



**CAROLINA GUSMÃO SOUZA**

**USO DE SÉRIES TEMPORAIS PARA O  
MAPEAMENTO DA CAFEICULTURA**

**LAVRAS – MG**

**2015**

**CAROLINA GUSMÃO SOUZA**

**USO DE SÉRIES TEMPORAIS PARA O MAPEAMENTO DA  
CAFEICULTURA**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Engenharia Florestal, área de concentração em Manejo Florestal, para a obtenção do título de Doutor.

Orientador

Dr. Luis Marcelo Tavares de Carvalho

**LAVRAS – MG**

**2015**

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

Souza, Carolina Gusmão.

Uso de séries temporais para o mapeamento da cafeicultura /  
Carolina Gusmão Souza. – Lavras : UFLA, 2015.  
162 p.

Tese(doutorado)–Universidade Federal de Lavras, 2015.  
Orientador(a): Luis Marcelo Tavares de Carvalho.  
Bibliografia.

1. Café. 2. Classificação de imagens. 3. Sensoriamento remoto. I.  
Universidade Federal de Lavras. II. Título.

**CAROLINA GUSMÃO SOUZA**

**USO DE SÉRIES TEMPORAIS PARA O MAPEAMENTO DA  
CAFEICULTURA**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Engenharia Florestal, área de concentração em Manejo Florestal, para a obtenção do título de Doutor.

APROVADA em 24 de fevereiro de 2015.

Dra. Margarete Marin Lodelo Volpato	EPAMIG
Dra. Helena Maria Ramos Alves	EMBRAPA
Dr. Luicano Teixeira de Oliveira	UFLA
Dra. Ludmila Zambaldi Suzuki	UFLA

Dr. Luis Marcelo Tavares de Carvalho  
Orientador

**LAVRAS – MG**  
**2015**

“Mesmo quando tudo parece  
desabar, cabe a mim decidir entre  
rir ou chorar, ir ou ficar, desistir ou  
lutar; porque descobri, no caminho  
incerto da vida, que o mais  
importante é decidir.”

Cora Coralina

Ao meu Senhor Jesus, que esteve comigo, me fortalecendo e capacitando em todo o tempo,

OFEREÇO

À minha família; minha mãe, Silene; minha avó Vina, meus irmãos, Juliana e Marcus e os meus sobrinhos, Ester, Raquel, Júlia, Vicente, Davi e Filipe, como eu amo vocês. E, em especial, ao meu amado esposo, Thiago. Você é minha inspiração e o meu orgulho.

Com muito amor,

DEDICO

## AGRADECIMENTOS

Em primeiro lugar, quero agradecer Àquele que sempre está com Sua mão sobre a minha vida, que me capacita e me dá forças para continuar, o meu Senhor Jesus.

Gostaria de agradecer também, à Universidade Federal de Lavras, ao Programa de Pós-Graduação em Engenharia Florestal, a todos os professores e, em especial, ao Laboratório de Estudos e Projetos em Manejo Florestal – LEMAF. Ao meu orientador, Luis Marcelo (Passarinho), pela paciência, disposição e boa vontade em me orientar e, principalmente, pelos ensinamentos. Eu aprendi muito contigo.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES, pela bolsa de estudos concedida.

Aos membros da banca, pela disponibilidade e por todas as sugestões para enriquecerem o trabalho.

Às pesquisadoras Rosângela Borém, Margarete Volpato, Helena Alves e Tatiana Vieira, pelo carinho com que me receberam desde o início do mestrado e também por me apresentarem este complexo, porém, delicioso mundo do mapeamento do café.

Gostaria de agradecer aos meus queridos colegas de laboratório, Célio, Lisiane, Nathália, Guilherme, Carol Nunes, Lucília e Daniel. Vocês, além de colegas, foram amigos, parceiros e tornaram os meus dias mais felizes. Obrigada pelo companheirismo diário e convivência harmoniosa. Às minhas queridas ICs, Query e Tisa, vocês foram os anjos dos processamentos de imagens (rsrsrs). Obrigada pela ajuda e pela dedicação de cada uma. Em especial, quero agradecer às minhas pimpolhas, Polly e Tassinha, pelo companheirismo, pelos muitos sábados e por algumas noites. Este trabalho também é de você; nossa parceria rendeu bons frutos.

Especialmente à minha família, que abriu mão da minha presença para que eu pudesse realizar um sonho e para que eu pudesse crescer como pessoa e como profissional. Obrigada por sonharem junto comigo e por serem a minha base. Eu AMO todos vocês!!!!

Ao meu amoroso esposo, por acreditar em mim, por me dar força para chegar até aqui. Você é uma inspiração para mim. Obrigada por entender que, por muitas vezes, não pude estar contigo, por me fazer crescer e, acima de tudo, pelo seu amor. Com você sou uma pessoa melhor. Também à família Souza, obrigada pelo cuidado, amor e carinho.

À minha, mais que especial, “família lavrense”, obrigada por fazerem a minha vida divertida e mais feliz. Nivaldo, Eduardo, Manuela, Fabiana, Simone, Amanda, Silvia, Rogério, Felipe, Ana, Isabel, Isaac, Jessé, Tida, Lara, Wesley, Lud, Elisa, Andressa, Adriane, Sheila, Josy, Guilherme, Cinthia, Patrícia, Natália e Livinha, vocês tornaram a minha caminhada mais amena. Cada passo dado junto com vocês me ajudou a concluir esta etapa. Vocês são muito especiais!!

Obrigada aos meus amigos que, mesmo de longe, sempre estiveram comigo e acreditaram no meu potencial. Amo cada um de forma ímpar e, não importa onde eu esteja, vocês sempre estarão em meu coração.

Ao Liliano e Geovani, cuja ajuda foi fundamental quando os softwares não ajudavam. À Ludmilla Zambaldi, pelas conversas e ideias compartilhadas.

Para todos aqueles que, direta ou indiretamente, contribuíram para esta vitória, o meu muito obrigada!!!!!!

## RESUMO

A cafeicultura representa uma das principais atividades agrícolas, com grande importância no Brasil e no mundo, sendo o estado de Minas Gerais o maior produtor de café do país. Estimar os dados básicos desta cultura corretamente é um desafio, uma vez que as informações obtidas são pouco detalhadas e o setor ainda é carente de dados precisos. As geotecnologias têm sido promissoras para suprir esta lacuna, avaliando de forma mais precisa a dinâmica da cafeicultura. Porém, o mapeamento dessas áreas ainda é uma tarefa difícil, uma vez que elas são muito complexas de serem mapeadas, apresentando uma alta confusão entre os alvos. Para suprir esta necessidade, este trabalho foi realizado com o objetivo geral de propor uma metodologia para o mapeamento da cafeicultura, por meio de variáveis multiespectrais e multitemporais. O estudo foi conduzido em duas áreas distintas do estado de Minas Gerais, uma na região sul e a outra na região centro-oeste. Primeiramente, foram realizadas classificações, utilizando imagens de alta resolução do satélite RapidEye, testando diferentes algoritmos de aprendizagem de máquina e a combinação de diferentes variáveis (espectrais, geométricas e texturais) no processo de classificação. Os resultados mostraram que o algoritmo *Support Vector Machine* obteve os melhores resultados nas classificações para todas as áreas, com acurácia global de 88,33%. As variáveis texturais, quando associadas às espectrais, melhoraram a acurácia da classificação, porém, não houve diferença significativa entre as classificações. Apesar de os resultados terem se mostrado com bons índices de acerto, ainda houve muita confusão entre as classes. Foi proposto um novo método de mapeamento, utilizando dados multitemporais como variáveis no processo de classificação. Os resultados mostraram que os índices de acerto utilizando as variáveis multitemporais, integrados a variáveis espectrais, apresentaram índices de acurácia global de 93,00% e diminuíram significativamente a confusão entre os alvos, tornando o processo de classificação mais preciso. A metodologia proposta neste estudo mostrou eficiência no mapeamento de áreas cafeeiras.

Palavras-chave: Café. Classificação de imagens. Sensoriamento remoto. Algoritmos de aprendizagem de máquina. Greenbown.

## ABSTRACT

Coffee is one of the main agricultural activities, with great importance in Brazil and in the world, being the State of Minas the largest coffee producer in the country. Estimate the basic data from this culture correctly is a challenge, once obtaining such information have little detail and the sector is still missing accurate information. Geotechnologies has been promising to fill this gap, evaluating more correctly the dynamics of coffee. However, the mapping of these areas is still a difficult task, since these areas are too complex to map, presenting a high confusion among the targets. To meet this need, the goal this study was to propose a methodology for mapping of coffee, by multispectral and multi-temporal variables. The study was conducted in two distinct areas, which are located in the State of Minas Gerais, the first one in the South region and the second in the Midwest region of the State. Firstly, classifications was performed, using high-resolution satellite imagery RapidEye, testing different machine learning algorithms and the combination of different variables (spectral, geometrical and textural) in the classification process. The results showed that the Suport Vector Machine algorithm achieved the best results in the rankings for all areas, with overall accuracy of 88.33%. The textural variables when associated with spectral, improved a little accuracy, however, there was not significant difference when the ratings were compared. Although the results have been shown with good levels of accuracy, yet there was much confusion between classes. To overcome this gap, we proposed a new method for mapping using data as variables multi-temporal in the classification process. The results showed that using the multi-temporal variables, integrated the spectral variables, obtained overall accuracy of 93% and reduced significantly the confusion among the targets, making more precise classification process. The methodology proposed in this study was efficient to map coffee areas.

Keyword: Coffee. Image Classification. Remote sensing. Machine learning algorithms. Greenbrown.

## LISTA DE SIGLAS

AAT	Annual Aggregated Time Series
AM	Aprendizado de máquina
BFAST	Breaks For Additive Seasonal and Trend
BRDF	Bidirectional Reflectance Distribution Function
CDR	Surface Reflectance Climate Data Record
CONAB	Companhia Nacional de Abastecimento
DT	Decision Tree
ETM+	Enhanced Thematic Mapper Plus
EVI	Enhanced Vegetation Index
HANTS	Harmonic Analysis of Time Series
IAV	Inter-Annual variability
IBGE	Instituto Brasileiro de Geografia e Estatística
KNN	K- Nearest Neighbor
LAI	Leaf Area Index
LandTrendr	Landsat-based detection of Trends in Disturbance and recovery
LEDAPS	The Landsat Ecosystem Disturbance Adaptive Processing System
LOESS	LOcally wEighted regreSion Smoother
MAXVER	Máxima Verossimilhança
MSS	Multispectral Scanner System
MODIS	Moderate Resolution Imaging Spectro-radiometer
NASA	National Aeronautics and Space Administration
NB	Naive Bayes
NDVI	Normalized Difference Vegetation Index
NIR	Banda do infravermelho próximo
OLI	Operational Land Imager

PCA	Principal Component Analysis
RBV	Return Beam Vidicon Camera
RED	Banda do vermelho
RF	Random Forest
SD	Standard Deviation
STARFM	Spatial and Temporal Adaptive Reflectance Fusion Model
STARS	Spectral-Temporal Analysis by Response Surface
STL	Seasonal-Trend Decomposition
STV	Short Term Variability
SVM	Support Vector Machine
TIRS	Thermal Infrared Sensor
TiSeg	Time-series Generator
TM	Thematic Mapper
USGS	United States Geological Survey
UTM	Universal Transversa de Mercator
6 S	Second Simulation of Satellite Signal in the Solar Spectrum

## SUMÁRIO

	<b>PRIMEIRA PARTE</b> .....	13
	Introdução geral .....	13
<b>1</b>	<b>INTRODUÇÃO</b> .....	13
1.1	Motivação .....	16
1.2	Objetivo geral.....	17
1.3	Contribuições .....	17
1.4	Organização da tese.....	17
	<b>REFERÊNCIAS</b> .....	20
<b>2</b>	<b>REFERENCIAL TEÓRICO</b> .....	23
2.1	Cultura cafeeira e sua importância.....	23
2.2	Mapeamento da cafeeicultura .....	24
2.3	Séries temporais.....	29
2.3.1	<b>BFAST</b> .....	32
2.3.2	<b>Greenbrown</b> .....	34
2.4	Fusão de imagens.....	38
2.4.1	<b>STARFM</b> .....	39
2.5	Dados de sensoriamento remoto.....	43
2.5.1	<b>RapidEye</b> .....	44
2.5.2	<b>Landsat</b> .....	45
2.5.3	<b>MODIS</b> .....	47
2.6	Métodos de classificação de imagens.....	48
2.6.1	Algoritmos de aprendizagem de máquina .....	50
2.6.1.1	Árvore de decisão.....	51
2.6.1.2	<i>Naive bayes</i> .....	53
2.6.1.3	<i>K-nearest neighbor</i> .....	55
2.6.1.4	<i>Random forest</i> .....	57
2.6.1.5	<i>Support vector machines</i> .....	58
2.7	Aplicações para cafeeicultura.....	60
	<b>REFERÊNCIAS</b> .....	62
	<b>SEGUNDA PARTE</b>	
	<b>ARTIGO 1</b> Diferentes algoritmos de aprendizagem de máquina e variáveis de sensoriamento remoto para o mapeamento da cafeeicultura.....	75
	<b>ARTIGO 2</b> Variáveis multitemporais para o mapeamento da cafeeicultura: uma nova abordagem metodológica.....	114
	<b>CONSIDERAÇÕES FINAIS</b> .....	160

## **PRIMEIRA PARTE**

### **Introdução Geral**

#### **1 INTRODUÇÃO**

A cafeicultura é uma atividade de grande importância no cenário internacional (GOMES; ROSADO, 2005; SOUZA et al., 2012), sendo um dos produtos básicos mais valiosos, durante muitos anos só superado, em valor, pelo petróleo, como fonte de divisas para os países em desenvolvimento (INTERNACIONAL COFFEE ORGANIZATION, 2013).

O Brasil se destaca como o maior produtor de café do mundo (INTERNACIONAL COFFEE ORGANIZATION, 2009), sendo responsável por cerca de 40% da produção mundial e por 25% das exportações de grãos verdes. A cafeicultura gera, aproximadamente, 7 milhões de empregos diretos e indiretos e tem uma riqueza anual de cerca de 10 bilhões de reais (EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA, 2013).

Esta cultura se destaca entre as principais culturas agrícolas do Brasil (EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA, 2013), tendo grande importância para a sua economia. O estado de Minas Gerais é o maior produtor de café do país, respondendo por cerca de 50% da produção nacional (INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA, 2009). As regiões sul e oeste de Minas Gerais são as maiores produtoras de café do estado, seguidas da região do Alto Paranaíba e do Triângulo Mineiro (BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA, 2009).

A região Sul de Minas é o mais importante centro de produção cafeeira do estado (EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA, 2013) e, no cenário atual, corresponde a cerca de 70% da renda dos produtores

rurais da região (BARTOLOZI, 2008). A cafeicultura foi inserida nesta região na década de 1850 e muitas cidades surgiram a partir das grandes fazendas (BARTOLOZI, 2008). Além da região sul, o Triângulo Mineiro tem municípios com grande produção de café, como é o caso de Patrocínio, Monte Carmelo e Araguari (SOUZA et al., 2012). O plantio dessa cultura, durante todos esses anos, acarretou em grandes mudanças no uso e na ocupação da terra nessas regiões.

Atualmente, no Brasil, a obtenção de informações referentes a esta cultura (área plantada, área colhida, produção, entre outros) é feita pela Companhia Nacional de Abastecimento (CONAB) e pelo Instituto Brasileiro de Geografia e Estatística (IBGE), que utilizam informações fornecidas pelos municípios, cooperativas, produtores rurais e representantes de órgãos públicos e privados (COMPANHIA NACIONAL DE ABASTECIMENTO, 2014). Entretanto, estes dados, segundo Souza et al. (2012), são susceptíveis a erros, podendo gerar resultados distintos da realidade.

Considerando a dimensão da cafeicultura para o país e por se tratar de um commodity, o monitoramento e o mapeamento dos plantios de café são essenciais para gerir os recursos e ter informações sobre a ocupação da cafeicultura em escala local, regional e nacional, e de forma mais precisa. Uma ferramenta útil empregada para monitorar as mudanças na superfície da Terra é o sensoriamento remoto orbital.

O uso e a cobertura da terra é altamente dinâmico e existe grande dificuldade em estabelecer parâmetros para mapear essas mudanças (DEFRIES; BELWARD, 2000; LAMBIN; GEIST, 2006). Além disso, a classificação usando imagens de sensoriamento remoto para esta cultura, muitas vezes, é uma tarefa difícil, devido à grande variabilidade espectral do café (ADAMI et al., 2009; VIEIRA et al., 2006) e por apresentar uma resposta espectral muito semelhante a diferentes tipos de vegetação, como floresta e pastagem

(SARMIENTO et al., 2014; MOREIRA; ADAMI; RUDORFF, 2004). Esta sobreposição impõe desafios práticos e teóricos para o uso operacional do sensoriamento remoto orbital (CARVALHO, 2001) e para o mapeamento de áreas cafeeiras.

Alguns trabalhos anteriores indicam que o mapeamento do café por imagens de sensoriamento remoto não apresenta índices muito elevados de acurácia na classificação. Moreira, Adami e Rudorff (2004) tentaram separar o café de outros usos, como mata e pastagem, por meio da informação espectral e análise de série temporal, em imagens Landsat TM, em Minas Gerais. Os resultados mostraram que a banda 4 do sensor foi a que melhor conseguiu separar os tipos de uso, porém, a cultura ainda foi bastante confundida com mata e pastagem. Cordeiro-Sancho e Sader (2007) combinaram análises de bandas espectrais e dados complementares de topografia para avaliar a precisão da classificação e identificar a natureza da confusão espectral entre café e outros tipos de cobertura da terra. Os resultados mostraram que a melhor classificação teve uma acurácia global de 65,00%, apresentando confusão com áreas de vegetação nativa.

Em alguns estudos, utilizando classificações baseadas em objeto, têm sido incorporados, além de dados espectrais, dados de textura e geométricos para ajudar no mapeamento da cafeicultura (SARMIENTO et al., 2014; MARUJO et al., 2013; SANTOS et al., 2012), porém, não se sabe se esses dados são realmente eficientes para distinguir os plantios de café de outras formações vegetacionais, como, por exemplo, vegetação nativa. Sendo assim, investigar a eficácia destas variáveis é essencial.

Devido aos problemas relativos à similaridade espectral, a ocorrência de plantações de café traz limitações para a classificação de remanescentes de florestas, dificultando a diferenciação destas duas formações vegetais. Uma alternativa para superar os desafios de mapeamento do café pode ser a utilização

do perfil espectro-temporal, ou seja, integrar dados multitemporais com os dados multiespectrais e, aliando estas informações, trabalhar com classificação orientada a objeto.

Poucos estudos utilizam dados de séries temporais para o mapeamento desta cultura. A maioria deles utiliza apenas períodos de observação mais curtos, de um a dois anos (MOREIRA; ADAMI; RUDORFF, 2004). Além disso, os estudos realizados no Brasil utilizam apenas classificações mais tradicionais para mapear o café, como classificação visual e supervisionada (TRABAQUINI et al., 2011; MACHADO et al., 2010; VIEIRA et al., 2009; MOREIRA et al., 2007). Em poucos estudos foi utilizada a classificação orientada a objeto para o mapeamento desta cultura (SANTOS et al., 2012), além de dados de índice de vegetação (CORDERO-SANCHO; SADER, 2007).

No intuito de melhorar a precisão da classificação de áreas cafeeiras, neste trabalho buscou-se implementar novas técnicas para o mapeamento do café, utilizando uma abordagem de classificação por objeto, que é pouco usada no mapeamento desta cultura. Buscou-se, também, identificar melhores variáveis para a separação entre classes. Além disso, foi utilizada uma abordagem multitemporal, caracterizando os perfis temporais dos diferentes tipos de cobertura da terra que mais se assemelham espectralmente com o café, utilizando esta abordagem como variáveis de classificação.

## **1.1 Motivação**

Diante da dificuldade em mapear automaticamente as áreas cafeeiras e pelo pouco conhecimento sobre informações básicas, existe, assim, a necessidade de desenvolver melhores métodos de classificação automática para esta cultura.

É importante salientar que este estudo se iniciou com a necessidade de quantificar, localizar e caracterizar corretamente os plantios de café.

## **1.2 Objetivo geral**

Diante do que foi exposto, o objetivo geral, neste trabalho, foi propor uma metodologia para mapeamento da cafeicultura, por meio de variáveis multiespectrais e multitemporais.

## **1.3 Contribuições**

Para obter um monitoramento mais preciso da cafeicultura, uma nova abordagem para o mapeamento desta cultura foi proposta, empregando uma metodologia que consiga separar corretamente os cafés de diferentes tipos de cobertura da terra, principalmente vegetação nativa e pastagem, de forma automática e, conseqüentemente, mais rápida. Nesse sentido, este trabalho contribui com uma alternativa mais promissora no mapeamento dessas áreas.

A identificação de forma mais correta das áreas de plantio de café é de suma importância, uma vez que uma informação mais precisa do tamanho das lavouras e de sua distribuição espacial pode contribuir para estimar a sua produção.

## **1.4 Organização da tese**

A tese está organizada da seguinte forma: na primeira parte, Referencial teórico, aborda-se o estado da arte do mapeamento da cafeicultura e os processos de classificação de imagens de sensoriamento remoto.

Na segunda parte, encontram-se os artigos, sendo o primeiro: Uso de diferentes algoritmos e variáveis para mapeamento da cafeicultura na região sudoeste do Brasil, no qual destaca-se o papel da cultura cafeeira em âmbito internacional, sendo o café considerado um dos produtos básicos mais importantes no mundo, hoje. Por se tratar de uma atividade tão significativa, o mapeamento, bem como o monitoramento destas lavouras, é essencial para entender a sua dinâmica espaço-temporal. Porém, mapear esta cultura não é uma tarefa fácil, uma vez que a similaridade espectral entre café e outros tipos de uso da terra é muito acentuada. Portanto, neste capítulo objetivou-se comparar diferentes algoritmos de aprendizagem de máquina usando distintos conjuntos de variáveis oriundas de dados de sensoriamento remoto e identificar as melhores combinações algoritmos-variáveis para o mapeamento de cafezais em três regiões do estado de Minas Gerais.

Neste sentido, pretendeu-se responder às seguintes questões: (1) Há diferença significativa entre as classificações utilizando diferentes combinações entre algoritmos de classificação e variáveis de entrada? (2) Quais conjuntos de variáveis são mais eficientes para classificar cafezais? (3) Quais são os algoritmos mais eficientes para classificar cafezais? Qual algoritmo é mais acurado para o mapeamento de cafezais em cada uma das três regiões em estudo? (4) Quais são as classes confundidas com cafezal durante o processo de classificação?

O segundo artigo, Uso de variáveis multitemporais para o mapeamento da cafeicultura: uma nova abordagem metodológica, constata-se que a classificação automática de áreas cafeeiras tem se mostrado um trabalho complexo. Obter boas acurácias de mapeamento para estas áreas, muitas vezes, é uma tarefa difícil, devido ao alto nível de confusão existente com outras formações vegetais. Várias metodologias já foram testadas, no intuito de sanar estas dificuldades, porém, nenhuma delas conseguiu obter altos índices de acerto

no mapeamento desta cultura, quando utilizados parâmetros de classificação automáticos. Sendo assim, o objetivo geral, neste trabalho, foi desenvolver um padrão metodológico para o mapeamento da cafeicultura por meio de dados multitemporais, aliado á classificação orientada a objeto. Nesse sentido, espera-se responder às seguintes questões: (i) Como o número de amostras de treinamento interfere no resultado da classificação? (ii) Variáveis temporais são eficazes na separabilidade do café? (iii) Qual a quantidade de variáveis necessária para aumentar a acurácia da classificação? (iv) Dados temporais Landsat TM são compatíveis para mapear imagens RapidEye?

Na terceira parte encontram-se as considerações finais com relação ao trabalho realizado.

## REFERÊNCIAS

- ADAMI, M. et al. Avaliação da exatidão do mapeamento da cultura do café no Estado de Minas Gerais. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 14., 2009, Natal. **Anais...** São José dos Campos: INPE, 2009. , p. 1-8.
- BARTOLOZI, L. Sul de Minas: café representa 70% da renda agrícola. **Coffee Break**, [S.l.], 2008. Disponível em: <<http://www.coffeebreak.com.br/ocafezal.asp?SE=6&ID=38>>. Acesso em: 20 set. 2014.
- CARVALHO, L. M. T. **Mapping and monitoring forest remnants: a multiscale analysis of spatio-temporal data**. 2001. 150 p. Tese (Doutorado em Sensoriamento Remoto) – Wageningen University, Wageningen, 2001.
- COMPANHIA NACIONAL DE ABASTECIMENTO. **Acompanhamento de safra brasileira: primeiro levantamento**. Brasília: CONAB, 2014.
- CORDERO-SANCHO, S.; SADER, S. A. Spectral analysis and classification accuracy of coffee crops using Landsat and a topographic-environmental model. **International Journal of Remote Sensing**, Basingstoke, v. 28, n. 7-8, p. 1577–1593, 2007.
- DEFRIES, R. S.; BELWARD, A. S. Global and regional land cover characterization from satellite data: an introduction to the special issue. **International Journal of Remote Sensing**, Basingstoke, v. 21, n. 6, p. 1083–1092, 2000.
- EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA. Histórico. **Embrapa Café**, Brasília, 2013. Disponível em: <<http://www22.sede.embrapa.br/cafe/unidade/historico.htm>>. Acesso em: 20 set. 2013.
- GOMES, M. F. M.; ROSADO, P. L. Mudança na produtividade dos fatores de produção da cafeicultura nas principais regiões produtoras do Brasil. **Revista de Economia e Sociologia Rural**, São Paulo, v. 43, n. 4, p. 353-378, out./dez. 2005.
- INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Banco de dados agregados: sistema de recuperação automática**. Brasília: IBGE, 2009. Disponível em: <<http://www.sidra.ibge.gov.br/>>. Acesso em: 20 set. 2013.

INTERNACIONAL COFFEE ORGANIZATION. **The world economic crisis and the coffee sector.** [S.l: sn], 2009. Disponível em: <<http://www.ico.org/documents/ed-2059e-economic-crisis.pdf>>. Acesso: 20 de setembro de 2013.

LAMBIN, E. F.; GEIST, H. (Eds.). **Land-use and land-cover change: local processes and global impacts.** Berlin: Springer-Verlag, 2006.

MACHADO, M. L. et al. Mapeamento de áreas cafeeiras (*Coffea arabica L.*) da Zona da Mata Mineira usando sensoriamento remoto. **Coffee Science**, Lavras, v. 5, n. 2, p. 113-122, maio/ago. 2010.

MARUJO, R. F. B. et al. Classificação orientada a objetos aplicada á cultivos cafeeiros em Três Pontas MG. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 16., 2013, Foz do Iguaçu. **Anais...** Foz do Iguaçu: INPE, 2013.

MOREIRA, M. A. et al. Tecnologia de informação: imagens de satélite para o mapeamento de áreas de café de Minas Gerais. **Informe Agropecuário**, Belo Horizonte, v. 28, n. 241, p. 27-37, nov./dez. 2007.

MOREIRA, M. A.; ADAMI, M.; RUDORFF, B. F. T. Análise espectral e temporal da cultura do café em imagens Landsat. **Pesquisa Agropecuária Brasileira**, Brasília, v. 39, n. 3, p. 223-231, mar. 2004.

SANTOS, J. A. et al. Multiscale classification of remote sensing images. **IEEE Transactions on Geoscience and Remote Sensing**, New York, v. 50, n. 10, p. 3764-3775, Oct. 2012.

SARMIENTO, C. M. et al. Comparação de classificadores supervisionados na discriminação de áreas cafeeiras em Campos Gerais – Minas Gerais. **Coffee Science**, Lavras, v. 9, n. 4, p. 546-557, 2014.

SOUZA, V. C. O. et al. Espacialização e dinâmica da cafeicultura mineira entre 1990 e 2008, utilizando técnicas de geoprocessamento. **Coffee Science**, Lavras, v. 7, n. 2, p. 122-134, 2012.

TRABAQUINI, K. et al. Caracterização de lavouras cafeeiras, utilizando técnicas de geoprocessamento e sensoriamento remoto, no município de Umarama – PR. **Ciência e Agrotecnologia**, Lavras, v. 35, n. 1, p. 35-44, jan./fev. 2011.

VIEIRA, T. G. C. et al. Análise espacial do parque cafeeiro da região de Três Pontas MG. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 14., 2009, Natal. **Anais...** São José dos Campos: MCT, 2009. p. 6361-6368.

VIEIRA, T. G. C. et al. Assessing and mapping changes, in space and time, of coffee lands of the state of Minas Gerais in Brazil. In: INTERNATIONAL GEOGRAPHICAL UNION COMMISSION ON GEOGRAPHICAL INFORMATION SYSTEMS, 2006, Vienna. **Anais...** Vienna: GICON, 2006.

## 2 REFERENCIAL TEÓRICO

### 2.1 Cultura cafeeira e sua importância

A planta do cafeeiro tem origem na África (GRANER; GODOY-JUNIOR, 1967). Depois de ser introduzido na Europa, em 1615, os holandeses disseminaram o café pelo mundo e as primeiras sementes chegaram ao Brasil em 1727, no estado do Pará (ECCARDI; SANDALJ, 2003). Com o passar dos anos, a cultura foi difundida para os estados do Maranhão, Bahia, Rio de Janeiro, São Paulo, Minas Gerais, Espírito Santo, Paraná e, posteriormente, Mato Grosso e Rondônia (GRANER; GODOY-JUNIOR, 1967).

O café (*Coffea* spp.) é uma planta perene, de porte arbustivo ou arbóreo (DAMATTA et al., 2007). Comercialmente, as espécies mais exploradas são *Coffea canephora* e *Coffea arabica*, responsáveis por 99% da produção comercial no mundo (DAMATTA et al., 2007). Cerca de 65% da produção mundial corresponde à *C. arabica* e o restante à *C. canephora* (DAMATTA et al., 2007). No Brasil, cerca de 80% da produção é de *C. arabica*, sendo as cultivares Catuaí e Mundo Novo as mais tradicionais; cerca de 20% é de *C. canephora*, espécie cultivada, principalmente, nos estados do Espírito Santo e de Rondônia (ECCADI; SANDALJ, 2003).

