



POLYANNE AGUIAR DOS SANTOS

**AVALIAÇÃO DO USO DE VARIÁVEIS
TEMPORAIS NA CLASSIFICAÇÃO DA
COBERTURA DA TERRA**

LAVRAS – MG

2014

POLYANNE AGUIAR DOS SANTOS

**AVALIAÇÃO DO USO DE VARIÁVEIS TEMPORAIS NA
CLASSIFICAÇÃO DA COBERTURA DA TERRA**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Ecologia Aplicada, área de concentração em Ecologia e Conservação de Paisagens Fragmentadas e Agrossistemas, para a obtenção do título de Mestre.

Orientador

Dr. Luís Marcelo Tavares de Carvalho

LAVRAS - MG

2014

**Ficha Catalográfica Elaborada pela Coordenadoria de Produtos e
Serviços da Biblioteca Universitária da UFLA**

Santos, Polyanne Aguiar dos.

Avaliação do uso de variáveis temporais na classificação da
cobertura da terra / Polyanne Aguiar dos Santos. – Lavras : UFLA,
2014.

80 p. : il.

Dissertação (mestrado) – Universidade Federal de Lavras, 2014.

Orientador: Luis Marcelo Tavares de Carvalho.

Bibliografia.

1. Algoritmos de aprendizagem de máquina. 2. Greenbrown. 3.
Série temporal sintética. I. Universidade Federal de Lavras. II.
Título.

CDD – 621.3678

POLYANNE AGUIAR DOS SANTOS

**AVALIAÇÃO DO USO DE VARIÁVEIS TEMPORAIS NA
CLASSIFICAÇÃO DA COBERTURA DA TERRA**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Ecologia Aplicada, área de concentração em Ecologia e Conservação de Paisagens Fragmentadas e Agrossistemas, para a obtenção do título de Mestre.

APROVADA em 18 de julho de 2014.

Dr. Luciano Teixeira de Oliveira UFLA

Dr. Marcelo de Carvalho Alves UFLA

Dr. Luís Marcelo Tavares de Carvalho
Orientador

LAVRAS - MG

2014

AGRADECIMENTOS

Agradeço a Deus por ter me concedido essa oportunidade de aprendizagem única, e por ter me guiado nos momentos difíceis, além disso, também agradeço por ter colocado pessoas especiais na minha vida.

Agradeço a minha família, que mesmo estando distante (quase 1.000 km), sempre estiveram presentes e me apoiaram incondicionalmente. Papi e Mami, eu amo vocês demais, obrigada por tudo!

As minhas irmãs, Roberta e Fernanda, que também me apoiaram e estiveram presente nessa caminhada. Ao meu cunhado Vinícius, que inúmeras vezes me acolheu e me levou e buscou na rodoviária de Belo Horizonte.

Agradeço a todos os meus tios e tias, primos e primas, minhas avós “Anas”, que diretamente ou indiretamente também participaram dessa conquista, principalmente em oração.

Agradeço ao meu orientador, Passarinho, pela oportunidade e todo aprendizado, e também pelos puxões de orelha...rs, pois com isso cresci e sou grata.

Agradeço ao Programa de Pós-graduação em Ecologia Aplicada que tanto me ensinou e que me proporcionou amizades verdadeiras.

A minha turma de mestrado, a única que fala esperanto e tem muito borogodó!

Agradeço especialmente a algumas pessoas que fizeram muita diferença nessa conquista, foram companheiras e amigas: Gui, Ernesto, Cris, Wallace, Lilian, Sassanha, Toru (esse “cabeçudo” que eu detesto!...rs), Lisi (minha mãe gaúcha), Tainá e Ananza (irmãs manaquenses), Tatau (sua amizade, apesar de recente, tem muito valor!) e Rafa.

Agradeço aos amigos que também conquistei no LEMAF, que na verdade me acolheram e me aceitaram mesmo sendo da Ecologia...rs. Pessoas

que estiveram presentes nos bons e maus momentos: Terrinha (meu companheiro, amigo,...), Tonica (Thais), Marcela, Eltinho, Elisa e Andressa (companheiras de bafões), Mari, Eva, Joyce, Thaisa, Ferraço, Ravi, Rafa, Tisa e Query (filhas e estagiárias maravilhosas) Nathália, Gustavo e Tonico Carlinhos (seguraram as pontas nos momentos em que a dissertação apertava).

Agradeço em especial ao meu laboratório, que inicialmente era formado por Célio (palmitinho que amo muito), Carol (minha mãe baiana), Lisi e Tássia (irmã branca). Essa família que sempre carregarei comigo, agradeço todo o apoio, todas as dicas, todas as conversas acadêmicas ou não, agradeço por aceitarem essa pretinha capixaba (eu) na vida de vocês de braços abertos. Amo vocês!

E por fim, agradeço a algumas pessoas que foram fundamentais para o desenvolvimento desse trabalho: Liliano (O Severino do LEMAF) que nos salvou instalando *softwares*, LINUX, formatando...

Ao Ronaldo (que nos ajudou a programar no LINUX), Feng Gao (por ter disponibilizado o algoritmo STARFM) e alguns que não conheci pessoalmente, mas que me ajudaram a desvendar os mistérios das análises de séries temporais: Jan Verbesselt, Matthias Forkel e Feng Tian.

Agradeço à Universidade Federal de Lavras, pela oportunidade concedida para a realização do mestrado.

Agradeço a todos que fizeram parte dessa caminhada, e que oraram e torceram pela minha conquista!

RESUMO

O monitoramento da superfície terrestre tem sido realizado por sensores remotos a bordo de satélites, o que possibilita obter informações sobre as alterações da paisagem ao longo do tempo. O uso de dados multitemporais auxilia a classificação da cobertura da terra e potencializa a sua caracterização levando em consideração aspectos fenológicos, bem como variações contínuas ou bruscas na superfície terrestre. Contudo, não se sabe quais as implicações de diferentes frequências temporais de imagens de sensoriamento remoto ao longo do tempo numa classificação. Tendo em vista a dificuldade de se adquirir imagens de qualidade para compor uma série temporal, é de grande importância avaliar qual frequência de imagens é necessária para obter uma classificação multitemporal precisa. Portanto, este trabalho tem por objetivo extrair informações de séries temporais de dados de sensoriamento remoto para melhorar a precisão da classificação da cobertura do solo, e assim contribuir com o entendimento de padrões temporais de diversas classes da cobertura do solo e indicar abordagens apropriadas para o mapeamento que será realizado durante a implementação do CAR. A área de estudo está localizada no Centro-Oeste do estado de Minas Gerais, e as análises foram limitadas a uma área de 625 km² referente a uma cena Rapideye (tile 2330313). Foram adquiridas todas as imagens disponíveis do sensor Landsat-5 TM (órbita/ponto 219/073) entre 2000 e 2011. Porém, como regiões tropicais apresentam grande incidência de nuvens e eventual indisponibilidade de imagens TM, foram utilizadas imagens sintéticas oriundas de fusão de imagens TM e MODIS. Para isso, foi adquirido um conjunto de 273 imagens do sensor MODIS (produto MOD13Q1). A fusão de imagens TM e MODIS foi realizada pelo algoritmo STARFM. As imagens TM originais e sintéticas compuseram séries temporais Estacional (quatro imagens/ano), Bimestral (seis imagens/ano), Mensal (12 imagens/ano) e Completa (23 imagens/ano). As séries temporais foram analisadas pelo pacote *greenbrown* e os resultados gerados foram incorporados como atributos nos algoritmos de aprendizagem de máquina. Os resultados mostraram que o uso de parâmetros gerados a partir do pacote *greenbrown*, mesmo quando usando imagens TM sintéticas, proporcionaram uma melhoria no desempenho dos classificadores, com destaque para os algoritmos SVM e RF. Além disso, não foram encontradas diferenças significativas entre as classificações geradas utilizando atributos de diferentes frequências temporais. Sendo assim, a classificação multitemporal da imagem Rapideye utilizando parâmetros extraídos de uma série temporal TM, se mostrou aplicável e promissora, visto que dados Rapideye não dispõem de imagens multidas suficientes.

Palavras-chave: Algoritmos de aprendizagem de máquina. *greenbrown*. Série temporal sintética.

ABSTRACT

The monitoring of land surface has been performed by remote sensors aboard satellites, which allows us to obtain information on changes of landscape over time. The use of multitemporal data aids in the classification of land cover and enhances its characterization considering phenological aspects, as well as continuous or sudden variations in the landscape. However, the implications of different temporal frequency of remote sensing images in a classification over time are unknown. Considering the difficulty of acquiring quality images for comprising a time series, it is highly important to evaluate which frequency is necessary to obtain a precise multitemporal classification. Therefore, this work had the objective of extracting information of time series remote sensing data to improve the classification precision of land cover, thus contributing with the understanding of temporal patterns of various land cover classes and indicating appropriate approaches for the mapping that will be performed during the implementation of the CAR. The study area is located in the Center-West of the state of Minas Gerais, Brazil, and the analyses were confined to an area of 625 km² referent to a Rapideye scene (tile 2330320). All images of the Landsat-5 TM sensor (orbit/point 219/073) available were acquired between 2000 and 2011. However, as tropical regions present large incidence of clouds and eventual unavailability of TM images, synthetic images derived from the fusion of TM and MODIS images were used. In order to do this, a set of 273 images of the MODIS sensor (product MOD13Q1) were acquired. The fusion of TM and MODIS was performed by the STARFM algorithm. The original and synthetic TM images comprised Stational (4 images/year), Bimonthly (6 images/per year), Monthly (12 images/year) and Complete (23 images/year) time series. The time series were analyzed by the *greenbrown* package and the generated results were incorporated as attributes in the machine learning algorithm. The results showed that the use of parameters generated from the *greenbrown* package, even when using synthetic TM images, provided an improvement in the performance of the classifiers, highlighting the SVM and RF algorithms. In addition, no significant differences were found between the classifications generated using different temporal frequency attributes. Thus, the multitemporal classification of the Rapideye image using parameters extracted from a TM time series, was shown to be applicable and promising, since the Rapideye data did not dispose of sufficient multidata images.

Keywords: Machine learning algorithms. *Greenbrown*. Synthetic time series.

LISTA DE FIGURAS

Figura 1	Representação gráfica do modelo de classificação <i>Naive Bayes</i> , em que é estimada a probabilidade de D_1, \dots, D_n pertencer à classe C_j	25
Figura 2	Representação gráfica do algoritmo Decision Tree.....	26
Figura 3	Representação gráfica do modelo de classificação <i>Random Forest</i>	28
Figura 4	Representação gráfica dos vetores de suportes utilizados para separar as classes num hiperplano com separação ótima	30
Figura 5	Componentes da série temporal estimados. O painel superior mostra uma série temporal NDVI com o seu valor médio (linha vermelha).....	37
Figura 6	Localização da área de estudo na região Sudeste do Brasil, com destaque para a cena Rapideye em composição RGB (R=banda 5; G=banda 4; B= banda 3)	39
Figura 7	Fluxograma dos procedimentos utilizados neste estudo	51
Figura 8	Comparação entre a reflectância da imagem TM real e predita nas abordagens por pixel e baseada em objeto.....	53
Figura 9	Classificação utilizando apenas variáveis espectrais da Rapideye e MLA's. (a) DT; (b) NB; (c) RF (d) SVM.....	57
Figura 10	Comparação das classificações com atributos temporais que apresentaram os mesmos valores de acurácia global e Kappa, mas resultaram em classificações diferentes; (a) DT –Completa, (b) DT-Mensal, (c) SVM-Mensal, (d) SVM-Bimestral.....	59

Figura 11	Comparação das melhores classificações: (a) DT-Bimestral, (b) NB-Mensal, (c) RF-Estacional e (d) SVM-Mensal; com destaque em algumas regiões que os algoritmos classificaram de forma diferente a cobertura da terra	63
Figura 12	Erros de comissão e omissão provenientes das matrizes de confusão para cada classe da cobertura da terra.....	65

LISTA DE TABELAS

Tabela 1	Pares de imagens MODIS e Landsat (sem nuvens) correspondentes utilizadas na predição de imagens Landsat sintéticas.....	44
Tabela 2	Atributos utilizados como <i>input</i> no processo de classificação dos algoritmos de aprendizagem de máquina	47
Tabela 3	Parâmetros específicos para cada algoritmo de aprendizagem de máquina utilizado na classificação	49
Tabela 4	Definição de elementos de matrizes utilizadas na equação 3	52
Tabela 5	Acurácia das classificações para cada MLA usando diferentes variáveis	55
Tabela 6	Comparações das classificações realizadas pelo teste McNemar com chi-quadrado tabulado $X^2=3,81$	61

LISTA DE SIGLAS

6S	Satellite Signal in the Solar Spectrum
ALOS	Advanced Land Observing Satellite
APP	Área de Preservação Permanente
AWiFS	Advanced Wide Field Sensor
BFAST	Breaks For Additive Seasonal and Trend
CAR	Cadastro Ambiental Rural
CART	Classification and Regression of Trees
CCDC	Continuous Change Detection and Classification
CDR	Surface Reflectance Climate Data Record
DT	Decision Tree
ETM+	Enhanced Thematic Mapper Plus
GEOBIA	Geographic Object Based Image Analysis
IAV	Inter-Annual variability
LEDAPS	Landsat Ecosystem Disturbance Adaptive Processing System
MAXVER	Máxima Verossimilhança
MLA	Machine Learning Algorithm
MODIS	Moderate Resolution Imaging Spectro-radiometer
MRT	MODIS Re-projection Tools
NB	Naïve Bayes
NDVI	Normalized Difference Vegetation Index
NIR	Banda do infravermelho próximo
OBIA	Object Based Image Analysis
OOB	Out-of-bag
RBF	Radial Basis Function
RED	Banda do vermelho
RF	Random Forest

RMSE	Root Mean Square Error
RNA	Redes Neurais Artificiais
SPOT	Satellite Pour l'Observation de la Terre
STARFM	Spatial and Temporal Adaptive Reflectance Fusion Model
STARS	Spectral–Temporal Analysis by Response Surface
STB	Série Temporal Bimestral
STC	Série Temporal Completa
STE	Série Temporal Estacional
STM	Série Temporal Mensal
STRS	Spectral Temporal Response Surface
STV	Short Term Variability
SVM	Support Vector Machine
TM	Thematic Mapper
USGS	United States Geological Survey
UTM	Universal Transversa de Mercator
VE	Variáveis Espectrais da Rapideye

SUMÁRIO

1	INTRODUÇÃO	14
2	REFERENCIAL TEÓRICO	18
2.1	Dados de sensoriamento remoto	18
2.1.1	Fusão de imagens	20
2.2	Classificação da cobertura terrestre	21
2.2.1	Algoritmos de classificação	23
2.2.2	Classificação multitemporal	31
3	BASE DE DADOS	38
3.1	Área de Estudo	38
3.2	Aquisição dos dados e pré-processamento	39
4	MÉTODOS	42
4.1	Predição de imagens sintéticas	42
4.1.1	Avaliação de imagens TM sintéticas	44
4.2	Análise das Séries Temporais	45
4.3	Classificação da cobertura da terra	46
4.3.1	Avaliação e acurácia das classificações	52
5	RESULTADOS E DISCUSSÃO	53
5.1	Acurácia das imagens sintéticas	53
5.2	Performances dos Algoritmos de Aprendizagem de Máquina	54
5.3	Analisando os atributos temporais na classificação	65
6	CONCLUSÕES	69
	REFERÊNCIAS	71

1 INTRODUÇÃO

A cobertura da superfície terrestre sofre naturalmente diversas alterações devido a aspectos peculiares de suas dinâmicas fenológicas, como também pela ação antrópica em uma variedade de escalas espaciais e temporais. Diferentes tipos de cobertura da superfície terrestre apresentam padrões temporais diferenciados, o que dificulta a caracterização de mudanças ao longo do tempo em paisagens muito heterogêneas.

O monitoramento da cobertura da terra ao longo do tempo fornece informações sobre as mudanças que ocorrem sazonalmente, gradualmente e abruptamente na superfície terrestre (JACQUIN; SHEEREN; LACOMBE, 2010; MARTÍNEZ; GILABERT, 2009; VERBESSELT et al., 2010a). As mudanças sazonais estão relacionadas aos padrões cíclicos (intra- e interanual) que influenciam a fenologia da vegetação. Contudo, as mudanças fenológicas da vegetação podem sofrer impactos, tais como desmatamento, queimadas ou inundações, que são caracterizadas como mudança abrupta. Mudanças graduais não são relacionadas com os ciclos fenológicos normais, trata-se de mudanças sutis que ocorrem na vegetação, incluindo secas, degradação, sucessão natural, pressão de pastejo e alterações do bioma induzidas pelo clima. O monitoramento dessas mudanças é um requisito fundamental para sistemas de fiscalização ambiental, mas, ao mesmo tempo, uma tarefa difícil devido à dificuldade de se executar programas de monitoramento em larga escala.

