approach and Texture Analysis for Change Detection in Very High Resolution Images. In: **Geoscience and Remote Sensing Symposium (IGARSS)**. Boston: IEEE International, 2008. p. IV663-IV666.

LEHMANN, C. E. R. et al. Deciphering the distribution of the savanna biome. **New Phytologist**, Hoboken, v. 191, n. 1, p. 197-209, 2011.

LIBONATI, R. et al. On a new coordinate system for improved discrimination of vegetation and burned areas using MIR/NIR information. **Remote Sensing of Environment**, New York, v. 115, n. 6, p. 1464-1477, 2011.

LIBONATI, R. et al. Retrieving middle-infrared reflectance using physical and empirical approaches: Implications for burned area monitoring. **IEEE Transactions on Geoscience and Remote Sensing**, Piscataway, v. 50, n. 1, p. 281-294, 2012.

LILLESAND, T. M.; KIEFER, R. W.; CHIPMAN, J. **Remote sensing and image interpretation**. 7º ed. Nova Jersey: Wiley, 2008.

LIU, H.; SETIONO, R. A probabilistic approach to feature selection - A filter solution. In: **13th International Conference on Machine Learning.** Bari: Morgan Kaufmann, 1996. p.319-327.

LU, D. et al. Change detection techniques. **International Journal of Remote Sensing**, Abingdon, England, v. 25, n. 12, p. 2365-2407, 2004.

LU, M. et al. Land cover change detection by integrating object-based data blending model of Landsat and MODIS. **Remote Sensing of Environment**, New York, v. 184, p. 374-386, 2016.

MACHADO, R. B. et al. **Estimativas de perda da área do Cerrado brasileiro**. Brasília, 2004. 23p.

MAKKEASORN, A.; CHANG, N. BIN; LI, J. Seasonal change detection of riparian zones with remote sensing images and genetic programming in a semi-arid watershed. **Journal of Environmental Management**, London, v. 90, n. 2, p. 1069-1080, 2009.

MAS, J. F.; FLORES, J. J. The application of artificial neural networks to the analysis of remotely sensed data. **International Journal of Remote Sensing**, Abingdon, England, v. 29, n. 3, p. 617-663, 2008.

MENINO, G. C. O. et al. Environmental heterogeneity and natural regeneration in riparian vegetation of the brazilian semi-arid region. **Edinburgh Journal of Botany**, v. 69, n. 1, p. 29-51, 2012.

MILLER, O.; PIKAZ, A.; AVERBUCH, A. Objects based change detection in a

pair of gray-level images. **Pattern Recognition**, Kidlington, v. 38, n. 11, p. 1976-1992, 2005.

MISHRA, N. B.; CREWS, K. A. Mapping vegetation morphology types in a dry savanna ecosystem: Integrating hierarchical object-based image analysis with Random Forest. **International Journal of Remote Sensing**, Abingdon, England, v. 35, n. 3, p. 1175-1198, 2014.

MOUNTRAKIS, G.; IM, J.; OGOLE, C. Support vector machines in remote sensing: A review. **ISPRS Journal of Photogrammetry and Remote Sensing**, Amsterdam, v. 66, n. 3, p. 247-259, 2011.

MUTANGA, O.; ADAM, E.; CHO, M. A. High density biomass estimation for wetland vegetation using worldview-2 imagery and random forest regression algorithm. **International Journal of Applied Earth Observation and Geoinformation**, Amsterdam, v. 18, n. 1, p. 399-406, 2012.

MYERS, N. et al. Biodiversity hotspots for conservation priorities. **Nature**, London, v. 403, n. 6772, p. 853-858, 2000.

OTUKEI, J. R.; BLASCHKE, T. Land cover change assessment using decision trees, support vector machines and maximum likelihood classification algorithms. **International Journal of Applied Earth Observation and Geoinformation**, Amsterdam, v. 12, n. SUPPL. 1, p. S27-S31, 2010.

PAL, M.; FOODY, G. M. Feature selection for classification of hyperspectral data by SVM. **IEEE Transactions on Geoscience and Remote Sensing**, Piscataway, v. 48, n. 5, p. 2297-2307, 2010.

PEEL, M. C.; FINLAYSON, B. L.; MCMAHON, T. A. Updated world map of the Köppen-Geiger climate classification. **Hydrology and Earth System Sciences**, Göttingen, v. 11, p. 1633-1644, 2007.

PELLETIER, C. et al. Assessing the robustness of Random Forests to map land cover with high resolution satellite image time series over large areas. **Remote Sensing of Environment**, New York, v. 187, p. 156-168, 2016.

PEREIRA, A. A. et al. Burned area mapping in the Brazilian Savanna using a one-class support vector machine trained by active fires. **Remote Sensing**, Basel, v. 9, n. 11, p. 1161, 2017.

PEREIRA, A. A. et al. Avaliação de índices espectrais para identificação de

áreas queimadas no cerrado utilizando dados Landsat TM. **Revista Brasileira de Cartografia**, v. 68, n. 8, p. 1665-1680, 2016.

PETITJEAN, F. et al. Analysing satellite image time series by means of pattern mining. In: FYFE, C. et al. **Intelligent Data Engineering and Automated Learning - IDEAL 2010**. Paisley: Springer, 2010. v. 6283, p. 45-52.

PONZONI, F. J.; SHIMABUKURO, Y. E.; KUPLICH, T. M. **Sensoriamento Remoto da Vegetação**. 2nd ed. Cubatão: Oficina dos Textos, 2012.