O café arábica produz uma bebida de qualidade superior e, portanto, é empregado na indústria do café torrado e moído, além de ter maior aceitação nos mercados consumidores (DAMATTA e RAMALHO, 2006; FERRÃO et al., 2007). O café canéfora, mais conhecido como robusta ou conilon, apresenta alto teor de cafeína e sólidos solúveis nos seus grãos. Com isso, produz uma bebida inferior, mas tem boa aceitação no mercado, devido aos menores preços (FERRÃO et al., 2007).

No comércio internacional, o café é um dos produtos básicos mais valiosos, durante muitos anos superado em valor apenas pelo petróleo, como fonte de divisas para os países em desenvolvimento (INTERNACIONAL COFEE ORGANIZATION, 2009). Sua importância é crucial para as economias e políticas de muitos países em desenvolvimento. O Brasil responde por cerca de 40% da produção mundial e, embora o café tenha diminuído sua participação nas exportações brasileiras ao longo do tempo, devido à diversificação dos produtos, ainda é um importante gerador de divisas para o país (EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA, 2013), com 25% das exportações mundiais de grãos verdes.

A cafeicultura é uma das mais importantes culturas agrícolas do Brasil (EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA, 2013) e de grande destaque na economia brasileira. Estima-se que a atividade relacionada ao café gere, no país, cerca de 7 milhões de empregos diretos e indiretos e uma riqueza anual de 10 bilhões de reais (EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA, 2013).

Minas Gerais se destaca como o maior produtor de café do país e é responsável por aproximadamente 50% da produção nacional (INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA, 2009). As regiões sul e oeste de Minas Gerais se destacam como as maiores produtoras de café do estado (INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA, 2009).

## **2.2 Mapeamento da cafeicultura**

A cafeicultura, apesar de sua importância econômica e social, carece de informações complementares para o seu sistema produtivo, principalmente em relação à sua extensão, distribuição espacial e o ambiente em que é cultivada.

Esta lacuna de informações restringe o planejamento desta atividade econômica, dificulta o planejamento ambiental, o planejamento de práticas de manejo e o controle da erosão, bem como as estimativas de produtividade que utilizam índices de penalização resultantes de adversidades climáticas e/ou fitossanitárias (MOREIRA et al., 2007). Todas essas informações são importantes para o planejamento estratégico do governo municipal, estadual e federal. No entanto, conhecer um sistema complexo, como o agroecossistema cafeeiro, por meio de métodos convencionais, é uma atividade extensa e de difícil desenvolvimento (MOREIRA; BARROS; RUDORFF, 2008). A utilização de imagens de satélites e técnicas de interpretação dessas imagens é, sem dúvida, a melhor maneira para obter a área plantada com esta cultura (MOREIRA et al., 2007).

Considerando as aplicações de dados de sensoriamento remoto no levantamento de dados sobre a cafeicultura, os primeiros trabalhos realizados foram por Velloso (1974), Velloso (1976) e Velloso e Souza (1978). Os resultados obtidos foram desanimadores, devido à baixa resolução espacial do sensor utilizado (MSS/Landsat) e à heterogeneidade da cultura cafeeira. A partir do lançamento do sensor “Thematic Mapper” (TM/Landsat 5), vários trabalhos foram desenvolvidos para mapear esta cultura (SARMIENTO et al., 2014; MARTÍNEZ-VERDUZCO; GALEANA-PIZAÑA; CRUZ-BELLO, 2012; TRABAQUINI et al., 2011, MACHADO et al., 2010; MOREIRA et al., 2010; VIEIRA et al., 2009; CORDERO-SANCHO; SADER, 2007; MOREIRA et al., 2007) e analisar o seu comportamento espectral (VIEIRA et al., 2006; MOREIRA et al., 2004; VIEIRA; ALVES; LACERDA, 2003).

Ainda hoje, o mapeamento desta cultura ainda é muito complexo, pois o cultivo do café no Brasil apresenta grande diversidade de práticas de manejo (com e sem irrigação, com e sem quebra-ventos, etc.), lavouras de diferentes idades, áreas, espaçamentos (tradicional, adensado e superadensado) e variedades (p. ex. Mundo Novo, Catuaí, Acaíá, Icatu e Caturra). Além disso,

diferentes condições de clima, relevo, iluminação e época do ano aumentam ainda mais essa diversidade, fazendo com que o café apresente uma significativa variação na resposta espectral gravada em imagens de sensoriamento remoto (Figura 1), dificultando sua identificação e a separação dos demais alvos presentes numa cena (MOREIRA et al., 2010). Isso faz com que o café apresente grande semelhança espectral com outros tipos de uso do solo, como, por exemplo, floresta e pastagem (MOREIRA et al., 2004). Estes autores mostraram que a cultura cafeeira, apesar de apresentar variações no comportamento espectral, por causa de fatores como espaçamento, idade da lavoura, sistema de época do ano, pode ser identificada e mapeada em imagens de satélites de média resolução espacial, com boa precisão de mapeamento, desde que o analista realize uma interpretação visual sobre os resultados da classificação feita no computador. Porém, o mapeamento desta cultura pode ser difícil, devido à grande variabilidade espectral desta cultura, tornando a identificação para o café mais difícil do que para outras culturas agrícolas (ADAMI et al., 2009; VIEIRA et al., 2006).

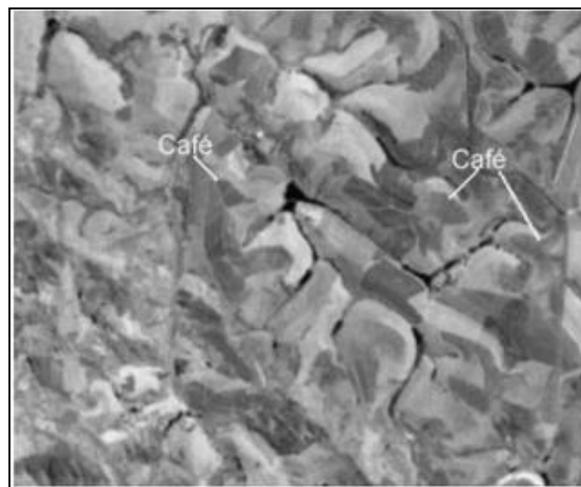


Figura 1 Plantios de café em uma imagem Landsat/TM

Para Machado et al. (2001), o uso de imagens Landsat para mapeamento de áreas cafeeiras da Zona da Mata, em Minas Gerais, mostrou-se restrito, apesar da pouca diversidade de uso da terra e cobertura vegetal encontrada na região. Os resultados da classificação supervisionada indicaram que a confusão na classificação aconteceu, principalmente, entre as plantações de café e as áreas de floresta. O baixo desempenho das classificações está principalmente relacionado à similaridade espectral e à topografia acidentada da região, além do consequente sombreamento das imagens e da fragmentação das lavouras de café, localizadas em áreas contíguas a florestas. Esperam-se melhores resultados quando da aplicação desta metodologia em condições de relevo mais suave e para regiões com áreas agrícolas mais extensas e homogêneas.

Moreira et al. (2007) mapearam as lavouras de café do estado de Minas Gerais, com imagens do sensor TM do satélite Landsat-5 e do sensor HRVIR do satélite SPOT 4. Para os mapas, a metodologia empregada foi fundamentada na classificação das imagens, utilizando interpretação visual das imagens. Por meio desta metodologia, os autores obtiveram informações de área plantada com café por municípios, macrorregiões, microrregiões e para todo o estado de Minas Gerais. O mesmo foi feito, em uma segunda etapa, para o estado de São Paulo (MOREIRA et al., 2010). Este tipo de estudo é de extrema importância para entender o padrão de distribuição desta cultura, além de obter dados mais precisos sobre o tamanho das lavouras e, assim, tentar estimar a produtividade.

Aliar dados ambientais topográficos ao mapeamento das áreas cafeeiras pode ser uma boa alternativa para melhorar a acurácia do mapeamento, além de dados de índice de vegetação (CORDERO-SANCHO; SADER, 2007). De acordo os autores, estes dados se mostraram importantes para conseguir índices de acurácia em torno de 85% a 90% no mapeamento do uso e cobertura da terra, em seu estudo realizado na Costa Rica, em que eles usaram classificação

supervisionada com combinações de bandas, índice de vegetação e dados topográficos.

Martínez-Verduzco, Galeana-Pizaña e Cruz-Bello (2012) utilizaram técnicas de geoprocessamento e sensoriamento remoto para mapear áreas cafeeiras no México. Na região montanhosa de Chiapas existe uma grande dificuldade em mapear café, devido ao fato de a floresta e esta cultura possuírem praticamente a mesma reposta espectral. Este conflito é acentuado em algumas áreas da região em que o relevo é mais acidentado. A metodologia utilizada por eles, aliando classificação pixel a pixel de imagens de alta resolução, mostrou boa acurácia para separação de café e floresta. Apesar das vantagens apresentadas por este método, mais estudos são necessários para atingir níveis mais elevados de precisão, de modo a discriminar floresta e café.

Utilizando algoritmos de aprendizagem de máquina, Sarmiento et al. (2014) mostraram resultados mais satisfatórios, mapeando o café em uma região de Minas Gerais. Os resultados obtidos mostraram que os índices de acerto destas classificações foram de 91,07% e 90,65%, para os algoritmos KNN e SVM, respectivamente. Porém, as áreas cafeeiras ainda foram confundidas, em parte, com floresta e pastagem.

Ainda hoje, a classificação por imagens de sensoriamento remoto para esta cultura, muitas vezes, é complexa, devido à grande variabilidade espectral do café (ADAMI et al., 2009; VIEIRA et al., 2006) e por apresentar uma resposta espectral análoga a diferentes tipos de vegetação, como vegetação nativa e pastagem (SARMIENTO et al., 2014; MOREIRA et al., 2004). Esta sobreposição estabelece desafios para o mapeamento de áreas cafeeiras.

### 2.3 Séries temporais

O uso e a ocupação da terra estão em constante mudança e sofrem alterações ao longo do tempo e do espaço (BENETT et al., 2006). Segundo Houghton e Goodale (2004), a maioria das mudanças no uso da terra afeta, principalmente, a vegetação e o solo de um ecossistema, e, como consequência, o estoque de carbono. Segundo os mesmos autores, as alterações podem ser grandes, como, por exemplo, a conversão de florestas em terras agrícolas ou o reflorestamento em terras nuas, ou, ainda, a substituição de espécies animais nativas por bovinos em pastagens naturais.

Estas mudanças no ecossistema ocorrem em uma variedade de escala espacial e temporal (COPPIN et al., 2004), devido a causas naturais e/ou antrópicas. Detectar estas mudanças é necessário para entender como estas alterações têm afetado o uso e a ocupação do solo. Segundo Singh (1989), a detecção de mudanças é o processo de identificação de diferenças no estado de um objeto ou fenômeno, ao observá-lo em diferentes momentos. Este método fornece a base para uma melhor compreensão das relações e interações entre os fenômenos naturais e humanos, para melhor gerenciar e usar os recursos (LU et al., 2004). Este processo envolve a aplicação de conjuntos de dados multitemporais para analisar quantitativamente os efeitos do fenômeno.

Porém, a maioria das metodologias aplicadas para a detecção de mudança prioriza apenas a diferença entre duas imagens, ou seja, análise bitemporal (COPPIN et al., 2004), o que impede a detecção de tendências em longo prazo (KENNEDY; COHEN; SCHROEDER, 2007). A análise bitemporal de detecção de mudanças tem dois problemas: a data de aquisição das imagens (devido a diferenças de sazonalidade) e o intervalo de tempo da mudança (resolução temporal) (COPPIN et al., 2004). Como consequência, a imagem produzida a partir da diferença entre as duas imagens gera resultados não muito

confiáveis, pois está sujeita a erros de registro das imagens, às condições atmosféricas nas datas de imageamento e à própria fenologia da vegetação (CARVALHO, 2008).

Para contornar estes entraves, alguns pesquisadores têm abordado o monitoramento dos ecossistemas por meio de séries temporais, a partir do qual é possível avaliar o perfil temporal e as tendências de mudanças (KENNEDY; COHEN; SCHROEDER, 2007; COPPIN et al., 2004). 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).

De acordo com Morettin e Toloí (2004), uma série temporal é qualquer conjunto de observações ordenadas no tempo. Segundo os mesmos autores, a análise de uma série temporal permite prever valores futuros da série e verificar a existência de tendências e periodicidades nos dados.

Uma grande variedade de procedimentos tem sido desenvolvida, ao longo das últimas duas décadas, para a detecção de alterações digitais (SALMON et al., 2011; VEBERSSELT et al., 2010a; COPPIN et al., 2004). Atualmente, o desenvolvimento de metodologias eficazes para a análise de dados multitemporais é uma das mais importantes e desafiadoras questões para a comunidade de sensoriamento remoto (BRUZZONE; SMITS; TILTON, 2003). Muitos trabalhos têm sido feitos utilizando séries temporais para detectar mudanças no uso da terra (VERBESSELT; ZEILEIS; HEROLD, 2012; OHMANN et al., 2012; SCHROEDER et al., 2011; STELLMES et al., 2010), mudanças na temperatura da superfície terrestre (JULIEN; SOBRINO; VERHOEF, 2006) e caracterização da fenologia da vegetação (OLIVEIRA et al., 2012; MARTÍNEZ; GILABERT, 2009; XIAO et al., 2006; JÖNSSON; EKLUNDH, 2002)

Vários estudos foram realizados, na última década, para melhorar as análises espectrais (DEMIR; PERSELLO; BRUZZONE, 2011; LANDGREBE, 2005) e temporais (DEMIR; BOVOLO; BRUZZONE, 2012; SMITS; BRUZZONE, 2004) e a compreensão de dados de sensoriamento remoto relacionados a mudanças na superfície da Terra. No entanto, esses estudos são, muitas vezes, limitados espectro ou temporalmente, seja pela análise do perfil espectral da imagem (por exemplo, análise multiespectral) (LANDGREBE, 2005) numa imagem em uma única data (LEE; ERSOY, 2007) ou restringindo a análise do perfil temporal da imagem (ou seja, a análise multitemporal) apenas para uma única camada espectral (por exemplo, um índice de vegetação) (SALMON et al., 2011; VEBERSSELT et al., 2010b). Além disso, poucos estudos que abordam a relação entre aspectos multitemporais e multiespectrais têm sido relatados na literatura científica (BRUZZONE; SMITS, 2002; COPPIN et al., 2004). Assim, existe um grande potencial para o desenvolvimento de métodos de sensoriamento remoto que integram e exploram ambas as informações, multiespectrais e multitemporais (COPPIN et al., 2004).

Além de serem utilizados para o monitoramento das mudanças na superfície da Terra, os dados de séries temporais também têm sido utilizados para o mapeamento de áreas agrícolas (MELLO et al., 2012; PUNIA; JOSHI; PORWAL, 2011; EPIPHANIO et al., 2010; LAMBIN; LINDERMAN, 2006). Mello et al. (2012) utilizaram dados multiespectrais e multitemporais para mapear áreas de cana-de-açúcar na região oeste do Brasil. Tais autores detectaram o comportamento espectro-temporal dessas lavouras e as mudanças ocorridas no uso da terra, obtendo excelentes resultados na classificação. Os dados oriundos de séries temporais, juntamente com os algoritmos de aprendizagem de máquina, são promissores para o mapeamento, pois auxiliam o processo de classificação destas áreas.

Alguns algoritmos têm sido desenvolvidos no intuito de extrair trajetórias espectro-temporais de mudanças na superfície da Terra, além de eliminar os ruídos que causam alterações na interpretação dos resultados, como, por exemplo, Landtrendr (do inglês *Landsat-based Detection of Trends in Disturbance and Recovery*) (KENNEDY; YANG; COHEN, 2010), TimeSat (JÖNSSON; EKLUNDH, 2004); TiSeg (COLDITZ et al., 2008), HANTS (do inglês *Harmonic Analysis of Time Series*) (VERHOEF; MENENTI; AZZALI, 1996), IDRISI (*Land Trend Modeller*), BFAST (do inglês *Breaks For Additive Seasonal and Trend*) (VERBESSELT et al., 2010a; VERBESSELT et al., 2010b) STARS (do inglês *Spectral-Temporal Analysis by Response Surface*) (MELLO et al., 2012; VIEIRA, 2000) e *Greenbrown* (FORKEL et al., 2013). A seguir, a descrição dos algoritmos que serão utilizados neste trabalho.

### 2.3.1 BFAST

O *Breaks For Additive Seasonal and Trend* (BFAST), proposto por Verbesselt et al. (2010a), é um algoritmo que tem a capacidade de decompor a série temporal em três componentes distintos: tendência ( $Z_t$ ), sazonalidade ( $S_t$ ) e ruído ( $\varepsilon_t$ ), utilizando métodos para detectar e caracterizar mudanças dentro da série temporal (Figura 2). Esta decomposição resulta na detecção de mudanças rápidas e graduais que acontecem dentro do componente de tendências ou mudanças fenológicas ocorrendo no componente sazonal (VERBESSELT et al., 2010a). Este é um modelo de decomposição aditivo que se encaixa de forma interativa na tendência linear por partes e modelo sazonal, dada pela expressão

$$Y_t = T_t + S_t + \varepsilon_t, \quad t = 1, \dots, n \quad (2)$$

em que  $Y_t$  são os dados observados no tempo;  $T_t$  é o componente de tendência;  $S_t$  é o componente sazonal e  $et$  é o restante componente ou ruído. Estas mudanças são detectadas por meio da identificação de pontos de quebra resultantes a partir do processo de decomposição BFAST.

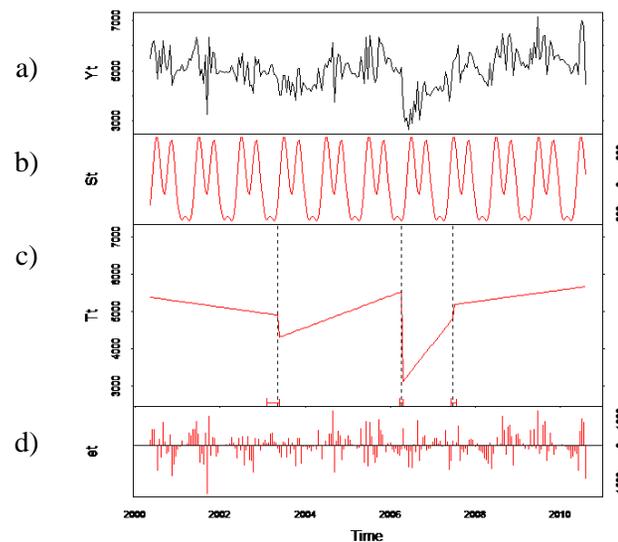


Figura 2 Simulação de uma série temporal MODIS NDVI. Em a) dados originais da série temporal; b) componente sazonal; c) componente de tendência, indicando uma mudança após o ano de 2004; d) componente restante (ruído)

Adaptado de Verbesselt et al. (2010a)

O BFAST emprega o procedimento *Seasonal-Trend decomposition* (STL) para decompor a série temporal, que fornece uma estimativa precisa e robusta de componentes de tendência e de sazonalidade, graças à sua capacidade de lidar com valores extremos ou a falta de valores dentro da série histórica (VERBESSELT et al., 2010a; VERBESSELT et al., 2010b).

Este procedimento é um método não paramétrico que, utilizando um modelo aditivo, de forma flexível, decompõe séries temporais. Este tipo de método implica que a magnitude das flutuações na série original resultante do padrão sazonal e a componente residual não seja afetada pelo nível da tendência (JACQUIN; SHEEREN; LACOMBE, 2010).

$$X_t = Z_t + S_t + \varepsilon_t \quad (3)$$

STL é um processo iterativo que utiliza *Locally weighted regression smoother* (Loess) para decompor os componentes da série temporal (VERBESSELT et al., 2010b). Com a formação da série temporal decomposta, a componente tendência terá grande relevância na detecção de distúrbio real, pois permite a determinação da direção da mudança, analisando o sinal de inclinação da tendência. Na Figura 2 mostra-se um exemplo do resultado obtido pelo algoritmo BFAST.

O BFAST foi utilizado neste estudo por apresentar uma fácil implementação (por meio do pacote BFAST do software livre R) e por ser um algoritmo mais robusto, podendo ser aplicado a qualquer tipo de sensor que, além de decompor a série nos componentes de sazonalidade e tendência, detecta mudanças graduais e abruptas na série.

### 2.3.2 Greenbrown

O Greenbrown é um algoritmo que avalia distintos métodos para estimação da tendência, extraíndo parâmetros de uma série temporal. Este algoritmo foi desenvolvido por Forkel et al. (2013). O método oferece uma gama de funções para analisar tendências e mudanças de tendências, bem como inferir sobre as propriedades da série: média, tendência, sazonalidade, variabilidade

interanual (*Inter-annual variability*, IAV) e variabilidade em curto prazo (*Short-Term Variability*, STV) e seus desvios padrões (IAV sd e STV sd). O cálculo destes parâmetros é feito pela decomposição da série (Figura 3).

1º) A média da série temporal é calculada.

2º) O parâmetro tendência é calculado com base no método *Annual Aggregated Time Series* (AAT), em que os valores mensais da série são agregados a valores anuais e, assim, o coeficiente de inclinação da reta da regressão linear é estimado, indicando uma tendência positiva ou negativa (sem calcular as mudanças abruptas).

3º) O IAV, que corresponde ao desvio padrão das anomalias (ruídos) anuais, é gerado pela subtração da média e da tendência dos valores originais da série.

4º) A sazonalidade é calculada pela subtração da média, da tendência e das anomalias da série temporal original.

5º) Por fim, a média, a tendência, as anomalias e o ciclo sazonal são subtraídos da série temporal, resultando as anomalias em curto prazo, ou seja, STV.

Além dessas propriedades estatísticas, este pacote também avalia outros métodos de tendência, sendo eles: método *Annual Aggregated Time Series*, ou AAT; *Season-Trend Model*, ou STM; *Mean anual cycle*, ou MAC e *Annual cycle based on singular spectrum analysis*, ou SSA. Estes métodos se diferenciam na resolução temporal da série NDVI, presença ou não do componente sazonal e em como a tendência é estimada (FORKEL et al., 2013).

No método AAT, a sazonalidade é agregada para valores anuais e a inclinação da reta de cada segmento de tendência é estimada por mínimos quadrados de uma regressão linear (FORKEL et al., 2013), em que a significância é estimada pelo teste de tendência Mann-Kendall Mann (1945) apud Forkel et al., (2013).

Para o método STM, a estimação da tendência e dos *breakpoints* é baseada no BFAST (VERBESSELT et al., 2010a), em que o valor de NDVI ( $y$ ) em um tempo  $t$  pode ser expresso por

$$y_t = \alpha_1 + \alpha_2 t + \sum_{j=1}^k \gamma_j \sin\left(\frac{2\pi jt}{f} + \delta_j\right) + \varepsilon_t \quad (4)$$

em que  $\alpha_1$  é o intercepto da reta de tendência;  $\alpha_2$  é a inclinação da reta de tendência;  $\gamma$  são as amplitudes;  $\delta$  são as fases;  $f$  é a frequência (número de observações em um ano, aqui  $f = 23$ );  $k$  são os termos harmônicos e  $\varepsilon$  é o erro (VERBESSELT et al., 2010a). A significância é estimada pelo teste-t (FORKEL et al., 2013).

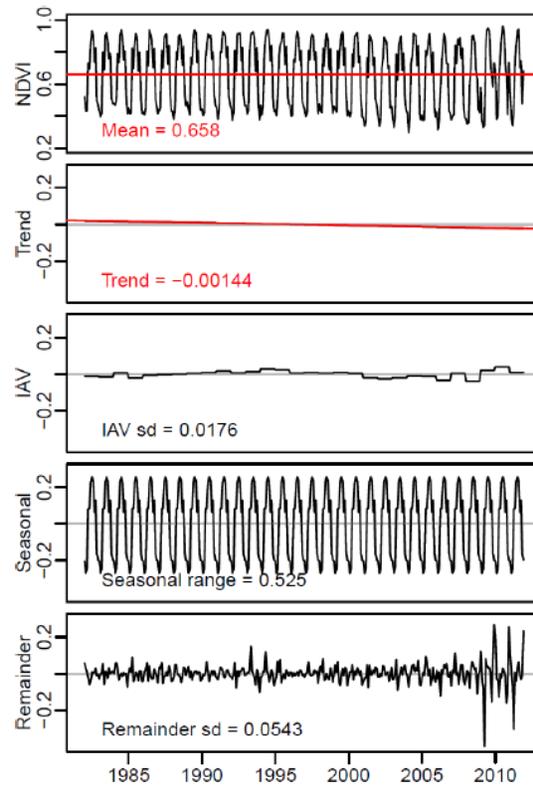


Figura 3 Componentes da série temporal estimados. O painel superior mostra uma série temporal NDVI com o seu valor de médio (linha vermelha). Os próximos painéis mostram a tendência estimada, a variabilidade interanual (IAV) (ou seja, anomalias anuais), sazonalidade (ou seja, a média de ciclo sazonal) e a variabilidade de curto prazo (componente residual), respectivamente. Soma de média, tendência, IAV, componente sazonal e residual compõem a série de temporal original

Fonte: Forkel et al. (2013)

Já nos métodos MAC e SSA, a sazonalidade é removida, ou seja, a tendência é estimada sem o componente sazonal e, assim, a inclinação da reta de tendência  $\alpha_2$  é estimada por mínimos quadrados ordinários (OLS).

$$a_t = \alpha_1 + \alpha_2 t + \varepsilon_t \quad (5)$$

em que  $a$  é a série ajustada sem o componente sazonal (FORKEL et al., 2013). A significância de cada segmento de tendência é estimada pelo teste de tendência Mann-Kendall (FORKEL et al., 2013). A diferença entre os métodos MAC e SSA está no ciclo sazonal que, no primeiro, é o ciclo anual médio, no qual se considera que o ciclo sazonal tem a mesma amplitude e frequência para todos os anos; no segundo, é o ciclo anual modulado, em que as amplitudes e as frequências variam ao longo dos anos (FORKEL et al., 2013)

Em todos os métodos, um mesmo algoritmo para a detecção de *breakpoints* foi utilizado, descrito por Bai e Perron (2003) e Zeileis et al. (2003), citados por Forkel et al. (2013). Primeiramente, o teste de *ordinary-least squares moving sum*, ou MOSUM, é aplicado para verificar a existência de *breakpoints* significativos na série; depois, o número ótimo de *breakpoints* e a sua localização na série são testados interativamente (FORKEL et al., 2013).

Este algoritmo foi escolhido para este estudo, pois é de fácil implementação (por meio do pacote “greenbrown” do software livre R). Como o BFAST, este algoritmo é robusto e pode ser aplicado a qualquer tipo de sensor. Além disso, ele disponibiliza dados matriciais de diversos componentes de sazonalidade e tendência, que podem ser utilizados como variáveis no processo de classificação.

## 2.4 Fusão de imagens

O processo de fusão de imagens de diferentes sensores e diferentes resoluções tem sido um importante aliado para solucionar problemas relacionados, principalmente, à contaminação por nuvens. Devido à escassez de

detalhamento espacial, temporal ou espectral de alguns sensores, algumas metodologias de fusão entre imagens foram desenvolvidas.

A fusão de imagens visa à integração de informação espacial e espectral (ACERBI-JUNIOR; CLEVERS; SCHAEPMAN, 2006) e temporal (GAO et al., 2006; HILKER et al., 2009a; ZHU et al., 2010) do mesmo sensor ou de diferentes sensores para produzir uma imagem de melhor qualidade do que a imagem de origem. Várias abordagens têm sido propostas na literatura para a fusão de imagens.

Os primeiros trabalhos realizados objetivaram fundir dados utilizando bandas multiespectrais e a banda pancromática, que tem melhor resolução espacial. Alguns autores usam métodos que utilizam *Intensity-Hue-Saturation* (IHS), transformada *wavelet* e análises de componentes principais (ACP) (LI; KWOK; WANG, 2002; TEGGI; CECCHI; SERAFINI, 2003; ACERBI-JUNIOR; CLEVERS; SCHAEPMAN, 2006). Estas técnicas são muito eficazes para originar imagens com alta resolução espacial (LI; KWOK; WANG, 2002).

No entanto, estes métodos não são capazes de melhorar a resolução espacial e temporal concomitantemente, o que não é ideal para estudos sobre a dinâmica da vegetação, já que o componente temporal é fundamental para estudos sobre a variabilidade inter e intra-anual (ZHU et al., 2010). Sendo assim, foi desenvolvido, por Gao et al. (2006), o algoritmo STARFM, que tem esta capacidade de aliar dados espaciais e temporais no processo de fusão de imagens.

#### **2.4.1 STARFM**

*Spatial and Temporal Adaptive Reflectance Fusion Model*, ou STARFM, proposto por Gao et al. (2006), é um algoritmo para fusão de imagens que prediz valores de reflectância da superfície a uma resolução espacial e

espectral de imagens dos sensores TM ou ETM+ do satélite Landsat, utilizando a frequência temporal de imagens MODIS (HILKER et al., 2009a).

As imagens TM/ETM<sup>+</sup> e MODIS são utilizadas para a fusão, com a vantagem de apresentarem grande similaridade entre os parâmetros orbitais de seus satélites, possuindo bandas em locais parecidos no espectro eletromagnético (GAO et al., 2006). Este algoritmo deve ser realizado, preferencialmente, em imagens Landsat TM/ETM<sup>+</sup> processadas pelo *The Landsat Ecosystem Disturbance Adaptive Processing System*, ou LEDAPS, o qual utiliza metodologia de correção atmosférica 6S baseada nas imagens MODIS (MASEK et al., 2006). Isso torna as imagens compatíveis para o processamento (GAO et al., 2006).

Este algoritmo prediz valores de reflectância da superfície de imagens Landsat a partir de pares de imagens Landsat (*fine-resolution*) e MODIS (tempo  $T_1$ ), de preferência de mesma data, e imagem MODIS (*coarse-resolution*), correspondente à data da imagem a ser predita (tempo  $T_2$ ) (GAO et al. 2006). O resultado final é uma imagem de alta resolução temporal e espacial (GAO et al. 2006).