No Brasil, está sendo implementado um novo código florestal, por meio da Lei 12.651 (BRASIL, 2012), que regulamenta o uso e manejo da terra em propriedades rurais e dispõe sobre a proteção da vegetação nativa. Tal lei designa áreas ambientalmente sensíveis como Áreas de Preservação Permanente (APPs) e Reservas Legais, além disso, tem por objetivo monitorar todo o

território brasileiro e com isso gerar informações ambientais que subsidiarão a fiscalização de remanescentes de vegetação nativa.

O monitoramento da superfície terrestre tem sido realizado por sensores remotos a bordo de satélites, porém há diferentes sensores que são empregados e adequados para cada tipo de estudo. Uma das vantagens de se utilizar imagens de sensoriamento remoto para o monitoramento de grandes áreas da cobertura da terra é a elevada taxa de repetição dos satélites (STELLMES et al., 2010), que possibilita obter informações sobre as alterações da paisagem ao longo do tempo. Além disso, o uso desses sensores tem baixo custo em relação ao monitoramento realizado em campo.

A análise multitemporal de imagens de sensoriamento remoto fornece informações importantes para o monitoramento ambiental, permitindo a análise de padrões complexos e a caracterização da dinâmica de cobertura da superfície terrestre (BRUZZONE; SMITS; TILTON, 2003). Desse modo, o uso de dados multitemporais auxilia a classificação da cobertura da terra e potencializa a sua caracterização, levando em consideração aspectos fenológicos, bem como variações contínuas ou bruscas na superfície terrestre.

A utilização de dados multitemporais integrados às metodologias de classificação da cobertura da terra tem contribuído para classificações da superfície terrestre (BRUZZONE; COSSU; VERNAZZA, 2002; CARRÃO; GONÇALVES; CAETANO, 2008; CONESE; MASELLI, 1991; JIA et al., 2014), melhorando a precisão dos mapeamentos de culturas agrícolas (MELLO et al., 2012; VIEIRA; MATHER; MCCULLAGH, 2000; VIEIRA et al., 2012), e em mapeamentos específicos da vegetação (COLSTOUN et al., 2003; OLIVEIRA; CARVALHO, 2010). Por isso, alguns pesquisadores têm abordado o monitoramento dos ecossistemas através de séries temporais, a partir do qual é possível avaliar o perfil temporal e as tendências de mudanças destes ambientes (COPPIN et al., 2004; VERBESSELT et al., 2010a). As técnicas baseadas em

dados multitemporais e multiespectrais, servem para detectar, identificar, mapear e monitorar mudanças nos ecossistemas, independentemente de seus agentes causadores (MARTÍNEZ; GILABERT, 2009).

O uso de dados multitemporais em classificações da cobertura da terra tem sido empregado adquirindo imagens multidas (CONESE; MASELLI, 1991; MELLO et al., 2012; VIEIRA; MATHER; MCCULLAGH, 2000; VIEIRA et al., 2012) e utilizando séries temporais densas de imagens dos sensores MODIS e/ou Landsat-TM (JIA et al., 2014; WATTS et al., 2011). Contudo, dentre os trabalhos realizados, nenhum avaliou quais as implicações de diferentes frequências temporais de imagens de sensoriamento remoto ao longo do tempo numa classificação. Assim como, não avaliou quais implicações do número de imagens e o intervalo temporal com que os dados multitemporais são amostradas podem influenciar os resultados da classificação.

Tendo em vista a dificuldade de se adquirir imagens de qualidade para compor uma série temporal, é de grande importância avaliar qual frequência de imagens é necessária para obter uma classificação multitemporal precisa. Portanto, este trabalho tem por objetivo extrair informações de séries temporais de dados de sensoriamento remoto para melhorar a precisão da classificação de cobertura do solo. Tendo como objetivos específicos: (i) comparar o desempenho de algoritmos de classificação digital de imagens de sensoriamento remoto utilizando atributos temporais e (ii) verificar o efeito da quantidade de imagens que compõem a série na classificação de cobertura do solo.

Neste trabalho, levantamos as seguintes questões científicas: (1) Imagens sintéticas são eficientes para derivar atributos temporais? (2) Há diferenças significativas entre as classificações utilizando séries de imagens com diferentes frequências temporais? (3) Os atributos temporais melhoram a separabilidade das classes e o desempenho dos algoritmos de classificação?

No âmbito do novo Código Florestal Brasileiro, visando o monitoramento de remanescentes nativas, combate ao desmatamento e planejamento ambiental e econômico, foi criado o CAR (Cadastro Ambiental Rural). O CAR tem a finalidade de integrar as informações ambientais e assim compor uma base de dados cadastrais, que será subsidiada pelo mapeamento de remanescentes de vegetação nativa e de áreas antropogênicas, classes caracterizadas por padrões temporais peculiares e geralmente diferentes. O presente trabalho irá contribuir com o entendimento de padrões temporais de diversas classes de cobertura do solo e indicar abordagens apropriadas para o mapeamento que será realizado durante a implementação do CAR.

2 REFERENCIAL TEÓRICO

2.1 Dados de sensoriamento remoto

Um sensor remoto é um dispositivo que captura dados sobre um objeto ou cena sem o contato direto com este. Uma vez que esses objetos (incluindo vegetação) têm suas características espectrais únicas (regiões de reflectância ou de emissão), eles podem ser identificados a partir de imagens de sensoriamento remoto de acordo com as suas características espectrais (XIE; SHA; YU, 2008b).

A capacidade técnica dos sensores a bordo de satélites depende de quatro resoluções: (a) espacial, que se refere ao tamanho do pixel da imagem gerada, sendo que o pixel corresponde a menor parcela imageada; (b) radiométrica, que se refere à intensidade da radiância da área de cada pixel que o sensor é capaz de medir, representado por valores numéricos ou nível de cinza, a depender da quantidade de bits de cada sensor (por exemplo, uma imagem de dois bits possui apenas quatro níveis de cinza, e uma imagem de oito bits, 256 níveis de cinza); (c) espectral, que se refere ao número de bandas do sensor, sua largura e posição em que estão situadas no espectro eletromagnético e (d) temporal, que corresponde à periodicidade com que o satélite revisita a mesma área (FIGUEIREDO, 2005; MENESES et al., 2012).

Conjuntos de dados com alta resolução espacial e temporal, como as imagens Rapideye, apresentam bom desempenho na caracterização da paisagem. Os dados Rapideye são gerados a partir de um conjunto de cinco satélites idênticos que coletam imagens diariamente em cinco bandas multiespectrais com a mesma resolução espacial - podendo ser 5 ou 6,5 metros, dependendo do produto (“Satellite Imagery Product Specifications,” 2013). Dentre as cinco bandas, ressalta-se a banda *red-edge*, que tem representado a vegetação de

maneira eficaz (TIGGES; LAKES; HOSTERT, 2013) devido à correlação ao teor de clorofila (MALATESTA et al., 2013). Contudo, essas imagens só foram disponibilizadas a partir de 2008. Portanto, para um monitoramento longo da superfície terrestre é desvantajoso. Nesse caso, o uso de sensores com mais observações ao longo do tempo e com resolução inferior é mais indicado.

Imagens provenientes do sensor MODIS (*Moderate Resolution Imaging Spectroradiometer*) têm sido bastante utilizadas em estudos de monitoramento da vegetação em escala local e regional (TIAN et al., 2013). Isso se deve principalmente à alta frequência temporal dos seus produtos, o que permite detectar mudanças em escala temporal, com ciclo de revisita de 1-2 dias (GAO et al., 2006). Porém, tal sensor apresenta uma baixa resolução espacial, pois é limitado a uma resolução mínima de 250 metros (XIN et al., 2013). Os produtos MODIS tem sido úteis em muitos trabalhos de monitoramento de mudanças no ecossistema em ampla escala, como o monitoramento de desmatamento na Amazônia (GALFORD et al., 2008a; MORTON et al., 2005), monitoramento de queimadas (GITAS, 2012) e monitoramento de secas (VERBESSELT; ZEILEIS; HEROLD, 2012). Todavia, imagens com baixa resolução espacial não fornecem informações detalhadas das dinâmicas da paisagem. Uma alternativa seria a utilização de dados da série de satélites Landsat, que também são utilizados para essa finalidade.

Os satélites Landsat TM/ETM+ (*Thematic Mapper/Enhanced Thematic Mapper Plus*) apresentam uma resolução espacial melhor que a do sensor MODIS, de 30 metros. Isso facilita a caracterização da cobertura da terra, bem como as mudanças que ocorrem. Além disso, tal sensor possui medições espectrais em todas as grandes porções do espectro eletromagnético solar, sendo eles: visível, infravermelho próximo, infravermelho médio e termal (COHEN; GOWARD, 2004), o que permite capturar melhor a complexidade da vegetação em escala local (TIAN et al., 2013). Bem como os dados MODIS, dados

TM/ETM+ também têm sido usados na detecção de distúrbios em florestas (BRESEE et al., 2004; HE et al., 2011), no mapeamento e monitoramento de vegetação aquática (LI et al., 2013), mapeamento de degradação induzida por pastagem em *habitats* semiáridos (THOMPSON et al., 2009), e no mapeamento de regeneração e distúrbios (KENNEDY et al., 2012). Porém, em regiões tropicais há uma grande ocorrência de nuvens, limitando o uso de dados do satélite Landsat. A presença de nuvens e as sombras de nuvens influenciam muito na análise dos dados, causando muitos problemas, incluindo equívocos na classificação da cobertura da terra ou até mesmo uma falsa detecção de mudança (ZHU; WOODCOCK, 2014). Desse modo, a resolução espacial das imagens TM/ETM+ é prejudicada. Uma solução para estes problemas são as técnicas de fusão de dados que permitem a integração das resoluções espacial e temporal.

2.1.1 Fusão de imagens

A fusão de imagens de diferentes sensores e resolução tem sido um importante aliado para solucionar problemas relacionados à contaminação por nuvens. Uma alternativa é combinar a resolução espacial da imagem TM/ETM+ com a frequência temporal de imagens do sensor MODIS (BHANDARI; PHINN; GILL, 2012; GAO et al., 2006; HILKER et al., 2009a; HILKER et al., 2009b; TIAN et al., 2013). O STARFM (*Spatial and Temporal Adaptive Reflectance Fusion Model*) é um algoritmo desenvolvido por (GAO et al., 2006) para prever reflectância de superfície diária da resolução espacial TM/ETM+ utilizando um ou mais pares de imagens TM/ETM+ e MODIS. Desse modo, é possível obter imagens TM/ETM+ com informações confiáveis das dinâmicas da vegetação, bem como possíveis alterações ao longo do tempo.

Uma janela móvel é utilizada a fim de evitar problemas causados tanto pelas diferenças entre as imagens TM/ETM+ e MODIS, quanto por mudanças

temporais (GAO et al., 2006). Sendo assim, é gerada uma função com diferentes pesos para cada pixel vizinho ao pixel central dentro da janela, sendo usados somente aqueles espectralmente similares e livres de nuvens.

A obtenção de imagens de qualidade é imprescindível na classificação da cobertura da terra, independente da resolução espacial e temporal, ou mesmo do tipo de sensor. Mas, além disso, são necessárias metodologias que sejam eficazes na separabilidade dos usos da terra, garantindo assim mapeamentos com melhores acurácias, e que, conseqüentemente, servem de subsídio para monitoramento da superfície terrestre.

2.2 Classificação da cobertura terrestre

As técnicas de classificação utilizadas na caracterização da superfície terrestre se desenvolveram consideravelmente, no sentido de melhorar a qualidade das classificações e o tempo de processamento. Além disso, a disponibilidade de imagens de alta resolução espacial tem aumentado nos últimos anos e os métodos de classificação baseados em pixel não são adequados para este tipo de imagem. As classificações adotavam uma abordagem baseada em pixel, ou seja, o aspecto espacial do objeto no mundo real e suas relações espaciais não eram modelados (HUSSAIN et al., 2013a). Por esta razão, as classificações dessa abordagem podem ter uma aparência “*salt-and-pepper*”, o que significa que há regiões com pixels isolados na classificação (SU et al., 2011). Diante disso, Malila (1980) utilizou a Análise de Vetor de Mudança em médias espectrais de pixels espectralmente homogêneos e contíguos visando uma melhor acurácia na caracterização da cobertura terrestre. Recentemente, houve o desenvolvimento da análise baseada em objeto (OBIA – *Object Based Image Analysis*), em que os pixels são agrupados em objetos ou segmentos com base em alguns critérios de homogeneidade (espectrais ou espaciais)

(WHITESIDE; BOGGS; MAIER, 2011). Posteriormente, Hay e Castilla (2008) argumentaram que o termo OBIA deveria ser chamado GEOBIA (*Geographic Object Based Image Analysis*), pois assim representaria uma subdisciplina das Ciências do Sistema de Informações Geográficas, e desse modo, dados de imagem representam porções da superfície da Terra. Uma das vantagens da segmentação é a eliminação do efeito “*salt-and-pepper*” associada à classificação baseada em pixel.

Alguns trabalhos têm comparado o desempenho de diferentes algoritmos de classificação usando a análise de imagem baseada em pixel e/ou baseada em objeto. Duro, Franklin, e Dubé (2012) utilizaram as duas abordagens na classificação de áreas agrícolas usando diferentes algoritmos e não encontraram diferença significativa com o mesmo algoritmo classificador. Todavia, Whiteside, Boggs e Maier (2011) encontraram diferenças significativas na precisão entre um classificador baseado em pixel e outro classificador com abordagem baseado em objeto, no mapeamento de savana tropical, mostrando que a escolha de abordagem (baseada em pixel e objeto) depende da região e objetivo do estudo.

A classificação da cobertura terrestre é uma tarefa complexa, especialmente em áreas vegetadas que apresentam características fenológicas específicas. O mapeamento da vegetação apresenta informações valiosas para a compreensão dos ambientes naturais e artificiais através da quantificação da cobertura vegetal do local para escalas globais em um determinado ponto do tempo ou durante um período contínuo (XIE; SHA; YU, 2008a). Por isso, são necessários algoritmos de classificação capazes de produzir mapeamentos com precisão e detalhamento, e assim, serem subsídios para programas de monitoramento como o CAR.

2.2.1 Algoritmos de classificação

Diversos classificadores têm sido usados a fim de melhorar a classificação da cobertura da terra, alguns supervisionados e outros não supervisionados. A abordagem não supervisionada é frequentemente usada em mapeamento temático sem nenhum conhecimento a priori sobre a área estudada, diferentemente da abordagem supervisionada que atribui novas unidades de amostragem para a classe a priori (XIE; SHA; YU, 2008a).

Dentre os classificadores não supervisionados, destaca-se o algoritmo Máxima Verossimilhança (MAXVER), que realiza a classificação a partir de um conhecimento a priori do número de classes de uso e cobertura do solo contido numa imagem (RUDORFF et al., 2004). Jia et al. (2014) utilizaram o MAXVER na classificação da cobertura do solo de imagem TM integrando atributos temporais, e apresentou desempenho satisfatório. Vieira, Mather e Mccullagh (2000) também obtiveram classificações satisfatórias utilizando o MAXVER na classificação de culturas agrícolas, porém, ressaltam que quando comparado com o algoritmo de Redes Neurais Artificiais (RNA), o mesmo é desvantajoso em relação ao tempo de processamento.

Recentemente, algoritmos de aprendizado de máquina (MLA - *Machine Learning Algorithm*) supervisionados surgiram como alternativas mais precisas e eficientes para algoritmos paramétricos convencionais (ROGAN et al., 2008). Este tipo de aprendizagem visa obter uma função (ou, em geral, um modelo), que pode identificar a classe a que pertencem os novos exemplos com base em um conjunto de dados de treinamento pré-definidos. Porém, é importante ressaltar que o desempenho dos algoritmos de classificação depende do local de estudo e do tipo de sensor utilizado.

Redes Neurais Artificiais (RNA) é um algoritmo de aprendizagem de máquina que consiste de uma rede de funções primitivas capazes de receber

múltiplas entradas que são avaliadas em termos do seu sucesso em discriminar determinada classe (CRACKNELL; READING, 2014). Diversos trabalhos têm utilizado o RNA em mapeamentos (FOODY, 2004; OLIVEIRA; CARVALHO, 2010; PEÑA et al., 2014; SHAO; LUNETTA, 2012; SU et al., 2011). Rogan et al. (2008) destacam o desempenho deste algoritmo no mapeamento de mudanças na cobertura terrestre em grandes áreas. No entanto, o RNA pode ser exigente em termos computacionais quando grandes conjuntos de dados são inseridos para formar a rede artificial (XIE; SHA; YU, 2008a).

Dentre outros MLA's, também destacamos o *Naive Bayes* (NB). O NB é um modelo de classificação probabilístico simples baseado no teorema de Bayes que pressupõe independência entre as classes (CRACKNELL; READING, 2014; WIELAND; PITTORE, 2014; WITTEN; FRANK; HALL, 2005), e a distribuição normal dos atributos para cada classe (WIELAND; PITTORE, 2014). Este algoritmo produz estimativas de probabilidade, que são muitas vezes mais úteis do que previsões simples (WITTEN; FRANK; HALL, 2005). Devido à natureza precisa do modelo de probabilidade, o classificador *Naive Bayes* pode treinar de forma muito eficiente em um ambiente de aprendizado supervisionado. O NB assume que a presença (ou ausência) de uma característica particular de uma classe não tem relação com a presença (ou ausência) de qualquer outra característica, dada a variável da classe (LI et al., 2013). O NB tem dois tipos de variáveis: a classe C_j e um conjunto de atributos preditivos $D = \{D_1, \dots, D_n\}$ (IBÁÑEZ; BIELZA; LARRAÑAGA, 2014). O classificador assume que os atributos têm distribuição independente, e, portanto estima a probabilidade de D estar na classe C_j (BISHOP, 2006). O *Naive Bayes* usa o teorema de Bayes como base, portanto temos:

$$p(C_j|D) = \frac{(D|C_j)p(C_j)}{p(D)} \quad (1)$$

em que $p(C_j|D)$ é a probabilidade da ocorrência de D estar na classe C_j (Figura 1).