POWELL, S. L. et al. Quantification of live aboveground forest biomass dynamics with Landsat time-series and field inventory data: A comparison of empirical modeling approaches. **Remote Sensing of Environment**, New York, v. 114, n. 5, p. 1053-1068, 2010.

RATTER, J. A.; RIBEIRO, J. F.; BRIDGEWATER, S. The Brazilian cerrado vegetation and threats to its biodiversity. **Annals of Botany**, Oxford, v. 80, n. 3, p. 223-230, 1997.

RIBEIRO, J. F.; WALTER, B. M. T. As principais fitofisionomias do bioma Cerrado. In: SANO, S. M.; ALMEIDA, S. P. de.; RIBEIRO, J. F. **Cerrado: Ecologia e flora**. 1st ed. Planaltina: Embrapa Cerrados; Embrapa Informação Tecnológica, 2008. p.152-212.

ROGAN, J. et al. Mapping land-cover modifications over large areas: A comparison of machine learning algorithms. **Remote Sensing of Environment**, New York, v. 112, n. 5, p. 2272-2283, 2008. Elsevier.

ROUSE, J. W.; HASS, R. H.; SCHELL, J. A.; DEERING, D. W. Monitoring vegetation systems in the great plains with ERTS. In: FREDEN, S. C.; MERCANTI, E. P.; FRIEDMAN, D. B. **Third Earth Resources Technology Satellite Symposium (ERTS)**. Greenbelt: Discipline Summary Reports, 1973. p. 309-317.

ROY, D. P. et al. Landsat-8: Science and product vision for terrestrial global change research. **Remote Sensing of Environment**, New York, v. 145, p. 154-172, 2014.

SANO, E. E. et al. Land cover mapping of the tropical savanna region in Brazil. **Environmental Monitoring and Assessment**, Dordrecht, v. 166, n. 1-4, p. 113-124, 2010.

SCHOWENGERDT, R. A. **Remote sensing: models and methods for image processing**. 3rd ed. Cambridge: Academic Press, 2007.

SCHULTZ, M. et al. Performance of vegetation indices from Landsat time series in deforestation monitoring. **International Journal of Applied Earth Observation and Geoinformation**, Amsterdam, v. 52, p. 318-327, 2016.

SCHWIEDER, M. et al. Mapping Brazilian savanna vegetation gradients with Landsat time series. **International Journal of Applied Earth Observation and Geoinformation**, Amsterdam, v. 52, p. 361-370, 2016.

SHALABY, A.; TATEISHI, R. Remote sensing and GIS for mapping and monitoring land cover and land-use changes in the Northwestern coastal zone of Egypt. **Applied Geography**, Kidlington, v. 27, n. 1, p. 28-41, 2007.

SHATAEE, S. et al. Forest attribute imputation using machine-learning methods and ASTER data: Comparison of k-NN, SVR and random forest regression algorithms. **International Journal of Remote Sensing**, Abingdon, England, v. 33, n. 19, p. 6254-6280, 2012.

SILVA, M. P. S. et al. Remote-sensing image mining: Detecting agents of land-use change in tropical forest areas. **International Journal of Remote Sensing**, Abingdon, England, v. 29, n. 16, p. 4803-4822, 2008.

SILVEIRA, E. M. O. da. et al. Assessment of geostatistical features for object-based image classification of contrasted landscape vegetation cover. **Journal of Applied Remote Sensing**, Bellingham, v. 11, n. 3, p. 1-15, 2017.

TEWKESBURY, A. P. A critical synthesis of remotely sensed optical image change detection techniques. **Remote Sensing of Environment**, New York, v. 160, p. 1-14, 2015.

TRANCOSO, R.; SANO, E. E.; MENESES, P. R. The spectral changes of deforestation in the Brazilian tropical savanna. **Environmental monitoring and assessment**, Dordrecht, v. 187, n. 1, p. 4145, 2015.

TRIGG, S.; FLASSE, S. An evaluation of different bi-spectral spaces for discriminating burned shrub-savannah. **International Journal of Remote Sensing**, Abingdon, England, v. 22, n. 13, p. 2641-2647, 2001.

VERMOTE, E. F. et al. Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. **Remote Sensing of Environment**, New

York, v. 185, p. 46-56, 2016.

VIEIRA, M. A. et al. Object Based Image Analysis and Data Mining applied to a remotely sensed Landsat time-series to map sugarcane over large areas. **Remote Sensing of Environment**, New York, v. 123, p. 553-562, 2012.

WITTEN, I. H.; FRANK, E.; HALL, M. A.; PAL, C. J. **Data mining : practical machine learning tools and techniques**. 4th ed. Burlington: Morgan Kaufmann, 2016.

WULDER, M. A. et al. Landsat continuity: Issues and opportunities for land cover monitoring. **Remote Sensing of Environment**, New York, v. 112, n. 3, p. 955-969, 2008.

XIAN, G.; HOMER, C. Updating the 2001 National Land Cover Database Impervious Surface Products to 2006 using Landsat Imagery Change Detection Methods. **Remote Sensing of Environment**, New York, v. 114, n. 8, p. 1676-1686, 2010.

XUE, J.; SU, B. Significant remote sensing vegetation indices: A review of developments and applications. **Journal of Sensors**, London, v. 2017, p. 1-17, 2017.

ZHU, Z. Change detection using landsat time series: A review of frequencies, preprocessing, algorithms, and applications. **ISPRS Journal of Photogrammetry and Remote Sensing**, Amsterdam, v. 130, p. 370-384, 2017.

ZHU, Z.; WOODCOCK, C. E. Object-based cloud and cloud shadow detection in Landsat imagery. **Remote Sensing of Environment**, New York, v. 118, p. 83-94, 2012.