Uma janela móvel é utilizada, a fim de evitar problemas causados tanto pelas diferenças entre as imagens TM/ETM<sup>+</sup>'s e MODIS (resoluções espaciais), quanto por mudanças temporais (mudanças no uso da terra, fenologia da vegetação ou *Bidirectional Reflectance Distribution Function*, BRDF). 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 (Figura 4). O algoritmo é dado por (GAO et al., 2006)

$$L\left(\frac{x_w}{2}, \frac{y_w}{2}, T_2\right) = \sum_{i=1}^w \sum_{j=1}^w W_{ij} (M(x_i, y_j, T_2) + L(x_i, y_j, T_1) - M(x_i, y_j, T_1)) \quad (6)$$

em que  $L\left(x_{\frac{w}{2}}, y_{\frac{w}{2}}, T_2\right)$  é o valor do pixel TM/ETM+ predito para o tempo 2 ( $T_2$ ),  $w$  é o tamanho da janela de busca e  $\left(x_{\frac{w}{2}}, y_{\frac{w}{2}}\right)$  é o pixel central dessa janela móvel.

O peso  $W_{ij}$  é calculado a partir de três medidas: diferença espectral entre MODIS e Landsat TM, diferença temporal entre MODIS  $T_1$  e MODIS  $T_2$  e distância entre o pixel central predito e o pixel vizinho candidato espectralmente semelhante, dessa forma determinando 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/ETM+ e MODIS no tempo  $T_1$ , respectivamente (HILKER et al., 2009a). O peso  $W_{ij}$  é calculado pela seguinte fórmula

$$W_{ij} = \left(1/C_{ij} / \sum_{i=1}^w \sum_{j=1}^w (1/C_{ij})\right) \quad (7)$$

Em alguns trabalhos já foi testado este algoritmo para a fusão de imagens e obtiveram-se resultados satisfatórios (GAO et al., 2006; HILKER et al., 2009b; TIAN et al., 2013).

Gao et al. (2006) apresentaram um novo método de fusão de imagens, em que os dados fusionados foram testados com imagens originais, para fim de comparação dos valores de reflectância destas imagens. Os resultados foram satisfatórios, mostrando que a reflectância das imagens sintéticas é predita com precisão.

Hilker et al. (2009b), testando o STARFM, mostraram que o algoritmo apresentou bons resultados na predição de imagens para a avaliação de mudanças sazonais em florestas boreais e sub-boreais. As imagens sintéticas mantiveram os detalhes espectrais, no entanto, quando analisaram as áreas de

distúrbio, o algoritmo apresentou menor eficácia nos valores de reflectância destas áreas.

Para compor uma série temporal densa, no intuito de avaliar a eficácia destas imagens, comparando com uma série temporal MODIS, Tian et al. (2013) também usaram o STARFM. Os resultados mostraram que a série temporal Landsat sintética foi compatível com a série temporal MODIS.

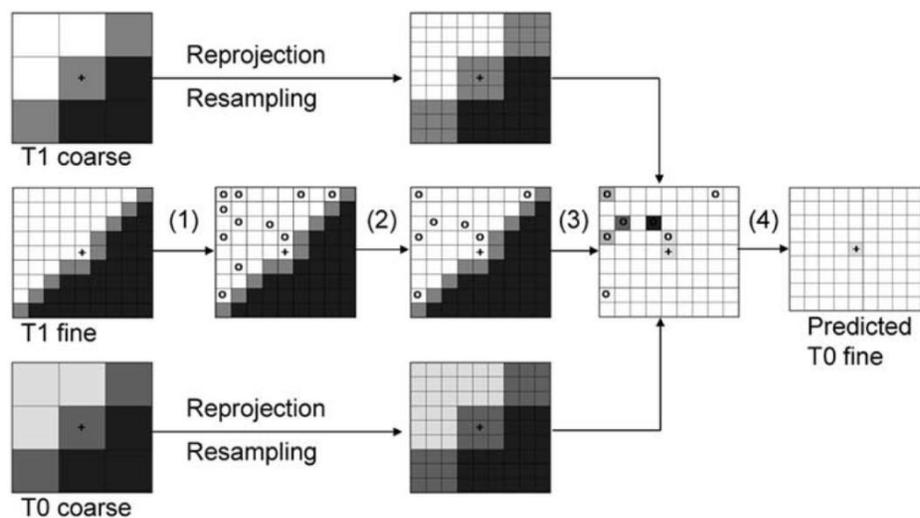


Figura 4 Esquema da metodologia empregada pelo STARFM. (passo 1) Dados de melhor resolução (*fine-resolution*) são primeiramente usados para procurar pixels (o) que são espectralmente similares ao pixel central (+) dentro da janela móvel. Estas amostras precisam ser filtradas para assegurar uma boa seleção (passo 2). Ambos os dados, de melhor e pior resolução, são, então, usados para determinar o peso das amostras (passo 3), de acordo com a função de peso descrita anteriormente (o pixel de cor preta possui o peso maior). Por último, o valor predito é computado com base no peso dos cinco pixels selecionados (passo 4)

Fonte: Gao et al. (2006)

## 2.5 Dados de sensoriamento remoto

O sensoriamento remoto pode ser entendido como o conjunto de atividades que permitem a obtenção de informações dos objetos que compõem a superfície terrestre, sem a necessidade de contato direto com os mesmos. Estas atividades envolvem a detecção, a aquisição e a análise (interpretação e extração de informações) da energia eletromagnética emitida ou refletida pelos objetos terrestres e registradas por sensores remotos (MORAES, 2002).

Os sensores remotos fazem parte do que se denomina “sistemas de sensoriamento remoto”. Estes sistemas são os veículos e os instrumentos necessários à coleta de dados para serem analisados pela comunidade científica e de usuários em geral (EPIPHANIO, 2002). Segundo o mesmo autor, os satélites são veículos colocados em órbita da Terra e que promovem continuamente a aquisição de dados relacionados às propriedades primárias dos objetos. Estes satélites estão em grandes altitudes (geralmente entre 600 e 1.000 km) e têm a capacidade de abranger em seu campo de visada uma grande porção de superfície terrestre. Ao mesmo tempo, como têm que orbitar ao redor da Terra, promovem uma cobertura que se repete ao longo do tempo, permitindo o acompanhamento da evolução das propriedades de reflexão ou emissão dos objetos e fenômenos.

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 à 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 2 bits possui apenas 4 níveis de cinza, e uma imagem de 8 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 (MENESES et al., 2012).

Atualmente, inúmeros satélites são utilizados para auxiliar no monitoramento e no mapeamento de grandes áreas, como, por exemplo, Terra, Landsat, Spot, Quickbird e RapidEye, entre outros.

### 2.5.1 RapidEye

Os satélites RapidEye foram projetados por Mac Donald Dettwiller e associados (MDA) de Richmond, Canadá e começaram a funcionar em agosto de 2008 (RAPIDEYE, 2011; ZENG; ZHANG; WANG, 2011).

Os satélites operam em uma órbita heliossíncrona a 630 km de altitude, cada um com um sensor do tipo *pushbroom* de cinco bandas multiespectrais e de 6,5 m de resolução espacial (RAPIDEYE, 2011; ZENG; ZHANG; WANG, 2011). As especificações do sistema RapidEye se encontram no Quadro 1.

As imagens RapidEye contam com a banda *Red Edge* (690-730 nm), específica para o monitoramento da atividade fotossintética da vegetação. Realizando um estudo para avaliar a contribuição da banda *Red Edge* em classificações de uso do solo, Schuster, Förster e Kleinschmit (2010) calcularam 24 índices espectrais nos quais foi incorporada a banda *Red Edge*. Os autores utilizaram os algoritmos classificadores *Support Vector Machine* e máxima verossimilhança e, como resultado, encontraram que, para ambos, houve uma sutil melhora na acurácia global da classificação quando se introduziu a banda *Red Edge* e os índices como inputs para a classificação.

Quadro 1 Especificações do satélite RapidEye

<b>Especificações do Sistema RapidEye</b>	
Número de satélites	5
Vida útil da plataforma	7 anos
Altitude de órbita	630 km (heliossíncrona)
Passagem pelo Equador	11:00 (aproximadamente)
Tipo de sensor	Multiespectral (pushbroom)
Bandas espectrais	Azul (440-510 nm) Verde (520-590 nm) Vermelho (630-685 nm) Red Edge (690-730 nm) Infravermelho próximo (760-850)
Campo de visada instantâneo (nadir)	6,5 m
Tamanho do pixel (ortoretificado)	5 m
Largura da faixa de imageamento	77 km
Armazenamento de dados	1.500 km de dados de imagem por órbita
Resolução temporal	Diária (fora do nadir)/5 dias e 12 horas (no nadir)
Capacidade de captura de imagens	4 milhões de km <sup>2</sup> por dia
Faixa dinâmica	12 bit

Adaptado de ZENG et al. (2011)

### 2.5.2 Landsat

O programa da série Landsat é o mais longo em imageamento espacial e o mais utilizado em pesquisas envolvendo uso da terra e atividade agrícola. Este satélite é bastante utilizado por possuir resolução espacial, espectral e temporal adequadas à identificação de alvos agrícolas e para monitorar as mudanças no uso da terra.

A série Landsat teve início em 1972 (COHEN; GOWARD, 2004; EPIPHANIO, 2002), sendo o primeiro veículo espacial implantado para o monitoramento terrestre, utilizado até hoje, com os sensores MSS e RBV a bordo dos satélites Landsat 1, 2 e 3. O sensor Thematic Mapper (TM), a bordo dos satélites Landsat 4 e Landsat 5, é discretizado em 8 bits e 7 bandas espectrais, sendo 3 na região do visível com resolução espacial de 30 m, 3 na

região do infravermelho próximo, também com 30 m e 1 na região do infravermelho termal, com resolução espacial de 120 m (Quadro 2).

Quadro 2 Especificações das bandas do sensor TM

<b>Banda espectral</b>	<b>Comprimento de onda (micrômetros)</b>	<b>Resolução espacial (metros)</b>	<b>Uso</b>
Banda 1 – azul-verde	0,45-0,52	30	Mapeamento batimétrico; distingue solo da vegetação, vegetação conífera de decídua.
Banda 2 – verde	0,52-0,61	30	Realça a vegetação, útil para avaliar o vigor da vegetação.
Banda 3 – vermelho	0,63-0,69	30	Realça os declives da vegetação.
Banda 4 – infravermelho	0,76-0,90	30	Realça o teor de biomassa e encostas.
Banda 5 – Infravermelho	1,55-1,75	30	Discrimina conteúdo de umidade do solo e da vegetação; capaz de penetrar nuvens finas.
Banda 6 – termal	10,4-12,5	120	Útil para mapeamento termal e estimativa da umidade do solo.
Banda 7 – Infravermelho	2,08-2,35	30	Útil para mapeamento de rochas alteradas hidrotermalmente, associadas com depósitos minerais.

Fonte: USGS (2014)

O sensor ETM+ (*Enhanced Thematic Mapper plus*), a bordo do Landsat 7, é semelhante ao TM, tendo, ainda, incluída, uma banda pancromática com resolução espacial de 15 m e a banda do infravermelho termal com resolução melhorada de 60 m. Atualmente, já está em operação o Landsat 8, utilizando o sensor *Operational Land Imager* (OLI) e o *Thermal Infrared Sensor* (TIRS) (NATIONAL AERONAUTICS AND SPACE ADMINISTRATION, 2012). As bandas espectrais do sensor OLI são similares ao sensor ETM+, porém,

aprimoradas em relação aos sensores anteriores. O Landsat 8 apresenta duas novas bandas espectrais, uma (Banda 1) na faixa do visível azul profundo, projetada especificamente para estudos hídricos e investigação da zona costeira e uma nova banda (Banda 9) para a detecção de nuvens densas (NATIONAL AERONAUTICS AND SPACE ADMINISTRATION, 2012). O instrumento TIRS tem duas bandas espectrais (Bandas 10 e 11) que são similares ao comprimento de onda coberto por uma única banda dos sensores anteriores TM e ETM+ (NATIONAL AERONAUTICS AND SPACE ADMINISTRATION, 2012).

Neste estudo foram utilizadas as imagens do sensor TM.

### **2.5.3 MODIS**

O sensor *Moderate Resolution Imaging Spectro-radiometer*, MODIS, também é bastante utilizado para estudos do sistema terrestres, sendo o mais abrangente a bordo dos satélites Terra e Aqua, lançados em 1999 e 2002, respectivamente (XIONG et al., 2009). O sensor MODIS foi desenvolvido e projetado para amplos estudos da Terra e suas mudanças temporais e geofísicas (XIONG et al., 2009). Consiste de um espectroradiômetro com 36 bandas, numa gama de variação espectral que vai de 0,4 a 14,4  $\mu\text{m}$  e resolução espacial de 250 m (2 bandas), 500 m (5 bandas) e 1.000 m (29 bandas). Apresenta resolução temporal de dois dias e resolução radiométrica de 12 bits (4.096 níveis de cinza), o que o torna o melhor instrumento orbital para o monitoramento de culturas agrícolas. Por questões de processamento, os dados são convertidos em 16 bits e assim disponibilizados para a comunidade científica (XIONG et al., 2009).

Segundo Soares et al. (2007), o MODIS gera produtos de alta qualidade, se comparado com outros sensores, graças à melhoria na qualidade dos detectores e no sistema de imageamento, maior número de bandas, além de

algoritmos específicos para a geração dos produtos. Outra vantagem do MODIS é a qualidade geométrica das imagens, aspecto fundamental para o monitoramento das modificações da superfície terrestre (SOARES et al., 2007). Os produtos MODIS fornecem base de dados para estudos relativos ao uso e à ocupação do solo, além de permitirem o monitoramento da cobertura vegetal do planeta, possibilitando identificar quais as mudanças da vegetação e suas relações com o clima e o tempo (JUSTICE et al., 2002).

Os produtos MODIS para ecossistemas terrestres disponíveis são: reflectância de superfície (MOD 09), temperatura de superfície (MOD 11), BRDF/Albedo (MOD 43), cobertura de neve e extensão de gelo (MOD 10 e 29) e índices de vegetação (MOD 13), em que dois índices de vegetação estão disponíveis, o *Normalized Difference Vegetation Index* (NDVI) e o *Enhanced Vegetation Index* (EVI), *Leaf Area Index* (LAI) (MOD 15), fotossíntese (MOD 17), cobertura da terra (MOD 12), anomalias térmicas (MOD 14) e conversão da cobertura vegetal (MOD 44) (SOARES et al., 2007). As especificações destes produtos estão listadas no Quadro 3.

Neste trabalho foram utilizadas as imagens do produto MOD13Q1.

## 2.6 Métodos de classificação de imagens

Os métodos de classificação de imagens de satélite podem ser agrupados em supervisionados e não supervisionados (TAN; LIM; JAFRI, 2011), isto vai depender de haver ou não presença de dados de verdade de campo inseridos como referência (XIE; SHA; YU, 2008).

A classificação não supervisionada é indicada para quando não se tem conhecimento das possíveis classes existentes na imagem, pois, neste caso, o próprio algoritmo de classificação define as classes e atribui cada um dos pixels a uma delas (TAN; LIM; JAFRI, 2011). Assim, os pixels cujas propriedades

espectrais são semelhantes formam grupos (ou clusters) no espaço multiespectral (RICHARDS, 1986). Na Figura 5 está ilustrado o agrupamento dos pixels em uma clusterização que utiliza somente duas bandas (5 e 7).

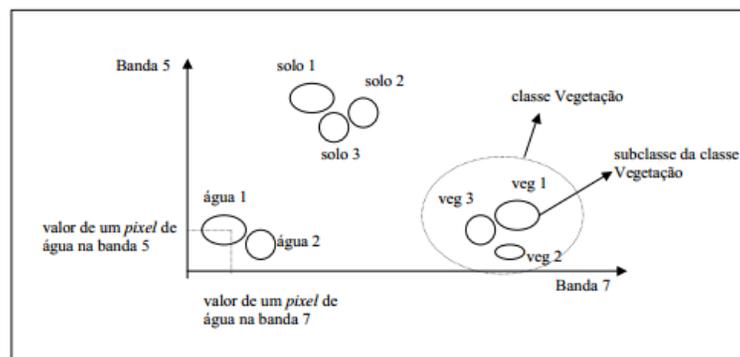


Figura 5 Classes e subclasses no espaço bidimensional

Adaptado de Richards (1986)

Nota-se que, em geral, uma determinada classe tem subclasses devido a influências topográficas, umidade ou características físico-químicas do alvo em questão (RICHARDS, 1986). No caso da vegetação, as subclasses podem surgir em função da diferença de umidade ou de biomassa (que difere conforme a idade das espécies vegetais), entre outras.

Na classificação supervisionada, as classes são definidas “a priori” e cada pixel da imagem é atribuído a uma determinada classe, conforme uma função discriminante preestabelecida (TAN; LIM; JAFRI, 2011). Por sua vez, os parâmetros que descrevem esta função são extraídos de um conjunto de amostras (denominado amostras de treinamento) que deve ser, também, previamente selecionado (SCHOWENGERDT, 2007).

Atualmente, existem muitos métodos de classificação automática disponíveis para imagens de sensoriamento remoto. Estes classificadores podem

ser divididos em paramétricos e não paramétricos. O modelo paramétrico utiliza parâmetros, como matriz de covariância e média, para definir as regras do classificador para cada classe, precisando de certa similaridade nas propriedades espectrais das classes. Portanto, é necessário que haja uma distribuição normal dos dados, para se obter um bom desempenho nas classificações (ANDRADE; FRANCISCO; ALMEIDA, 2013). O mais comumente usado é o MAXVER. Já os classificadores não paramétricos podem ser aplicados com distribuições que não obedecem aos parâmetros da curva normal (ANDRADE; FRANCISCO; ALMEIDA, 2013). Os mais comumente usados são árvore de decisão e *Support Vector Machines* (SVM). Alguns autores têm trabalhado com algoritmos de aprendizagem de máquina (AM) para o mapeamento do uso da terra. Poucos estudos trabalharam com este tipo de algoritmo para o mapeamento do café (SARMIENTO et al., 2014; ANDRADE et al., 2013; MARUJO et al., 2013; SANTOS et al., 2012).

### **2.6.1 Algoritmos de aprendizagem de máquina**

A aprendizagem de máquina (AM) é uma área de pesquisa cujo objetivo é o desenvolvimento de sistemas computacionais capazes de aprender e adquirir conhecimento de forma automática (MONARD; BARANAUSKA, 2003). Um algoritmo de aprendizado é um programa computacional capaz de tomar decisões baseadas em soluções de problemas anteriores. Os sistemas de aprendizado têm características que permitem sua classificação de acordo com a linguagem de descrição, o modo, o paradigma e a forma de aprendizado (CRACKNELL; READING, 2014). Técnicas de AM são cada vez mais empregadas para tratar problemas, por sua capacidade de aprender automaticamente a partir de grandes volumes de dados e produzir hipóteses úteis (MONARD; BARANAUSKA, 2003; CRACKNELL; READING, 2014).

Os algoritmos de aprendizagem de máquina (AM) são bem difundidos na comunidade científica e há uma vasta literatura sobre este assunto (DUDA; HART; STORK, 2000; BREIMAN, 2001; MONARD; BARANAUSKA, 2003; SCHNEIDER et al., 2009; SANTOS et al., 2012; ANDRADE et al., 2013; SANTOS, 2013; XU et al., 2013; CRACKNELL; READING, 2014; LI et al., 2014; SARMIENTO et al., 2014). Algumas técnicas de algoritmo de AM existentes são: as árvores de decisão, redes neurais, *Random Forest*, *Support Vector Machines*, *K-Nearest Neighbor* e *Naive Bayes*, entre outros. Neste estudo foram abordados cinco classificadores diferentes.

### 2.6.1.1 Árvore de decisão

Classificadores de árvore de decisão são utilizados com sucesso em diversas áreas do conhecimento, como diagnósticos médicos, reconhecimento de caracteres e sensoriamento remoto, entre outros. É um classificador não paramétrico e o processo de construção da árvore é apresentado por Breiman (1986).

A ideia básica envolvida em qualquer abordagem de múltiplos estágios é quebrar processos mais complexos em um conjunto de decisões mais simples, com base em um conjunto de testes em cada ramo (ou nó) da árvore (FRIEDL; BRODLEY, 1997), ou seja, a árvore de decisão consiste de uma hierarquia de nós internos e externos que são conectados por ramos. A árvore é composta de um nó chamado raiz (formado a partir de todos os dados), um conjunto de nós internos (*splits*) e um conjunto de nós terminais (folhas) (Figura 6).

Segundo Friedl e Brodley (1997), os classificadores em árvore têm vantagens expressivas para os problemas de classificação em sensoriamento remoto por causa da sua flexibilidade, por ser um método intuitivo, simples e pela sua eficiência computacional. Ainda segundo os autores, as árvores de

decisão são métodos estritamente não paramétricos e não requerem suposições sobre a distribuição dos dados de entrada. Além disso, esses métodos lidam com as relações não lineares entre as características e as classes e são capazes de lidar tanto com dados de entrada numéricos como categóricos. Vários trabalhos utilizam árvore de decisão para a classificação de imagens em sensoriamento remoto (PUNIA; HI; PORWAL, 2011; OTUKEI; BLASCHKE, 2010; SCHNEIDER; FRIEDL; POTERE, 2010; SCHNEIDER; FRIEDL; POTERE, 2009; FRIEDL; BRODLEY, 1997) e mostraram bons resultados de acerto no mapeamento de uso e ocupação do solo.

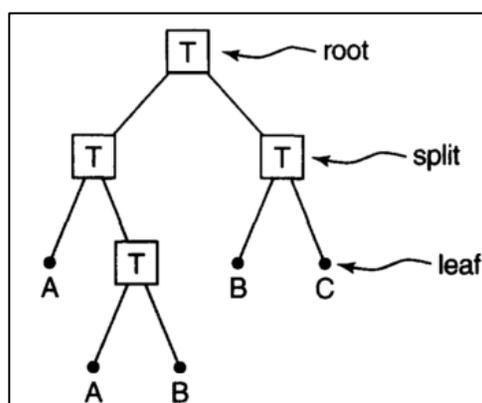


Figura 6 Classificador árvore de decisão. Testes (T) são aplicados para dividir os dados em grupos menores. As folhas A, B e C referem-se à classe atribuída a cada observação

Schneider, Friedl e Potere (2009) realizaram um mapeamento da distribuição da área urbana global, utilizando dados do satélite MODIS e utilizaram o algoritmo de classificação supervisionada por árvore de decisão que teve um desempenho bastante satisfatório de acurácia na classificação. O mapa obteve uma exatidão global de 93%, em nível de pixel e um elevado nível de acerto em escala municipal ( $R^2 = 0,90$ ). Em um estudo realizado em Délhi,

Índia, Punia, Hi e Porwal (2011) elaboraram um amplo mapeamento de uso e cobertura da terra, utilizando dados multitemporais e uma classificação com este algoritmo, apresentando bons resultados de acurácia, com um índice de acerto de 91%. Os dados temporais, juntamente com boas amostras de treinamento para o algoritmo, mostraram boa eficácia em separação das classes.

### 2.6.1.2 *Naive Bayes*

*Naive Bayes* (NB) é um algoritmo de aprendizagem estatístico recomendado como um classificador de nível base para a comparação com outros algoritmos (GUYON et al., 2009). É um modelo de classificação probabilística simples que assume a distribuição normal do espaço de características para cada classe (WIELAND; PITTORE, 2014). Isto significa que toda a função de distribuição de dados pressupõe independência entre as classes (CRACKNELL; READING, 2014; WIELAND; PITTORE, 2014) e a distribuição normal dos atributos para cada classe (WIELAND; PITTORE, 2014). O NB assume que a presença (ou ausência) de uma determinada classe não está relacionada com a presença (ou a ausência) de qualquer outra característica, dada a variável de classe (LI et al., 2013). Devido à natureza exata do modelo de probabilidade, este classificador pode ser eficientemente treinado em um ambiente de aprendizagem supervisionada.

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, 1995). A seguir, uma descrição matemática detalhada do modelo de classificação.

$$p(C_j|D) = \frac{(D|C_j)p(C_j)}{p(D)} \quad (8)$$

em que  $p(C_j|D)$  é a probabilidade da ocorrência de  $D$  estar na classe  $C_j$  (Figura 7).

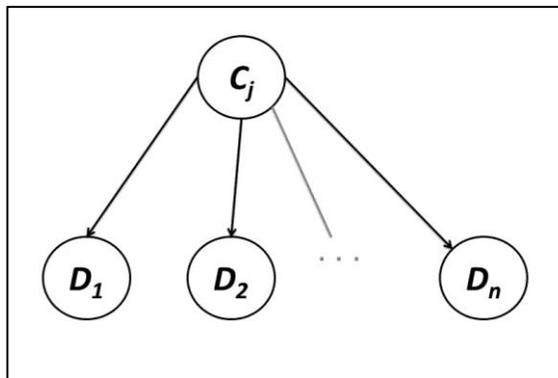


Figura 7 Representação gráfica do modelo de classificação *Naive Bayes*, em que é estimada a probabilidade de  $D_1, \dots, D_n$  pertencer à classe  $C_j$ .

Cracknell e Reading (2014), utilizando imagens Landsat TM e Palsar para mapear áreas geológicas, mostraram que o algoritmo NB não obteve os melhores resultados de acurácia. Para mapear áreas urbanas, Wieland e Pittore (2014), usando o algoritmo NB, obtiveram bom desempenho em algumas configurações, mas este algoritmo não tinha a capacidade de generalizar e apresentou comportamento instável sob variados cenários de treinamento para testes. Entretanto, Li et al. (2013) geraram uma classificação de diferentes tipos de zonas úmidas utilizando imagem Landsat TM e apresentaram bons índices de acurácia, usando este classificador.

### 2.6.1.3 *K-nearest neighbor*

O algoritmo *K-nearest neighbor* (KNN) foi desenvolvido por Fukunaga e Narendra (1975), sendo em uma técnica empregada no reconhecimento de padrões, baseando-se na técnica do vizinho mais próximo (*nearest neighbor*), e utiliza os ‘k’ vizinhos mais próximos do padrão de consulta, em vez de apenas o vizinho mais próximo (XU et al., 2013). O software exige a seleção de valores para o parâmetro k, sendo este o número de vizinhos considerado durante a classificação (XU et al., 2013).

KNN é um classificador em que o aprendizado é baseado na analogia. O conjunto de treinamento é formado por vetores n-dimensionais e cada elemento deste conjunto representa um ponto no espaço n-dimensional (DUDA; HART; STORK, 2000) (Figura 8).

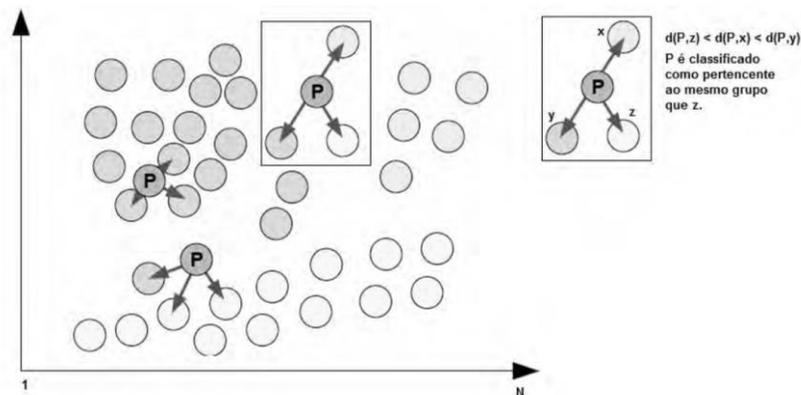


Figura 8 Classificação de objetos utilizando o KNN com  $k = 3$ . Cada cor representa uma classe e os círculos marcados com um P representam objetos desconhecidos a serem rotulados

Dado que os pixels (amostras) das imagens de sensoriamento remoto estão em algum espaço de características, o KNN usa uma medida de distância para definir a proximidade de uma amostra com a outra (DUDA; HART; STORK, 2000; SANTOS, 2013; CRACKNELL; READING, 2014; SARMIENTO et al., 2014). Este algoritmo classifica novas amostras de acordo com as  $K$  amostras mais próximas do conjunto de treinamento (SANTOS, 2013). Como exemplo, sendo  $\psi = \{1, \dots, C\}$  o conjunto de  $C$  classes existentes,  $x = \{x_1, \dots, x_n\} \in \mathcal{R}^{dxn}$  um conjunto de treinamento em um espaço  $d$ -dimensional e  $y = \{y_1, \dots, y_n\} \in \psi^n$  seus respectivos rótulos (classes). Dado um conjunto de teste não rotulado  $t = \{t_1, \dots, t_n\} \in \mathcal{R}^{dxn}$ , para classificá-lo, o algoritmo KNN calcula a distância de  $t_i$  para toda amostra no conjunto de treino  $x$ . A distância é determinada por uma métrica de similaridade, como, por exemplo, a distância euclidiana. Então, um rótulo  $y_i \in \psi$  é atribuído à  $t_i$ , considerando a classe mais frequente dentre as  $K$  amostras de treino mais próximas à  $t_i$ . Mitchell (1977) aponta algumas desvantagens no uso deste algoritmo, como baixa precisão em espaços de características muito elevadas e encontrar o valor ideal de  $K$ .

Gil et al. (2014), avaliando diversos classificadores para o mapeamento de espécies invasoras, nos Açores, em Portugal, utilizando imagens multiespectrais, demonstraram que os melhores resultados obtidos de acurácia foram com o KNN. Mapeando áreas cafeeiras em Campos Gerais, região sul do estado de Minas Gerais, Sarmiento et al. (2014) mostraram que o algoritmo KNN apresentou bons índices para estas áreas, com, aproximadamente, 90,00% de acerto. Todavia, Wieland e Pittore (2014) afirmam que o algoritmo KNN mostrou performances instáveis para classificar áreas urbanas.

### 2.6.1.4 *Random Forest*

A técnica *Random Forest* utiliza um conjunto de árvores de decisão  $\{DT(x, \theta_m), m = 1, \dots, B\}$ , em que  $\theta_m$  representa vetores aleatórios independentes identicamente distribuídos e  $x$  um padrão de entrada. Cada árvore dentro do conjunto é treinada em um subconjunto das amostras de treinamento original; além disso, a regra de divisão em cada *split* é determinada usando uma única característica (aleatoriamente selecionada) das variáveis de entrada. O número de variáveis é um parâmetro definido pelo usuário (muitas vezes dito ser o único parâmetro ajustável), porém, o algoritmo não é sensível a ele (GISLASON; BENEDIKTSSON; SVEINSSON, 2004; BREIMAN, 2001) (Figura 9).