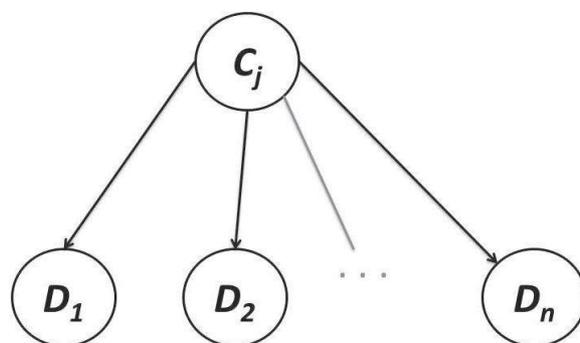


Figura 1 Representação gráfica do modelo de classificação *Naive Bayes*, em que é estimada a probabilidade de D_1, \dots, D_n pertencer à classe C_j

Em classificações resultantes dos trabalhos de Cracknell e Reading (2014) e Shiraishi et al. (2014a), utilizando imagem TM e ALOS, o NB não apresentou os melhores resultados de acurácia. Todavia, Li et al. (2013) gerou uma classificação dos diferentes tipos de zonas úmidas utilizando imagem TM e obteve uma acurácia satisfatória.

Outro algoritmo de MLA é o *Decision Tree* (DT). O DT é um algoritmo que apresenta vantagens consideráveis para os problemas de classificação de imagens de sensoriamento remoto por causa da sua flexibilidade, simplicidade intuitiva e computacional (FRIEDL; BRODLEY, 1997). A árvore de decisão é um processo de classificação divisível que repetitivamente particiona um

conjunto de dados em subdivisões baseadas em testes definidos em cada ramo (ou nó) na árvore (MUCHONEY; WILLIAMSON, 2001). Cada nó representa um teste em uma série de valores de atributos, e cada ramo representa um resultado do teste, enquanto que as folhas da árvore representam as classes (HUSSAIN et al., 2013b) (Figura 2). O algoritmo trabalha de cima para baixo, procurando em cada etapa um atributo que melhor separa as classes, e, em seguida, recursivamente processa os subproblemas que resultam da divisão (WITTEN; FRANK; HALL, 2005). Após dividir recursivamente os dados de treino em partes menores, o DT continua até que as folhas da árvore contenham apenas os casos de uma classe ou até que a separação não proporcione qualquer melhoria na relação de ganho, desta forma ajusta os dados (COLSTOUN et al., 2003).

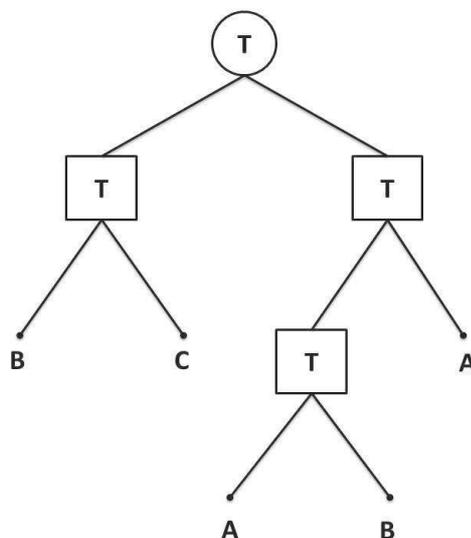


Figura 2 Representação gráfica do algoritmo Decision Tree

Nota: Testes (T) são aplicados para dividir os dados em grupos menores. As folhas A, B e C referem-se à classe atribuída a cada observação

Alguns trabalhos têm aplicado o DT em mapeamentos de grandes áreas, como o de Vieira et al. (2012), que o aplicaram no mapeamento de cana de açúcar utilizando dados multitemporais TM, ou mesmo em áreas menores (DURO; FRANKLIN; DUBÉ, 2012) utilizando o SPOT-5 no mapeamento de paisagens agrícolas. Punia, Joshi e Porwal (2011) encontraram bons resultados com o uso do DT na classificação da cobertura da terra utilizando imagens multitemporais AWiFS. Uma desvantagem desse algoritmo é que a árvore pode ter muitas ramificações, o que torna a interpretação difícil (HUSSAIN et al., 2013b).

O *Random Forest* (RF) é um algoritmo de aprendizado de máquina desenvolvido por Breiman (2001) a fim de melhorar o algoritmo CART (*Classification and Regression of Trees*) pela combinação de um grande conjunto de árvores. O RF emprega uma operação de “ensacamento”, ou seja, agregação *bootstrap* - procedimento não paramétrico para estimar parâmetros em geral e as taxas de erro em particular, em que um número de árvores (*n_{tree}*) é construído com base em um subconjunto aleatório de amostras provenientes dos dados de treinamento (ADELABU et al., 2013).

Primeiramente, o RF constrói muitas árvores de classificação binárias usando várias amostras *bootstrap* com substituições extraídas das observações originais. Essas árvores de classificação contribuem com um voto e a classificação correta é determinada pelo voto da maioria de todas as árvores da floresta (ADAM et al., 2014). As amostras que não são amostras de *bootstrap* são chamadas de amostras “fora-do-saco” (termo comumente utilizado é o OOB - “*out-of-bag*”), e, a partir dessas amostras, é calculado o erro para todas as árvores (ADELABU et al., 2013). Posteriormente, em cada nó, um determinado número de entrada de variáveis (*m_{try}*) é escolhido de forma aleatória a partir de um subconjunto aleatório de atributos e a melhor separação é calculada utilizando apenas este subconjunto de atributos (ADAM et al., 2014; DURO;

FRANKLIN; DUBÉ, 2012; HASTIE; TIBSHIRANI; FRIEDMAN, 2009)
(Figura 3).

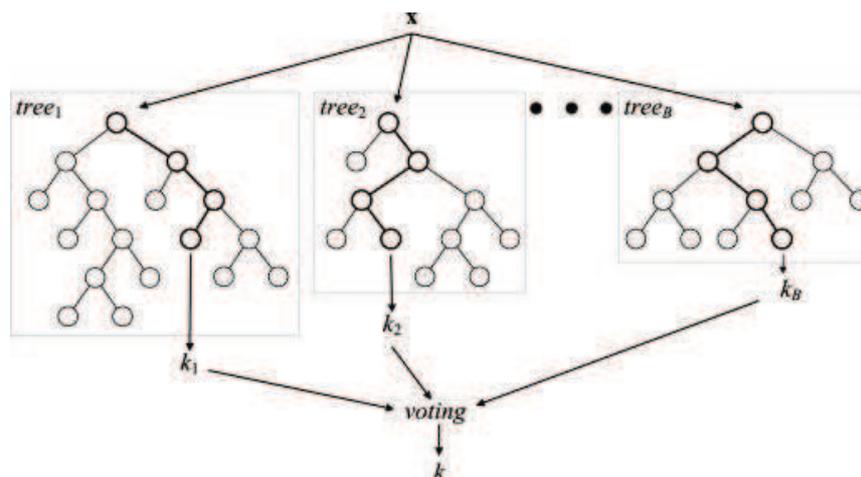


Figura 3 Representação gráfica do modelo de classificação *Random Forest*.

Nota: Sendo x os dados de entrada (amostras de treinamento) que são classificadas por todas as árvores ($tree_1, tree_2, \dots, tree_B$) que contribuem para a classificação com um voto, se correta

O RF não considera todas as variáveis em cada nó para determinar o melhor limite de divisão, mas um subconjunto aleatório do conjunto original dos atributos (PUISSANT; ROUGIER; STUMPF, 2014). Em vários trabalhos, o RF tem mostrado alto desempenho (CRACKNELL; READING, 2014; PUISSANT; ROUGIER; STUMPF, 2014; WATTS et al., 2011; ZHU; WOODCOCK; OLOFSSON, 2012). Shiraishi et al. (2014) obtiveram a melhor classificação de usos florestados e não florestados dentre outros algoritmos. Este é um resultado promissor tendo em vista o monitoramento de áreas vegetadas.

Outro algoritmo de aprendizagem de máquina que tem se destacado é o *Support Vector Machine* (SVM), pois seu desempenho é adequado para aplicações de sensoriamento remoto, em que uma quantidade limitada de dados

de referência é muitas vezes fornecida (MOUNTRAKIS; IM; OGOLE, 2011). Tal algoritmo foi proposto por Vapnik (1998) e tem por objetivo gerar limites de decisão que determine a separação ótima das classes, para isso, seleciona entre o número infinito de limites de decisão lineares o que minimiza o erro de generalização (PAL; MATHER, 2005). Usando dados de treinamentos com k números de amostras, representado por $\{X_i Y_i\}$, $i = 1, \dots, k$ em que $X \in R^N$ e está num espaço N -dimensional e $y \in \{-1, +1\}$ é uma classe rotulada (OTUKEI; BLASCHKE, 2010; PRADHAN, 2013). Muitos hiperplanos podem ser ajustados para separar as classes, mas apenas um hiperplano com separação ótimo é usado (FOODY et al., 2006). Os subconjuntos de pontos que ficam próximos do hiperplano são usados para medir a margem máxima ente os pontos e tais são denominados vetores de suportes (MOUNTRAKIS; IM; OGOLE, 2011; PAL; MATHER, 2005) (Figura 4).

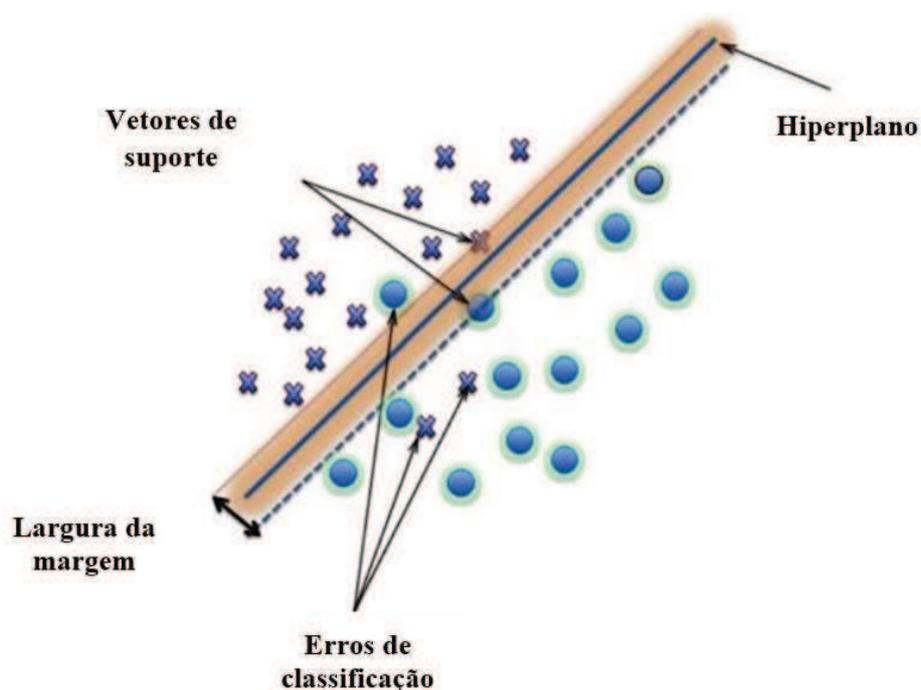


Figura 4 Representação gráfica dos vetores de suportes utilizados para separar as classes num hiperplano com separação ótima

Fonte: Mountrakis, Im e Ogole (2011).

Nota: O comprimento da margem é definido pelos vetores de suporte, porém, podem ocorrer casos de erros de classificação

A implementação de um SVM linear assume que os dados são linearmente separáveis no espaço, porém, há sobreposição de dados devido à difícil separabilidade linear e que afetará os padrões de precisão da classificação. Por isso, é importante escolher uma função que consiga encontrar a separabilidade das classes no hiperplano. Uma das funções mais utilizadas é a função Kernel RBF (*Radial Basis Function*), que tem apresentado melhores resultados (PRADHAN, 2013).

Há uma gama de trabalhos que incluem aplicações do SVM no sensoriamento remoto, como descrito por Mountrakis, Im e Ogole (2011). Em

classificações de imagens de sensoriamento remoto, vários trabalhos apresentaram classificações precisas graças ao desempenho do SVM (ADAM et al., 2014; ALCANTARA et al., 2012; CRACKNELL; READING, 2014; DURO; FRANKLIN; DUBÉ, 2012; PAL; MATHER, 2005; SHAO; LUNETTA, 2012).

A atuação desses algoritmos de aprendizagem de máquina na classificação da cobertura da terra também está relacionada aos atributos que são incorporados na classificação, por exemplo, dos dados temporais. No monitoramento da cobertura da terra, dados temporais podem melhorar a classificação, pois as informações temporais e fenológicas da cobertura da terra são úteis na separabilidade das classes (CARRÃO; GONÇALVES; CAETANO, 2008; VIEIRA et al., 2012).

2.2.2 Classificação multitemporal

A classificação multitemporal é conhecida como uma combinação de bandas a partir de mais de uma data, estação, ou ano (LANGLEY; CHESHIRE; HUMES, 2001), ou seja, envolve o emprego de múltiplas datas, com a aquisição de duas ou mais imagens, além da manipulação de conjuntos de dados maiores de imagens co-registradas. Dentre os trabalhos realizados utilizando dados multitemporais (incluindo séries temporais) a fim de potencializar a precisão dos mapeamentos, destacamos algumas metodologias:

- a) **STRS** (*Spectral Temporal Response Surface*): proposto por Vieira, Mather e Mccullagh (2000), tal método explora a utilização de uma superfície de resposta espectro-temporal para caracterizar o comportamento de um pixel ao longo do tempo para cada faixa de frequência. Neste método, cada amostra de dados de treinamento é

representada como um conjunto de pontos em um espaço tridimensional, os eixos representam o tempo, comprimento de onda do espectro eletromagnético e o valor de reflectância de diferentes sensores. Assim, um conjunto de amostras de treinamento representa o mesmo tipo de cobertura da terra medido em pontos ao longo de um período, e pode ser visualizado como pontos situados sobre uma superfície. Uma superfície analítica ajustada para esta distribuição pode ser parametrizada e seus coeficientes utilizados como atributos no processo de classificação de imagens de sensoriamento remoto.

- b) **STARS** (*Spectral–Temporal Analysis by Response Surface*): Mello et al. (2012) utilizando os conceitos do STRS desenvolveu o STARS. Tal método sintetiza todas as informações de um conjunto de dados multitemporais-multiespectrais de uma imagem de sensoriamento remoto para representar a variação espectral ao longo do tempo, definindo uma única imagem multicoeficiente. O STARS também permite a utilização dos dados de imagem a partir de diferentes sensores, com diferentes bandas de onda espectrais e intervalos de tempo irregulares. Além disso, diferentes opções de modelo podem ser utilizadas para ajustar as superfícies de resposta de acordo com a aplicação dos dados. O STARS é tratado como um método de processamento de imagem genérica que pode ser usada não só para a classificação, mas também para outras aplicações, como a compreensão da mudança de cobertura da terra.
- c) **CCDC** (*Continuous Change Detection and Classification*): tal algoritmo foi desenvolvido por (ZHU; WOODCOCK; OLOFSSON, 2012) a fim de melhorar tanto a detecção de alterações contínuas como a classificação da cobertura da terra com a utilização do domínio temporal dos dados Landsat. O uso do termo "contínuo" se

refere à capacidade do algoritmo de detectar mudanças cada vez que uma nova imagem é adicionada (monitoramento contínuo). Em vez de classificar as imagens Landsat originais como os métodos convencionais fariam, o algoritmo CCDC utiliza os coeficientes dos modelos de séries temporais, gerados pelo ajuste do algoritmo, como entrada para a classificação de cobertura do solo. Esses componentes fornecem informações temporais sobre os padrões intra-anuais (ou sazonalidade), também mede diferenças ou tendências interanuais, além do RMSE (*Root Mean Square Error*) e valor espectral no centro de cada modelo de séries temporais para cada banda Landsat. Ao classificar os coeficientes do modelo de séries temporais, este algoritmo pode fornecer um tipo de cobertura do solo durante todo o período de tempo para cada modelo de séries temporais. Contudo, este algoritmo é computacionalmente caro e precisa de muito armazenamento de dados, além disso, requer alta frequência temporal de observações sem nuvens.