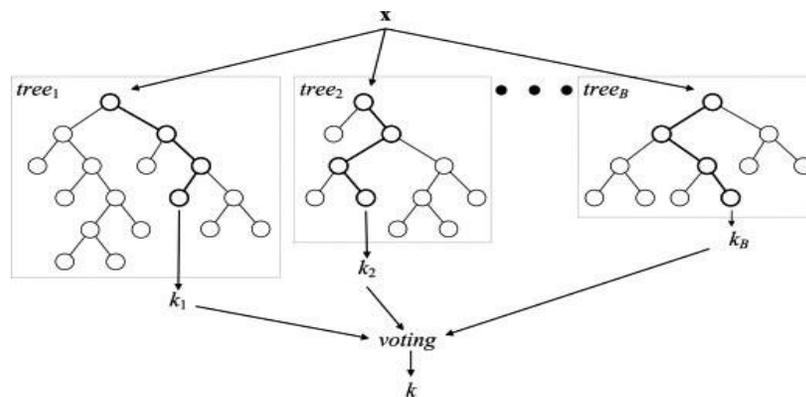


Figura 9 Representação gráfica do modelo de classificação *Random Forest*, sendo  $x$  os dados de entrada (amostras de treinamento) que são classificadas por todas as árvores ( $tree_1, tree_2, \dots, tree_B$ ) que contribuem para a classificação com um voto, se correta

Desde os primeiros trabalhos com árvore de decisão (BREIMAN, 2001), o *Random Forest* tem mostrado resultados promissores em diversas áreas, como genética (DIAZ-URIARTE; ANDRES, 2006), ecologia (CUTLER et al., 2007) e

sensoriamento remoto (STUMPF; KERLE, 2011; WATTS et al., 2009; GISLASON; BENEDIKTSSON; SVEINSSON, 2006).

Guo et al. (2011) enumeram algumas vantagens desse classificador, tais como rápida execução, não sofre com *overfitting* e é eficiente, mesmo com grandes bases de dados, como, por exemplo, dados de LIDAR. Ele não requer suposições sobre a distribuição dos dados, o que é interessante quando se utilizam características de diferentes tipos ou escalas. Essas são características que o tornam adequado para classificações em sensoriamento remoto e uma interessante ferramenta para ser combinada com a abordagem orientada a objetos.

Stumpf e Kerle (2011) aplicaram o método de *Random Forest* como suporte à abordagem orientada a objeto, para a classificação de deslizamentos de terra em imagens de altíssima resolução espacial (Quickbird, IKONOS, Geoeye-1 e fotografias aéreas), no Haiti, na China, na França e na Itália. Os autores observaram que, empregando-se, aproximadamente, 20% dos dados como dados de treinamento para o algoritmo, obtiveram, em valores de exatidão, entre 73% e 87%, para as áreas afetadas pelos deslizamentos de terra.

#### **2.6.1.5 Support Vector Machines**

Os *Support Vector Machines* (SVM) são um conjunto de algoritmos de aprendizagem utilizados para a classificação e regressão. Como os classificadores de árvore de decisão, o SVMs também são classificadores não paramétricos. A teoria do SVM foi originalmente proposta por Vapnik e Chervonenkis (1971) e, mais tarde, discutida em detalhe por Vapnik (1999). O sucesso deste algoritmo depende de quão bem o processo é treinado. A maneira mais fácil para treinar o SVM é usando classes linearmente separáveis (Figura 10).

De acordo com Osuna et al. (1997), o treinamento de dados com o número  $k$  de amostras é representado como  $\{X_i, y_i\}$ ,  $i = 1, \dots, k$ , em que  $X \in \mathbb{R}^N$  é um espaço  $n$ -dimensional e  $y \in \{-1, 1\}$  é um rótulo de classe. Em seguida, essas classes são consideradas linearmente separáveis se existe um vector  $W$  perpendicular ao hiperplano linear (que determina a direção do plano discriminador) e uma escala  $b$  mostrando o deslocamento da discriminação do hiperplano a partir da origem. Para as duas classes, ou seja, a classe 1 representada como  $-1$  e classe 2 representada como  $+1$ , dois hiperplanos podem ser utilizados para discriminar os pontos de dados nas respectivas classes. São expressos como

$WX_i + b \geq +1$  para todos em que  $y = +1$ , ou seja, um número para a classe 1

$WX_i + b \leq -1$  para todos em que  $y = -1$ , ou seja, um número para a classe 2.

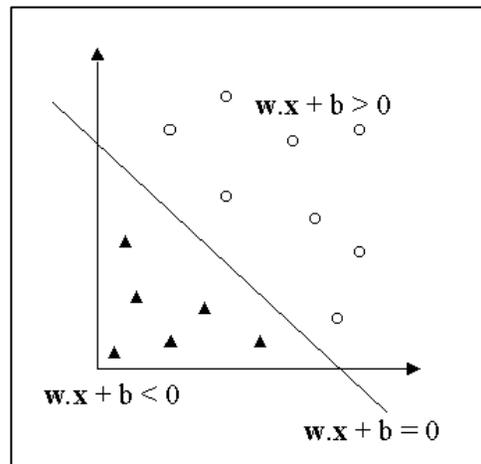


Figura 10 Conjunto de dados linearmente separáveis e um possível hiperplano separador

Adaptado de Huang, Davis e Townshed (2002)

Os dois hiperplanos são selecionados de modo a não só maximizar a distância entre as duas classes de dados, mas também para não incluir quaisquer

pontos entre elas (HUANG; DAVIS; TOWNSHED, 2002). O objetivo geral é descobrir em que classe os novos pontos de dados cairão. Em geral, os SVM são relatados para produzir resultados de maior precisão, em comparação com as abordagens tradicionais, mas o resultado depende da função de kernel utilizada, da escolha de parâmetros do kernel e do método utilizado para SVM gerado (HUANG ; DAVIS; TOWNSHED, 2002).

Huang, Davis e Townshed (2002) compararam a precisão da classificação baseada em pixel usando quatro diferentes algoritmos de classificação: *support vector machines*, árvores de decisão, rede neural e o classificador de máxima verossimilhança (MLC). Os resultados mostraram que a precisão do classificador *support vector machines* superou os outros três algoritmos de classificação. Pal et al. (2005) compararam a precisão de *support vector machines* e *random forest*, usando imagem Landsat 7 ETM+ em uma classificação baseada em pixel, obtendo bons resultados para as duas abordagens de classificação. Otakei e Blaschke (2010) mapearam o uso da terra, comparando os algoritmos de classificação *support vector machines*, árvore de decisão e máxima verossimilhança, usando imagens Landsat ETM+ de dois anos diferentes, em que a acurácia do mapeamento dos algoritmos ultrapassou 85% de acerto.

## **2.7 Aplicações para cafeicultura**

Áreas cafeeiras são extremamente difíceis de serem mapeadas, principalmente pela grande variabilidade espectral e semelhança espectral que esta cultura tem com outros tipos de cobertura da terra, como é o caso de áreas de vegetação nativa e pastagem (VIEIRA et al., 2006; SARMIENTO et al., 2014; MOREIRA et al., 2004).

Algoritmos de AM têm sido amplamente utilizados em sensoriamento remoto, principalmente no mapeamento do uso e da cobertura da terra (DUDA; HART; STORK, 2000; PAL et al., 2005; GISLASON; BENEDIKTSSON; SVEINSSON, 2006; OTUKEI; BLASCHKE, 2010; XU et al., 2013; CRACKNELL; READING, 2014; LI et al., 2014; WIELEND; PITTORE, 2014), apresentando resultados acurados no processo de classificação de imagens. Porém, para o mapeamento de áreas cafeeiras, estes algoritmos foram pouco utilizados (SANTOS et al., 2012; ANDRADE et al., 2013; SARMIENTO et al., 2014) , podendo ser uma alternativa promissora para a classificação dessas áreas.

Além disso, novas abordagens para o mapeamento desta cultura podem ser promissoras na separação entre diferentes classes de uso e cobertura da terra, como a utilização de dados oriundos de séries temporais. Poucos estudos utilizaram estas variáveis no processo de classificação (JIA et al., 2014; ZHU; WOODCOCK, 2014), obtendo bons índices de acurácia no mapeamento da cobertura da terra. No entanto, estas variáveis ainda não foram utilizadas no mapeamento de áreas cafeeiras.

Sendo assim, aliar algoritmos de AM, juntamente com variáveis provenientes de séries temporais, pode auxiliar de forma positiva o mapeamento do café, conseguindo separar adequadamente os diferentes tipos de uso e cobertura da terra.

## REFERÊNCIAS

ACERBI-JUNIOR, F. W.; CLEVERS, J. G. P. W.; SCHAEPMAN, M. E. The assessment of multi-sensor image fusion using wavelet transforms for mapping the Brazilian Savanna. **International Journal of Applied Earth Observation and Geoinformation**, Enschede, v. 8, n. 4, p. 278-288, Dec. 2006.

ADAMI, M. et al. Avaliação da exatidão do mapeamento da cultura do café no Estado de Minas Gerais. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 14., 2009, Natal. **Anais...** São José dos Campos: INPE, 2009. p. 1-8.

ANDRADE, A. C.; FRANCISCO, C. N.; ALMEIDA, C. M. Desempenho de classificadores paramétrico e não paramétrico na classificação da fisionomia vegetal. **Revista Brasileira de Cartografia**, Rio de Janeiro, v. 2, n. 65, p. 227-242, mar./abr. 2013.

ANDRADE, L. N. et al. Aplicação de redes neurais artificiais classificação de áreas cafeeiras da região de Machado (MG). **Coffee Science**, Lavras, v. 8, n. 1, p. 78-80, 2013.

BAI, J.; PERRON, P. Computation and analysis of multiple structural change models. **Journal of Applied Econometrics**, Chichester, v. 18, n. 1, p. 1–22, Jan./Feb. 2003.

BENNETT, M, et al. Inositol pyrophosphates: metabolism and signaling. **Cellular and Molecular Life Sciences**, Basel, v. 63, n. 5, p. 552-64, Mar. 2006.

BISHOP, C. M. **Neural networks for pattern recognition**. Oxford: Oxford University Press, 1995.

BREIMAN, L. Bagging predictors. **Machine Learning**, Boston, v. 24, n. 2, p. 123–140, Aug. 1996.

BREIMAN, L. Random Forests. **Machine Learning**, Boston, v. 45, n. 1, p. 5–32, Oct. 2001.

BRUZZONE, L.; SMITS, P. C. **Analysis of multi-temporal remote sensing images: proceedings of multitemp 2001**. Singapore: World Scientific, 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. Detecção de modificações na cobertura do solo. In: CARVALHO, L. M. T.; SCOLFORO, J. R.(Ed.). **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, June 2004.
- COLDITZ, R. R. et al. TiSeG : a flexible software tool for time-series generation of MODIS data utilizing the quality assessment science data set. **IEEE Transactions on Geoscience and Remote Sensing**, New York, v. 46, n. 10, p. 3296-3308, Oct. 2008.
- 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.
- CORDERO-SANCHO, S.; SADER, S. A. Spectral analysis and classification accuracy of coffee crops using Landsat and a topographic-environmental model. **International Journal of Remote Sensing**, Basingstoke, v. 28, n. 7-8, p. 1577–1593, 2007.
- 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.
- CUTLER, D. R. et al. Random forest for classification in ecology. **Ecology**, Washington, v. 88, n. 11, p. 2783–2792, Nov. 2007.

- DAMATTA, F. M. et al. Ecophysiology of coffee growth and production. **Brazilian Journal of Plant Physiology**, Londrina, v. 19, n. 4, p. 485-510, Oct./Dec. 2007.
- DAMATTA, F. M.; RAMALHO, J. D. C. Impacts of drought and temperature stress on coffee physiology and production: a review. **Brazilian Journal of Plant Physiology**, Londrina, v. 18, n. 1, p. 55-81, Jan./Mar. 2006.
- DEMIR, B.; BOVOLO, F.; BRUZZONE, L. Detection of land-cover transitions in multitemporal remote sensing images with active-learning-based compound classification. **IEEE Transactions on Geoscience and Remote Sensing**, New York, v. 50, n. 5, p. 1930–1941, May 2012.
- DEMIR, B.; PERSELLO, C.; BRUZZONE, L. Batch-mode active-learning methods for the interactive classification of remote sensing images. **IEEE Transactions on Geoscience and Remote Sensing**, New York, v. 49, n. 3, p. 1014–1031, Mar. 2011.
- DIAZ-URIARTE, R.; ANDRES, S. A. de. Gene selection and classification of microarray data using random forest. **BMC Bioinformatics**, London, v. 7, n. 3, p. 01-13, Jan. 2006.
- DUDA, R.; HART, P.; STORK, D. **Pattern classification and scene analysis**. 2. ed. New York: John Wiley & Sons, 2000.
- ECCARDI, F.; SANDAJL, V. **O café: ambientes e diversidade**. Rio de Janeiro: Casa da Palavra, 2003.
- EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA. Histórico. **Embrapa Café**, Brasília, 2013. Disponível em: <<http://www22.sede.embrapa.br/cafe/unidade/historico.htm>>. Acesso em: 20 set. 2013.
- EPIPHANIO, J. C. N. **Satélites de sensoriamento remoto**. São José dos Campos: INPE, 2002. Disponível em: <[http://www.dsr.inpe.br/vcsr/files/capitulo\\_2.pdf](http://www.dsr.inpe.br/vcsr/files/capitulo_2.pdf)>. Acesso em: 22 set. 2013.
- EPIPHANIO, R. D. V. et al. Estimating soybean crop areas using spectral–temporal surfaces derived from MODIS images in Mato Grosso, Brazil. **Pesquisa Agropecuária Brasileira**, Brasília, v. 45, n. 1, p. 72–80, jan. 2010.

FERRÃO, M. A. G. et al. Origem, dispersão geográfica, taxonomia e diversidade genética de *Coffea canephora*. In: FERRÃO, R. G. et al. (Ed.). **Café conilon**. Vitória: Incaper, 2007. p. 66-91.

FORKEL, M. et al. Trend change detection in NDVI time series: effects of inter-annual variability and methodology. **Remote Sensing**, Nova Zelândia, 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**, Elmsford, v. 61, n. 3, p. 399-409, Sept. 1997.

FUKUNAGA, K.; NARENDRA, P. M. A branch and bound algorithms for computing k-nearest neighbors. **IEEE Transactions on Computers**, New York, v. 24, n. 7, p. 750-753, July 1975.

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. 3287-3319, Oct. 2006.

GIL, A. et al. Using aster multispectral imagery for mapping woody invasive species in Pico da Vara natural reserve (Azores Islands, Portugal). **Revista Árvore**, Viçosa, v. 38, n. 3, p. 391-401, maio/jun. 2014.

GISLASON, P. O.; BENEDIKTSSON, J. A.; SVEINSSON, J. R. Random forests for land cover classification. **Pattern Recognition Letters**, Amsterdam, v. 27, n. 4, p. 294–300, Mar. 2006.

GISLASON, P. O.; BENEDIKTSSON, J. A.; SVEINSSON, J. R. Random forests classification of multisource remote sensing and geographic data. In: *Geoscience and Remote Sensing Symposium*, 4., 2004, Reykjavik. **Proceedings...** Reykjavik: IEEE, 2004.

GRANER, E. A.; GODOY-JUNIOR, C. **Manual do cafeicultor**. São Paulo: Universidade de São Paulo, 1967.

GUO, L. et al. Relevance of airborne lidar and multispectral image data for urban scene classification using random forests. **ISPRS Journal of Photogrammetry and Remote Sensing**, Amsterdam, v. 66, n. 1, p. 56–66, Jan. 2011.

GUYON, I. et al. A practical guide to model selection. In: MARIE, J. (Ed.). **Proceedings of the machine learning summer school**. Canberra: Springer, 2009. p. 27-43.

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.

HOUGHTON, R. A.; GOODALE, C. L. **Effects of land-use change on the carbon balance of terrestrial ecosystems**. Washington: American Geophysical Union, 2004.

HUANG, C.; DAVIS, L. S.; TOWNSHED, J. R. G. An assessment of support vector machines for land cover classification. **International Journal of Remote Sensing**, Basingstoke, v. 23, n. 4, p. 725-749, 2002.

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

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Banco de dados agregados: sistema de recuperação automática**. Brasília: IBGE, 2009. Disponível em: <<http://www.sidra.ibge.gov.br/>>. Acesso em: 20 set. 2013.

INTERNACIONAL COFFEE ORGANIZATION. **The world economic crisis and the coffee sector**. [S.l.: sn], 2009. Disponível em: <<http://www.ico.org/documents/ed-2059e-economic-crisis.pdf>>. Acesso: 20 de setembro de 2013.

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, supl. 1, p. 03-10, 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.

JÖNSSON, P.; EKLUNDH, L. Seasonality extraction by function fitting to time-series of satellite sensor data. **IEEE Transactions on Geoscience and Remote Sensing**, New York, v. 40, n. 8, p. 1824-1832, Aug. 2002.

JULIEN, Y.; SOBRINO, J. A.; VERHOEF, W. Changes in land surface temperatures and NDVI values over Europe between 1982 and 1999. **Remote Sensing of Environment**, New York, v. 103, n. 1, p. 43-55, July 2006.

JUSTICE, C. O. et al. An overview of MODIS Land data processing and product status. **Remote Sensing of Environment**, New York, v. 83, n. 1-2, p. 3-15, Nov. 2002.

KENNEDY, R. E.; COHEN, W. B.; SCHROEDER, T. A. Trajectory-based change detection for automated characterization of forest disturbance dynamics. **Remote Sensing of Environment**, New York, v. 110, n. 3, p. 370-386, Oct. 2007.

KENNEDY, R. E.; YANG, Z.; COHEN, W. B. Detecting trends in forest disturbance and recovery using yearly landsat time series: 1. LandTrendr — temporal segmentation algorithms. **Remote Sensing of Environment**, New York, v. 114, n. 12, p. 2897-2910, 15 Dec. 2010.

LAMBIN, E. F.; LINDERMAN, M. Time series of remote sensing data for land change science. **IEEE Transactions on Geoscience and Remote Sensing**, New York, v. 44, n. 7, p. 1926–1928, July 2006.

LANDGREBE, D. A. Multispectral land sensing: where from, where to? **IEEE Transactions on Geoscience and Remote Sensing**, New York, v. 43, n. 3, p. 414–421, Mar. 2005.

LEE, J.; ERSOY, O. K. Consensual and hierarchical classification of remotely sensed multispectral images. **IEEE Transactions on Geoscience and Remote Sensing**, New York, v. 45, n. 9, p. 2953–2963, Sept. 2007.

LI, C. et al. Comparison of classification algorithms and training sample sizes in urban land classification with landsat thematic mapper imagery. **Remote Sensing**, Nova Zelândia, v. 6, n. 2, p. 964–983, Jan. 2014.

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, n. 1, p. 1–16, Jan. 2013.

LI, S.; KWOK, J. T.; WANG, Y. Using the discrete wavelet transform to merge Landsat TM and SPOT panchromatic images. **Information Fusion**, Amsterdam, v. 3, n. 1, p. 17–23, Mar. 2002.

LU, D. et al. Change detection techniques. **International Journal of Remote Sensing**, Basingstoke, v. 25, n. 12, p. 2365–2407, June 2004.

MACHADO, M. L. et al. Mapeamento de áreas cafeeiras (*Coffea arabica L.*) da Zona da Mata Mineira usando sensoriamento remoto. **Coffee Science**, Lavras, v. 5, n. 2, p. 113-122, maio/ago. 2010.

MACHADO, M. L. et al. Caracterização do meio físico de agroecossistemas cafeeiros da Zona da Mata de Minas Gerais usando técnicas de sensoriamento remoto e sistemas de informação geográficas. In: SIMPÓSIO DE PESQUISA DOS CAFÉS DO BRASIL, 2., 2001, Vitória. **Anais...** Vitória: Embrapa Café, 2001. p. 2399-2409.

MARTÍNEZ, B.; GILBERT, 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.

MARTÍNEZ-VERDUZCO, G. C.; GALEANA-PIZAÑA, J. M.; CRUZ-BELLO, G. M. Coupling community mapping and supervised classification to discriminate shade coffee from natural vegetation. **Applied Geography**, Oxford, v. 34, p. 1-9, May 2012.

MARUJO, R. F. B. et al. Classificação orientada a objetos aplicada á cultivos cafeeiros em Três Pontas MG. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 16., 2013, Foz do Iguaçu. **Anais...** Foz do Iguaçu: INPE, 2013.

MASEK, J. G. et al. A landsat surface reflectance dataset for North America, 1990-2000. **IEEE Geoscience and Remote Sensing Letters**, New York, v. 3, n. 1, p. 68-72, 2006.

MELLO, M. P. et al. STARS: a new method for multitemporal remote sensing. **IEEE Transactions on Geoscience and Remote Sensing**, New York, v. 51, n. 4, p. 1-17, Apr. 2012.

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

MITCHELL, T. Version spaces: a candidate elimination approach to rule learning. Proceedings In: INTERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE, 15., 1997, Massachusetts. **Proceedings...** Massachusetts: Academic Press, 1977.

MONARD, M. C.; BARANAUSKA, J. A. Indução de regras e árvores de decisão. In: REZENDE, S. O. (Org.). **Sistemas inteligentes-fundamentos e aplicações**. Barueri: Manole, 2003. p. 115-139.

MORAES, E. C. **Sensoriamento remoto no estudo da vegetação:** diagnosticando a Mata Atlântica. In: RUDORFF, B. F. T. et al. Curso de uso do sensoriamento remoto no estudo do meio ambiente. São José dos Campos: INPE, 2002. Cap. 8, p. 27-42.

MOREIRA, M. A. et al. Geotecnologias para mapear lavouras de café nos estados de Minas Gerais e São Paulo. **Engenharia Agrícola**, Jaboticabal, v.30, n.6, p.1123-1135, nov./dez. 2010.

MOREIRA, M. A. et al. Tecnologia de informação: imagens de satélite para o mapeamento de áreas de café de Minas Gerais. **Informe Agropecuário**, Belo Horizonte, v. 28, n. 241, p. 27-37, nov./dez. 2007.

MOREIRA, M. A.; ADAMI, M.; RUDORFF, B. F. T. Análise espectral e temporal da cultura do café em imagens Landsat. **Pesquisa Agropecuária Brasileira**, Brasília, v. 39, n. 3, p. 223-231, mar. 2004.

MOREIRA, M. A.; BARROS, M. A.; RUDORFF, B. F. T. Geotecnologias no mapeamento da cultura do café em escala municipal. **Sociedade & Natureza**, Uberlândia, v. 20, n. 1, p. 101-110, jun. 2008.

MORETTIN, P. A.; TOLOI, C. M. C. Preliminares. In: MIGON, H. **Análise de séries temporais**. São Paulo: Edgard Blücher, 2004. p. 1-20.

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION. About landsat. **Landsat Science**, Washington, 2012. Disponível em: <<http://landsat.gsfc.nasa.gov/about/tm.html>>. Acesso em: 24 out. 2014.

OHMANN, J. L. et al. Mapping change of older forest with nearest-neighbor imputation and Landsat time-series. **Forest Ecology and Management**, Amsterdam, v. 272, p. 13-25, May 2012.

OLIVEIRA, T. et al. NDVI time series for mapping phenological variability of forests across the Cerrado Biome. In: ZHANG, X. **Phenology and climate**. Croatia: InTech, 2012. Cap. 14, p. 253-272.

OSUNA, E. E. et al. **Support vector machines: training and applications**. Massachusetts: Massachusetts Institute of Technology and Artificial Intelligence Laboratory, 1997.

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. 27–31, Feb. 2010.

PAL, M. et al. Random forest classifier for remote sensing classification. **International Journal of Remote Sensing**, Basingstoke, v. 26, n. 1, p. 217–222, Jan. 2005.

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.

RAPIDEYE, A. G. RapidEye standard image specifications: version 3.2. [S.l: s.n], 2011. Disponível em: <[http://www.rapideye.de/upload/RE\\_Product\\_Specifications\\_ENG.pdf](http://www.rapideye.de/upload/RE_Product_Specifications_ENG.pdf)>. Acesso em: 15 set. 2013.

RICHARDS, J. A. **Remote sensing digital image analysis: an introduction**. New Yourk, Springer-Verlag, 1986.

SALMON, B. P. et al. Unsupervised land cover change detection: meaningful sequential time series analysis. *IEEE Journal of Selected Topics in Applied Earth Observations*, Washington, v. 4, n. 2, p. 327–335, June 2011.

SANTOS, A. B. **Combinação de múltiplas abordagens de classificação para interpretação de imagens hiperespectrais de sensoriamento remoto**. 2013. 103 p. Dissertação (Mestrado Ciências da Computação) - Universidade Federal de Minas Gerais, Belo Horizonte, 2013.

SANTOS, J. A. et al. Multiscale classification of remote sensing images. **IEEE Transactions on Geoscience and Remote Sensing**, New York, v. 50, n. 10, p. 3764-3775, Oct. 2012.

SARMIENTO, C. M. et al. Comparação de classificadores supervisionados na discriminação de áreas cafeeiras em Campos Gerais – Minas Gerais. **Coffee Science**, Lavras, v. 9, n. 4, p. 546- 557, out./dez. 2014.

SCHNEIDER, A.; FRIEDL, M. A.; POTERE, D. A new map of global urban extent from MODIS satellite data. **Environmental Research Letters**, Easton, v. 4, p. 1-11, 2009.

SCHNEIDER, A.; FRIEDL, M. A.; POTERE, D. Mapping global urban areas using MODIS 500-m data: new methods and datasets based on ‘urban ecoregions’. **Remote Sensing of Environment**, New York, v. 114, n. 8, p. 1733–1746, Aug. 2010.

SCHOWENGERDT, R. A. **Remote sensing models and methods for image processing**. San Diego: Academic Press, 2007.

SCHROEDER, T. A. et al. Mapping wildfire and clearcut harvest disturbances in boreal forests with Landsat time series data. **Remote Sensing of Environment**, New York, v. 115, n. 6, p. 1421-1433, June 2011.

SCHUSTER, C.; FÖRSTER, M.; KLEINSCHMIT, B. Evaluation of the RapidEye Red edge channel for improving land-use classifications. In: KOHLHOFER, G.; FRANZEN, M. (Hrsg.). TAGUNGSBAND DREILÄNDERTAGUNG OVG, DGPF UND SGPF, 30., 2010, Österreich. **Wissenschaftlich-Technische Jahrestagung der DGPF**. Österreich: Technische Universität Wien, 2010.

SINGH, A. Digital change detection techniques using remotely sensed data. **International Journal of Remote Sensing**, Basingstoke, v. 10, n. 6, p. 989–1003, 1989.

SMITS, P.; BRUZZONE, L. **Analysis of multi-temporal remote sensing images**: proceedings of multitemp 2003. Singapore: World Scientific, 2004.

SOARES, J. V. et al. **O sensor MODIS e suas aplicações ambientais no Brasil**. São José dos Campos: Parêntese, 2007.

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.

STUMPF, A.; KERLE, N. Object-oriented mapping of land slides using random forests. **Remote Sensing of Environment**, New York, v. 115, n. 10, p. 2564–2577, Oct. 2011.

TAN, K. C.; LIM, H. S.; JAFRI, M. Z. M. Comparison of neural network and maximum likelihood classifiers land cover classification using landsat multispectral data. In: INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS CONFERENCE, 2., 2011, Langkawi. **Proceedings...** Langkawi: IEEE, p. 241-244, 2011.

TEGGI, S.; CECCHI, R.; SERAFINI, F. TM and IRS-1C-PAN datafusion using multiresolution decomposition methods based on the ‘a trous’ algorithm. **International Journal of Remote Sensing**, Basingstoke, v. 24, n. 6, p. 1287–1301, 2003.

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**, New York, v. 5, n. 9, p. 4255-4279, Sept. 2013.

TRABAQUINI, K. et al. Caracterização de lavouras cafeeiras, utilizando técnicas de geoprocessamento e sensoriamento remoto, no município de Umuarama – PR. **Ciência e Agrotecnologia**, Lavras, v. 35, n. 1, p. 35-44, jan./fev. 2011.

VAPNIK, W. N. An overview of statistical learning theory. **IEEE Transactions of Neural Networks**, Amsterdam, v. 10, n. 5, p. 988–999, Sept. 1999.

VAPNIK, W. N.; CHERVONENKIS, A. Y. On the uniform convergence of the relative frequencies of events to their probabilities. **Theory of Probability and its Applications**, Philadelphia, v. 16, n. 2, p. 264–280, 1971.

VELLOSO, M. H. **Coffe inventory through orbital imagery**. Rio de Janeiro: IBC, 1974.

VELLOSO, M. H.; SOUZA, D. D. **Sistema automático de inventário cafeeiro**. Rio de Janeiro: IBC, 1976.

VELLOSO, M. H.; SOUZA, D. D. **Trabalho experimental de inventariação automática de cafezais utilizando imagens orbitais e o equipamento image-100**. Rio de Janeiro: IBC, 1978.

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, Aug. 2012.

VERHOEF, W.; MENENTI, M.; AZZALI, S. A colour composite of NOAA–AVHRR–NDVI based on time series analysis (1981–1992). **International Journal of Remote Sensing**, Basingstoke, v. 17, n. 2, p. 231-235, 1996.

VIEIRA, C. A. O. **Accuracy of remotely sensing classification of agricultural crops: a comparative study**. 2000. 352 p. Thesis (Doctor) - University of Nottingham, Nottingham, 2000.

VIEIRA, T. G. C. et al. Análise espacial do parque cafeeiro da região de Três Pontas MG. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 14., 2009, Natal. **Anais...** São José dos Campos: MCT, 2009. p. 6361-6368.

VIEIRA, T. G. C. et al. Crop Parameters for the evaluation of the spectral response of coffee (*Coffea Arabica L.*) areas within the State of Minas Gerais, Brazil. **Coffee Science**, Lavras, v. 1, n. 2, p. 111-118, jul./dez. 2006.

VIEIRA, T. G. C.; ALVES, H. M. R.; LACERDA, M. P. Parâmetros culturais para avaliação do comportamento espectral da cultura do café (coffea arábica) em Minas Gerais. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 11., 2003, Belo Horizonte. **Anais...** Belo Horizonte: Embrapa, 2003. p. 247-254.

WATTS, J. D. et al. Monitoring of cropland practices for carbon sequestration purposes in north central Montana by Landsat remote sensing. **Remote Sensing of Environment**, New York, v. 113, n. 9, p. 1843–1852, Sept. 2009.

WIELAND, M.; PITTORE, M. Performance evaluation of machine learning algorithms for urban pattern recognition from multi-spectral satellite images. **Remote Sensing**, Nova Zelândia, v. 6, n. 4, 2912–2939, Mar. 2014.

XIAO, X. et al. Detecting leaf phenology of seasonally moist tropical forests in South America with multi-temporal MODIS images. **Remote Sensing of Environment**, New York, v. 103, n. 4, p. 465-473, Aug. 2006.

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, 2008.

XIONG, L. et al. Deficient SUMO attachment to Flp recombinase leads to homologous recombination-dependent hyperamplification of the yeast 2 microm circle plasmid. **Molecular Biology of the Cell**, Bethesda, v. 20, n. 4, p. 1241-51, Feb. 2009.

XU, Y. et al. Coarse to fine K nearest neighbor classifier. **Pattern Recognition Letters**, North-Holland, v. 34, p. 980-986, Feb. 2013.

ZEILEIS, A. et al. Testing and dating of structural changes in practice. **Computational Statistics & Data Analysis**, Amsterdam, v. 44, n. 1-2, p. 109–123, Oct. 2003.

ZENG, Y.; ZHANG, J.; WANG, G.; LI, Y. RapidEye satellite image quality analysis and solutions for its true color composition. In: INTERNATIONAL WORKSHOP, 2011, Xiamen. **Proceedings...** Xiamen: IEEE, 2011.

ZHU, X. et al. An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions. **Remote Sensing of Environment**, New York, v. 114, n. 11, p. 2610-2623, Nov. 2010.

## SEGUNDA PARTE

### ARTIGO 1

(Artigo submetido de acordo as normas da revista PAB)

**Diferentes algoritmos de aprendizagem de máquina e variáveis de sensoramento remoto para o mapeamento da cafeicultura**

**Carolina Gusmão Souza <sup>(1)</sup>, Luis Carvalho <sup>(1)\*</sup>, Polyanne Aguiar <sup>(1)</sup> e Tássia Borges Arantes <sup>(1)</sup>**

<sup>(1)</sup>Departamento de Ciência Florestal, Universidade Federal de Lavras;  
carolinagusmaosouza@gmail.com, passarinho@dcf.ufla.br,  
polyanneaguiar@gmail.com, barantes@gmail.com.

\*Autor de correspondência: Luis Carvalho, passarinho@dcf.ufla.br

### RESUMO

A cafeicultura é uma das principais culturas agrícolas do Brasil, portanto, realizar o mapeamento e o monitoramento desta cultura é fundamental. Porém, mapear esta cultura utilizando imagens de sensoramento remoto não é uma tarefa fácil, pois existe grande similaridade espectral com outros tipos de cobertura da terra. Sendo assim, este trabalho foi realizado com o objetivo de comparar o uso de diferentes variáveis e algoritmos de classificação para o mapeamento de áreas cafeeiras. O trabalho foi desenvolvido em três áreas ambientalmente diferentes e bastante significativas na produção de café. Foram utilizados 5 algoritmos de aprendizagem de máquinas e 7 combinações de variáveis espectrais, texturais e geométricas associadas ao processo de classificação. Um total de 105 classificações foram realizadas, sendo 35 classificações para cada uma das áreas. Todas as classificações que não usaram variáveis espectrais não resultaram em bons índices de acurácia. Nas três áreas, o algoritmo que apresentou as melhores acurácias foi o *Support vector machine*,

com acurácia global de 85,33% em Araguari, 87,00% em Carmo de Minas e 88,33% em Três Pontas. Os piores resultados foram encontrados para o algoritmo *Random Forest* em Araguari, com acurácia global de 76,66% e para o *Naive Bayes* em Carmo de Minas e Três Pontas, com 76,00% e 82,00% de acerto. Nas três áreas, variáveis texturais, quando associadas às espectrais, melhoraram a acurácia da classificação. O *Support vector machine* apresentou o melhor desempenho para as três áreas.

**Termos para indexação:** classificação do café; sensoriamento remoto; algoritmos de aprendizagem de máquina.

### **Use of different algorithms and variables for coffee mapping in the southwest region of Brazil**

#### **ABSTRACT**

Coffee is one of the main crops in Brazil, therefore, performing the mapping and monitoring of this culture is essential. However, map this culture is not an easy task, because of their high spectral similarity with other types of land cover. Thus, the objective of this study was to compare the use of different variables and classification algorithms for coffee area classification. The study was conducted in three areas, environmentally different. We use 5 machine learning algorithms and 7 combinations of variables, using spectral, textural and geometric variables associated with the classification process. A total of 105 maps were made. All ratings that have not used spectral variables don't achieved good levels of accuracy. In all three areas, the algorithm that presented the best accuracies was the Support Vector Machine with overall accuracy 85.33% in Araguari, 87.00% in Carmo de Minas and 88.33% in Três Pontas. The worst results were found by Random Forest algorithm in Araguari, with 76.66% accuracy and Naive Bayes in Carmo de Minas and Três Pontas, with 76.00% and 82.00%. In all three areas, textural variables when associated with spectral, improved the classification accuracy. The Support vector machine showed the best performance for the three areas.

**Index terms:** coffee mapping; remote sensing; machine learning algorithms.

## 1 INTRODUÇÃO

A cafeicultura é uma importante atividade econômica no panorama internacional (OIC, 2014). A maior movimentação ocorre nos Estados Unidos, que comercializa cerca de 23 milhões de sacas/ano (OIC, 2014). O café é uma das principais culturas agrícolas do Brasil, com grande importância para a economia do país (CONAB, 2014).

Minas Gerais se destaca como o maior estado produtor de café, com mais de 50,00% da produção nacional (CONAB, 2014). Projeções futuras indicam um crescimento de 2,00% ao ano, em relação ao total de café que é consumido hoje em dia (OIC, 2014). Sendo assim, torna-se fundamental conhecer a distribuição espacial da atividade cafeeira para prever e planejar seu crescimento, bem como a estratégia de comercialização da sua produção de forma eficiente e otimizada.

Tecnologias e sistemas associados ao sensoriamento remoto têm sido amplamente empregadas para mapear e monitorar áreas agrícolas (Velo, 1974;Moreira et al., 2004; Cordero-Sancho & Sader, 2007; Li et al., 2014). Existem, no entanto, diversas dificuldades que envolvem o processo de derivar informações úteis a partir de imagens de sensoriamento remoto. Um dos complicadores é a heterogeneidade de paisagens dos ambientes tropicais (Li et al., 2014) que, conseqüentemente, aumenta a complexidade das cenas retratadas nas imagens, em termos de materiais constituintes. Essa complexidade, por sua vez, aumenta a confusão espectral entre os diferentes tipos de cobertura da terra com respostas espectrais semelhantes, como é o caso das áreas cobertas por cafezais e por remanescentes de vegetação nativa (Moreira et al., 2004; Cordero-Sancho & Sader, 2007; Adami et al., 2009).

Estas limitações afetam diretamente o mapeamento dos cafezais devido à ampla variabilidade espectral, temporal e espacial das lavouras cafeeiras

(Vieira et al., 2007; Adami et al., 2009), tornando a definição de um padrão de identificação para o café mais difícil do que para outras culturas agrícolas. Segundo Adami et al. (2009), os diversos métodos de planejamento do plantio, como espaçamento e sistema de cultivo, apresentam similaridade espectral com áreas de mata nativa e com outras culturas. O mapeamento pode ser dificultado, ainda, em regiões montanhosas (Santos et al., 2012; Andrade et al., 2013). Estas áreas sombreadas aumentam ainda mais a variabilidade espectral da cultura nas imagens de sensoriamento remoto (Santos et al., 2012).

O mapeamento de áreas agrícolas usando dados de sensoriamento remoto, especialmente aquelas relacionadas à cafeicultura, já vem sendo pesquisado desde a década de 1970 (Veloso, 1974), porém, os resultados deste primeiro estudo não permitiam uma discriminação apropriada dos cafezais em relação a outros tipos de uso. Em Minas Gerais, Moreira et al. (2004), Vieira et al. (2007), Machado et al. (2010), Santos et al. (2012) e Andrade et al. (2013) trabalharam mais recentemente no mapeamento e na caracterização desta cultura e reportaram resultados promissores, porém, apresentando muita confusão entre os alvos. Andrade et al. (2013) conduziram uma classificação automática para mapear uma área cafeeira na região de Machado, MG. A área foi dividida em duas, uma com relevo mais suave e a outra com o relevo mais movimentado. Os resultados mostraram índice Kappa inferiores a 0,60. Cordero-Sancho & Sader (2007) mapearam áreas de café na Costa Rica e analisaram a combinação de bandas espectrais, além de dados complementares, a fim de avaliar a precisão do mapeamento desta cultura e de outros tipos de cobertura da terra. O maior índice de acurácia global foi de 65,00%, tendo a separação espectral entre floresta e café não sido bem sucedida. Avaliando a importância de variáveis espectrais, geométricas e texturais para o mapeamento de áreas urbanas, Wieland et al. (2014), demonstraram que as variáveis mais importantes no processo de

classificação foram as espectrais, seguidas das texturais e, por fim, das geométricas.

Estudos realizados na América Latina utilizaram apenas métodos de classificação mais tradicionais, como a classificação baseada em pixel e classificadores paramétricos padrões (Adami et al., 2009; Martínez-Verduzco et al., 2012). Poucos estudos têm utilizado a classificação baseada em objetos combinados com novos algoritmos de classificação (Santos et al., 2012), como os algoritmos de aprendizagem de máquina (AM).

Algoritmos de aprendizagem de máquina, como *Support vector machine* (SVM), mostraram bons resultados na acurácia do mapeamento do café (Santos et al., 2012; Sarmiento et al., 2014). Porém, poucos estudos trabalharam com este algoritmo para o mapeamento do café no Brasil (Santos et al., 2012) e Marujo et al. (2013). Além disso, alguns algoritmos de AM que vêm sendo muito utilizados em classificação de imagens ainda não foram utilizados para o mapeamento da cafeicultura, como é o caso dos algoritmos *Random Forest* (RF), *Decision Tree* (DT) e *Naive Bayes* (NB). Estes algoritmos têm demonstrado excelente desempenho na análise de bases de dados de sensoriamento remoto que apresentam grande complexidade (Li et al., 2013).

Alguns estudos também têm incorporado variáveis geométricas e texturais para auxiliar o mapeamento de lavouras cafeeiras, mas ainda não se sabe se estas variáveis são eficientes para distinguir as plantações de café de outros tipos de uso da terra, como vegetação e pastagem (Gomez et al., 2010; Santos et al., 2012; Marujo et al., 2013). Estudos analisando estas variáveis separadamente e em conjunto ainda não foram realizados para o mapeamento desta cultura.

É importante salientar, ainda, que existe uma escassez de trabalhos utilizando imagens de alta resolução para o estudo da cafeicultura (Marujo et al., 2013; Sarmiento et al., 2014) e para os demais mapeamentos do uso da terra. Na

maioria dos casos são utilizadas imagens de média resolução, como, por exemplo, imagens dos sensores TM e ETM+ (Wieland et al., 2014; Pradhan et al., 2013; Otukey & Blaschke, 2010).

Sendo assim, considerando a importância da produção cafeeira para a economia do país, a dificuldade intrínseca de mapeamento desta cultura e o fato de ainda não existir um mapa oficial dos parques cafeeiros do Brasil, fica evidente a necessidade de mais estudos para subsidiar o desenvolvimento de métodos apropriados e confiáveis de mapeamento das lavouras de café.

O estudo apresentado aqui visa contribuir para o conhecimento das peculiaridades do mapeamento de cafezais, analisando o desempenho de algoritmos inovadores e de variáveis que retratam características espectrais, geométricas e texturais na classificação digital de imagens de alta resolução espacial. Este trabalho foi realizado com o objetivo principal de comparar algoritmos de AM usando diferentes conjuntos de variáveis derivadas de dados de sensoriamento remoto e identificar as melhores combinações algoritmos-variáveis para o mapeamento de cafezais em três regiões de Minas Gerais.

Nesse sentido, pretendeu-se responder às seguintes questões: (1) Há diferença significativa entre as classificações utilizando diferentes combinações entre algoritmos de classificação e variáveis de entrada? (2) Quais conjuntos de variáveis são mais eficientes para classificar cafezais? (3) Quais são os algoritmos mais eficientes para classificar cafezais? Qual algoritmo é mais acurado para o mapeamento de cafezais em cada uma das três regiões em estudo? (4) Quais são as classes confundidas com cafezal durante o processo de classificação?

## **2 MATERIAIS E MÉTODOS**

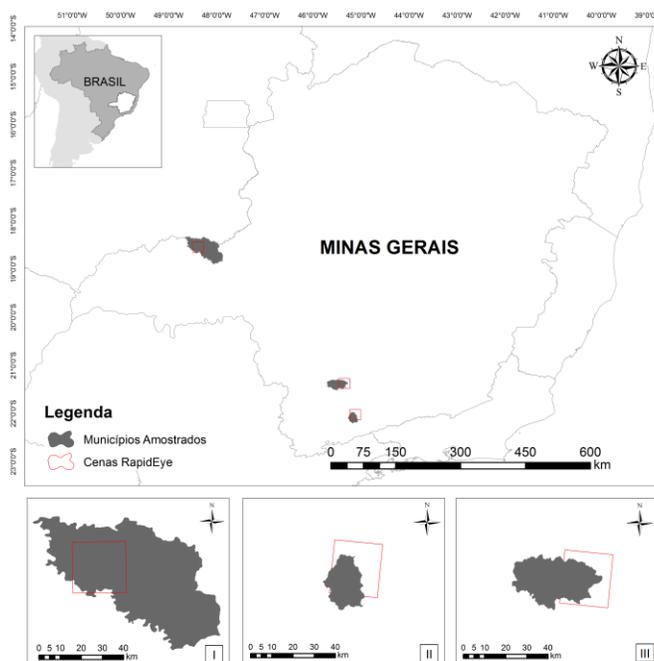
### **2.1 Área de estudo e dados de sensoriamento remoto**

As áreas de estudo estão localizadas no estado de Minas Gerais, Brasil. Foram escolhidas três áreas distintas, doravante denominadas áreas I, II e III: a primeira corresponde à cena RapidEye 2230526, retratada em 14/06/2010, que cobre parte do município de Araguari (área I), região oeste do estado. As outras duas áreas estão localizadas na região sul do estado e correspondem às cenas RapidEye 2328914, retratada em 22/07/2010 e 2329213, retratada em 18/06/2010, cobrindo parte dos municípios de Carmo de Minas (área II) e Três Pontas (área III), respectivamente (Figura 1). Na região sul, o bioma predominante é Mata Atlântica e, na região oeste, o bioma Cerrado.

O clima no município de Araguari, de acordo com a classificação de Köppen, é do tipo Cwa, tropical de altitude, com temperatura média de 21 °C e índice pluviométrico anual de 1.400 mm. A altitude média é de 1.013 m, onde são encontradas as formas tabulares, e as atividades principais são a agricultura de grãos (soja e milho) e a cafeicultura (IBGE, 2009).

O município de Carmo de Minas tem altitude média de 960 m, temperatura média anual de 19,1 °C e índice pluviométrico médio anual de 1.568 mm (IBGE, 2009). O clima, segundo a classificação de Köppen, é Cwb, subtropical de altitude, com temperatura média anual de 17 °C. A base da sua economia é a agricultura, destacando-se a cafeicultura e a pecuária (IBGE, 2009).

O município de Três Pontas tem altitude média de 905 m. O clima, de acordo com a classificação de Köppen, é tropical de altitude. A temperatura média anual de 18 °C e média anual de pluviosidade é de 1.440 mm. A principal atividade econômica da região é a cafeicultura (IBGE, 2009).



**Figura 1** Mapa de localização da área de estudo. Na região oeste do estado está o município de Araguari (área I) e na região sul, os municípios de Carmo de Minas (área II) e Três Pontas (área III)

Estas regiões foram escolhidas por serem áreas com fitofisionomias e características ambientais diferentes e também por estarem localizadas em municípios bastante representativos na produção de café no cenário mineiro.

As imagens selecionadas para este estudo são do satélite RapidEye, as quais têm 5 m de resolução espacial, resolução radiométrica de 16 bits e resolução espectral de cinco bandas. As imagens RapidEye utilizadas neste estudo foram adquiridas com correções geométricas e radiométricas (nível 3A).

## **2.2 Estratégia de classificação**

Todas as imagens RapidEye foram segmentadas utilizando-se o algoritmo multirresolução, do software eCognition Developer, em que foram testados visualmente diferentes parâmetros de escala, compacidade e forma, a fim de verificar quais parâmetros de segmentação foram mais adequados, considerando as áreas de café e suas variações na paisagem. Depois das análises, os parâmetros de segmentação escolhidos foram: escala 250 e, para forma e compacidade, foi utilizado o peso 0,5 para ambas. Estas características foram escolhidas por apresentarem boa separação das áreas de café para as três áreas de estudo.

Foram definidas cinco classes para as classificações, de acordo com as suas características espectrais, sendo elas: vegetação nativa: áreas de formações florestais densas e florestas de galeria, e formações de cerrado; Café 1: lavouras em idade não produtiva, no início do estágio de crescimento; Café 2: lavouras em estágio de crescimento intermediário; Café 3: lavouras com idade superior a 3 anos; Pastagem: áreas de pastagens naturais e formadas; Outros usos 1: áreas com culturas anuais em diversos estágios de desenvolvimento; floresta de produção; Outros usos 2: áreas urbanas e benfeitorias, áreas de solo exposto, áreas de queimadas; Corpos d'água: rios, córregos e represas, lagoas naturais e artificiais.

## **2.3 Amostras de treinamento**

Para cada área, foram selecionados, no mínimo, 10% do total de objetos gerados pela segmentação para servirem de amostras de treinamento. Isso gerou um total de 751 amostras para Araguari, 1.058 para Carmo de Minas e 939 para Três Pontas. As amostras de treinamento foram coletadas de maneira uniforme por toda imagem. A escolha dos objetos foi feita utilizando-se interpretação visual e de forma criteriosa; objetos com mais de um tipo de cobertura da terra

não foram utilizados. Em cada uma das áreas, as amostras escolhidas foram as mesmas para todas as classificações.

## **2.4 Processo de classificação**

### **2.4.1 Algoritmos de classificação**

A fim de avaliar o desempenho de diferentes métodos de classificação da cobertura da terra, foram utilizados os seguintes algoritmos: *decision tree* (DT), *naive bayes* (NB), *random forest* (RF), *support vector machine* (SVM) e *K-nearest neighbor* (KNN).

Estes algoritmos foram utilizados por estarem disponíveis para qualquer usuário e serem de fácil utilização. Os parâmetros empregados pelos algoritmos de classificação estão discriminados na Tabela 1. Todas as fontes de códigos são pacotes oriundos do software RStudio. Para escolher os parâmetros a serem usados no treinamento dos algoritmos, foram realizados testes preliminares usando apenas variáveis espectrais e escolhidas as combinações de parâmetros que apresentaram maior precisão durante os ajustes dos modelos.

**Tabela 1** Conjunto de parâmetros utilizados em cada algoritmo e sua fonte de códigos

Algoritmo	Sigla	Fonte de códigos	Tipos de parâmetros	Descrição dos parâmetros	Parâmetros
<i>Decision tree</i>	DT	rpart	cp	É usado para controlar o tamanho da árvore de decisão e para selecionar o tamanho ideal da árvore.	0.001
			maxdepth	Define a profundidade máxima de qualquer nó da árvore final.	30
<i>K-nearest neighbor</i>	KNN	eCognition	k	São identificados por uma medida de distância que compara os vetores de características da instância não marcada e o conjunto de instâncias de treinamento fornecido ao classificador. Depois que uma lista de vizinhos mais próximos é obtida, a previsão é baseada em votação (maioria ou distância ponderada).	1
<i>Naive bayes</i>	NB	eCognition	laplace	Suaviza dados categóricos, com o propósito de evitar que o cálculo da probabilidade seja igual à zero.	0
<i>Random forest</i>	RF	randomForest	mtry	Número de variáveis amostradas aleatoriamente como candidatos em cada divisão. Note-se que os valores padrões são diferentes para classificação.	2, 4, 6, 8, 10, 12, 14, ..., 50
			ntree	Número de árvores.	1000
<i>Support vector machine</i>	SVM	Kernlab	$\sigma$	O parâmetro $\sigma$ define a largura da função de kernel.	0.02
			kernel	Define qual função kernel é utilizada no treinamento e na predição dos dados.	rbfdot - Radial Basis Function
			C	Ajusta a sensibilidade da margem de decisão de vetores de suporte erroneamente classificados.	1

#### **2.4.2 Variáveis usadas na classificação**

Foram testados três conjuntos de variáveis diferentes, espectrais, texturais e geométricas, selecionados de acordo com a maioria dos trabalhos que realizam mapeamento do café (Moreira et al., 2004; Cordeiro-Sancho & Sader, 2007; Gomez et al., 2010; Machado et al., 2010; Santos et al., 2012; Marujo et al., 2013). Estas variáveis foram utilizadas separadamente e combinadas entre elas, totalizando sete combinações diferentes. Alguns algoritmos de AM mostram as variáveis mais importantes utilizadas para o processo de classificação. Para selecionar as melhores variáveis utilizadas, foi utilizado o ranking obtido pelo *random forest* e pelo *decision tree*.

As variáveis espectrais usadas foram: média e desvio padrão de cada banda; NVDI e SAVI; brilho e máxima diferença.

Para derivar informações de textura, foram utilizadas as métricas conhecidas como *grey level cooccurrence matrix*, ou GLCM, desenvolvidas por Haralick et al. (1973). As variáveis de textura GLCM foram entropia, homogeneidade, média, desvio padrão e correlação.

As variáveis geométricas estão associadas aos objetos derivados da segmentação da imagem. Neste trabalho foram utilizadas as variáveis baseadas em Wieland et al. (2014) e van der Werff & van der Meer (2008). As variáveis usadas foram: área, comprimento, largura, comprimento de borda, largura de borda, razão comprimento/largura, assimetria, densidade, circularidade, compacidade, índice de forma e borda, e rugosidade.

#### **2.5 Amostras de acurácia e análise estatística**

Com a finalidade de verificar a acurácia das classificações, foram coletados 300 pontos como dados de validação, para cada uma das áreas. Na área I, os pontos coletados foram divididos da seguinte forma: 60 como pastagem, 90 como vegetação, 70 como outros usos, 65 como café e 15 como

água. Na área II: 95 como pastagem, 80 como vegetação, 40 como outros usos, 70 como café e 15 como água. Na área III: 75 como pastagem, 70 como vegetação, 40 como outros usos, 100 como café e 15 como água. Esses pontos foram obtidos por meio de uma amostragem estratificada aleatória, em que cada estrato foi representado por uma categoria de classificação (Martínez-Verduzco et al., 2012). A conferência foi realizada a partir de visitas a campo e do Google Earth. Foram utilizados o coeficiente *kappa* e a acurácia global como critério de avaliação.

Para comparar os resultados das acurácias foi utilizado o teste de McNemar (Foody, 2004), um teste não paramétrico e que avalia a significância estatística das diferenças entre as duas classificações, baseando-se em matrizes de confusão de duas dimensões (Tabela 2). Neste teste, a atenção é focada na distinção binária entre a alocação das classes corretamente ou incorretamente. O teste de McNemar é baseado no teste estatístico normal padronizado (Foody, 2004).

**Tabela 2** Matriz de confusão usada no teste de McNemar

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

### 3 RESULTADOS E DISCUSSÃO

Neste trabalho avaliou-se a efetividade do uso de diversos classificadores, *decision tree* (DT), *naive bayes* (NB), *random forest* (RF), *support vector machine* (SVM) e *K-nearest neighbor* (KNN), bem como de diferentes variáveis para a classificação de áreas cafeeiras, visando melhorar a separabilidade entre as classes. As variáveis utilizadas são apresentadas no texto como *s* (espectrais), *t* (texturais) e *g* (geométricas).

Um total de 105 mapas foram gerados para as três áreas (35 para cada uma delas). Mapas que apresentaram índice Kappa inferior a 0,65 não foram considerados nas análises subsequentes. Todas as classificações que usaram somente variáveis texturais e geométricas, ou seja, sem os dados espectrais, apresentaram acurácia inferior a 0,65. Sendo assim, foram consideradas na apresentação e na discussão dos resultados apenas as classificações que incluíam variáveis espectrais (Tabela 3).

**Tabela 3** Valores de acurácia global, índice kappa e porcentagem de acerto para cada classe de uso. Estes valores estão dispostos para cada área, cada algoritmo usado e cada conjunto de variáveis

Área	Algoritmos	Variáveis	Índices		Acurácia (%)									
			Global (%)	Kappa	Vegetação		Pastagem		Café		Outros usos		Água	
					Produtor	Usuário	Produtor	Usuário	Produtor	Usuário	Produtor	Usuário	Produtor	Usuário
I	DT	stg	81,33	0,75	84,44	96,20	91,67	60,44	92,31	78,95	57,14	97,56	86,67	100,00
		s	82,66	0,77	86,67	95,12	91,67	59,14	93,95	87,14	58,57	97,62	86,67	100,00
		sg	81,33	0,75	84,44	93,83	91,67	60,44	92,31	81,08	57,14	97,56	86,67	100,00
		st	81,33	0,75	84,44	95,00	91,67	60,44	90,77	79,73	58,57	97,62	86,67	100,00
	KNN	stg	83,00	0,77	91,11	91,11	93,33	58,33	90,77	96,72	52,86	97,37	100,00	100,00
		s	81,66	0,76	88,89	95,24	91,67	55,56	93,85	92,42	50,00	94,59	93,33	100,00
		sg	81,66	0,76	91,11	89,13	93,33	57,14	90,77	96,72	48,57	97,14	93,33	100,00
		st	82,33	0,76	91,11	88,17	88,33	58,24	89,23	98,31	58,57	93,18	86,67	100,00
	NB	stg	79,66	0,73	93,33	92,31	95,00	54,29	90,77	98,33	44,29	86,11	53,33	100,00
		s	83,66	0,78	88,89	96,39	91,67	62,50	98,46	91,43	57,14	85,11	80,00	100,00
		sg	83,00	0,77	92,22	92,22	95,00	58,78	93,85	98,39	52,86	92,50	73,33	100,00
		st	81,33	0,75	92,22	94,32	91,67	56,70	95,38	98,41	51,43	81,82	53,33	100,00
	RF	stg	78,33	0,71	78,89	97,26	93,33	53,33	93,85	83,56	50,00	94,59	80,00	100,00
		s	76,66	0,69	70,00	95,45	91,67	56,70	95,38	73,81	52,86	92,50	86,67	100,00
		sg	77,33	0,70	78,89	88,75	91,67	55,56	87,69	82,61	52,86	92,50	80,00	100,00
		st	80,00	0,74	78,89	98,61	91,67	56,70	95,38	82,67	54,29	90,48	93,33	100,00
	SVM	stg	82,66	0,77	91,11	96,47	93,33	55,45	92,31	92,31	50,00	94,59	100,00	100,00
		s	85,33	0,80	93,33	94,38	93,33	62,22	95,38	95,38	57,14	95,24	93,33	100,00
		sg	80,33	0,74	91,11	95,35	81,67	52,69	95,38	96,88	47,14	78,57	100,00	100,00
		st	84,66	0,80	93,33	94,38	91,67	61,11	95,38	95,38	55,71	92,86	93,33	100,00
	DT	stg	84,33	0,78	71,25	89,06	98,95	80,34	91,43	81,01	67,50	93,10	73,33	100,00
		s	84,33	0,78	75,00	85,71	98,95	80,34	85,71	83,33	67,50	96,43	80,00	92,31
		sg	84,33	0,78	75,00	85,71	98,95	80,34	85,71	83,33	67,50	96,43	80,00	92,31
		st	84,33	0,78	71,25	89,06	98,95	80,34	91,43	81,01	67,50	93,10	73,33	100,00
KNN	stg	85,00	0,79	71,25	98,95	95,79	76,47	97,14	83,95	67,50	90,00	80,00	100,00	
	s	85,33	0,80	73,75	93,65	97,89	78,81	91,43	84,21	67,50	90,00	86,67	100,00	
	sg	84,00	0,78	78,75	88,73	94,74	81,08	91,43	86,49	60,00	77,42	73,33	84,62	
	st	84,66	0,79	71,25	96,61	94,74	76,92	98,57	85,19	65,00	83,97	80,00	100,00	
NB	stg	77,66	0,70	53,75	100,00	87,37	84,69	98,57	71,88	65,00	53,06	80,00	85,71	
	s	79,33	0,72	53,75	93,48	92,63	85,44	95,71	72,04	70,00	60,87	80,00	100,00	
	sg	81,33	0,75	58,75	95,92	93,68	86,41	97,14	73,12	72,50	67,44	73,33	91,67	
	st	76,00	0,68	50,00	100,00	88,42	82,45	98,57	74,19	57,50	46,00	80,00	80,00	
RF	stg	87,00	0,82	71,25	100,00	97,89	81,58	98,57	85,19	75,00	83,33	80,00	100,00	
	s	85,33	0,80	71,25	93,44	72,50	80,17	92,86	84,42	72,50	85,29	80,00	100,00	
	sg	84,66	0,79	72,50	95,08	96,84	80,00	94,29	84,62	67,50	77,14	73,33	100,00	
	st	85,66	0,80	71,25	100,00	97,89	78,81	98,57	85,19	65,00	81,25	80,00	100,00	
SVM	stg	86,66	0,82	76,25	93,85	96,84	83,64	91,43	85,33	75,00	81,08	86,67	100,00	
	s	82,00	0,75	71,25	85,07	97,89	77,50	82,86	82,86	62,50	83,33	100,00	100,00	
	sg	85,00	0,79	78,75	87,50	97,89	83,04	87,14	89,29	62,50	89,29	86,67	76,47	
	st	85,33	0,80	73,75	96,72	96,84	76,03	95,71	87,01	62,50	89,29	86,67	100,00	

**Tabela 3, conclusão**

Área	Algoritmos	Variáveis	Índices		Acurácia (%)									
			Global (%)	Kappa	Vegetação		Pastagem		Café		Outros usos		Água	
					Produtor	Usuário	Produtor	Usuário	Produtor	Usuário	Produtor	Usuário	Produtor	Usuário
III	DT	stg	85,66	0,80	78,57	100,00	96,00	79,12	92,00	85,98	62,50	89,29	86,67	68,42
		s	84,00	0,78	74,29	92,86	96,00	79,12	91,00	81,98	62,50	89,29	80,00	85,71
		sg	84,00	0,78	74,29	92,86	96,00	79,12	91,00	81,98	62,50	89,29	80,00	85,71
		st	85,66	0,80	78,57	100,00	96,00	79,12	92,00	85,98	62,50	89,29	86,67	68,42
		stg	85,66	0,80	77,14	94,74	93,33	81,40	96,00	81,36	62,50	92,59	80,00	100,00
		s	86,00	0,81	77,14	91,53	96,00	81,82	94,00	83,19	65,00	96,30	80,00	92,31
	KNN	sg	82,66	0,76	78,57	87,30	93,33	76,92	89,00	81,65	55,00	91,67	92,31	92,31
		st	86,00	0,81	78,57	98,21	90,67	79,07	99,00	83,19	60,00	88,89	80,00	100,00
		stg	82,00	0,75	64,29	100,00	90,67	78,16	94,00	80,34	67,50	75,00	80,00	80,00
		s	84,00	0,78	65,71	97,87	93,33	83,33	95,00	78,51	75,00	81,08	73,33	100,00
		sg	84,66	0,79	68,57	94,12	96,00	81,82	93,00	80,87	72,50	87,88	80,00	92,31
		st	82,00	0,75	60,00	100,00	93,33	79,55	95,00	77,87	65,00	78,79	86,67	86,67
	NB	stg	86,00	0,81	77,14	100,00	98,67	76,29	93,00	84,55	62,50	92,59	80,00	100,00
		s	85,00	0,79	74,29	94,55	97,33	77,66	92,00	82,14	65,00	96,30	80,00	100,00
		sg	85,00	0,79	72,86	94,44	98,67	77,08	92,00	82,88	65,00	96,30	80,00	100,00
		st	85,00	0,79	72,86	98,11	98,67	77,66	92,00	80,70	65,00	92,59	80,00	100,00
		stg	87,66	0,83	78,57	96,49	96,00	80,90	99,00	88,39	62,50	83,33	80,00	100,00
		s	86,33	0,81	80,00	93,33	96,00	80,00	96,00	84,96	57,50	95,83	80,00	92,31
	SVM	sg	86,33	0,81	81,43	91,94	92,00	83,13	96,00	86,49	62,50	80,65	80,00	92,31
		st	88,33	0,84	78,57	96,49	100,00	81,52	99,00	87,61	60,00	96,00	80,00	92,31

### 3.1 Desempenho dos algoritmos de classificação

As classificações geradas apresentaram desempenhos diferentes, tanto em relação ao algoritmo de classificação usado, quanto pelas variáveis empregadas. O algoritmo SVM apresentou os melhores resultados para as três áreas, sendo que na área II o algoritmo RF apresentou acurácia igual ao algoritmo SVM (Tabela 3). Os piores resultados foram obtidos pelos algoritmos RF na área I e NB nas áreas II e III. Quian et al. (2015), comparando diferentes algoritmos de AM, mostraram que o SVM apresentou os melhores desempenhos de classificação, corroborando os resultados encontrados neste estudo. O desempenho do algoritmo SVM está relacionado à escolha dos parâmetros de Kernel, no qual o Kernel RBF é o mais recomendado pela literatura e apresenta os melhores resultados de acurácia (Pradhan, 2013). Neste estudo, utilizou-se o Kernel RBF para as classificações utilizando SVM e isso pode ter colaborado para que este algoritmo apresentasse os melhores resultados de classificação.