Como visto anteriormente, dados multitemporais têm sido usados como atributos de entrada para algoritmos de classificação melhorando consideravelmente a precisão dos mapeamentos da cobertura do solo. Muitas análises têm sido desenvolvidas a fim de melhorar o entendimento acerca do comportamento temporal da cobertura da terra. (CRUZ, 2008; GALFORD et al., 2008b; MARTÍNEZ; GILABERT, 2009) usaram transformada *wavelets* para caracterizar os perfis temporais de usos da terra. O método utilizado gera séries temporais suavizadas que auxiliam no entendimento das características fenológicas da superfície terrestre. Verbesselt et al. (2010a) propôs um algoritmo denominado BFAST (*Breaks For Additive Seasonal and Trend*), que decompõe séries temporais de maneira iterativa em três componentes: tendência (Z_t),

sazonalidade (S_t) e ruído (ε_t), utilizando métodos para detectar e caracterizar mudanças dentro da série temporal. Esta decomposição resulta na detecção de mudanças rápidas e graduais que acontecem dentro do componente de tendência ou mudanças fenológicas ocorrendo no componente sazonal.

Uma importante ferramenta endossada às análises de séries temporais são os índices de vegetação, por exemplo, o NDVI (*Normalized Difference Vegetation Index*), que é capaz de extrair informações do ecossistema. Esse índice se baseia no fato de que a radiação vermelha é utilizada pelas plantas durante a fotossíntese, enquanto a radiação infravermelha é refletida, resultando em uma baixa reflectância na banda vermelha e elevada reflectância na banda do Infravermelho (GAMON et al., 1995). Por isso, a análise de tendências de séries temporais NDVI é adequada para o monitoramento sutil de alterações da vegetação em longo prazo, bem como as mudanças interanuais das dinâmicas da vegetação (MARTÍNEZ et al., 2011).

Forkel et al. (2013) avaliaram o desempenho de diferentes métodos para detecção de tendências e *breakpoints* (pontos de interrupção na série temporal) em séries temporais NDVI. A utilização de tais métodos permitiu uma avaliação crítica da tendência, bem como compreender o componente sazonal. Forkel et al. (2013) desenvolveram um pacote denominado *greenbrown* a partir do qual é possível extrair parâmetros de uma série temporal e assim inferir sobre as propriedades da série, como descrito na Figura 5. Para computar tais parâmetros algumas etapas são realizadas:

- a) A média da série temporal NDVI é calculada;
- b) Posteriormente, os valores mensais são calculados para valores anuais e a tendência é calculada, mas sem computar os pontos de interrupção. Desse modo, é obtido o declive da tendência anual da série NDVI ao longo do comprimento total da série temporal;

- c) Para estimar a variabilidade interanual, o desvio padrão e amplitude das anomalias anuais são calculados. A média da série temporal e o componente de tendência derivada do passo (2) são subtraídos dos valores anuais da série para derivar os valores anuais de tendência removida e média centrada (anomalias anuais). Se o declive da tendência não for significativa ($p > 0,05$), apenas a média é subtraída. O desvio padrão e o intervalo das anomalias anuais são calculados como medidas para a variabilidade interanual da série temporal;
- d) No passo seguinte, é estimado o intervalo do ciclo sazonal. A média, o componente de tendência e as anomalias anuais são subtraídos da série temporal original para calcular uma série temporal ajustada. Com base nesta série temporal o ciclo sazonal é estimado como média do ciclo sazonal e a variação é então computada;
- e) E por último, o desvio padrão e o intervalo das anomalias de curto prazo são calculados. Anomalias de curto prazo são calculadas subtraindo a média, o componente de tendência, as anomalias anuais e a média do ciclo sazonal a partir da série temporal original. Como resultado, obtém-se o componente residual da série. O desvio padrão do componente residual da série temporal é uma medida de variabilidade de curto prazo, ou seja, a variabilidade intra-anual.

É possível obter parâmetros importantes para a caracterização da cobertura terrestre, e assim, melhorar a separabilidade das classes, pois, as variações temporais dos diferentes usos da terra apresentam comportamentos de sazonalidade, tendência e variabilidade intra- e interanual peculiares. Sendo assim, os dados gerados pelo pacote *greenbrown* são atributos temporais capazes de aprimorar as classificações da cobertura da terra juntamente com os algoritmos de aprendizagem de máquina. No caso do mapeamento que dará

suporte ao CAR, as classificações serão realizadas em imagens com alta resolução espacial - Rapideye. Tais imagens não apresentam uma densa série temporal, isto é, não possuem imagens multidas que possam ser usadas como parâmetro temporal. Neste caso, é necessário utilizar dados multitemporais de imagens TM a fim de derivar atributos temporais que reforcem a precisão das classificações.

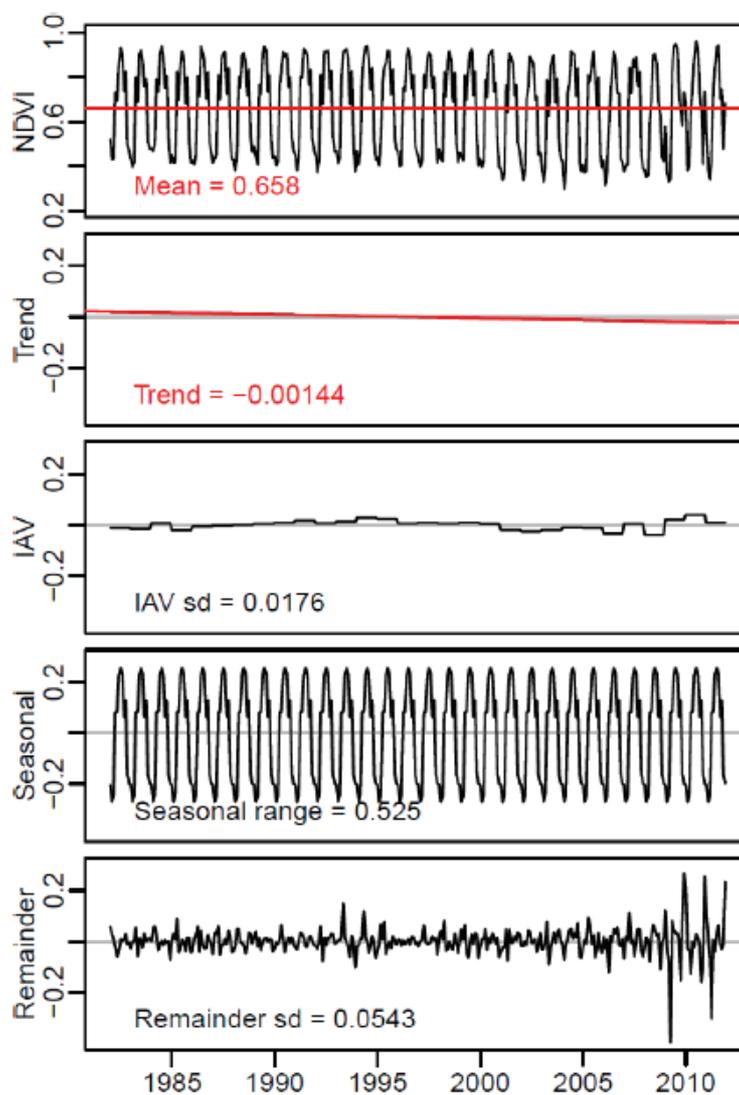


Figura 5 Componentes da série temporal estimados. O painel superior mostra uma série temporal NDVI com o seu valor médio (linha vermelha)

Fonte: Forkel et al. (2013)

Nota: Os próximos painéis mostram a tendência estimada, a variabilidade interanual (IAV) (ou seja, anomalias anuais), sazonalidade (ou seja, a média de ciclo sazonal) e a variabilidade de curto prazo (componente residual), respectivamente. A soma de média, tendência, IAV, componente sazonal e residual compõem a série de temporal original

3 BASE DE DADOS

3.1 Área de Estudo

A área de estudo está localizada no Centro-Oeste do estado de Minas Gerais, região Sudeste do Brasil, e as análises foram limitadas a uma área de 625 km² referente a uma cena Rapideye (Figura 6). Tal cena Rapideye está centralizada em 18° 52' 44.33" de latitude Sul e 45° 27' 37.70" de longitude Oeste. Os solos da região são geralmente bem drenados e ácidos (OLIVEIRA-FILHO, 2006). A vegetação predominante na área é o Cerrado (savana brasileira) (CARVALHO, 2008), que apresenta vários tipos de fitofisionomias (OLIVEIRA et al., 2012). Em termos gerais, pode ser considerada uma transição entre florestas e pastagens, apresentando reduções graduais da biomassa e densidade arbóreo/arbustivas na paisagem, possuindo plantas com sistemas radiculares profundos como estratégia de sobrevivência em épocas de seca (GOMIDE et al., 2012). Estima-se que o Cerrado teve cerca de 48% da sua vegetação suprimida entre os anos 2002 a 2010 de um total de dois milhões de quilômetros quadrados (MINISTÉRIO DO MEIO AMBIENTE, 2011), portanto, essa região apresenta uma paisagem bastante heterogênea.

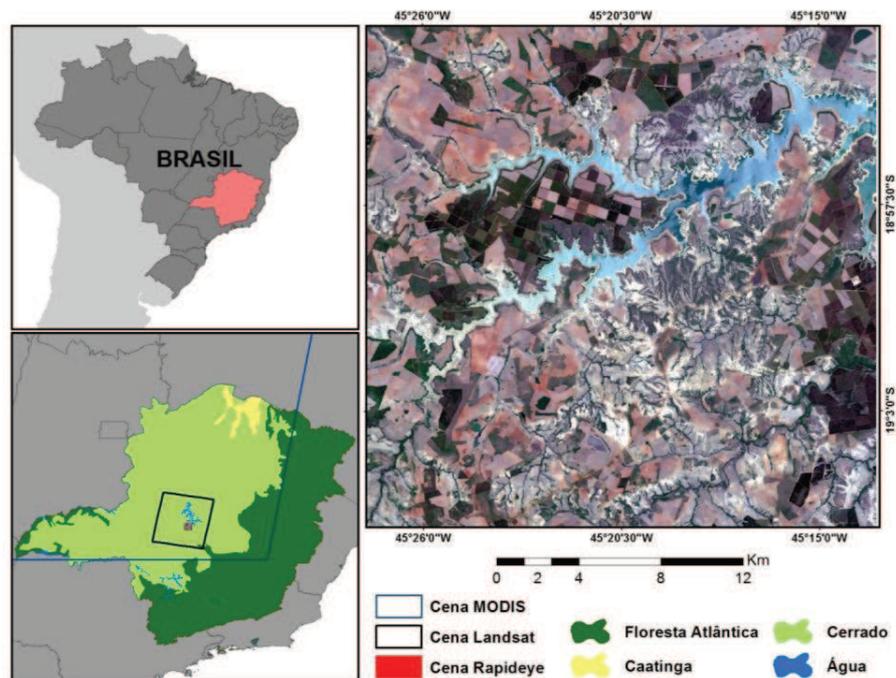


Figura 6 Localização da área de estudo na região Sudeste do Brasil, com destaque para a cena Rapideye em composição RGB (R=banda 5; G=banda 4; B= banda 3)

3.2 Aquisição dos dados e pré-processamento

Para este estudo, utilizou-se uma imagem Rapideye (tile 2330313) de 11 outubro de 2012. O conjunto de satélites Rapideye coleta imagens multiespectrais e com resolução espacial de 5m. As cinco bandas incluem azul (440-510 nm), verde (520-590 nm), vermelho (630-685 nm), vermelho-edge (690-730 nm) e NIR (760-850 nm). Esta imagem representa um conjunto de dados de alta resolução espacial que será utilizado como base de detalhamento espacial deste trabalho.

Todas as imagens disponíveis do sensor Landsat-5 TM (órbita/ponto 219/073) foram adquiridas entre 2000 e 2011, totalizando 214 imagens. As imagens TM possuem dados de reflectância em sete bandas, incluindo azul (450-520 nm), verde (520-610 nm), vermelho (630-690 nm), infravermelho próximo (760-900 nm), infravermelho médio (1550-1750 nm), infravermelho distante (2080-2350 nm) e a termal (10400-12500 nm). Com exceção da banda termal que apresenta resolução espacial de 120 metros, todas as demais possuem resolução espacial de 30 metros.

Para este estudo foi utilizado o produto CDR (*Reflectance Climate Data Record Landsat Surface*) disponível no portal United States Geological Survey (USGS)(<http://earthexplorer.usgs.gov/>). O produto Landsat CDR é gerado a partir do *software* especializado chamado LEDAPS (*Landsat Ecosystem Disturbance Adaptive Processing System*). O LEDAPS gera dados sobre refletância e temperatura de superfície, localização de nuvens e sombras de nuvens, solo e água, a partir de informações como vapor de água, ozônio, altura geopotencial, espessura óptica de aerossóis e elevação do terreno. Esse conjunto de dados serve de entrada para o algoritmo 6S (*Satellite Signal in the Solar Spectrum*), que resulta em produtos CDR (MASEK, 2009), além disso, o produto é ortorretificado.

As imagens TM foram utilizadas para compor as séries temporais de média resolução espacial. No entanto, como regiões tropicais apresentam grande incidência de nuvens e eventual indisponibilidade de imagens TM, foi necessária a utilização de imagens sintéticas oriundas de fusão de dados.

Além das imagens TM, também foi adquirido um conjunto de 273 imagens do sensor MODIS a bordo do satélite Terra (produto MOD13Q1) no portal USGS (<https://mrtweb.cr.usgs.gov/>), cena h13v10 entre 2000 e 2011. Esse produto fornece a composição NDVI e dados de reflectância nas bandas do infravermelho próximo (NIR) e vermelho. O MOD13Q1 apresenta uma

resolução espacial de 250 metros (VERBESSELT et al., 2010b), considerado de baixa resolução espacial no presente estudo. Esse produto representa um mosaico de pixels sem nuvens observados nas imagens diárias durante um período de 16 dias (ANDERSON; SHIMABUKURO, 2007).

Para atender os pré-requisitos para a realização da fusão de imagens pelo algoritmo STARFM, ambas as bandas dos dados MODIS, NIR e vermelho (RED), foram reprojctadas para a projeção Universal Transversa de Mercator (UTM), fuso 23S. Para isso, utilizou-se o *software* MODIS Re-projection Tools (MRT). Além disso, estas bandas foram reamostradas para 30 m de resolução espacial utilizando uma abordagem do vizinho mais próximo. Em todas as imagens foram feitos subsetsna extensão do polígono correspondente à imagem TM.

Como observado, a resolução espacial dos conjuntos de dados utilizados neste estudo diminui das imagens Rapideye para as imagens MODIS. No entanto, o comportamento da resolução temporal é inverso. As imagens MODIS utilizadas representam observações a cada 16 dias, as imagens TM distribuídas de forma irregular ao longo do tempo e a imagem Rapideye representando uma única observação.

4 MÉTODOS

4.1 Predição de imagens sintéticas

Neste estudo, foi utilizado o algoritmo STARFM, proposto por Gao et al. (2006) para fusão de imagens TM/ETM+ e MODIS. O STARFM emprega uma janela móvel que incorpora informações adicionais dos pixels vizinhos das imagens TM/ETM+ e MODIS usando uma função de ponderação definida como a combinação espectral, temporal e espacial das imagens (GAO et al., 2006). Pixels espectralmente similares e livres de nuvens dentro da janela móvel são usados para computar a reflectância, desse modo, garantindo que informações corretas de pixels vizinhos serão utilizadas (GAO et al., 2006). O algoritmo prediz valores de pixels utilizando cenas TM e MODIS adquiridas em um Tempo 1 (T_1), e uma ou mais cenas MODIS referente ao dia da predição da imagem sintética Tempo 2 (T_2) (HILKER et al., 2009b).

O algoritmo é dado por:

$$\begin{aligned}
 L\left(x_{\frac{w}{2}}, y_{\frac{w}{2}}, T_2\right) &= \sum_{i=1}^w \sum_{j=1}^w W_{ij} (M(x_i, y_j, T_2) + L(x_i, y_j, T_1) \\
 &\quad - M(x_i, y_j, T_1))
 \end{aligned} \tag{2}$$

em que $L\left(x_{\frac{w}{2}}, y_{\frac{w}{2}}, T_2\right)$ é o valor do pixel TM predito, w é o tamanho da janela de busca e $\left(x_{\frac{w}{2}}, y_{\frac{w}{2}}\right)$ é o pixel central da janela móvel. O peso W_{ij} é calculado a partir da diferença espectral entre MODIS e TM, diferença temporal entre MODIS T_1 e MODIS T_2 e a distância entre o pixel central predito e o pixel

vizinho candidato espectralmente similar. Assim, é determinado o quanto cada pixel vizinho (x_i, y_j) irá contribuir para estimar o valor de reflectância do pixel central (GAO et al., 2006). $M(x_i, y_j, T_2)$ é a reflectância da imagem MODIS em x_i, y_j observada no T_2 . Por fim, $L(x_i, y_j, T_1)$ e $M(x_i, y_j, T_1)$ são as reflectâncias das imagens TM e MODIS no tempo T_1 , respectivamente (HILKER et al., 2009b). Para a realização da fusão das imagens, a cobertura de nuvens das imagens TM foi verificada no intuito de avaliar quais imagens deveriam ser totalmente preditas ou parcialmente preditas, ou seja, somente em locais com nuvens. As imagens com mais de 80% de nuvens foram totalmente preditas. As imagens TM livres de nuvens foram pareadas com imagens MODIS, levando em consideração a diferença mínima entre as datas de imageamento dos sensores (Tabela 1). As máscaras de nuvens e sombras de nuvens foram utilizadas para substituir os valores reais de nuvens e sombras por valores de reflectância preditos pelo STARFM.