Na área I, os índices de acerto global das classificações variaram entre 76,66% e 85,33% (Tabela 3). As melhores classificações foram geradas usando o algoritmo SVM, com os conjuntos de variáveis *s* (acurácia global = 85,33% e índice Kappa = 0,80) e *st* (acurácia global = 84,66% e índice Kappa = 0,80). Os piores resultados encontrados foram usando o algoritmo RF, com as variáveis *s* (acurácia global = 76,66% e índice Kappa = 0,69), *sg* com acurácia global de 77,33% e índice kappa de 0,70, *stg* (acurácia global = 78,33% e índice Kappa = 0,71) e usando o algoritmo NB com as variáveis *stg* (acurácia global = 79,66% e índice Kappa = 0,73). Todos os outros algoritmos usados, independente do conjunto de variáveis, obtiveram resultados de acurácia global e índice kappa superiores. Resultados distintos foram encontrados por alguns autores (Duro et al., 2012; Gislason et al., 2006; Pal, 2005) os quais mostraram que o RF vem obtendo bons desempenhos nas classificações, inclusive quando comparado a

outros algoritmos de classificação (Li et al., 2014; Gislason et al., 2006; Pal, 2005).

Na área II, a porcentagem de acerto global ficou entre 76,00% e 87,00% (Tabela 3). As melhores classificações foram geradas utilizando-se os algoritmos SVM e RF e o mesmo conjunto de variáveis *stg*, com acurácia global de 87,00% e 86,66%, respectivamente. No entanto, o índice kappa foi idêntico nas duas classificações (0,82). O pior desempenho foi obtido pelo algoritmo NB, com os conjuntos de variáveis *st* (acurácia global = 76,00% e índice Kappa = 0,69), *stg* (acurácia global = 77,66%) e *s* com 79,33% de acerto. Todas as classificações usando o algoritmo NB, independente das variáveis utilizadas, apresentaram os menores índices de acerto, quando comparadas aos outros algoritmos.

Outros estudos indicaram que ambos os algoritmos RF e SVM podem alcançar resultados de acurácia global semelhantes e que são tipicamente maiores do que aqueles obtidos utilizando algoritmos como DT, como mostrado por Pal (2005), que relatou que tanto SVM quanto RF produziram precisões de classificação semelhantes. Gislason et al. (2006) demonstraram que os modelos baseados RF alcançaram precisões de classificação mais elevadas do que as produzidas por DT. Estes resultados são diferentes dos relatados por Li et al. (2013), usando imagens Landsat TM, em que o classificador NB obteve um desempenho um pouco acima do SVM (86,60% NB e 85,90% SVM), porém, estes autores trabalharam mapeando vegetação aquática.

Amostras mais homogêneas e em grande quantidade podem melhorar o desempenho de classificação do algoritmo NB, tornando-o mais preciso, uma vez que este algoritmo é sensível ao tamanho e à uniformidade das amostras de treinamento, uma vez que utiliza amostras de treinamento para estimar os valores dos parâmetros para a distribuição dos dados (Qian et al., 2015).

Na área III, os índices de acerto global variaram de 82,00% a 88,33% (Tabela 3). As melhores classificações obtidas foram geradas utilizando-se o

classificador SVM, para todos os conjuntos de variáveis usadas. O índice de acerto mais alto foi alcançado usando as variáveis *st* com acurácia global de 88,33% e índice kappa de 0,84, seguido de *stg* (acurácia global = 87,66% e índice Kappa = 0,83). Os menores índices de acerto foram obtidos usando o algoritmo NB, com as variáveis *stg* e *st*, ambas com acerto de 82,00%. Resultados melhores foram obtidos utilizando-se o algoritmo DT com as variáveis *sg* e *s*, todas com o mesmo índice de acerto (acurácia global = 84,00% e índice Kappa = 0,78). Li et al. (2014) também encontraram resultados semelhantes, realizando o mapeamento de uso da terra numa região da China. Eles mostraram que os classificadores SVM e RF apresentaram bons resultados na separabilidade entre as classes. O algoritmo DT pode mostrar desvantagens, pois a árvore pode conter muitas ramificações, o que torna a interpretação da classificação difícil (Hussain et al., 2013).

Na área II foi verificada a maior variação entre as acurácias (11,00%), provavelmente pelo fato de esta área apresentar uma configuração fisiográfica bem diferente das demais. Esta região é muito íngreme, com relevo bastante acidentado, apresentando algumas áreas sombreadas na imagem, o que pode levar interpretações distintas entre os algoritmos de classificação (Andrade et al., 2013a). Já na área I, a variação foi menor (8,67%), porém, as diferenças nas classificações podem ser explicadas pela configuração da paisagem. As maiores confusões foram encontradas na classe outros usos, que foi bastante confundida com a classe pastagem em todos os algoritmos usados. A área III foi a que apresentou menor amplitude entre as acurácias (6,33%), além de mostrar os maiores índices de acerto das três áreas. Este resultado era esperado devido ao fato de a região apresentar uma estrutura fisiográfica relativamente menos complexa, com áreas mais homogêneas e com relevo suave ondulado predominante. Estes fatores podem contribuir para o bom desempenho na classificação (Bertoldo, 2008).

Alguns trabalhos usando algoritmos de AM mostraram que o desempenho destes classificadores tem sido superior, quando comparado ao dos demais algoritmos de classificação de imagens (Li et al., 2014; Pradhan, 2013; Otakei & Blaschke, 2010). Entretanto, alguns algoritmos de AM apresentam melhores resultados que outros, como é o caso do SVM e RF (Wieland et al., 2014; Li et al., 2014; Duro et al., 2012; Gislason et al. 2006). Wieland et al. (2014) apresentaram resultados promissores no uso de diferentes algoritmos de AM para a classificação em diferentes sensores, no qual os algoritmos SVM e RF apresentaram os melhores desempenhos de classificação, e NB com os piores resultados. Estes resultados corroboram com os obtidos neste estudo, no qual as melhores acurácias foram obtidas pelos algoritmos SVM e RF, nas áreas II e III.

Analisando-se as comparações entre as melhores classificações (Tabela 4), observa-se que os algoritmos utilizados não interferem na acurácia, visto que não houve diferenças significativas na comparação entre os mesmos. Na área I é possível observar que as maiores diferenças foram em relação ao algoritmo RF; na área II, as maiores diferenças envolvem o algoritmo NB e, na área III, as maiores diferenças envolvem o algoritmo SVM.

**Tabela 4** Matriz de comparações entre as melhores classificações de cada algoritmo de AM. As comparações foram realizadas pelo teste de McNemar com chi-quadrado tabulado  $X^2=3,81$ ; a) área I; b) área II; c) área III

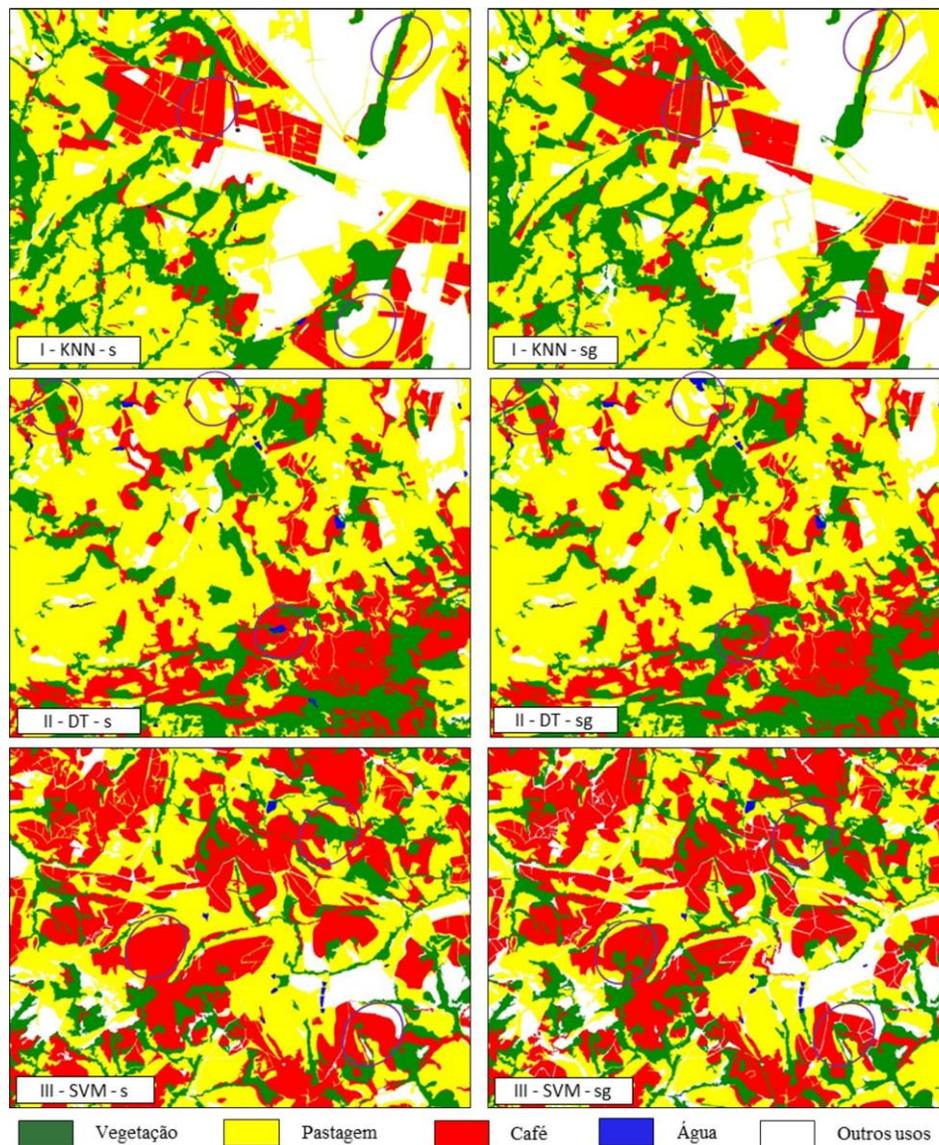
a)	DT	KNN	NB	RF	SVM
DT	*				
KNN	0	*			
NB	0,06	0	*		
RF	0,43	0,54	0,81	*	
SVM	0,24	0,41	0,06	1,31	*

b)	DT	KNN	NB	RF	SVM
DT	*				
KNN	0,02	*			
NB	0,54	0,80	*		
RF	0,42	0,24	1,92	*	
SVM	0,32	0,16	1,70	0	*

c)	DT	KNN	NB	RF	SVM
DT	*				
KNN	0,08	*			
NB	0,06	0,10	*		
RF	0	0	0,10	*	
SVM	0,42	0,32	0,80	0,32	*

É importante salientar que, mesmo não apresentando diferença significativa, as classificações têm diferenças entre si, como é possível observar na Figura 2. Na área I, as classificações apresentaram a mesma acurácia (índice Kappa = 0,75; acurácia global 81,33%) usando o algoritmo DT, com os atributos *stg*, *sg* e *st*, além das classificações utilizando o conjunto de variáveis *s* e *sg* (índice Kappa = 0,76; acurácia global 81,66%), usando o algoritmo KNN. No entanto, algumas áreas na imagem foram classificadas de forma diferente. Isto pode ser observado também na área II, com os quatro conjuntos de classificação (índice Kappa = 0,74; acurácia global 84,33%), usando o algoritmo DT e na área III isso ocorre para, pelo menos, dois conjuntos de variáveis em cada classificador usado. Estes resultados mostram que, mesmo possuindo o mesmo índice de acerto, as classificações são diferentes. Portanto, além das análises estatísticas e de acurácia, é fundamental uma interpretação visual das classificações, para selecionar a que melhor separou as classes de uso.

Algumas classificações mostraram diferenças significativas, quando comparadas com o mesmo conjunto de variáveis, porém, diferenciando o algoritmo usado (Tabela 5). Estes resultados foram observados em duas áreas (I e II), nas quais as classificações obtiveram a maior amplitude entre as acurácias. Na área I, as únicas classificações que mostraram diferença significativa foram aquelas usando variáveis espectrais, entre os algoritmos RF e SVM, pois elas mostraram grande diferença na acurácia (Tabela 3). Nesta área, o algoritmo RF foi o que apresentou as menores acurácias, independente do conjunto de variáveis usados. A área II foi a que mostrou maior número de diferenças significativas entre as classificações. As classificações usando as variáveis *stg* somente apresentaram diferenças significativas utilizando o algoritmo NB, comparadas com RF e com SVM (Tabela 5). Comparando o algoritmo NB com os demais, utilizando as variáveis *st*, todas as classificações foram diferentes, significativamente. As demais comparações entre algoritmos não mostraram diferença significativas, entretanto, é possível observar que as maiores variações ocorrem quando se compara NB aos demais classificadores. Todas as classificações na área III não mostraram diferenças significativas. Resultados semelhantes foram encontrados por Duro et al. (2012) e Quian et al. (2015), que também não verificaram diferenças significativas usando DT, KNN, RF e SVM.



**Figura 2** Diferença entre classificações com o mesmo índice de acerto. a) Área I, algoritmo KNN e variável  $s$  ( $s$ ); b) Área I, algoritmo KNN e variável  $sg$ ; c) Área II, algoritmo DT e variável  $s$ ; d) Área II, algoritmo DT e variável  $sg$ ; e) Área III, algoritmo SVM e variável  $s$ ; f) Área III, algoritmo SVM e variável  $sg$

**Tabela 5** Matriz de comparações usando o mesmo conjunto de variáveis, porém, diferenciando o algoritmo usado. As comparações foram realizadas pelo teste McNemar com chi-quadrado tabulado  $X^2=3,81$ . Em negrito, o que foi significativo; a) área I; b) área II; c) área III

a)	DT-stg	KNN-stg	NB-stg	RF-stg	SVM-stg
DT-stg	*				
KNN-stg	0,16	*			
NB-stg	0,16	0,66	*		
RF-stg	0,54	1,30	0,10	*	
SVM-stg	0,10	0,00	0,54	1,12	*

a)	DT-st	KNN-st	NB-st	RF-st	SVM-st
DT-st	*				
KNN-st	0,06	*			
NB-st	0,00	0,06	*		
RF-st	0,10	0,32	0,10	*	
SVM-st	0,66	0,32	0,16	1,30	*

b)	DT-stg	KNN-stg	NB-stg	RF-stg	SVM-stg
DT-stg	*				
KNN-stg	0,02	*			
NB-stg	2,66	3,22	*		
RF-stg	0,42	0,24	<b>5,22</b>	*	
SVM-stg	0,32	0,32	<b>4,86</b>	0,00	*

b)	DT-st	KNN-st	NB-st	RF-st	SVM-st
DT-st	*				
KNN-st	0,00	*			
NB-st	<b>4,16</b>	<b>4,50</b>	*		
RF-st	0,10	0,06	<b>5,60</b>	*	
SVM-st	0,06	0,02	<b>5,22</b>	0,00	*

c)	DT-stg	KNN-stg	NB-stg	RF-stg	SVM-stg
DT-stg	*				
KNN-stg	0,00	*			
NB-stg	0,80	0,80	*		
RF-stg	0,00	0,00	0,96	*	
SVM-stg	0,24	0,24	1,92	0,16	*

c)	DT-st	KNN-st	NB-st	RF-st	SVM-st
DT-st	*				
KNN-st	0,00	*			
NB-st	0,80	0,96	*		
RF-st	0,06	0,10	0,42	*	
SVM-st	0,42	0,32	2,40	0,80	*

a)	DT-s	KNN-s	NB-s	RF-s	SVM-s
DT-s	*				
KNN-s	0,06	*			
NB-s	0,06	0,24	*		
RF-s	2,16	1,50	2,94	*	
SVM-s	0,42	0,80	0,16	<b>4,50</b>	*

a)	DT-sg	KNN-sg	NB-sg	RF-sg	SVM-sg
DT-sg	*				
KNN-sg	0,00	*			
NB-sg	0,16	0,10	*		
RF-sg	0,96	1,12	1,92	*	
SVM-sg	0,06	0,42	0,42	0,54	*

b)	DT-s	KNN-s	NB-s	RF-s	SVM-s
DT-s	*				
KNN-s	0,06	*			
NB-s	1,50	2,16	*		
RF-s	0,06	0,00	2,16	*	
SVM-s	0,32	0,66	0,42	0,66	*

b)	DT-sg	KNN-sg	NB-sg	RF-sg	SVM-sg
DT-sg	*				
KNN-sg	0,00	*			
NB-sg	0,54	0,42	*		
RF-sg	0,00	0,02	0,66	*	
SVM-sg	0,02	0,06	0,80	0,00	*

c)	DT-s	KNN-s	NB-s	RF-s	SVM-s
DT-s	*				
KNN-s	0,24	*			
NB-s	0,00	0,24	*		
RF-s	0,06	0,06	0,06	*	
SVM-s	0,32	0,00	0,32	0,10	*

c)	DT-sg	KNN-sg	NB-sg	RF-sg	SVM-sg
DT-sg	*				
KNN-sg	0,10	*			
NB-sg	0,02	0,24	*		
RF-sg	0,06	0,32	0,00	*	
SVM-sg	0,32	0,80	0,16	0,10	*

Estes resultados colaboram para inferirmos que o algoritmo RF, na área I, e NB, nas áreas II e III, apresentaram os resultados mais distintos entre as

classificações. Os demais classificadores mostraram um comportamento instável nas três áreas, em determinados momentos mostrando melhores desempenhos e, em outros, piores. Já o algoritmo SVM foi o mais eficiente por apresentar resultados acurados para as três áreas analisadas, mesmo usando diferentes conjuntos de variáveis, portanto, sendo recomendado para a classificação de áreas cafeeiras.

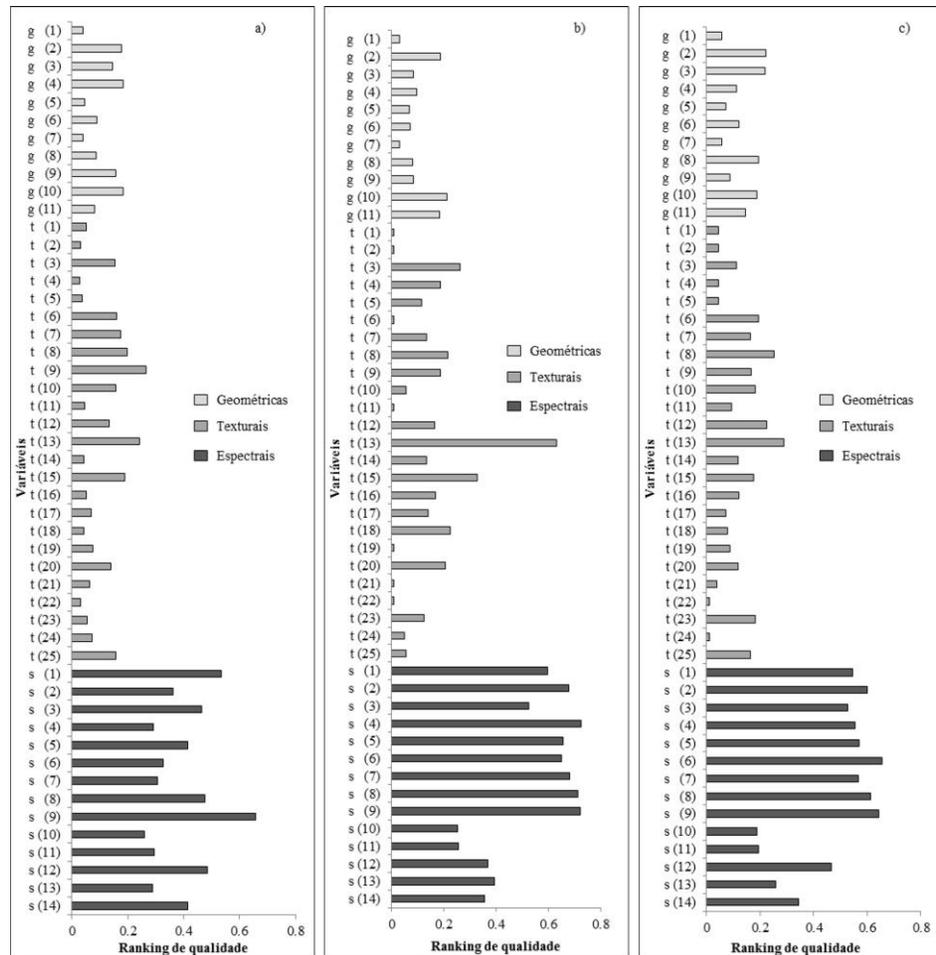
### **3.2 Variáveis mais eficientes para separação das classes**

Um total de 14 características espectrais, 25 texturais e 11 geométricas foi utilizado como variáveis para as classificações. Estes conjuntos foram testados individualmente e em conjunto para identificar qual a contribuição e a eficiência destas variáveis no processo de classificação. As classificações geradas utilizando somente variáveis texturais e geométricas apresentaram índices de acerto muito baixos, valores discrepantes em relação às classificações geradas utilizando variáveis espectrais. Todas as classificações geradas usando o conjunto de variáveis espectrais apresentaram bons índices de acurácia, acima de 75% de acerto (Tabela 3), para as três áreas. Todas as classificações mostram que as características espectrais são as mais importantes, seguidas por características texturais, enquanto as características geométricas mostram a menor importância para a separabilidade das classes (Figura 3).

As variáveis espectrais podem discriminar melhor as diferenças no comportamento entre os alvos da superfície terrestre (Araújo e Melo, 2010), enquanto as variáveis de textura e geometria são secundárias, auxiliando no processo de classificação.

Isso demonstra como os dados espectrais foram fundamentais para uma boa acurácia nas classificações, em todas as áreas estudadas. Resultados semelhantes foram encontrados por Wieland et al. (2014), estudando o comportamento destas variáveis para a classificação de áreas urbanas, em que os

dados espectrais se destacaram pela melhor separabilidade das classes, seguidos das variáveis de textura.



**Figura 3** Pontuação das variáveis usadas no processo de classificação, obtidas com os algoritmos RF e DT. a) I; b) II; c) III. Variáveis: s(1) Brilho; s(2) Máxima diferença; s(3) Médiab1; s(4) Média b2; s(5) Médiabandab5; s(6) Médiab3; s(7) Médiab4; s(8) NDVI; s(9) SAVI; s(10) Desviopadrãob1; s(11) Desviopadrãob2; s(12) Desviopadrãob5; s(13) Desviopadrãob3; s(14) Desviopadrãob4; g(1) Assimetria; g(2) Índice de borda; g(3)

Comprimento da borda; g(4) Compacidade; g(5) Densidade; g(6) Comprimento; g(7) Comprimento/Largura; g(8) Área; g(9) Circularidade; g(10) Índice de Forma; g(11) Largura; t(1) GLCMCorrelaçãob1; t(2) GLCMCorrelaçãob2; t(3) GLCMCorrelaçãob5; t(4) GLCMCorrelaçãob3; t(5) GLCMCorrelaçãob4; t(6) GLCMEntropiab1; t(7) GLCMEntropiab2; t(8) GLCMEntropiab5; t(9) GLCMEntropiab3; t(10) GLCMEntropiab4; t(11) GLCMHomogeneidadeb1; t(12) GLCMHomogeneidadeb2; t(13) GLCMHomogeneidadeb5; t(14) GLCMHomogeneidadeb3; t(15) GLCMHomogeneidadeb4; t(16) GLCMMédiab1; t(17) GLCMMédiab2; t(18) GLCMMédiab5; t(19) GLCMMédiab3; t(20) GLCMMédiab4; t(21) GLCMDesviopadrãob1; t(22) GLCMDesviopadrãob2; t(23) GLCMDesviopadrãob5; t(24) GLCMDesviopadrãob3; t(25) GLCMDesviopadrãob4

Na área I, a menor acurácia foi obtida utilizando-se o conjunto de variável *s*, porém, utilizando alguns classificadores com este mesmo conjunto de variáveis (*s*), obtiveram-se as melhores classificações, como foi o caso de DT, NB e SVM (Tabela 3). Utilizando as variáveis texturais, os melhores resultados foram obtidos utilizando-se os classificadores RF e KNN. Os resultados na área II mostraram que melhores classificações foram geradas utilizando-se o conjunto de variáveis *stg* e *st*, em todos os classificadores utilizados. Estes resultados também foram percebidos na área III, em que as melhores classificações foram oriundas do conjunto de variáveis que continham dados texturais para todos os classificadores, exceto para o classificador NB, cuja melhor classificação foi com o conjunto de variáveis *sg*. Nas três áreas, os melhores resultados foram obtidos utilizando-se os conjuntos de variáveis *stg* ou *st*, para todos os algoritmos testados, mostrando que variáveis texturais podem ser importantes na classificação.

Para as três áreas, em relação aos conjuntos de variáveis, não houve diferença estatisticamente significativa entre as classificações, independente do algoritmo de classificação utilizado (Tabela 6). As diferenças significativas foram observadas somente quando não foram utilizadas variáveis espectrais nas classificações. As características de textura podem auxiliar muito no processo de classificação de imagens de sensoriamento remoto, principalmente quando se lida com áreas de grande heterogeneidade espectral (Ruiz et al., 2004). Segundo os mesmos autores, as características intrínsecas dos diversos objetos na superfície terrestre podem apresentar uma boa alternativa para distinguir as diferentes classes de uso. Porém, como foi observado neste estudo, para melhorar a classificação, estas variáveis precisam estar associadas a variáveis espectrais.

**Tabela 6** Matriz de comparações usando o mesmo algoritmo de classificação, porém, diferenciando o conjunto de variável usada. As comparações foram realizadas pelo teste de McNemar com chi-quadrado tabulado  $X^2=3,81$ ; a) área I; b) área II; c) área III

a)				
DT	stg	s	st	sg
stg	*			
s	0,11	*		
st	0,00	0,11	*	
sg	0,00	0,11	0,00	*
NB	stg	s	st	sg
stg	*			
s	0,96	*		
st	0,17	0,33	*	
sg	0,67	0,03	0,17	*
SVM	stg	s	st	sg
stg	*			
s	0,43	*		
st	0,24	0,03	*	
sg	0,33	1,5	1,13	*
b)				
DT	stg	s	st	sg
stg	*			
s	0,00	*		
st	0,00	0,00	*	
sg	0,00	0,00	0,00	*
NB	stg	s	st	sg
stg	*			
s	0,16	*		
st	0,16	0,66	*	
sg	0,80	0,24	1,70	*
SVM	stg	s	st	sg
stg	*			
s	1,30	*		
st	0,10	0,66	*	
sg	0,16	0,54	0,00	*
c)				
DT	stg	s	st	sg
stg	*			
s	0,16	*		
st	0,00	0,16	*	
sg	0,16	0,00	0,16	*
NB	stg	s	st	sg
stg	*			
s	0,24	*		
a)				
KNN	stg	s	st	sg
stg	*			
s	0,11	*		
st	0,03	0	*	
sg	0,11	0,03	0,03	*
RF	stg	s	st	sg
stg	*			
s	0,17	*		
st	0,17	0,67	*	
sg	0,06	0,03	0,43	*
b)				
KNN	stg	s	st	sg
stg	*			
s	0,00	*		
st	0,00	0,02	*	
sg	0,06	0,10	0,02	*
RF	stg	s	st	sg
stg	*			
s	0,16	*		
st	0,10	0,00	*	
sg	0,32	0,02	0,06	*
c)				
KNN	stg	s	st	sg
stg	*			
s	0,00	*		
st	0,00	0,00	*	
sg	0,54	0,66	0,66	*
RF	stg	s	st	sg
stg	*			
s	0,06	*		

**Tabela 6, conclusão**

DT					c)	KNN				
stg	s	st	sg	sg		stg	s	st	sg	
st	0,00	0,24	*		st	0,10	0,00	*		
sg	0,42	0,02	0,42	*	sg	0,06	0,00	0,00	*	
SVM										
stg	s	st	sg	sg						
stg	*									
s	0,10	*								
st	0,02	0,24	*							
sg	0,10	0,00	0,24	*						

Já as variáveis geométricas não foram muito eficazes na separação entre classes, sobretudo entre objetos como café, floresta e pastagem, principalmente porque os objetos destas classes, muitas vezes, apresentarem praticamente o mesmo formato, tamanho, assimetria, etc., o que pode ter dificultado uma melhor separação pelos algoritmos. Quando foram usados juntamente com variáveis espectrais (*sg*), a acurácia melhorou bastante, porém, a maioria ficou abaixo dos índices de acerto de conjuntos, como *s*, *st* e *stg*. O uso de múltiplas medidas de forma permite uma melhor discriminação entre os objetos e melhora a classificação de imagens (Van der Werff & Van der Meer, 2008), porém, estes autores trabalharam com áreas espectralmente semelhantes, o que tornou as variáveis geométricas importantes no processo de classificação, diferentemente deste estudo, no qual se trabalhou com áreas espectralmente heterogêneas. Além disso, segundo Witten et al. (2011), as características redundantes ou irrelevantes que fornecem pouca informação para uma classificação específica podem ter um efeito negativo sobre modelos de AM e podem levar a uma diminuição na acurácia da classificação. Portanto, as variáveis de geometria podem não ter desempenhado um papel tão importante na processo de separação das classes de cobertura da terra.

Durante a fase de treinamento de um classificador, os próprios algoritmos de AM já selecionam as variáveis importantes e ignoraram as irrelevantes ou redundantes (Witten et al., 2011). Como visto na Figura 3, as

classificações geradas nas três áreas mostraram que as variáveis mais expressivas foram as espectrais, seguidas das texturais e, por último, as geométricas, em todos os conjuntos de variáveis utilizados (*s*, *st*, *sg* e *stg*).

Todas as áreas mostraram os índices de vegetação NDVI e SAVI como uma das variáveis mais importantes para separabilidade nas classificações. Estes índices estão entre os principais utilizados no mapeamento da cobertura da terra (Machado et al., 2010). Diferentemente dos resultados obtidos por Sarmiento et al. (2014), a data de aquisição das imagens pode ter influenciado positivamente a capacidade dos índices em separar as classes de cobertura da terra, uma vez que, embora estas apresentem uma resposta espectral semelhante, a aquisição das imagens no período seco contribuiu para que as classes café e pastagem apresentassem níveis de biomassa verde distintos e, conseqüentemente, apresentando valores diferentes de NDVI e SAVI.

As 10 primeiras variáveis mais significativas na área I e III foram todas espectrais; já na área II, as 10 melhores foram, na grande maioria, espectrais, porém, a variável GLCMHomogeneidadeb5 (banda do infravermelho próximo) foi importante no processo de classificação. Isso mostra a importância destas variáveis no processo de classificação usando imagens de sensoriamento remoto. Segundo Souza et al. (2009), as variáveis de textura descrevem padrões de suavidade, rugosidade e regularidade dos alvos, sendo características importantes para reconhecer e classificar objetos.

É possível observar que, em todas as áreas, as maiores diferenças foram geradas quando se utilizaram variáveis texturais em algum momento das classificações, exceto na área II, usando KNN, na qual a melhor classificação foi obtida usando o conjunto de variáveis *s*. Nesse sentido, os resultados mostraram que as variáveis texturais, quando usadas combinadas com as espectrais, podem trazer alguns benefícios à separabilidade das classes, porém, não representam diferença estatisticamente significativa. Ruiz et al. (2004) também atestaram isso

em um estudo realizado em áreas florestais e urbanas, no qual as variáveis de textura forneceram uma alternativa para auxiliar as variáveis espectrais para a classificação de unidades florestais com um alta heterogeneidade espectral, ou quando as classes são definidas pelas diferenças na densidade da vegetação nativa.

Além disso, é importante ressaltar que muitas comparações não apresentaram nenhuma diferença estatística. Assim, é possível afirmar que, independente do conjunto de variáveis usadas, é imprescindível que as variáveis espectrais estejam presentes no processo de classificação de imagens de sensoriamento remoto.

### **3.3 Melhores variáveis para separação da classe café**

Nas três áreas, os conjuntos de variáveis apresentaram performances diferentes para a classificação do café, para todos os algoritmos utilizados. O maior índice de acerto, na acurácia do produtor, foi verificado para a área III, com 99,00% e o menor para a área II, com índice acerto de 82,86% (Tabela 3).

O maior índice de acerto obtido na área I foi verificado utilizando-se o conjunto *s* (98,46%), usando o algoritmo NB, enquanto o menor acerto foi usando a variável *sg*, usando o algoritmo RF, com 87,69% de acurácia (Tabela 3). A maioria dos conjuntos de variáveis *s* foi dos que geraram melhores índices de acerto em todos os algoritmos, porém, é importante salientar que quanto maior foi o índice de acerto desta classe, menores foram os índices de acerto para a classe vegetação nativa, sendo confundida com a classe café. O algoritmo RF usando o conjunto de variáveis *s* apresentou bastante confusão para as áreas de vegetação nativa, que foram classificadas, em sua maioria, como café. Isto pode ter ocorrido devido ao fato de a área ser composta por cerrado, o que pode ter causado uma confusão espectral maior com áreas cafeeiras. Todos os outros

conjuntos de variáveis obtiveram índices intercalados, variando o percentual de acerto para a classe café.

Na área II, o conjunto de variáveis *st* e *stg* foi o que mostrou melhores índices de acerto para a classe café (98,57%), em todos os classificadores testados, enquanto os menores índices foram obtidos utilizando-se a variável *s*, seguido da *sg* (Tabela 3). A classe vegetação nativa também foi confundida com a classe café, mostrando índices de acerto bem baixos quando usado o algoritmo NB para todos os conjuntos de variáveis empregados.

A área III foi a que apresentou os maiores índices de acerto para o café; todas as classificações obtiveram índices acima de 89,00%. Os menores índices foram obtidos utilizando o conjunto de variáveis *sg*, em todos os classificadores usados. Nesta área também foi verificado que a classe vegetação nativa foi confundida com a classe café, obtendo índices mais baixos de acerto.

Outros estudos realizados mapeando café indicaram resultados semelhantes, no qual áreas de café foram bem classificadas e áreas de vegetação nativa apresentaram índices de acerto mais baixos (Martínez-Verduzco et al., 2012; Andrade et al., 2013b; Sarmiento et al., 2014).

É possível observar, em sua maioria, as melhores variáveis para separar o café foram aquelas que continham dados texturais, porém, sempre estando associadas aos dados espectrais. Alguns trabalhos mostram que variáveis texturais podem auxiliar na melhoria da qualidade do mapeamento em áreas cafeeiras (Marujo et al., 2013; Santos et al., 2012), assim como alguns índices de vegetação, como o NDVI (Cordero-Sancho & Sader, 2007). Este fato também ocorreu neste estudo, avaliando áreas distintas, o que mostra que as variáveis texturais podem favorecer a classificação de áreas cafeeiras. Neste caso, a melhoria na qualidade das classificações pode ser decorrente dos algoritmos utilizados e não das variáveis usadas, uma vez que a maioria dos resultados obtidos utilizando algoritmos de AM produziu bons índices de acerto.

A classe café foi bem classificada nas três áreas, para acurácia do produtor, porém, algumas áreas de vegetação nativa foram classificadas como café, apresentando índices de acerto menores, principalmente nas áreas II e III (Tabela 3). Na área I, os maiores erros na classe vegetação nativa foram encontrados para o classificador RF, para todos os conjuntos de variáveis utilizados. Para as áreas II e III, os índices mais baixos foram obtidos utilizando o classificador NB, também independente do conjunto de variáveis empregadas. Analisando a acurácia do usuário, observou-se que a classe café apresentou índices mais baixos, sendo confundida com a classe vegetação nativa. Estes resultados mostram que, mesmo usando um conjunto de variáveis mais robusto e diferentes algoritmos para classificação, a vegetação e o café ainda são confundidos no processo de classificação. Adami et al. (2009) e Moreira et al. (2004) afirmam que o mapeamento de áreas cafeeiras, muitas vezes, é uma tarefa difícil, pois há uma grande confusão espectral entre os diferentes tipos de cobertura com respostas espectrais semelhantes, como é o caso da vegetação nativa e do café. De acordo com Li et al. (2014), o desempenho dos algoritmos pode ser pior em áreas mais complexas, devido à heterogeneidade espectral existente.

#### 4 CONCLUSÕES

Neste trabalho avaliaram-se a eficácia do uso de diferentes algoritmos de aprendizagem de máquina e a importância de diferentes conjuntos de variáveis para o mapeamento da cafeicultura, em três áreas ambientalmente distintas. De acordo com os resultados apresentados, pode-se chegar às seguintes conclusões:

1) nas áreas I e II, houve diferença significativa entre algumas classificações que utilizaram o mesmo conjunto de variáveis, porém, diferenciando o algoritmo usado. Entretanto, para todas as áreas, as classificações que utilizaram diferentes combinações de variáveis não mostraram diferença significativa entre elas;

2) os conjuntos de variáveis mais eficientes para classificar cafezais foram os espectrais nas três áreas. Apesar de as classificações não apresentaram diferença estatisticamente significativa, os mapas produzidos com a combinação de variáveis espectrais e texturais mostraram uma melhor discriminação entre plantações de café e vegetação nativa, no presente estudo;

3) os algoritmos mais eficientes para classificar cafezais, no presente estudo, foram SVM e RF, na área II e SVM, nas áreas I e III. No entanto, o algoritmo SVM foi o mais robusto, apresentando os melhores resultados para todas as áreas analisadas, usando diferentes conjuntos de variáveis;

4) durante o processo de classificação, a classe mais confundida com a classe café foi a classe vegetação nativa. As áreas de pastagem não apresentaram confusão com as áreas de café.

Porém, é necessário que haja mais testes para melhorar ainda mais a separabilidade entre as classes vegetação e café, uma vez que foram muito confundidas. Mais estudos estão sendo realizados para avaliar atributos temporais e outros sensores para melhorar o mapeamento de café em Minas Gerais.

## REFERÊNCIAS

ADAMI, M.; MOREIRA, M. A.; BARROS, M. A.; Avaliação da exatidão do mapeamento da cultura do café no estado de Minas Gerais. In: XIV Simpósio Brasileiro de Sensoriamento Remoto, 2009, Natal. **Anais. INPE**, 2009, p. 1–8.

ANDRADE, A. C.; FRANCISCO, C. N.; ALMEIDA, C. M. Desempenho de classificadores paramétrico e não paramétrico na classificação da fisionomia vegetal. **Revista Brasileira de Cartografia**, v. 65, n. 2, p. 227–242, 2013<sup>a</sup>.

ANDRADE, L. N.; VIEIRA, T. G. C.; LACERDA, W. S.; VOLPATO, M. M. L.; DAVIS-JUNIOR, C. A. Aplicação de redes neurais artificiais na classificação de áreas cafeeiras em Machado - MG. **Coffee Science**, v. 8, n. 1, p. 78–90, 2013b.

ARAUJO, T. P.; MELLO, F. M. Processamento de imagens digitais: razão entre bandas. **Geociências**, v. 1, p. 121-131, 2010.

BERTOLDO, M. A. **Caracterização edafambiental da cefeicultura na região de Três Pontas, Minas Gerais**. 2008. 161 p. Tese (Doutorado) – Universidade Federal de Lavras, Lavras.

COMPANHIA NACIONAL DE ABASTECIMENTO - CONAB. **Acompanhamento de safra brasileira**: Primeiro levantamento, 2014. 21p. (CONAB).

CORDERO-SANCHO, S.; SADER, S. A. Spectral analysis and classification accuracy of coffee crops using Landsat and a topographic-environmental model. **International Journal of Remote Sensing**, v. 28, n. 7, p. 1577–1593, 2007.

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**, v. 118, p. 259–272, 2012.

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

GISLASON, P. O.; BENEDIKTSSON, J. A.; SVEINSSON, J. R. Random Forests for land cover classification. **Pattern Recognition Letters**, v. 27, n. 4, p. 294–300, 2006.

GOMEZ, C.; MANGEAS, M.; PETIT, M.; CORBANE, C.; HAMON, P.; HAMON, S.; DE KOCHKO, A.; LE PIERRES, D.; PONCET, V.; DESPINOY, M. Use of high-resolution satellite imagery in an integrated model to predict the distribution of shade coffee tree hybrid zones. **Remote Sensing of Environment**, v. 114, n. 11, p. 2731–2744, 2010.

HARALICK, R. M.; SHANMUGAM, K.; DINSTEN, I. Textural features for image classification. **IEEE Transactions on Systems, Man, and Cybernetics**, v. 3, n. 6, p. 610–621, 1973.

HUSSAIN, M.; CHEN, D.; CHENG, A.; WEI, H.; STANLEY, D. Change detection from remotely sensed images: From pixel-based to object-based approaches. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 80, p. 91–106, 2013.

LANDIS, J. R.; KOCH, G. G. The Measurement of observer agreement for categorical data data for categorical of observer agreement the measurement. **Biometrics**, v. 33, n. 1, p. 159–174, 1977.

LI, C.; WANG, J.; WANG, L.; HU, L.; GONG, P. Comparison of classification algorithms and training sample sizes in urban land classification with Landsat Thematic Mapper imagery. **Remote Sensing**, v. 6, n. 2, p. 964–983, 2014.

LI, F.; LI, C.; XIAO, B.; WANG, Y. Mapping large-scale distribution and changes of aquatic vegetation in Honghu Lake , China , using multitemporal satellite imagery. **Journal of Applied Remote Sensing**, v. 7, p. 1–16, 2013.

MACHADO, M. L.; ALVES, H. M. R.; VIEIRA, T. G. C.; FERNANDES-FILHO, E. I.; LAERDA, M, P. C. Mapeamento de áreas cafeeiras (*Coffea arabica* L.) da Zona da Mata Mineira usando sensoriamento remoto. **Coffee Science**, v. 5, n. 2, p. 113–122, 2010.

MARTÍNEZ-VERDUZCO, G. C.; GALEANA-PIZAÑA, J. M.; CRUZ-BELLO, G. M. Coupling community Mapping and supervised classification to discriminate Shade coffee from Natural vegetation. **Applied Geography**, v. 34, p. 1–9, 2012.

MARUJO, R. F. B.; VOLPATO, M. M. L.; VIEIRA, T. G. C.; ALVES, H. M. R.; RIBEIRO, M. B. P. Classificação orientada a objetos aplicada a cultivos cafeeiros em Três Pontas - MG.. In: XVI Simpósio Brasileiro de Sensoriamento Remoto, 2013, Foz do Iguaçu. **Anais**. INPE, 2013, p. 1338–1345.

MOREIRA, M. A.; ADAMI, M.; THEODOR, F. Análise espectral e temporal da cultura do café em imagens Landsat Spectral and temporal behavior analysis of coffee crop in Landsat images. **Pesquisa Agropecuária Brasileira**, v. 39, n. 3, p. 223–231, 2004.

NOVACK, T.; ESCH, T.; KUX, H.; STILLA, U. Machine Learning comparison between WorldView-2 and QuickBird-2-simulated imagery regarding object-based urban land cover classification. **Remote Sensing**, v. 3, n. 12, p. 2263–2282, 2011.

ORGANIZAÇÃO INTERNACIONAL DE CAFÉ - OIC. **Promoção e desenvolvimento de mercado**. 1 p. (OIC). Disponível em: <http://www.ico.org/>. Acesso em: 25/11/2014.

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**, v. 12, p. S27–S31, fev. 2010.

PAL, M. Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, v. 26, n. 1, p. 217–222, jan. 2005.

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**, v. 51, p. 350–365, fev. 2013.

RUIZ, L. A.; FDEZ-SARRÍA, A.; RECIO, J. A. Texture feature extraction for classification of remote sensing data using wavelet decomposition: A comparative study. In: 20th ISPRS Congress, 2004, **Proceedings**. 2004, p. 1–6.

SANTOS, J. A.; GOSSELIN, P. H.; PHILIPP-FOLIGUET, S.; TORRES, R. S.; FALCÃO, A. X. Multiscale classification of remote sensing images. **IEEE Transactions on Geoscience and Remote Sensing**, v. 50, n. 10, p. 3764–3775, 2012.

SARMIENTO, C. M.; RAMIREZ, G. M.; COLTRI, P. P.; SILVA, L. F. L.; NASSUR, A. C.; SOARES, J. F. Comparação de classificadores supervisionados na discriminação de áreas cafeeiras em Campos Gerais – Minas Gerais. **Coffee Science**, v. 9, n. 4, p. 546-557, 2014.

VAN DER WERFF, H. M. A.; VAN DER MEER, F. D. Shape-based classification of spectrally identical objects. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 63, n. 2, p. 251–258, 2008.

VELOSO, M.H. **Coffe inventory through orbital imagery**. Rio de Janeiro: Instituto Brasileiro do Café, 1974. 20p. (SR-525).

VIEIRA, T. G. C.; ALVES, H. M. R.; BERTOLDO, M. A.; SOUZA, V. C. O. Geotechnologies in the assessment of land use changes. **Coffee Science**, v. 2, n. 2, p. 142–149, 2007.

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

WITTEN, I. H.; FRANK, E.; HALL, M. A. **Data mining: Practical machine learning tools and techniques**, 2011. p. 629.

## ARTIGO 2

(Artigo preparado de acordo com as normas da revista PAB)

### **Variáveis multitemporais para o mapeamento da cafeicultura: uma nova abordagem metodológica**

**Carolina Gusmão Souza<sup>1</sup>, Luis Marcelo Carvalho<sup>1,\*</sup> Tássia Borges Arantes<sup>1</sup>, Polyanne Aguiar<sup>1</sup>**

<sup>(1)</sup>Departamento de Ciência Florestal, Universidade Federal de Lavras;  
carolinagusmaosouza@gmail.com, passarinho@dcf.ufla.br,  
tassiabarantes@gmail.com

\*Autor de correspondência: Luis Carvalho, passarinho@dcf.ufla.br

### RESUMO

A cafeicultura é uma das principais atividades agrícolas do Brasil e desempenha papel importante no cenário socioeconômico brasileiro e internacional. Mapear e monitorar estas áreas de forma precisa é essencial. Sendo assim, é importante desenvolver técnicas que cumpram este papel, uma vez que as metodologias existentes para o mapeamento desta cultura ainda trazem muita confusão entre alvos. Portanto, o objetivo, neste trabalho, foi propor uma nova metodologia para o mapeamento de áreas cafeeiras por meio de dados multitemporais. As variáveis temporais utilizadas foram extraídas de uma série temporal NDVI Landsat, formada por imagens originais e imagens sintéticas. Foi utilizado o algoritmo BFAST para a extração de dados dos perfis temporais de café, pastagem e vegetação nativa. Para a classificação, foram usadas variáveis multitemporais derivadas do pacote *greenbrown*. Os resultados mostraram que o café, a pastagem e a vegetação nativa têm comportamentos temporais distintos, o que facilita a utilização deles dados como variáveis de entrada para o mapeamento. As classificações utilizando as variáveis temporais, associadas a dados espectrais, obtiveram altos índices de acurácia global com 93,00% de acerto. Quando utilizados somente os dados temporais, as classificações ainda

mostraram um percentual de acerto acima de 80%. As variáveis estatísticas mostraram os melhores resultados para o mapeamento, aumentando o índice de acerto global entre as classificações. O número de amostras de treinamento não teve influência no índice de acerto das classificações. Dados oriundos de séries temporais do pacote *greenbrown* foram eficientes no mapeamento de áreas cafeeiras, diminuindo a confusão entre os alvos e tornando o processo de classificação mais preciso.

**Termos de indexação:** sensoriamento remoto; *greenbrown*; imagens sintéticas; STARFM.

### **Multi-temporal variables for coffee mapping: A new methodological approach**

#### **ABSTRACT**

Coffee is one of the main agricultural activities in Brazil and plays an important role in the Brazilian and international socio-economic scenario. Map and monitor these areas accurately is essential. so it is important to develop techniques that fulfill this role, since the existing methodologies for mapping of this culture still bring a lot of confusion among targets. Therefore, the aim of this work is to propose a new methodology for mapping coffee areas through multi-temporal data. Temporal variables used were extracted from a Landsat NDVI time series consists of original and synthetic pictures. BFAST was using the algorithm for data extraction of temporal profiles of coffee, pasture and native vegetation. For classification were used multi-temporal variables derived from *greenbrown* package. The results showed that coffee, grass land and native vegetation have different temporal behavior, which facilitates use this data as input variables for the mapping. The maps using the temporal variables associated with spectral data, obtained high accuracy rates with 93.00%. When only used temporal data, the ratings also showed a hit percentage above 80.00% accuracy. The statistical variables fared best for mapping, increasing the success rate between the classifications. The number of training samples had no influence on the ratings hit rating. Derived from time series data *greenbrown* package were efficient to map coffee areas, reducing confusion between the targets and making more precise classification process.

**Index terms:** remote sensing; *greenbrown*; images synthetic; STARFM

## 1 INTRODUÇÃO

O mapeamento de áreas agrícolas tem se tornado importante, uma vez que estas áreas representam boa parte do desenvolvimento econômico de diversas regiões. No Brasil, a agricultura é responsável por 23% do PIB (Brasil, 2014), sendo o café umas das principais fontes de exportação do país, desempenhando papel importante no cenário socioeconômico brasileiro e internacional (CONAB, 2014). Dessa forma, desenvolver técnicas que façam o monitoramento e o mapeamento dessas áreas de forma precisa é fundamental.

Várias técnicas para o mapeamento desta cultura já foram utilizadas por diversos autores (Adami et al., 2009; Cordero-Sancho & Sader, 2007; Martínez-Verduzco et al., 2012; Marujo et al., 2013; Moreira et al., 2004; Santos et al., 2012; Sarmiento et al., 2014; Vieira et al., 2006, 2007), como classificação visual (Machado et al., 2010; Moreira et al., 2007; Trabaquini et al., 2011), classificação supervisionada usando abordagem por pixel (Cordero-Sancho & Sader, 2007; Martínez-Verduzco et al., 2012), abordagem baseada em objetos (Marujo et al., 2013; Santos et al., 2012; Sarmiento et al., 2014), algoritmos de aprendizagem de máquina (Andrade et al., 2013; Santos et al., 2012; Sarmiento et al., 2014), usando diferentes variáveis (Santos et al., 2012; Sarmiento et al., 2014; Souza et al., 2009). Porém, a maioria não alcançou resultados satisfatórios de acurácia (Andrade et al., 2013; Cordero-Sancho & Sader, 2007), uma vez que esta cultura é bastante confundida com outros tipos de vegetação, como pastagem e, principalmente, com vegetação nativa (Adami et al., 2009), quando utilizam-se técnicas de classificação automática.

Uma alternativa para o melhoramento do mapeamento da cafeicultura pode estar no uso de dados provenientes de séries temporais, por se tratar de uma opção para a avaliação da dinâmica da superfície terrestre, pois o componente temporal possibilita observar padrões complexos e analisar a dinâmica de cobertura da superfície (Bruzzone et al., 2003). Em poucos

trabalhos utilizaram-se estes dados como variáveis no processo de classificação da cobertura da terra (Jia et al., 2014; Mello et al., 2012; Santos, 2014), no qual se mostraram bem promissores no mapeamento. Porém, para a classificação de áreas cafeeiras, estas variáveis ainda não foram utilizados.

Dados provenientes de sensoriamento remoto de alta, como imagens Landsat, são muito importantes para o monitoramento da cobertura da terra porque as informações temporais ou fenológicas contidas nos dados de séries temporais oriundas destes dados são muito úteis para o mapeamento, especialmente para a classificação da cobertura vegetal (Jia et al., 2013; Xiao et al., 2002). Dados de índices de vegetação extraídos de séries temporais (ex.: NDVI e EVI) são bastante utilizados por apresentar uma boa descrição do crescimento da vegetação, e a forma dos perfis de crescimento da vegetação representada por estes índices contém informações sobre o tipo de vegetação analisada (Xiao et al., 2002; Brown et al., 2013). No entanto, a frequente contaminação por nuvens em imagens torna difícil a aquisição de séries temporais de alta resolução espacial e temporal de dados de sensoriamento remoto (Zhang et al., 2013). Assim, a classificação da cobertura da terra usando séries temporais foca mais em imagens com baixa resolução espacial, enquanto as imagens com resolução espacial melhor são utilizadas na análise da cobertura terrestre apenas com um único ou com menos dados de sensoriamento remoto (Gong et al., 2013).

Pape & Franklin (2008) afirmam que, para um bom mapeamento de habitats e monitoramento terrestre, é necessário que haja uma boa relação entre resolução espacial e temporal. Uma forma possível de atingir esse objetivo é por meio da fusão de dados de sensores com diferentes características espaciais e temporais (Hilker et al., 2013). Em geral, a fusão de dados reúne informações de multissatélites para gerar informações com alta resolução espacial e temporal (Gao et al., 2006).

Alguns autores utilizaram variáveis multitemporais no processo de classificação e obtiveram bons resultados usando estas variáveis no mapeamento da cobertura da terra. Mello et al. (2012) demonstraram uma nova abordagem no mapeamento de cana-de-açúcar, usando variáveis temporais provenientes do método STARS. Dois modelos foram testados em modelagem espectro-temporal de superfícies respostas. O modelo PTS apresentou superfícies suaves que podem ser eficazes para descrever a mudança gradual, como fenologia. Em contraste, o modelo CS apresentou superfícies mais nítidas e definidas, úteis para a caracterização de mudanças abruptas. As classificações obtiveram resultados excelentes, com índices de acerto de 99,00%.

Jia et al. (2014) utilizaram variáveis temporais de dados fusionados para melhorar o mapeamento da cobertura da terra. Características temporais foram extraídas de dados fusionados e adicionadas para melhorar a precisão da classificação. Os resultados indicaram que as características temporais extraídas de dados de imagens com resolução mais grosseira têm efeito significativo na melhoria da precisão da classificação de imagens com resolução melhor, especialmente para os diferentes tipos de vegetação. Santos (2014) avaliou 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 a compreensão do componente sazonal, usando imagens originais e sintéticas. Os dados temporais provenientes destas análises foram utilizados como variáveis no processo de classificação. Os resultados obtidos mostraram que estes dados proporcionaram uma melhoria significativa nas classificações, obtendo índices de acurácia de mais de 90,00%.

Nesse sentido, utilizar o perfil espectral ao longo do tempo, juntamente com uma abordagem de classificação orientada a objeto, integrando os dados temporais e espectrais no processo de classificação, pode ser uma boa alternativa

para bons resultados no mapeamento da cafeicultura. O uso de séries temporais para o mapeamento desta cultura ainda não foi empregado; a maioria dos trabalhos utiliza apenas períodos de observação mais curtos, de um a dois anos (Moreira et al., 2004). Além disso, ainda foi visto, na literatura, o uso de dados temporais como variáveis de entrada no processo de classificação de áreas cafeeiras.

Sendo assim, este trabalho foi realizado com o objetivo de desenvolver um padrão metodológico para mapeamento da cafeicultura por meio de dados multitemporais, aliado à classificação orientada a objeto. Nesse sentido, espera-se responder às seguintes questões: (i) Como o número de amostras de treinamento interfere no resultado da classificação? (ii) Variáveis temporais são eficazes na separabilidade do café? (iii) Qual a quantidade de variáveis necessária para aumentar a acurácia da classificação? (iv) Dados temporais Landsat TM são compatíveis para mapear imagens RapidEye?

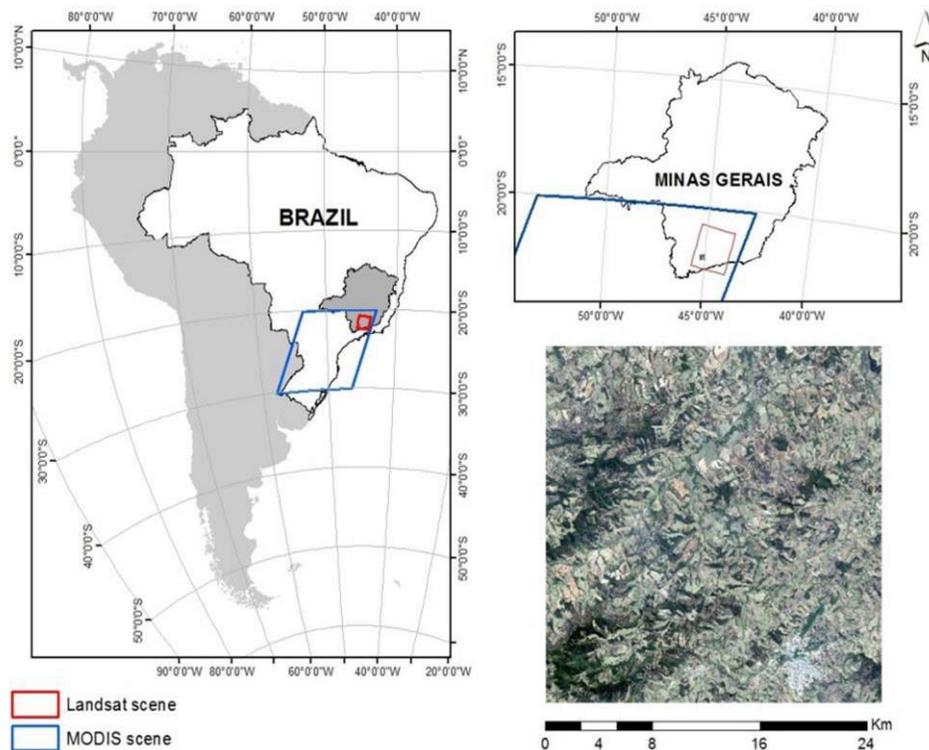
Desenvolver uma metodologia com maior precisão é importante, uma vez que resultados mais precisos e com uma rapidez maior sobre o estado e o tamanho destas áreas são fundamentais, pois estes dados podem servir como base para o monitoramento dessas áreas.