Tabela 1 Pares de imagens MODIS e Landsat (sem nuvens) correspondentes utilizadas na predição de imagens Landsat sintéticas

MODIS T1	Landsat T1	MODIS T2	TM Sintética
2008097	2008103	2008001	2008007*
2008097	2008103	2008017	2008023
2008097	2008103	2008033	2008039
2008097	2008103	2008049	2008055
2008097	2008103	2008065	2008071
2008097	2008103	2008081	2008087
2008193	2008199	2008097	2008103
2008193	2008199	2008113	2008119
2008193	2008199	2008129	2008135
2008193	2008199	2008145	2008151
2008193	2008199	2008161	2008167*
2008193	2008199	2008177	2008183
2008273	2008231	2008241	2008247
2008321	2008279	2008257	2008263
2008321	2008279	2008289	2008295
2008321	2008279	2008305	2008311*
2008321	2008279	2008321	2008327
2008321	2008279	2008337	2008343
2008321	2008279	2008353	2008359

Nota: Note-se que algumas cenas Landsat (negrito) já tinham as nuvens mascaradas.

* Imagem não disponível

4.1.1 Avaliação de imagens TM sintéticas

A fim de avaliar a qualidade das imagens geradas pelo STARFM e comprovar sua usabilidade em análises de séries temporais, a reflectância observada em imagens reais e sintéticas foram comparadas em duas abordagens: baseada em pixel e baseada em objeto. Para a realização da abordagem baseada em pixel, foram feitos subsets das imagens TM reais e sintéticas (1000 x 1000

pixels) nas bandas NIR, vermelho e no índice NDVI. Já para a abordagem baseada em objeto, foi realizada a segmentação da imagem Landsat e a média da reflectância dos 117.884 objetos foi computada para as bandas NIR vermelho e índice NDVI, tanto para imagens reais quanto para as sintéticas. Para avaliar a correlação entre as imagens reais e sintéticas, foram realizadas regressões simples entre os valores reais e sintéticos para cada abordagem.

4.2 Análise das Séries Temporais

As imagens TM tanto originais quanto sintéticas que compuseram as séries temporais eram compostas de dados provenientes do cálculo do NDVI (*Normalized Difference Vegetation Index*). Esse índice é uma relação baseada na diferença de reflectância espectral entre as bandas do vermelho (ρ_{RED}) e infravermelho próximo (ρ_{NIR}) realçando os sinais de vegetação (SONNENSCHNEIN et al., 2011), definido como:

$$NDVI = \frac{(\rho_{NIR} - \rho_{RED})}{(\rho_{RED} + \rho_{NIR})} \quad (3)$$

O NDVI foi calculado utilizando as imagens TM preditas e reais, assim, totalizando 270 imagens de 2000 a 2011. Foram construídas séries com diferentes frequências temporais: Estacional (quatro imagens/ano – aproximadamente a cada três meses), Bimestral (seis imagens/ano), Mensal (12 imagens/ano) e Completa (23 imagens/ano), para que assim fosse analisado a influência da frequência temporal na classificação.

As análises das séries foram realizadas pela metodologia proposta por Forkel et al. (2013). Essa metodologia foi implementada usando o pacote denominado *greenbrown* (<http://greenbrown.r-forge.r-project.org/>), disponível

para o *software* R (R CORE TEAM, 2013). Este pacote apresenta uma gama de funções para analisar tendências e mudanças de tendências em séries temporais. Neste estudo, computamos as propriedades estatísticas das séries temporais de imagens TM. Esse procedimento gera sete camadas de informação, sendo elas: Média, Tendência, Variabilidade Interanual (IAV), desvio padrão de IAV, Variação Sazonal, Variabilidade Intra-Anual (STV) e desvio padrão de STV.

Foram feitos subsets referentes ao tamanho do tile da cena Rapideye a partir das séries temporais (Estacional, Bimestral, Mensal e Completa), e em seguida foram realizadas as análises com o pacote *greenbrown*. As camadas de informação resultantes foram incorporadas como atributos temporais no processo de classificação.

4.3 Classificação da cobertura da terra

A abordagem adotada no trabalho é a baseada em objetos, portanto, a cena Rapideye foi segmentada em duas etapas: (1) *multi-resolution image segmentation* com escala de 180, shape 0.3 e compactness 0.6; (2) *spectral difference algorithm* de 180. Todos esses parâmetros foram testados visando obter uma segmentação com maior caracterização dos objetos. A etapa (2) consiste num processo de ressegmentação da etapa (1), em que objetos vizinhos são fundidos de acordo com o valor médio da intensidade da banda e um limiar máximo definido pelo usuário (ECOGNITION DEVELOPER, 2011), resultando em 44.779 objetos.

Posteriormente, foram calculados: média, desvio padrão, brilho e máxima diferença para cada objeto em cada banda da imagem Rapideye, e o NDVI, os quais foram exportados juntamente com os demais atributos, provenientes dos resultados das análises temporais. Os atributos associados aos objetos foram subsequentemente usados como dados de entrada para os

algoritmos de classificação. A Tabela 2 mostra os atributos utilizados como dados de entrada para o processo de classificação, lembrando que os atributos temporais foram gerados para as diferentes frequências de séries temporais.

Tabela 2 Atributos utilizados como *input* no processo de classificação dos algoritmos de aprendizagem de máquina

Tipo	Origem	Nome
Espectral	Rapideye	Brilho
		Desvio padrão (b1, b2, b3, b4, b5)
		Máxima diferença
		Média (b1, b2, b3, b4, b5)
		NDVI
Temporal	Análise da série temporal TM –pacote <i>greenbrown</i>	Média
		Tendência
		IAV
		Desvio padrão IAV
		Variação sazonal
		STV
		Desvio padrão STV

Nota: b1=banda 1; b2= banda 2; b3= banda 3; b4= banda 4; b5= banda 5.

As semelhanças espectrais da cobertura da terra mostram a dificuldade de mapear áreas heterogêneas, principalmente áreas antropogênicas vegetadas que se assemelham com áreas remanescentes nativas. Neste estudo, foi testada a separabilidade entre as classes: Agricultura, Área Alagada, Corpos d'Água, Floresta Plantada, Pastagem, Remanescente (que se refere às áreas nativas) e Solo Exposto.

O mesmo conjunto de objetos segmentados foi utilizado para cada algoritmo de classificação, bem como as mesmas amostras de treinamento. Foram coletadas 1.205 amostras de treinamento, abrangendo todas as variações

espectrais das classes da terra: Agricultura= 10 objetos; Área alagada= 115 objetos; Corpos d'Água= 85 objetos; Floresta Plantada= 75 objetos; Pastagem= 220 objetos; Remanescente= 395 objetos, e Solo Exposto= 305 objetos.

A fim de avaliar o desempenho de diferentes métodos de classificação da cobertura da terra foram utilizados os seguintes algoritmos de aprendizagem de máquina: *Decision Tree* (DT), *Naive Bayes* (NB), *Random Forest* (RF) e *Support Vector Machine* (SVM). Os parâmetros utilizados pelos algoritmos de classificação neste trabalho estão descritos na Tabela 3. Todas as fontes de códigos são pacotes oriundos do *software* R (R CORE TEAM, 2013). Alguns parâmetros variam com o conjunto de dados utilizados para treinar o algoritmo, pois tais parâmetros são calculados a partir dos dados fornecidos como treinamento, como o $mtry$ e o σ . Como foram utilizados diferentes atributos temporais, obteve-se parâmetros diferentes em algumas classificações utilizando o mesmo MLA, no caso, o RF e o SVM.

Para cada algoritmo de classificação, foi gerado uma matriz de confusão usando os dados de treinamento, assim, um parâmetro ótimo foi escolhido para cada MLA levando em consideração o maior valor de acurácia global gerado no processo. Com isso, apenas os modelos com melhores valores de acurácia foram usados no processo de classificação.

Tabela 3 Parâmetros específicos para cada algoritmo de aprendizagem de máquina utilizado na classificação

Algoritmo	Fonte de códigos	Tipo de parâmetro	Descrição do parâmetro	Parâmetro
Decision Tree	rpart	cp	É usado para controlar o tamanho da árvore de decisão e para selecionar o tamanho ideal de árvore	0.001
		maxdepth	Define a profundidade máxima de qualquer nó da árvore final	30
Naive Bayes	e1071	laplace	Suaviza dados categóricos com o propósito de evitar que o cálculo da probabilidade seja igual a zero	0
Random Forest	randomForest	mtry	Número de variáveis amostradas aleatoriamente como candidatos em cada divisão. Note-se que os valores padrão são diferentes para classificação	2 (classificação utilizando atributos mensais), 4
		ntree	Número de árvores	1000
Support Vector Machine	e1071	σ	O parâmetro σ define a largura da função de kernel	0.02, 0.03 (classificação utilizando atributos estacionais)
		kernel	Define qual função kernel é utilizada no treinamento e predição dos dados.	rbfdot - Radial Basis Function
		C	Ajusta a sensibilidade da margem de decisão de vetores de suporte erroneamente classificados	1

A Figura 7 demonstra todo o procedimento realizado desde pré-processamento dos dados até o processo de classificação que resultou nos mapas temáticos.

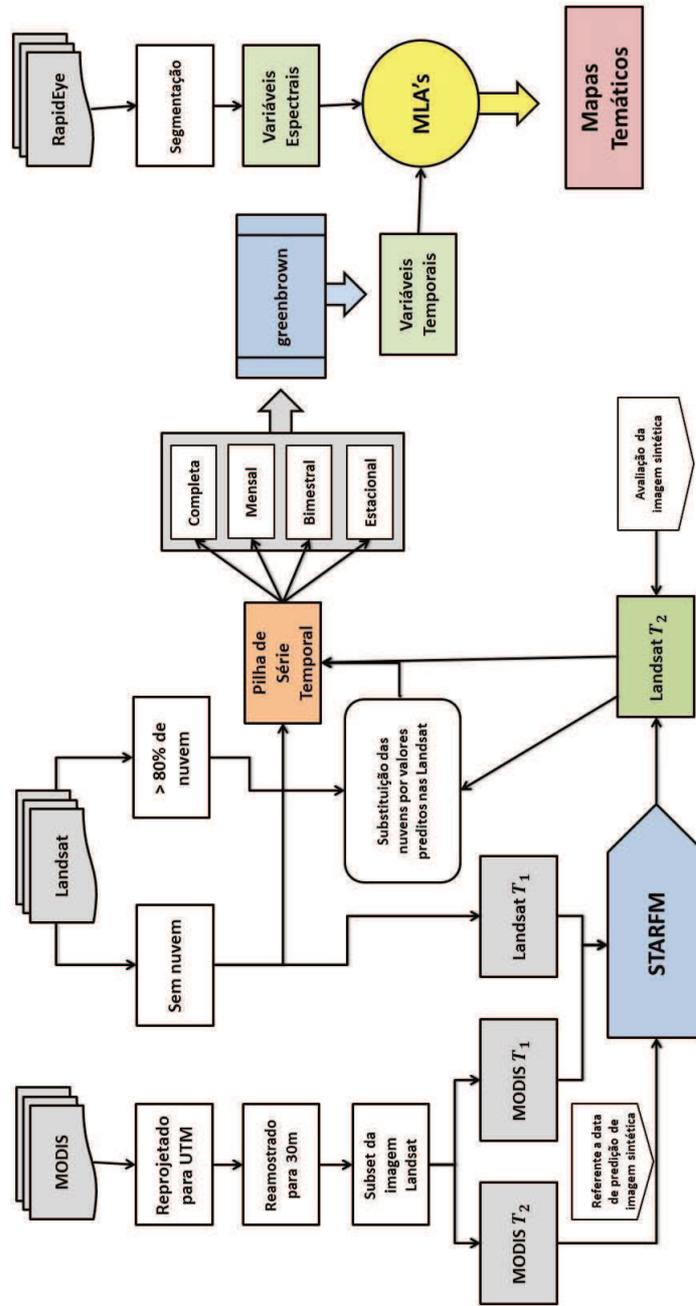


Figura 7 Fluxograma dos procedimentos utilizados neste estudo

4.3.1 Avaliação e acurácia das classificações

Com intuito de verificar a acurácia das classificações, foram coletados separadamente 315 pontos como dados de teste. Essas amostras foram selecionadas aleatoriamente e conferidas com o auxílio do Google Earth (<http://earth.google.com/>), sendo assim, a acurácia das classificações foi avaliada somente por interpretação visual. Posteriormente, foi utilizado o coeficiente Kappa e acurácia global como critério de avaliação.

Os valores de acurácia das imagens classificadas foram comparados usando o Teste McNemar (FOODY, 2004; MAS; GAO; PACHECO, 2010). Tal teste é não paramétrico e avalia a significância estatística das diferenças entre as duas classificações, baseando-se em matrizes de confusão de duas dimensões (Tabela 4). A equação do teste pode ser expressa por:

$$x^2 = \frac{(f_{12} - f_{21})^2}{f_{12} + f_{21}} \quad (4)$$

A avaliação do teste é baseada na distribuição chi-quadrado (x^2). Assim, o valor de chi-quadrado calculado é comparado com o valor de chi-quadrado tabelado indicando sua significância estatística (FOODY, 2004; MAS; GAO; PACHECO, 2010).

Tabela 4 Definição de elementos de matrizes utilizadas na equação 3

Classificação 2		
Classificação 1	Correto	Incorreto
Correto	f_{11}	f_{12}
Incorreto	f_{21}	f_{22}

5 RESULTADOS E DISCUSSÃO

5.1 Acurácia das imagens sintéticas

A avaliação da acurácia das imagens TM sintéticas foi realizada usando duas abordagens: por pixel e baseada em objeto. De modo geral, a abordagem baseada em objeto apresentou melhores valores de R^2 do que a baseada em pixel, porém a banda RED na abordagem baseada em pixel ($R^2= 0.951$) apresentou correlação um pouco melhor que a mesma banda na abordagem baseada em objeto ($R^2= 0.9491$), como visto na Figura 8.

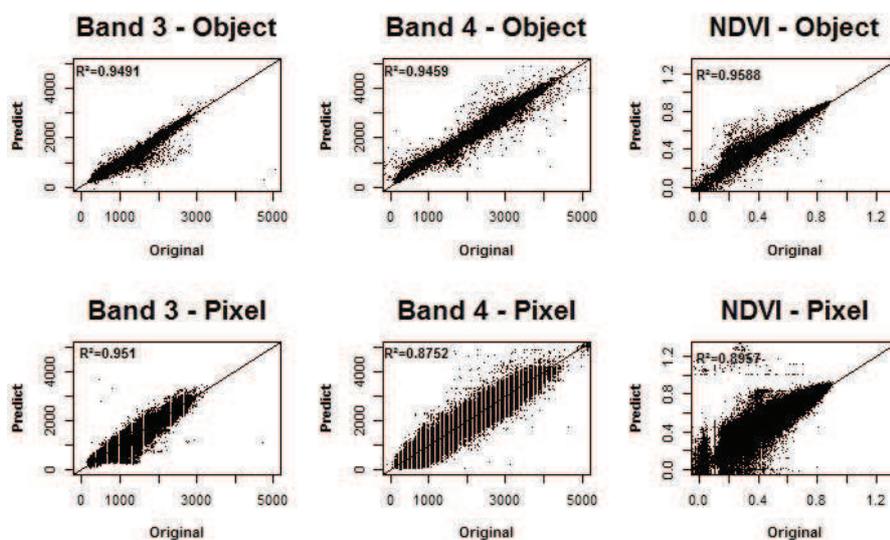


Figura 8 Comparação entre a reflectância da imagem TM real e predita nas abordagens por pixel e baseada em objeto

As análises de correlação das imagens sintéticas apresentaram resultados similares aos de análises realizadas por Bhandari, Phinn e Gill (2012), Walker et al. (2012), Watts et al. (2011). Em tais trabalhos, que adotaram apenas a

abordagem baseado em pixel, também houve maior correlação na predição da banda RED. No presente estudo, foi verificado, ainda, que a correlação entre a imagem sintética e a imagem original aumenta usando a abordagem baseada em objeto, mostrando que as imagens preditas podem ser usadas com segurança para derivar atributos temporais importantes para o processo de classificação e monitoramento da cobertura da terra.

5.2 Performances dos Algoritmos de Aprendizagem de Máquina

As classificações geradas pelos algoritmos de aprendizagem de máquina tiveram performances variadas tanto em relação ao algoritmo de classificação quanto pelas variáveis utilizadas. A Tabela 5 mostra os resultados da acurácia global e coeficiente Kappa, em que os melhores desempenhos foram dos algoritmos SVM e RF. O SVM apresentou acurácia global variando entre 76,51% e 92,06%, e o RF obteve acurácia global variando entre 77,78% e 90,48%, enquanto que os outros algoritmos tiveram performances inferiores. O NB apresentou acurácia global variando entre 68,25% e 86,03%, e o DT variando entre 71,75% e 86,98%.

Tabela 5 Acurácia das classificações para cada MLA usando diferentes variáveis

DT	Parâmetros	Acurácia	
		Global (%)	Kappa
DT	STC + VE	85,40	0,81
	STM + VE	85,40	0,81
	STB + VE	86,98	0,84
	STE + VE	84,13	0,80
	VE	71,75	0,65
NB	STC + VE	85,40	0,82
	STM + VE	86,03	0,83
	STB + VE	84,44	0,81
	STE + VE	85,08	0,81
	VE	68,25	0,60
RF	STC + VE	88,89	0,86
	STM + VE	89,84	0,87
	STB + VE	89,21	0,86
	STE + VE	90,48	0,88
	VE	77,78	0,72
SVM	STC + VE	91,43	0,89
	STM + VE	92,06	0,90
	STB + VE	92,06	0,90
	STE + VE	90,16	0,88
	VE	76,51	0,70

Nota: Em negrito estão os melhores resultados. (STC= Série temporal completa; STM= Série temporal mensal; STB= Série temporal bimestral; STE= Série temporal estacional; VE= Variáveis espectrais da Rapideye).

Analisando os resultados por algoritmo de classificação, os parâmetros temporais incorporados à classificação tiveram comportamentos diferentes quanto aos atributos temporais, visto que esses tinham frequência de observações diferentes (Completa, Mensal, Bimestral e Estacional). O classificador DT obteve melhor resultado utilizando variáveis temporais bimestrais (Acurácia global=86,98%, Kappa=0,84). Já o NB obteve melhor

resultado utilizando variáveis temporais mensais (Acurácia global=86,03%, Kappa=0,83). O SVM apresentou valores iguais de acurácia global=92,06% e Kappa=0,90 utilizando variáveis temporais mensais e bimestrais, sendo esses os melhores resultados desempenhados pelo algoritmo. Contudo, o RF se mostrou melhor utilizando variáveis temporais estacionais (Acurácia global=90,48%, Kappa=0,88). Já as classificações sem as variáveis temporais apresentaram desempenhos inferiores em todos os algoritmos de aprendizagem de máquina em termo de acurácia.

Dentre as classificações utilizando apenas as variáveis espectrais da Rapideye, a melhor acurácia foi apresentada pelo algoritmo RF, com 77,78% de acurácia global e Kappa= 0,72. Porém, o SVM também obteve um bom desempenho e bastante semelhante, com acurácia global de 76,51% e Kappa= 0,70. Os algoritmos DT e NB apresentaram novamente acurácias inferiores, o DT apresentou acurácia global de 71,75% e Kappa= 0,65, enquanto que o NB apresentou o pior desempenho, com acurácia global de 68,25% e Kappa= 0,60. A Figura 9 mostra as classificações utilizando apenas as variáveis espectrais da Rapideye para cada algoritmo de aprendizagem de máquina.

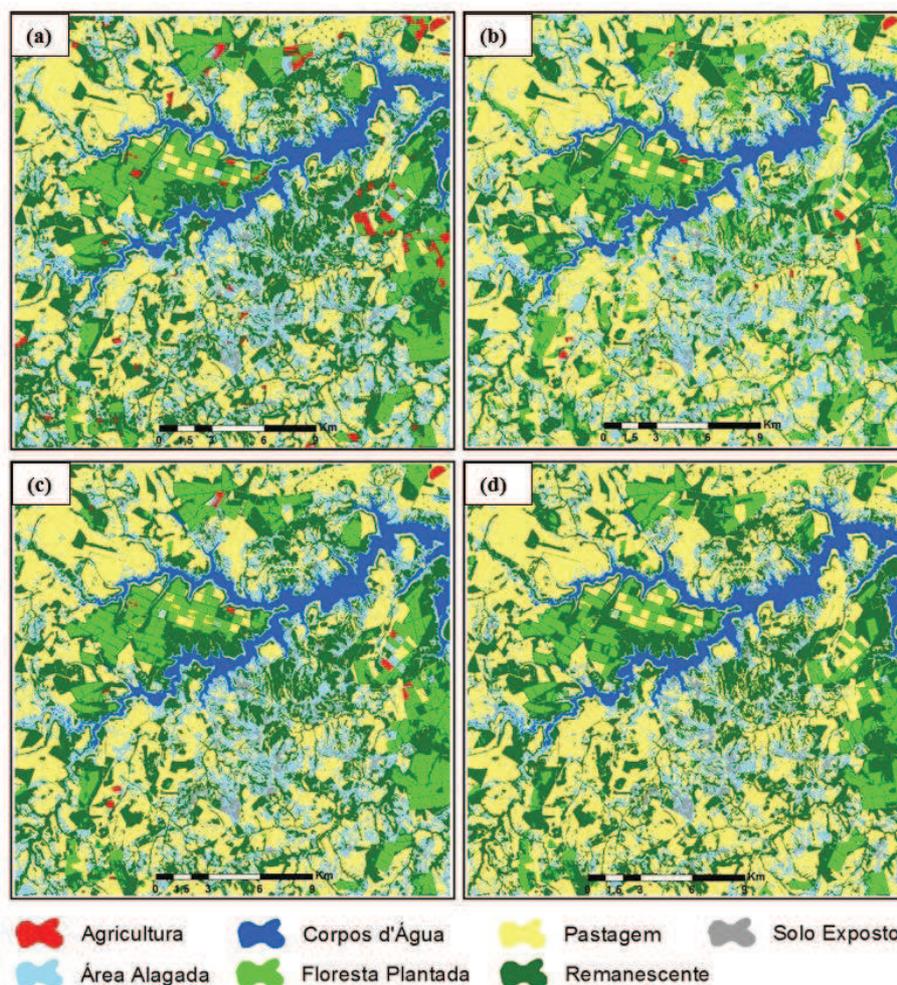


Figura 9 Classificação utilizando apenas variáveis espectrais da Rapideye e MLI's. (a) DT; (b) NB; (c) RF (d) SVM

Duro, Franklin e Dubé (2012) compararam a precisão da classificação de três algoritmos de aprendizagem de máquina (DT, RF, e SVM) usando imagens SPOT-5, tais autores também encontraram um melhor desempenho do SVM e RF em relação ao DT. Shao e Lunetta (2012) também compararam algoritmos de classificação (DT, RNA e SVM) e as melhores performances

foram encontradas nas classificações derivadas do SVM. Li et al. (2013), no entanto, ao utilizarem os algoritmos NB, DT e SVM no mapeamento de vegetação aquática usando imagens multitemporais TM, encontraram a melhor acurácia da classificação gerada pelo NB (86,8%), seguida pelo DT (86,6%), diferentemente dos nossos resultados que apenas demonstram um desempenho similar desses dois algoritmos em todas as classificações. O DT e NB utilizando atributos temporais apresentaram valores médios de acurácia global próximos (DT: Acurácia global média=85,48%; NB: Acurácia global média=85,24%), o que sugere que esses algoritmos têm desempenho semelhante.

É importante ressaltar que o SVM obteve, tanto para as variáveis temporais mensais quanto para as bimestrais, valores iguais de acurácia (92,06%) e de Kappa (0,90). Ao analisar as duas classificações, ambas apresentaram erros de classificação das classes Remanescente, Pastagem e Floresta Plantada. A Figura 10c-d, destaca alguns desses erros de classificação, no qual o item 1 mostra uma área de Floresta Plantada que na classificação mensal foi colocada como pastagem, e no item 2, mostra uma área de Floresta Plantada que na classificação bimestral foi colocada como Remanescente. O mesmo também ocorreu entre as classificações com variáveis temporais completa e mensal usando DT (Acurácia global=85,40%, Kappa=0,81), também foram analisado os erros de classificação e foram encontrado erros distribuídos em todas as classes, em destaque na Figura 10a-b nos itens 1-3.

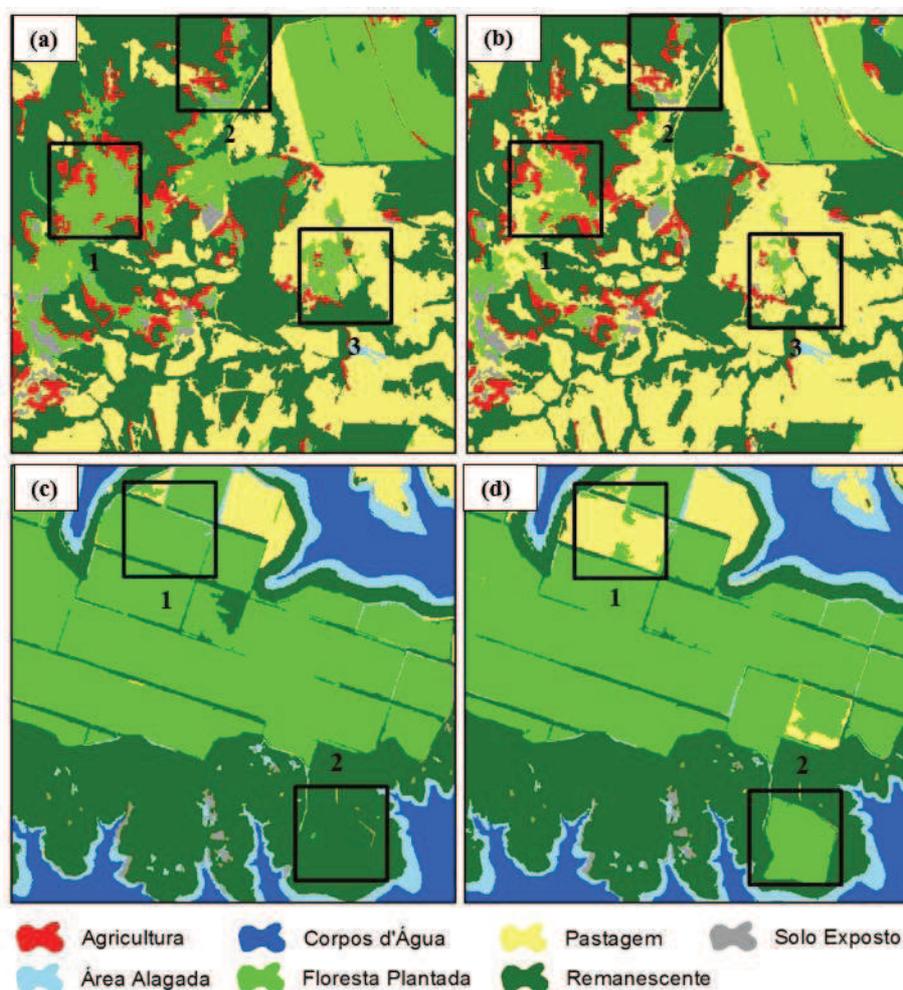


Figura 10 Comparação das classificações com atributos temporais que apresentaram os mesmos valores de acurácia global e Kappa, mas resultaram em classificações diferentes; (a) DT –Completa, (b) DT-Mensal, (c) SVM-Mensal, (d) SVM-Bimestral

Os resultados mostraram que as frequências de observações das séries temporais não interferem fortemente na acurácia das classificações, bem como no desempenho dos algoritmos, visto que não houve diferenças significativas entre o tipo de variáveis temporais (Completa, Mensal, Bimestral e Estacional)

utilizando o mesmo algoritmo (Tabela 6). Havendo apenas diferenças significativas quando comparada uma classificação com variáveis espectrais e outra com atributos temporais do mesmo algoritmo. A classificação RF com atributos temporais oriundos de uma série completa quando comparada com a classificação RF com apenas as variáveis espectrais da Rapideye apresentaram a menor diferença significativa ($X^2=7,78$), ou seja, mesmo não utilizando atributos temporais o RF obteve um bom desempenho, como discutido anteriormente.

Tabela 6 Comparações das classificações realizadas pelo teste McNemar com chi-quadrado tabulado $X^2=3,81$

	DT-C	DT-M	DT-B	DT-E	DT-VE	NB-C	NB-M	NB-B	NB-E	NB-VE
DT-C	*					*				
DT-M	0	*				0,03	*			
DT-B	0,04	0,16	*			0,06	0,16	*		
DT-E	0,10	0,10	0,51	*		0,01	0,06	0,03	*	
DT-VE	11,74	11,74	14,63	9,66	*	18,51	19,91	16,51	17,83	*

	RF-C	RF-M	RF-B	RF-E	RF-VE	SVM-C	SVM-M	SVM-B	SVM-E	SVM-VE
RF-C	*					*				
RF-M	0,06	*				0,03	*			
RF-B	0,01	0,03	*			0,03	0	*		
RF-E	0,16	0,03	0,10	*		0,10	0	0,23	*	
RF-VE	7,78	9,17	8,23	10,16	*	14,03	15,24	15,24	11,74	*

	DT-b	NB-b	RF-b	SVM-b
DT-b	*			
NB-b	0,06	*		
RF-b	0,77	1,25	*	
SVM-b	1,63	2,3	0,65	*

Nota: Em negrito estão os melhores resultados. (-C= Série temporal completa; -M= Série temporal mensal; -B= Série temporal bimestral; -E= Série temporal estacional; -VE= Variáveis espectrais da Rapideye; -b= Melhor classificação realizada pelo ML.A).

A maior diferença significativa entre classificações com e sem atributos temporais foi verificada entre as classificações NB com atributos temporais mensais e sem atributos temporais ($X^2=19,61$), comprovando que os atributos temporais potencializaram a precisão das classificações geradas pelo algoritmo NB. Avaliando as classificações com apenas variáveis espectrais executadas por diferentes classificadores, encontramos diferenças significativas entre os algoritmos NB e RF ($X^2=4,29$) e NB e SVM ($X^2=5,71$), isso se deve à baixa acurácia apresentada pelo NB. Apesar de o DT ter apresentado a classificação com a menor acurácia, quando usando apenas variáveis espectrais, tal classificação não apresentou diferença significativa com o RF ($X^2=2,29$), o SVM ($X^2=1,43$) e o NB ($X^2=0,77$). É importante ressaltar que o RF e o SVM apresentaram menor valor de chi-quadrado ($X^2=0,03$), sugerindo que tais classificações são semelhantes. Resultado similar também foi encontrado por Duro, Franklin e Dubé (2012). Ao compararmos o melhor resultado de cada algoritmo, não se obteve diferenças significativas entre as classificações usando o teste McNemar. Sendo assim, as melhores classificações de cada algoritmo (DT-Bimestral, NB-Mensal, RF-Estacional e SVM-Mensal) são estatisticamente semelhantes, porém ao compararmos visualmente, pode se verificar algumas diferenças nas classificações, como destacado na Figura 11, percebe-se que o DT gerou uma classificação mais dissimilar dentre as outras, pois confundiu áreas de Remanescentes com a Agricultura (Figura 11a – item 3) e áreas de Floresta Plantada com Remanescente (Figura 11a – item 4). A melhor classificação foi gerada pelo NB, que mesmo apresentando algumas confusões de áreas de Floresta Plantada com Pastagem (Figura 11b – item 2) não confundiu áreas Remanescentes com Floresta Plantada (Figura 11b – item 4), bem como o SVM (Figura 11d – item 1). O RF mostrou-se eficiente para classificar Florestas Plantadas e Remanescentes (Figura 11c – item 2-4), apresentando apenas algumas confusões em determinadas áreas como destacado na Figura 11c – item 1.

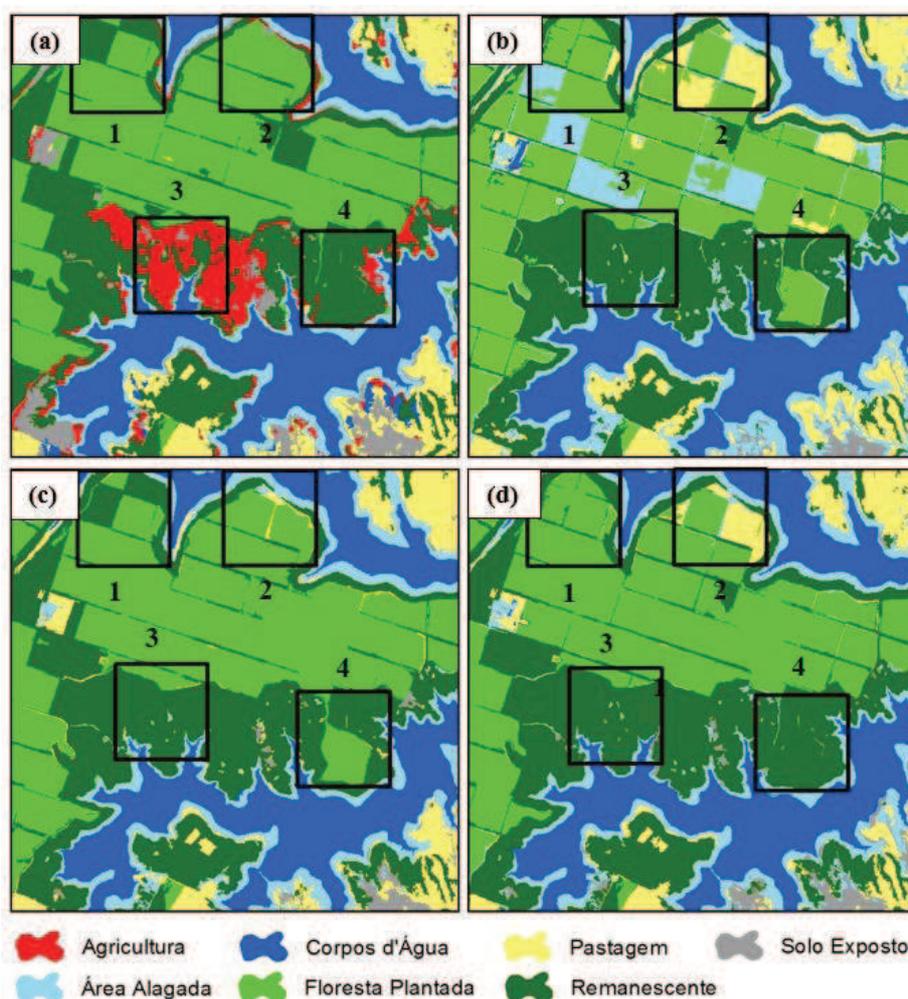


Figura 11 Comparação das melhores classificações: (a) DT-Bimestral, (b) NB-Mensal, (c) RF-Estacional e (d) SVM-Mensal; com destaque em algumas regiões que os algoritmos classificaram de forma diferente a cobertura da terra

Com intuito de avaliar o desempenho dos algoritmos por classe, foi calculado o erro de omissão e comissão a partir da matriz de confusão de cada classificação. Os erros de comissão mostram os pontos que foram indevidamente

incluídos em uma classe, já os erros de omissão mostram os pontos que pertencendo a uma determinada classe, não foram incluídos nela. A Figura 12 mostra os erros de omissão e comissão para cada classe da cobertura da terra originado a partir das matrizes de confusão provenientes das classificações para todos os algoritmos de aprendizagem de máquina. De modo geral, as menores percentagens de erros foram apresentadas pelos algoritmos SVM e RF, pois obtiveram menores erros de omissão e erros de comissão, como esperado, por apresentarem classificações mais precisas. A classe Agricultura não apresentou erro de comissão nas classificações geradas pelo SVM e RF utilizando os atributos temporais, todavia, ao utilizar apenas as variáveis espectrais o RF apresentou erro de comissão de 50%. O NB mostrou maior erro de comissão ao classificar a classe Área Alagada em todas as classificações (Erro de comissão variando entre 44,44% a 70,73%), mas o erro de omissão só apresentou maiores valores na classificação sem atributos temporais. Isso pode ter ocorrido porque a classe Área alagada apresenta uma sazonalidade natural que foi detectada utilizando os atributos temporais. As classes Corpos d'Água, Pastagem e Solo Exposto obtiveram erros de omissão e comissão menores com o uso o algoritmo SVM.

Como o foco desse trabalho é servir de subsídio para o mapeamento do CAR, é importante obter melhores classificações para a classe Remanescente. Tal classe obteve menores erros de comissão nas classificações oriundas do NB (Erro de comissão variando entre 0% a 28,57%), sendo o maior erro de comissão atrelado à classificação sem atributo temporal. Mas em se tratando de erro de omissão, o SVM se destacou (Erro de omissão variando entre 4,17% a 11,11%), sendo também o maior erro de comissão atrelado à classificação sem atributo temporal.

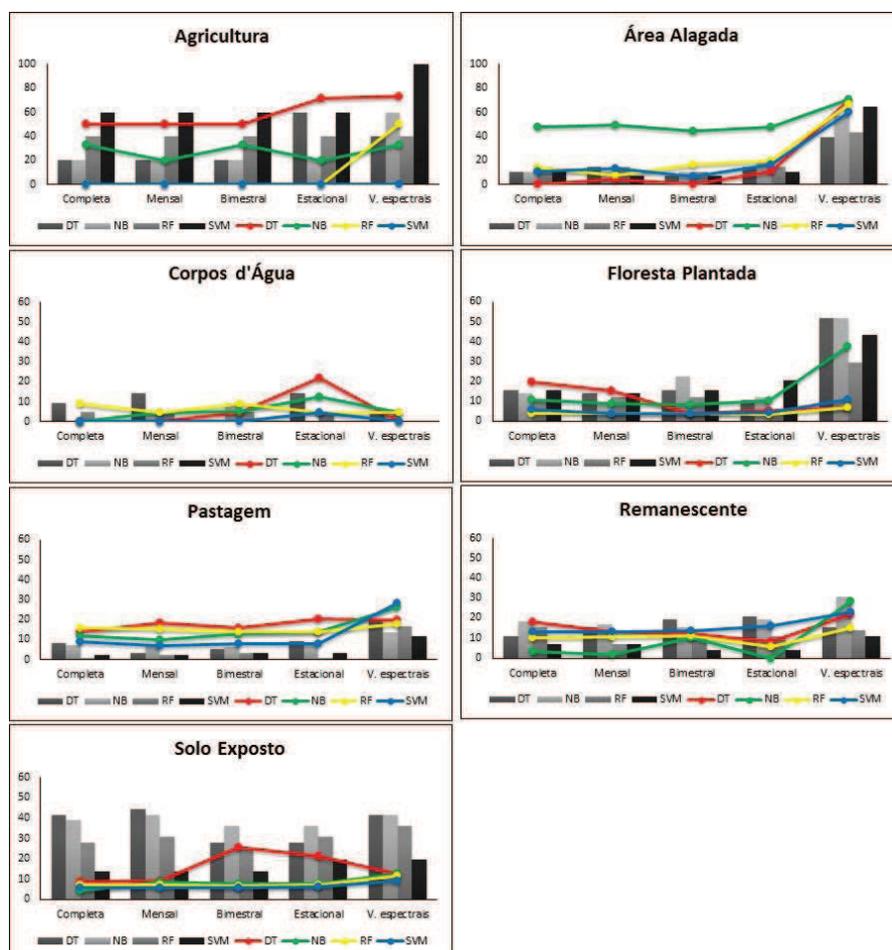


Figura 12 Erros de comissão e omissão provenientes das matrizes de confusão para cada classe da cobertura da terra

Legenda: Barras verticais representam o erro de omissão (%); Linhas com marcadores representam o erro comissão (%).

5.3 Analisando os atributos temporais na classificação

De forma geral, os atributos temporais tiveram um efeito significativo quando utilizados em conjuntos com variáveis espectrais. Mesmo utilizando imagens preditas para compor as séries que posteriormente foram processadas pelo pacote *greenbrown*, tais imagens foram úteis para derivar os parâmetros temporais utilizados na classificação, assim como relatado por Jia et al. (2014), Watts et al. (2011).

Os resultados encontrados mostram que o uso de dados temporais derivados do pacote *greenbrown* melhora a classificação da cobertura da terra, independente do algoritmo de aprendizagem de máquina utilizado. Bem como, o uso de dados originados de uma série temporal TM se mostrou eficiente para caracterizar objetos derivados de imagens de alta resolução espacial, como a Rapideye.

O uso de variáveis temporais é uma alternativa para potencializar a precisão das classificações. Zhu e Woodcock (2014) também utilizaram componentes gerados de modelos de séries temporais (média, padrões intra- e interanual, sazonalidade) como entradas para a classificação de cobertura do solo, e obtiveram uma acurácia global de 90,20% com o RF, resultado semelhante ao encontrados neste trabalho (90,48%).

Alguns trabalhos utilizaram imagens multitemporais e não apresentaram resultados positivos na classificação, como Carrão, Gonçalves e Caetano (2008), Langley, Cheshire e Humes (2001), porém em tais trabalhos os componentes fenológicos da cobertura da terra não foram detectados, diferentemente deste trabalho que conseguiu caracterizar usos da terra com fenologia característica como os Remanescentes de vegetação. O uso de imagens multidas também deve levar em consideração as características do ambiente, pois caso contrário, aspectos fenológicos não são captados na classificação. Por exemplo, no trabalho desenvolvido por Vieira et al. (2012) os autores utilizaram apenas quatro imagens TM num período de 2000 a 2001, todavia, essa amostragem

considerou as dinâmicas fenológicas e assim, apresentaram resultados satisfatórios no mapeamento da cana-de-açúcar.

Os resultados demonstraram que as classificações utilizando dados do pacote *greenbrown* de uma série temporal estacional, bimestral e completa, apresentaram resultados similares. Ao compararmos as classificações com diferentes atributos temporais e o mesmo algoritmo não houve diferença significativa em nenhuma classificação, e mesmo comparando as melhores classificações de cada algoritmo também não houve diferença significativa. No caso do SVM, tal algoritmo apresentou $X^2=0$ quando comparado os atributos temporais bimestral e estacional com os atributos temporais mensal. Desse modo, utilizar uma série temporal estacional em regiões onde há indisponibilidade de imagens TM é uma alternativa viável para mapeamentos e monitoramento da cobertura da terra com melhor precisão, pois o uso desses dados temporais acoplado a imagens de alta resolução espacial, como a Rapideye, garante classificações mais acuradas.

Os algoritmos DT e RF apresentam quais atributos foram usados no processo de classificação. No caso da série completa, ambos os algoritmos, usaram o desvio padrão da variabilidade intra- e interanual e a média, e somente o DT utilizou a variação sazonal. Já o RF também incorporou as variáveis de variabilidade intra- e interanual. Considerando atributos oriundos da série temporal mensal, o algoritmo DT utilizou os mesmos atributos temporais da série completa, já o RF incorporou um novo atributo, a variação sazonal, e não utilizou a variabilidade intra-anual. Utilizando os atributos da série temporal bimestral, o DT incorporou a variabilidade intra-anual e o RF incorporou novamente variabilidade intra-anual e não utilizou a variação sazonal. Em se tratando dos atributos estacionais, tanto o DT quanto o RF utilizaram os mesmos atributos usados na classificação da série bimestral. Portanto, dentre as variáveis

temporais geradas pelo pacote *greenbrown* apenas a tendência não foi utilizada como atributo na classificação desses algoritmos.

O componente sazonal oriundo de séries temporais de índices de vegetação pode representar mudanças na cobertura do solo, bem como respostas fenológicas (EASTMAN et al., 2013). A mudança sazonal relaciona-se com os padrões intra- e interanuais da fenologia. Desse modo, podem melhorar o mapeamento e monitoramento de muitos processos característicos da cobertura terrestre e que são importantes para a gestão de recursos naturais (VOGELMANN et al., 2012). Por isso, o uso de atributos sazonais na classificação potencializa a distinção das classes de cobertura da terra, principalmente aqueles vegetados, como Remanescentes e Florestas Plantadas. A separabilidade dessas classes é muito importante para gerar dados que irão alimentar o CAR, visto que representam condições diferenciadas quanto às restrições e possibilidades de alterações do uso do solo.

6 CONCLUSÕES

Neste trabalho, avaliou-se a efetividade do uso de variáveis temporais na classificação da cobertura da terra. A indisponibilidade de imagens TM para compor uma série temporal foi contornada utilizando o algoritmo STARFM para a predição de imagens TM sintéticas. Desse modo, foram construídas séries temporais com diferentes frequências de observação, para que assim fossem obtidas as variáveis temporais utilizadas como atributos nos MLA's. As principais conclusões que podem ser derivadas desse trabalho são:

- a) O uso de imagens sintéticas para compor séries temporais é válido, pois as características temporais da superfície terrestre foram detectadas nas análises das séries temporais pelo pacote *greenbrown*;
- b) O uso de parâmetros gerados a partir do pacote *greenbrown*, mesmo quando usando imagens TM sintéticas, proporcionaram uma melhoria no desempenho dos classificadores. Esta possibilidade representa um avanço para as regiões tropicais que apresentam pouca disponibilidade de imagens sem nuvens;
- c) Os desempenhos dos MLA's na classificação da cobertura da terra apresentaram bons resultados, com destaque para os algoritmos SVM e RF;
- d) Os atributos temporais melhoraram a acurácia global das classificações, bem como o coeficiente Kappa;
- e) Neste trabalho não foram encontradas diferenças significativas entre as classificações geradas utilizando atributos de diferentes frequências temporais (Completa, Mensal, Bimestral e Estacional), nem mesmo utilizando diferentes MLA's;

- f) A utilização de uma série temporal estacional é válida, pois mesmo apresentando poucas imagens por ano, podem ser obtidos resultados semelhantes aos de uma série temporal completa. Utilizando poucas imagens por ano para gerar atributos temporais, dependendo da região de estudo, não haverá necessidade de utilizar imagens sintéticas;
- g) O uso de atributos temporais melhora a separabilidade entre as classes, principalmente Remanescente e Floresta Plantada, que apresentam características espectrais similares, porém temporalmente são distintas;
- h) A classificação multitemporal da imagem Rapideye utilizando parâmetros extraídos de uma série temporal TM, se mostrou aplicável e promissora, visto que dados Rapideye não dispõem de imagens multidas suficientes;

Neste trabalho, alguns classificadores se mostraram mais precisos para classificar determinadas classes, enquanto que outros foram mais eficientes para classificar outras classes. Sendo assim, em trabalhos futuros pretende-se avaliar quais MLA's devem ser utilizados para separar classes específicas da cobertura do solo em um processo de classificação híbrida, ou seja, o uso de diferentes MLA's num mesmo processo de classificação visando melhorar a acurácia e robustez da classificação da superfície terrestre.

REFERÊNCIAS

- ADAM, E. et al. Land-use/cover classification in a heterogeneous coastal landscape using RapidEye imagery: evaluating the performance of random forest and support vector machines classifiers. **International Journal of Remote Sensing**, Basingstoke, v. 35, n. 10, p. 3440–3458, May 2014.
- ADELABU, S. et al. Exploiting machine learning algorithms for tree species classification in a semiarid woodland using RapidEye image. **Journal of Applied Remote Sensing**, Oxford, v. 7, n. 1, p. 073480, Nov. 2013.
- ALCANTARA, C. et al. Mapping abandoned agriculture with multi-temporal MODIS satellite data. **Remote Sensing of Environment**, New York, v. 124, p. 334–347, Sept. 2012.
- ANDERSON, L. O.; SHIMABUKURO, Y. E. Classificação espectro-temporal de formações vegetais. In: RUDORFF, B. F. T.; SHIMABUKURO, Y. E.; CEBALLOS, J. C. (Ed.). **O sensor MODIS e suas aplicações ambientais no Brasil**. São José dos Campos: Parêntese, 2007. p. 159-171.
- BHANDARI, S.; PHINN, S.; GILL, T. Preparing landsat image time series (LITS) for monitoring changes in vegetation phenology in Queensland, Australia. **Remote Sensing**, Lower Hutt, v. 4, n. 12, p. 1856–1886, June 2012.
- BISHOP, C. M. **Pattern recognition and machine learning**. Cambridge: Springer, 2006.
- BLACKBRIDGE DELIVERING THE WORLD. **Satellite imagery product specifications**: version 6.0.[S.l.: s.n.], 2013.
- BRASIL. Lei nº 12.651, de 25 de maio de 2012. Dispõe sobre a proteção da vegetação nativa; altera as Leis nos 6.938, de 31 de agosto de 1981, 9.393, de 19 de dezembro de 1996, e 11.428, de 22 de dezembro de 2006; revoga as Leis nos 4.771, de 15 de setembro de 1965, e 7.754, de 14 de abril de 1989, e a Medida Provisória no 2.166-67, de 24 de agosto de 2001; e dá outras providências. **Diário Oficial da União**, Brasília, 28 maio 2012. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2012/lei/112651.htm>. Acesso em: 08 maio 2014.
- BREIMAN, L. Random forests. **Machine Learning**, Boston, v. 45, n. 1, p. 5–32, 2001.

BRESEE, M. K. et al. Disturbance and landscape dynamics in the Chequamegon National Forest Wisconsin, USA, from 1972 to 2001. **Landscape Ecology**, Dordrecht, v. 19, n. 3, p. 291–309, May 2004.

BRUZZONE, L.; COSSU, R.; VERNAZZA, G. Combining parametric and non-parametric algorithms for a partially unsupervised classification of multitemporal remote-sensing images. **Information Fusion**, Amsterdam, v. 3, n. 4, p. 289–297, Dec. 2002.

BRUZZONE, L.; SMITS, P. C.; TILTON, J. C. Foreword special issue on analysis of multitemporal remote sensing images. **IEEE Transactions on Geoscience and Remote Sensing**, New York, v. 41, n. 11, p. 2419–2422, Nov. 2003.

CARRÃO, H.; GONÇALVES, P.; CAETANO, M. Contribution of multispectral and multitemporal information from MODIS images to land cover classification. **Remote Sensing of Environment**, New York, v. 112, n. 3, p. 986–997, Mar. 2008.

CARVALHO, L. M. T. de. Detecção de modificações na cobertura do solo. In: CARVALHO, L. M. T.; SCOLFORO, J. R. **Inventário florestal de Minas Gerais: monitoramento da flora nativa 2005 – 2007**. Lavras: Editora da UFLA, 2008. p. 2–24.

COHEN, W. B.; GOWARD, S. N. Landsat's role in ecological applications of remote sensing. **Bioscience**, Washington, v. 54, n. 6, p. 535–545, 2004.

COLSTOUN, E. C. B. de et al. National Park vegetation mapping using multitemporal Landsat 7 data and a decision tree classifier. **Remote Sensing of Environment**, New York, v. 85, n.3, p. 316–327, May 2003.

CONESE, C.; MASELLI, F. Use of multitemporal information to improve classification performance of TM scenes in complex terrain. **ISPRS Journal of Photogrammetry and Remote Sensing**, Amsterdam, v. 46, n. 4, p. 187–197, Aug. 1991.

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

CRACKNELL, M. J.; READING, A. M. Geological mapping using remote sensing data: a comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information. **Computers & Geosciences**, New York, v. 63, p. 22–33, Feb. 2014.

CRUZ, R. C. **Uso de funções de transferência e da transformada Wavelet aplicadas a sinais meteorológicos da Floresta Nacional de Caxiuanã**. Pará: Universidade Federal do Pará, 2008.

DURO, D. C.; FRANKLIN, S. E.; DUBÉ, M. G. A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. **Remote Sensing of Environment**, New York, v. 118, p. 259–272, Mar. 2012.

EASTMAN, J. et al. Global trends in seasonality of normalized difference vegetation index (NDVI), 1982–2011. **Remote Sensing**, Lower Hutt, v. 5, n. 10, p. 4799–4818, Sept. 2013.

ECOGNITION DEVELOPER. Using e Cognition® developer in quickmap mode. In: ECOGNITION DEVELOPER. **User guide**. München: Trimble Germany GmbH, 2011. p. 27-39.

FIGUEIREDO, D. **Conceitos básicos de sensoriamento remoto**. Brasília: CONAB, 2005. Disponível em: <http://www.conab.gov.br/conabweb/download/SIGABRASIL/manuais/conceitos_sm.pdf>. Acesso em: 10 set. 2011.

FOODY, G. M. Thematic Map Comparison : Evaluating the statistical significance of differences in classification accuracy. **Photogrammetric Engineering & Remote Sensing**, Falls Church, v. 70, n. 5, p. 627–633, 2004.

FOODY, G. M. et al. Training set size requirements for the classification of a specific class. **Remote Sensing of Environment**, New York, v. 104, n. 1, p. 1–14, Sept. 2006.

FORKEL, M. et al. Trend change detection in NDVI time series: effects of inter-annual variability and methodology. **Remote Sensing**, Lower Hutt, v. 5, n. 5, p. 2113–2144, May 2013.

FRIEDL, M. A.; BRODLEY, C. E. Decision tree classification of land cover from remotely sensed data. **Remote Sensing of Environment**, New York, v. 61, n.3, p. 399–409, Sept. 1997.

GALFORD, G. L. et al. Wavelet analysis of MODIS time series to detect expansion and intensification of row-crop agriculture in Brazil. **Remote Sensing of Environment**, New York, v. 112, n. 2, p. 576–587, Feb. 2008a.

GALFORD, G. L. et al. Wavelet analysis of MODIS time series to detect expansion and intensification of row-crop agriculture in Brazil. **Remote Sensing of Environment**, New York, v. 112, n. 2, p. 576–587, Feb. 2008b.

GAMON, J. A. et al. Relationships between NDVI, canopy structure, and photosynthesis in three Californian vegetation types. **Ecological Applications**, Tempe, v. 5, n. 1, p. 28–41, Feb. 1995.

GAO, F. et al. On the blending of the landsat and MODIS surface reflectance: predicting daily Landsat surface reflectance. **IEEE Transactions on Geoscience and Remote Sensing**, New York, v. 44, n. 8, p. 2207–2218, Aug. 2006.

GITAS, I. Z. Burned area mapping and post-fire monitoring of a Mediterranean forest using NDVI time-series of low resolution imagery and the BFAST method / Mapeamento de área queimada e monitoramento pós-fogo de floresta na região Mediterrânea a partir de série tempor. **Ambiência**, Guarapuava, v. 8, n. 4, p. 461–470, Oct. 2012.

GOMIDE, L. R. et al. Sustainable forest management of native vegetation remnants in Brazil. In: MARTÍN-GARCIA, J.; DIEZ, J. J. **Sustainable forest management: case studies**. Croatia: Intech, 2012. p. 77–96.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The elements of statistical learning: data mining, inference, and prediction**. 2. ed. New York: Springer, 2009.

HAY, G. J.; CASTILLA, G. Geographic object-based image analysis (GEOBIA). In: BLASCHKE, T.; LANG, S.; HAY, G. J. (Ed.). **Object-based image analysis: spatial concepts for knowledge-driven remote sensing applications**, New York: Springer, 2008. Chap. 1, p. 81–92.

HE, L. et al. Normalized algorithm for mapping and dating forest disturbances and regrowth for the United States. **International Journal of Applied Earth Observation and Geoinformation**, Enschede, v. 13, n. 2, p. 236–245, Apr. 2011.

HILKER, T. et al. A new data fusion model for high spatial- and temporal-resolution mapping of forest disturbance based on Landsat and MODIS. **Remote Sensing of Environment**, New York, v. 113, n. 8, p. 1613–1627, Aug. 2009a.

HILKER, T. et al. Generation of dense time series synthetic landsat data through data blending with MODIS using a spatial and temporal adaptive reflectance fusion model. **Remote Sensing of Environment**, New York, v. 113, n. 9, p. 1988–1999, Sept. 2009b.

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, June 2013a.

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, June 2013b.

IBÁÑEZ, A.; BIELZA, C.; LARRAÑAGA, P. Cost-sensitive selective naive Bayes classifiers for predicting the increase of the h-index for scientific journals. **Neurocomputing**, Amsterdam, v. 135, p. 42–52, July 2014.

JACQUIN, A.; SHEEREN, D.; LACOMBE, J.-P. Vegetation cover degradation assessment in Madagascar savanna based on trend analysis of MODIS NDVI time series. **International Journal of Applied Earth Observation and Geoinformation**, Enschede, v. 12, suppl. 1, p. S3–S10, Feb. 2010.

JIA, K. et al. Land cover classification of finer resolution remote sensing data integrating temporal features from time series coarser resolution data. **ISPRS Journal of Photogrammetry and Remote Sensing**, Amsterdam, v. 93, p. 49–55, July 2014.

KENNEDY, R. E. et al. Spatial and temporal patterns of forest disturbance and regrowth within the area of the Northwest Forest Plan. **Remote Sensing of Environment**, New York, v. 122, p. 117–133, July 2012.

LANGLEY, S. K.; CHESHIRE, H. M.; HUMES, K. S. A comparison of single date and multitemporal satellite image classifications in a semi-arid grassland. **Journal of Arid Environments**, London, v. 49, n. 2, p. 401–411, Oct. 2001.

LI, F. et al. Mapping large-scale distribution and changes of aquatic vegetation in Honghu Lake, China, using multitemporal satellite imagery. **Journal of Applied Remote Sensing**, Oxford, v. 7, 2013.

MALATESTA, L. et al. Vegetation mapping from high-resolution satellite images in the heterogeneous arid environments of Socotra Island (Yemen). **Journal of Applied Remote Sensing**, Oxford, v. 7, n. 1, p. 073527, 2 Aug. 2013.

MALILA, W. A. Change vector analysis : an approach for detecting forest changes with landsat. In: SYMPOSIUM MACHINE PROCESSING OF REMOTELY SENSED DATA. **Proceedings...** Indiana: IEEE, 1980.

MARTÍNEZ, B. et al. Characterizing land condition variability in Ferlo, Senegal (2001–2009) using multi-temporal 1-km Apparent Green Cover (AGC) SPOT vegetation data. **Global and Planetary Change**, Amsterdam, v. 76, n. 3-4, p. 152–165, Apr. 2011.

MARTÍNEZ, B.; GILABERT, M. A. Vegetation dynamics from NDVI time series analysis using the wavelet transform. **Remote Sensing of Environment**, New York, v. 113, n. 9, p. 1823–1842, Sept. 2009.

MAS, J.-F.; GAO, Y.; PACHECO, J. A. N. Sensitivity of landscape pattern metrics to classification approaches. **Forest Ecology and Management**, Amsterdam, v. 259, n. 7, p. 1215–1224, Mar. 2010.

MASEK, J. Automated registration and orthorectification package for Landsat and Landsat-like data processing. **Journal of Applied Remote Sensing**, Oxford, v. 3, n. 1, p. 033515, Mar. 2009.

MELLO, M. P. et al. **STARS** : a new method for multitemporal remote sensing. São José dos Campos: Inpe, 2012.

MENESES, P. R. et al. **Introdução ao processamento de imagens de sensoriamento remoto**. Brasília: Editora da UnB, 2012.

MINISTÉRIO DO MEIO AMBIENTE. **Monitoramento do desmatamento nos biomas brasileiros por satélite: monitoramento do bioma cerrado 2009-2010**. Brasília: IBAMA, 2011.

MORTON, D. C. et al. Rapid assessment of annual deforestation in the Brazilian Amazon using MODIS data. **Earth Interactions**, London, v. 9, n. 8, p. 1–22, June 2005.

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, May 2011.

MUCHONEY, D.; WILLIAMSON, J. A gaussian adaptive resonance theory neural network classification algorithm applied to supervised land cover mapping using multitemporal vegetation index data. **Transactions on Geoscience and Remote Sensing**, New York, v. 39, n. 9, p. 1969–1977, 2001.

OLIVEIRA, T. et al. NDVI Time Series for Mapping Phenological Variability of Forests Across the Cerrado Biome. In: MONTGOMERY, R. A. **Phenology and Climate change**. Minnesota: Department of Forest Resources, 2012. p. 253–272.

OLIVEIRA, T. de A.; CARVALHO, L. de. Mapping deciduous forests by using time series of filtered MODIS NDVI and neural networks. **Cerne**, Lavras, v. 16, n. 2, p. 123–130, abr./jun. 2010.

OLIVEIRA-FILHO, A. T. de. Definição e delimitação de domínios e subdomínios das paisagens naturais do estado de Minas Gerais. In: SCOLFORO, J. R.; CARVALHO, L. M. T. (Ed.). **Mapeamento e inventário da flora e dos reflorestamentos de Minas Gerais**. Lavras: Editora da UFLA, 2006. Cap. 2, p. 21–35.

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**, Enschede, v. 12, Supl. 1, p. S27–S31, fev. 2010.

PAL, M.; MATHER, P. M. Support vector machines for classification in remote sensing. **International Journal of Remote Sensing**, Basingstoke, v. 26, n. 5, p. 1007–1011, Mar. 2005.

PEÑA, J. et al. Object-based image classification of summer crops with machine learning methods. **Remote Sensing**, Lower Hutt, v. 6, n. 6, p. 5019–5041, 30 May 2014.

PRADHAN, B. A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. **Computers & Geosciences**, New York, v. 51, p. 350–365, Feb. 2013.

PUISSANT, A.; ROUGIER, S.; STUMPF, A. Object-oriented mapping of urban trees using Random Forest classifiers. **International Journal of Applied Earth Observation and Geoinformation**, Enschede, v. 26, p. 235–245, Feb. 2014.

PUNIA, M.; JOSHI, P. K.; PORWAL, M. C. Decision tree classification of land use land cover for Delhi, India using IRS-P6 AWiFS data. **Expert Systems with Applications**, New York, v. 38, n. 5, p. 5577–5583, May 2011.

R CORE TEAM. **R**: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2013. Disponível em: <<http://www.R-project.org/>>. Acesso em: 10 dez. 2013.

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, May 2008.

RUDORFF, C. M. et al. Classificação de soja pelo método de superfícies de resposta espectro-temporal. In: RUDORFF, B. F. T.; SHIMABUKURO, Y. E.; CEBALLOS, J. C. (Org.). **O sensor modis e suas aplicações ambientais no BRASIL**. São José dos Campos: [s.n.], 2004. p. 111–121.

SHAO, Y.; LUNETTA, R. S. Comparison of support vector machine, neural network, and CART algorithms for the land-cover classification using limited training data points. **ISPRS Journal of Photogrammetry and Remote Sensing**, Amsterdam, v. 70, p. 78–87, June 2012.

SHIRAISHI, T. et al. Comparative assessment of supervised classifiers for land use – land cover classification in a tropical region using time-series PALSAR mosaic data. **IEEE Geoscience & Remote Sensing Society**, Canadá, v. 7, n. 4, p. 1186–1199, Apr. 2014.

SONNENSCHNEIN, R. et al. Differences in landsat-based trend analyses in drylands due to the choice of vegetation estimate. **Remote Sensing of Environment**, New York, v. 115, n. 6, p. 1408–1420, June 2011.

STELLMES, M. et al. Dryland observation at local and regional scale: comparison of landsat TM/ETM+ and NOAA AVHRR time series. **Remote Sensing of Environment**, New York, v. 114, n. 10, p. 2111–2125, Oct. 2010.

SU, M.-C. et al. Mapping multi-spectral remote sensing images using rule extraction approach. **Expert Systems with Applications**, New York, v. 38, n. 10, p. 12917–12922, Sept. 2011.

THOMPSON, M. et al. Mapping grazing-induced degradation in a semi-arid environment: a rapid and cost effective approach for assessment and monitoring. **Environmental Management**, New York, v. 43, n. 4, p. 585–96, Apr. 2009.

TIAN, F. et al. Mapping and evaluation of NDVI trends from synthetic time series obtained by blending landsat and MODIS data around a coalfield on the loess plateau. **Remote Sensing**, Lower Hutt, v. 5, n. 9, p. 4255–4279, 2 set. 2013.

TIGGES, J.; LAKES, T.; HOSTERT, P. Urban vegetation classification: Benefits of multitemporal RapidEye satellite data. **Remote Sensing of Environment**, New York, v. 136, p. 66–75, Sept. 2013.

VAPNIK, V. N. **Statistical learning theory**. New York: Wiley-Interscience, 1998.

VERBESSELT, J. et al. Detecting trend and seasonal changes in satellite image time series. **Remote Sensing of Environment**, New York, v. 114, n. 1, p. 106–115, Jan. 2010a.

VERBESSELT, J. et al. Phenological change detection while accounting for abrupt and gradual trends in satellite image time series. **Remote Sensing of Environment**, New York, v. 114, n. 12, p. 2970–2980, Dec. 2010b.

VERBESSELT, J.; ZEILEIS, A.; HEROLD, M. Near real-time disturbance detection using satellite image time series. **Remote Sensing of Environment**, New York, v. 123, p. 98–108, ago. 2012.

VIEIRA, C.; MATHER, P.; MCCULLAGH, M. The spectral-temporal response surface and its use in the multi-sensor, multi-temporal classification of agricultural crops. **International Archives of Photogrammetry and Remote Sensing**, Amsterdam, v. 33, n. 1980, p. 582–589, 2000.

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, Aug. 2012.

VOGELMANN, J. E. et al. Monitoring gradual ecosystem change using Landsat time series analyses: Case studies in selected forest and rangeland ecosystems. **Remote Sensing of Environment**, New York, v. 122, p. 92–105, July 2012.

WALKER, J. J. et al. Evaluation of landsat and MODIS data fusion products for analysis of dryland forest phenology. **Remote Sensing of Environment**, New York, v. 117, p. 381–393, Feb. 2012.

WATTS, J. D. et al. Improved classification of conservation tillage adoption using high temporal and synthetic satellite imagery. **Remote Sensing of Environment**, New York, v. 115, n. 1, p. 66–75, Jan. 2011.

WHITESIDE, T. G.; BOGGS, G. S.; MAIER, S. W. Comparing object-based and pixel-based classifications for mapping savannas. **International Journal of Applied Earth Observation and Geoinformation**, Enschede, v. 13, n. 6, p. 884–893, Dec. 2011.

WIELAND, M.; PITTORE, M. Performance Evaluation of Machine Learning Algorithms for Urban Pattern Recognition from Multi-spectral Satellite Images. **Remote Sensing**, Lower Hutt, v. 6, n. 4, p. 2912–2939, Mar. 2014.

WITTEN, I. H.; FRANK, E.; HALL, M. A. **Data mining: practical machine learning tools and techniques**. 3.ed. Burlington: Morgan Kaufmann, 2005.

XIE, Y.; SHA, Z.; YU, M. Remote sensing imagery in vegetation mapping: a review. **Journal of Plant Ecology**, Oxford, v. 1, n. 1, p. 9–23, Apr. 2008a.

XIE, Y.; SHA, Z.; YU, M. Remote sensing imagery in vegetation mapping: a review. **Journal of Plant Ecology**, Oxford, v. 1, n. 1, p. 9–23, Apr. 2008b.

XIN, Q. et al. Toward near real-time monitoring of forest disturbance by fusion of MODIS and Landsat data. **Remote Sensing of Environment**, New York, v. 135, p. 234–247, Aug. 2013.

ZHU, Z.; WOODCOCK, C. E. Continuous change detection and classification of land cover using all available Landsat data. **Remote Sensing of Environment**, New York, v. 144, p. 152–171, Mar. 2014.

ZHU, Z.; WOODCOCK, C. E.; OLOFSSON, P. Continuous monitoring of forest disturbance using all available landsat imagery. **Remote Sensing of Environment**, New York, v. 122, p. 75–91, July 2012.