## **2 MATERIAIS E MÉTODOS**

### **2.1 Área de estudo**

A área de estudo está localizada na região sul do estado de Minas Gerais, Brasil e corresponde a um subset da cena Landsat TM órbita/ponto 218/75 (Figura 1). A vegetação é caracterizada por dois biomas brasileiros, Mata Atlântica e Cerrado (Carvalho, 2008). O clima da região é uma transição entre Cwa (clima temperado úmido com inverno seco e verão quente) e Cwb (clima temperado úmido com inverno seco e verão moderadamente quente), segundo a classificação de Köppen (Sá Júnior, 2009). O local de estudo é uma região de transição entre os biomas citados, que são considerados hotspots da biodiversidade e encontram-se bastante antropizados (Metzger, 2009). Dessa forma, a área é caracterizada por uma grande diversidade de configurações ambientais.

A vegetação nativa predominante na região é floresta estacional semidecidual montana (Veloso et al., 1991). A vegetação remanescente nativa é bastante fragmentada, sendo a parte antropizada composta, principalmente, por áreas de pastagem e por plantios de café.



**Figura 1** Localização da área de estudo, região sul do estado de Minas Gerais, Brasil, com a localização das cenas Landsat TM 218/75 e MODIS h13v11. Em detalhe o subset da RapidEye onde se conduziu o estudo

## 2.2 Dados das imagens e pré-processamento

Para este estudo, foram utilizadas imagens de três sensores: uma imagem RapidEye (tile 2328914) de 19 de setembro de 2012. *Imagens thematic mapper* (TM), satélite Landsat, órbita/ponto 218/75 e *Moderate resolution imaging spectroradiometer* (MODIS), cena h13v11, satélite Terra.

As imagens do sensor RapidEye representam um conjunto de dados de alta resolução espacial, que foi utilizado como base de detalhamento espacial deste trabalho. As imagens deste sensor têm 5 m de resolução espacial,

resolução radiométrica de 16 bits e resolução espectral de 5 bandas. A imagem utilizada foi fornecida pelo Ministério do Meio Ambiente.

Todas as imagens Landsat TM, disponíveis, entre 2000 e 2011, foram adquiridas junto ao portal U. S. Geological Survey (<http://earthexplorer.usgs.gov/>), totalizando 195 imagens. Estas imagens têm resolução espacial de 30 m, resolução radiométrica de 8 bits e resolução espectral de 7 bandas. Foi utilizado o produto *Reflectance climate data record landsat surfasse* (CDR). Este produto é gerado a partir do software especializado chamado *Landsat Ecosystem Disturbance Adaptive Processing System*, ou LEDAPS, que 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 (Gao, 2009). Além disso, o produto é ortoretificado. Estas imagens foram utilizadas para compor a série temporal NDVI, juntamente com as imagens Landsat sintéticas (ver itens 2.3 e 2.4).

Também foi adquirido um conjunto de 273 imagens do sensor MODIS *composite*, (produto MOD13Q1), no portal Modis Reproject Tools Web (<https://mrtweb.cr.usgs.gov/>), do mesmo período das imagens Landsat TM. As imagens MODIS consistem em um mosaico de pixels sem nuvens, observados nas imagens diárias do MODIS, em um período de 16 dias (Anderson & Shimabukuro, 2007). O produto MOD13Q1 tem resolução espacial de 250 m para as bandas NIR e RED, as quais foram utilizadas para a predição das respectivas bandas Landsat e para compor a série temporal NDVI MODIS (ver itens 2.3 e 2.4). As bandas NIR e RED das imagens MODIS foram reprojctadas para Universal Transverse Mercator (UTM), datum WGS-84 zona 23S, usando o software MODISRe-projection Tools (MRT). Além disso, o pixel foi

reamostrado para 30 m, seguindo os mesmos parâmetros das imagens Landsat TM. Em todas as imagens MODIS foram feitos subsets do tamanho da imagem Landsat TM.

### 2.3 Predição de imagens sintéticas

Para a predição de imagens foi utilizado o algoritmo *Spatial and Temporal Adaptive Reflectance Fusion Model* (STARFM), proposto por Gao et al. (2006), para a fusão das imagens Landsat TM e MODIS. Este procedimento é necessário, uma vez que frequentes coberturas de nuvens podem comprometer a análise dos dados e, aliado a isso, algumas imagens Landsat TM não estão disponíveis para download em determinadas datas, dificultando o uso contínuo destas imagens. Já as imagens MODIS, apesar de uma resolução espacial menor e fornecer dados orbitais há menos tempo, têm uma resolução temporal melhor (1 ou 2 dias), o que melhora significativamente a composição da série temporal.

O algoritmo STARFM prediz valores de reflectância da superfície de imagens Landsat a partir de pares de imagens Landsat (*fine-resolution*) e MODIS (tempo  $T_1$ ), de preferência de mesma data, e imagem MODIS (*coarse-resolution*), correspondentes à data da imagem a ser predita (tempo  $T_2$ ) (Gao et al. 2006). O resultado final é uma imagem de alta resolução temporal e espacial (Gao et al. 2006).

O algoritmo é dado por Gao et al. (2006).

$$L\left(\frac{x_w}{2}, \frac{y_w}{2}, t_2\right) = \sum_{i=1}^w \sum_{j=1}^w \sum_{k=1}^n W_{ijk} * (M(x_i, y_j, t_2) + L(x_i, y_j, t_1) - M(x_i, y_j, t_1)) \quad (1)$$

em que  $L\left(\frac{x_w}{2}, \frac{y_w}{2}, T_2\right)$  é o valor do pixel TM predito para o tempo 2 ( $T_2$ ),  $w$  é o tamanho da janela móvel e  $\left(\frac{x_w}{2}, \frac{y_w}{2}\right)$  é o pixel central dessa janela móvel.

O peso  $W_{ijk}$  é calculado a partir de três medidas: diferença espectral entre MODIS e Landsat 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 semelhante, dessa forma determinando o quanto cada pixel vizinho 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., 2009a). Neste estudo, foi utilizada uma área de 750 m X 750 m como *moving window*.

Todas as imagens Landsat TM que tinham nuvem, independente da quantidade, foram preditas pelo algoritmo STARFM. A predição foi feita para as bandas NIR e RED, para posterior cálculo do NDVI. Foi feito um mosaico das imagens Landsat TM originais e sintéticas, para que os valores preditos substituíssem apenas os locais com cobertura de nuvem. Sendo assim, nos lugares onde não havia nuvens na imagem, foram mantidos os valores originais. As imagens que tinham mais de 80% de nuvens foram descartadas, sendo substituídas por imagens totalmente preditas.

Para a predição, optou-se por usar imagens LandsatTM totalmente livres de nuvem ou com menos de 20%, adquiridas nas datas mais próximas das imagens a serem preditas. A imagem MODIS  $T_1$  foi aquela cuja data estava mais próxima à data da imagem Landsat $T_1$ , e a MODIS  $T_2$ , correspondente à imagem Landsat a ser predita (Quadro 1).

**Quadro 1** Exemplo de imagens MODIS e TM em  $T_1$  e  $T_2$  usadas para a predição de imagens TM em  $T_2$ , para o ano de 2010

<b>MODIS <math>T_1</math></b>	<b>TM <math>T_1</math></b>	<b>MODIS <math>T_2</math></b>
28/07/2010	01/08/2010	01/01/2010
02/02/2010	06/02/2010	17/01/2010
09/05/2010	13/05/2010	02/02/2010
02/02/2010	06/02/2010	18/02/2010
28/07/2010	01/08/2010	06/03/2010
28/07/2010	01/08/2010	22/03/2010
28/07/2010	01/08/2010	07/04/2010
09/05/2010	13/05/2010	23/04/2010
28/07/2010	01/08/2010	25/05/2010
28/07/2010	01/08/2010	10/06/2010
28/07/2010	01/08/2010	26/06/2010
28/07/2010	01/08/2010	12/07/2010
14/09/2010	18/09/2010	13/08/2010
16/10/2010	20/10/2010	14/09/2010
16/10/2010	20/10/2010	30/09/2010
16/10/2010	20/10/2010	16/10/2010
16/10/2010	20/10/2010	01/11/2010

### 2.3.1 Avaliação de imagens Landsat sintéticas

Para avaliar a eficácia do algoritmo STARFM na predição de imagens e comprovar sua eficácia em análises de séries temporais, os valores observados nas imagens originais e preditas foram comparados por meio de duas abordagens, uma baseada em objeto e outra baseada em pixel. Para a abordagem baseada em objeto a imagem Landsat TM original foi segmentada, por meio do algoritmo multirresolução, do software eCognition Developer, tendo como parâmetros de escala 250 e forma e compacidade de 0,5, para ambas. Os mesmos segmentos foram utilizados na imagem sintética. Foi realizada a predição de uma imagem sem nuvem para o ano de 2010, com o objetivo de comparar os valores originais (dos objetos e dos pixels) com os valores preditos.

Foram feitas regressões lineares simples para avaliar o coeficiente de determinação das bandas 3 e 4 e imagem NDVI.

#### 2.4 Série temporal NDVI

A série temporal Landsat TM foi construída por imagens originais e preditas. Para montar a série temporal, todas as imagens Landsat TM, originais e preditas, foram transformadas em índice *normalized difference vegetation index* (NDVI). Esse índice é baseado em uma relação da diferença entre as bandas do vermelho ( $\rho_{RED}$ ) e infravermelho próximo ( $\rho_{NIR}$ ), realçando os sinais de vegetação (Sonnenschein et al., 2011), em que valores mais próximos de 1 indicam áreas mais vegetadas, e variam de -1 a 1. O NDVI é o índice de vegetação mais utilizado em estudos sobre vegetação (Sonnenschein et al., 2011; Verbesselt et al., 2010). Este índice é dado por

$$NDVI = (\rho_{NIR} - \rho_{RED}) / (\rho_{NIR} + \rho_{RED}) \quad (2)$$

em que  $\rho_{NIR}$  e  $\rho_{RED}$  indicam os valores de reflectância nas regiões das bandas NIR e RED, respectivamente (Rouse, 1974).

##### 2.4.1 Análise da série temporal

No intuito de verificar se dados temporais são eficientes na separabilidade das classes, foi avaliado o comportamento das assinaturas temporais de diferentes tipos de cobertura da terra. Para isto foi usado o algoritmo *breaks for additive seasonal and trend* (BFAST), desenvolvido por Verbesselt et al. (2010). Para utilizar os parâmetros derivados de dados multitemporais no processo de classificação das imagens foi utilizado o algoritmo Greenbrown, desenvolvido por Forkelet et al. (2013).

#### 2.4.1.1 BFAST

O BFAST é um algoritmo capaz de decompor a série temporal em três componentes separadamente: tendência ( $T_t$ ), sazonalidade ( $S_t$ ) e ruído ( $et$ ), utilizando métodos para detectar e caracterizar mudanças dentro da série temporal (Verbesselt et al., 2010). Sua fórmula pode ser expressa por

$$Y_t = T_t + S_t + et, t = 1, \dots, n \quad (3)$$

em que  $Y_t$  são os dados observados no tempo  $t$ ;  $T_t$  é o componente de tendência;  $S_t$  é o componente sazonal e  $et$  é o restante do componente ou ruído. As mudanças são detectadas por meio da identificação de pontos de quebra resultantes a partir do processo de decomposição BFAST.

Este algoritmo utiliza o procedimento *seasonal-trend decomposition* (STL) para decompor a série temporal, o qual fornece uma estimativa precisa e robusta de componentes de tendência e de sazonalidade, graças à sua capacidade de lidar com valores extremos ou à falta de valores dentro da série histórica (Jacquin, Sheeren and Lacombe, 2010). STL é um processo iterativo que utiliza *LOcally wEighted regreSsion Smoother* (Loess) para decompor os componentes da série temporal (Verbesselt et al., 2010). Com a formação da série temporal decomposta, o componente tendência terá grande relevância na detecção de distúrbio real, pois permite a determinação da direção da mudança analisando o sinal de inclinação da tendência. Para a aplicação deste método, foi utilizado o pacote “BFAST” disponível no programa RStudio (R Development Core Team, 2014).

#### 2.4.1.2 Greenbrown

Neste estudo, foi utilizada a metodologia proposta por Forkel et al., (2013), que permite avaliar diferentes métodos para a estimação da tendência.

Essa metodologia foi implementada usando o pacote denominado *greenbrown* (<http://greenbrown.r-forge.r-project.org/>), disponível para o software R (R Core Team, 2014). Este pacote apresenta uma gama de funções para analisar tendências e mudanças de tendências em séries temporais. Neste estudo, computaram-se as propriedades estatísticas das séries temporais de imagens Landsat 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. Além destas propriedades estatísticas também analisaram-se outros métodos de tendência, sendo eles: *annual aggregated time series* (AAT), *season-trend model* (STM), *mean anual cycle* (MAC) e *annual cycle based on singular spectrum analysis* (SSA). Estes métodos se diferenciam na resolução temporal da série NDVI, presença ou não do componente sazonal e em como a tendência é estimada (Forkel et al., 2013).

No método AAT, a sazonalidade é agregada para valores anuais e a inclinação da reta de cada segmento de tendência é estimada por mínimos quadrados de uma regressão linear (Forkel et al., 2013), em que a significância é estimada pelo teste de tendência Mann-Kendall (Mann, 1945), citado por Forkel et al. (2013).

No método STM, a estimação da tendência e dos *breakpoints* é baseada no BFAST (Verbesselt, et al., 2010), em que o valor de NDVI ( $y$ ) em um tempo  $t$  pode ser expresso por

$$y_t = \alpha_1 + \alpha_2 t + \sum_{j=1}^k \gamma_j \sin\left(\frac{2\pi j t}{f} + \delta_j\right) + \varepsilon_t \quad (4)$$

em que  $\alpha_1$  é o intercepto da reta de tendência;  $\alpha_2$  é a inclinação da reta de tendência;  $\gamma$  são as amplitudes;  $\delta$  são as fases;  $f$  é a frequência (número de

observações em um ano, aqui  $f = 23$ );  $k$  são os termos harmônicos e  $\varepsilon$  é o erro (Verbesselt et al., 2010). A significância é estimada pelo teste-t (Forkel et al., 2013).

Para os métodos MAC e SSA, a sazonalidade é removida, ou seja, a tendência é estimada sem o componente sazonal e, assim, a inclinação da reta de tendência  $\alpha_2$  é estimada por mínimos quadrados ordinários (OLS).

$$a_t = \alpha_1 + \alpha_2 t + \varepsilon_t \quad (5)$$

em que  $a$  é a série ajustada sem o componente sazonal (Forkel et al., 2013). A significância de cada segmento de tendência é estimada pelo teste de tendência Mann-Kendall (Forkel et al., 2013). A diferença entre os métodos MAC e SSA está no ciclo sazonal que, no primeiro, é o ciclo anual médio, no qual considera-se que o ciclo sazonal tem a mesma amplitude e frequência para todos os anos e, no segundo, é o ciclo anual modulado, em que as amplitudes e frequências variam ao longo dos anos (Forkel et al., 2013).

Em todos os métodos, um mesmo algoritmo para a detecção de *breakpoints* foi utilizado, descrito por Bai & Perron (2003) e Zeileis et al., (2003) citados por Forkel et al. (2013). Primeiramente, o teste de *ordinary-least squares moving sum*, ou MOSUM, é aplicado para verificar a existência de *breakpoints* significativos na série. Depois, o número ótimo de *breakpoints* e a sua localização na série são testados interativamente (Forkel et al., 2013).

## 2.5 Processo de classificação

A classificação foi realizada em uma imagem RapidEye do ano de 2012 e, sendo assim, as variáveis espectrais foram provenientes desta imagem. Foi utilizada uma abordagem baseada em objeto, portanto, a imagem foi segmentada utilizando-se o algoritmo multirresolução, com escala de 250, forma 0,3 e

compacidade 0,5. A segmentação foi realizada pelo software eCognition Developer. Posteriormente, foram calculados: média, desvio padrão, brilho e máxima diferença para cada objeto, em cada uma das cinco bandas da imagem, os quais foram exportados juntamente com as demais variáveis, provenientes dos resultados das análises temporais. As variáveis associadas aos objetos foram subsequentemente utilizadas como dados de entrada para o algoritmo de classificação (Tabela 2), lembrando que as variáveis temporais foram geradas para os diferentes conjuntos de amostras de treinamento (ver item 2.5.1).

Neste estudo foi testada a separabilidade entre as classes Vegetação nativa: áreas de formações florestais densas e florestas de galeria, e formações de cerrado; Café 1: lavouras em idade não produtiva, no início do estágio de crescimento; Café 2: lavouras em estágio de crescimento intermediário; Café 3: lavouras com idade superior a 3 anos; Pastagem: áreas de pastagens naturais e formadas; Outros usos 1: áreas com culturas anuais em diversos estágios de desenvolvimento; floresta de produção; Outros usos 2: áreas urbanas e benfeitorias, áreas de solo exposto, áreas de queimadas; Corpos d'água: rios, córregos e represas, lagoas naturais e artificiais.

Todas as classificações foram realizadas por meio do algoritmo de aprendizagem de máquina *Support Vector Machine* (SVM). Este algoritmo foi escolhido por apresentar bons resultados em diversos trabalhos (Duro et al., 2012; Li et al., 2014; Wieland & Pittore, 2014).

### **2.5.1 Amostras de treinamento**

No intuito de estimar um número ideal de amostras de treinamento para classificação, foram testadas diferentes quantidades de amostras. Foram utilizados quatro conjuntos distintos, sendo eles: (A) 500 amostras; (B) 700 amostras; (C) 1.000 amostras e (D) 1.500 amostras. Estes números foram escolhidos de acordo com o recomendado por Niel et al. (2005), que aconselham

que o tamanho de amostra de treinamento não deve ser menor do que 10 a 30 vezes o número de bandas. Nesse sentido, foram utilizados intervalos que contemplassem estes números.

### **2.5.2 Variáveis usadas na classificação**

No intuito de verificar a eficiência de diferentes variáveis na classificação, foram utilizadas diferentes conjuntos de variáveis: espectrais e temporais (Tabela 1). As variáveis espectrais foram utilizadas neste estudo de acordo com as variáveis usadas no capítulo 3, que mostraram a importância destas variáveis na classificação de imagens de sensoriamento remoto para áreas cafeeiras. As variáveis temporais foram escolhidas de acordo com a metodologia proposta por Forkel et al. (2013), que disponibiliza distintas variáveis temporais que permitem diferenciar o comportamento de diferentes tipos de uso da terra.

A fim de avaliar a importância das variáveis temporais e quais delas eram mais importantes no processo de separação das classes, foram testados diferentes conjuntos de variáveis temporais: (1) GetStatistic; (2) GetStatistic+STM; (3) GetStatistic+STM+AAT; (4) GetStatistic+STM+AAT+MAC; (5) GetStatistic+STM+AAT+MAC+SSA; (6) Melhores variáveis. Estes dados foram testados de duas maneiras, a primeira usando, juntamente com os dados temporais, dados espectrais da RapidEye e a segunda, usando somente os dados temporais.

**Tabela 1** Variáveis utilizadas como input no processo de classificação do algoritmo SVM. (b1=banda 1; b2= banda 2; b3= banda 3; b4= banda 4; b5= banda 5)

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 - <i>greenbrown</i>	GetStatistic (Média, Tendência, IAV, Desvio padrão IAV, Variação sazonal, STV, Desvio padrão STV)
		AAT
		STM
		MAC
		SSA

## 2.6 Amostras de acurácia e análise estatística

Com o objetivo de conferir a acurácia das classificações, foram coletados 400 pontos como dados de teste. Os pontos foram divididos da seguinte forma: 60 como pastagem, 90 como vegetação, 70 como outros usos, 65 como café e 15 como água. Os pontos foram obtidos por meio de uma amostragem estratificada aleatória, em que cada estrato foi representado por uma categoria de classificação (Martínez-Verduzco et al., 2012). A conferência dos pontos foi feita por meio de visitas a campo e a partir de dados obtidos pelo Google Earth. Foram utilizados o coeficiente Kappa e a acurácia global como critérios de avaliação.

Os resultados das acurácias foram comparados usando o teste de McNemar (Foody, 2004), um teste não paramétrico que pondera a significância entre as diferenças de duas, baseando-se em uma matriz de confusão de duas dimensões (Foody, 2004) (Tabela 2).

**Tabela 2** Definição de elementos de matrizes utilizadas na equação 6 e 7

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

Este é um teste não paramétrico e a atenção é focada na distinção binária entre a alocação das classes correta ou incorretamente. O teste de McNemar é baseado no teste estatístico normal padronizado (Foody, 2004; Mas et al., 2010). É expresso por

$$Z = \frac{f_{12} - f_{21}}{\sqrt{f_{12} + f_{21}}} \quad (6)$$

O quadrado de Z segue uma distribuição chi-quadrado com um grau de liberdade. A equação do teste pode ser expressa por

$$Z^2 = \frac{(f_{12} - f_{21})^2}{f_{12} + f_{21}} \quad (7)$$

A avaliação do teste é baseada na distribuição chi-quadrado ( $Z^2$ ). Assim, o valor de chi-quadrado calculado é comparado com o valor de chi-quadrado tabelado, indicando sua significância estatística (Foody, 2004; Mas et al., 2010). Neste estudo foi usado o chi-quadrado nas análises.

### **3 RESULTADOS E DISCUSSÃO**

#### **3.1 Avaliação das imagens Landsat sintéticas**

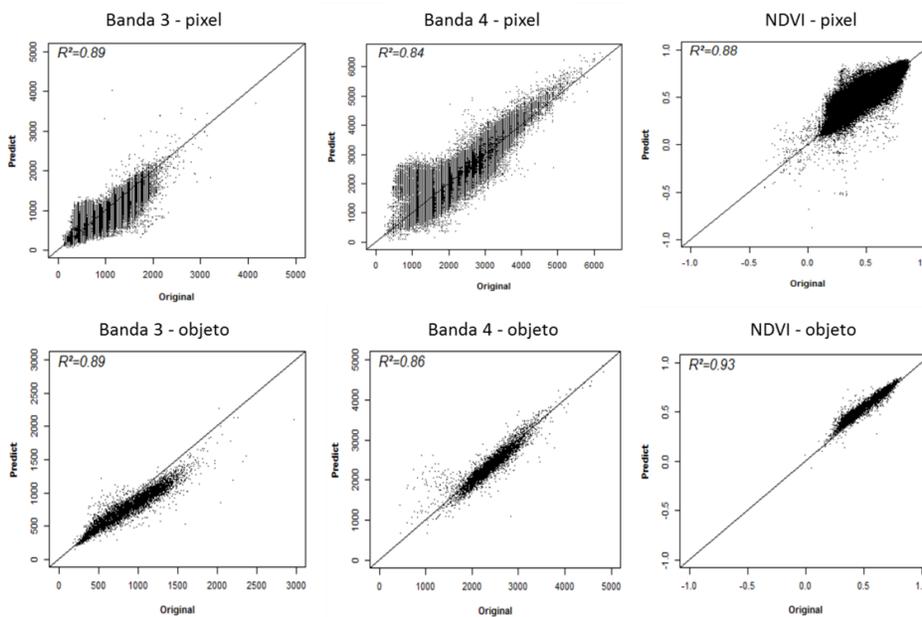
Um total de 195 imagens Landsat TM estavam disponíveis para download e 77 imagens não estavam disponíveis para download. Das imagens disponíveis, apenas 16 estavam completamente sem nuvens, o que corresponde a 8,2% do total. Cerca de 25%, ou seja, 46 imagens estavam totalmente ou com mais de 70% cobertas por nuvens. Portanto, foram preditas um total de 256, totalmente ou parcialmente, para compor a série temporal.

Com o objetivo de testar a eficácia de imagens sintéticas na análise de séries temporais, foram feitas regressões lineares para a imagem de 2010. As regressões foram realizadas para os valores de reflectância das bandas 3 e 4 e índice NDVI, comparando a imagem original e a respectiva imagem sintética.

A avaliação da imagem Landsat sintética foi realizada utilizando duas abordagens: a primeira baseada em pixel e a segunda baseada em objeto. Em todas as avaliações, a abordagem baseada em objeto apresentou melhores valores de coeficiente de determinação ( $R^2$ ) que a abordagem por pixel (Figura 2). Resultados semelhantes foram encontrados por Arantes (2014), na mesma área de estudo e por Santos (2014), trabalhando em uma área de Cerrado em Minas Geras, em que a abordagem por objeto apresentou melhores resultados que a abordagem por pixel.

Os melhores resultados na abordagem por pixel foram obtidos para a banda 3, seguido da banda NDVI. Para a abordagem por objeto, os melhores resultados foram para a banda NDVI, seguida da banda 3. A correlação das imagens mostrou resultados similares aos encontrados por Arantes (2014) e Santos (2014), que realizaram análises de correlação baseada em objetos, em que os valores de NDVI foram maiores que os das demais bandas. Em relação à

abordagem por pixel, os resultados obtidos também foram semelhantes aos encontrados por Arantes (2014); Santos (2014); Bhandari et al.(2012); Walker et al.(2012) e Watts et al. (2011), em que a correlação foi maior para a banda RED. Porém, todas as bandas mostraram bons resultados de correlação, confirmando a eficiência dessas imagens no uso para análises em séries temporais. Hilker et al. (2009) e Gao et al. (2006) mostraram, em estudos anteriores, uma alta correlação entre valores de pixels originais e preditos em todos os comprimentos de ondas observados. Sendo assim, pode-se afirmar que as imagens sintéticas são eficientes na substituição de imagens com nuvens.



**Figura 2** Comparação entre a reflectância da imagem Landsat TM original com a predita, nas abordagens por pixel e baseadas em objeto, referente a uma imagem do ano de 2010

Os padrões espaciais relacionados com diferentes tipos de cobertura da terra foram bem demonstrados na imagem sintética. O algoritmo STARFM tem

a característica de manter um alto nível de detalhe espacial nas imagens preditas, incluindo as áreas mais heterogêneas e complexas, com mais tipos de cobertura da terra (Hilker et al., 2009).

A imagem sintética analisada apresentou alguns valores de pixels que foram preditos erroneamente, com valores de background (-9999). Isso ocorreu tanto para a banda 3 quanto para a banda 4. Consequentemente, ao gerar a imagem NDVI, alguns pixels obtiveram valores discrepantes, com valores maiores que 1 e menores que -1. Segundo Arantes (2014), isto pode ocorrer devido ao fato de estes pixels se encontrarem em lugares sombreados na imagem original, apresentando valores de reflectância muito baixos na imagem Landsat TM em  $T_1$ , mas um valor não muito baixo na imagem MODIS em  $T_1$ , pois o pixel MODIS é maior, e o valor de reflectância é a soma das reflectâncias de diferentes alvos da superfície.

Os dados da predição podem ser influenciados pela qualidade imagens MODIS ( $T_1$  ou  $T_2$ ) (Arantes, 2014), uma vez que a qualidade dos dados preditos podem ser afetados em função da presença de nuvens nas imagens MODIS em  $T_2$ , principalmente em comprimentos de ondas mais curtos (Hilker et al., 2009). Portanto, é importante usar bons dados MODIS como input para predição.

A autocorrelação das imagens mostrou que é possível usar imagens sintéticas, principalmente utilizando uma abordagem baseada em objeto, uma vez que os resultados obtidos foram melhores.

## **3.2 Análise da série temporal**

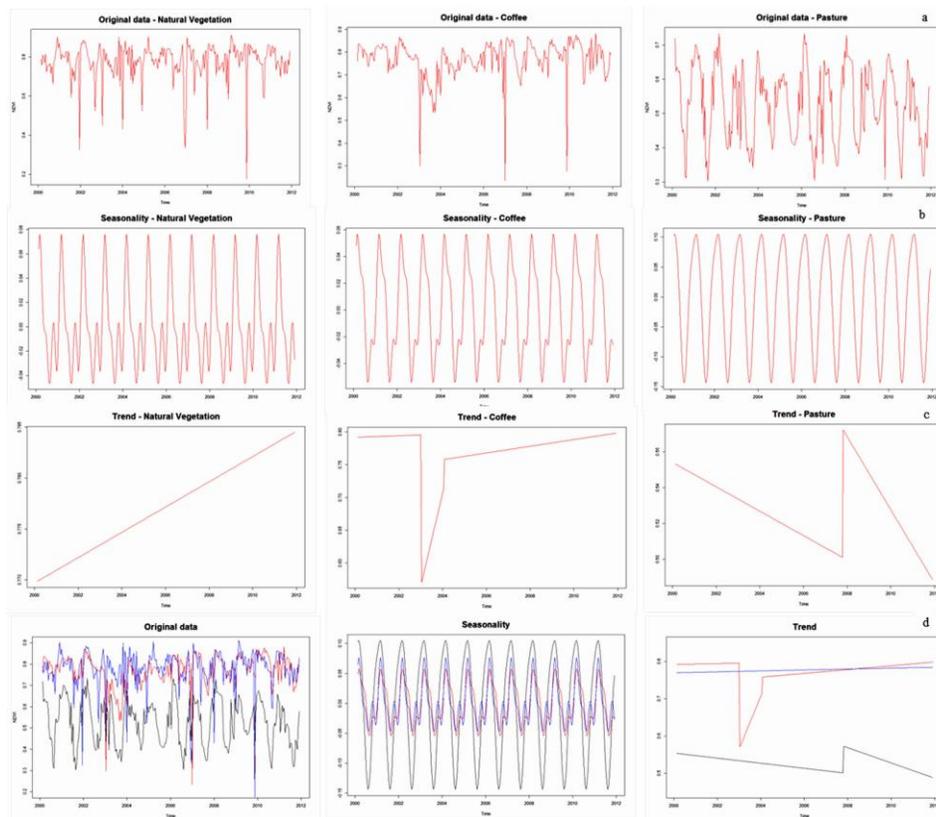
### **3.2.1 BFAST**

Por meio da análise da decomposição da série temporal, é possível observar que o comportamento dos diferentes tipos de cobertura da terra são distintos ao longo do tempo (Figura 3). O componente de tendência obteve *breakpoints* para a classe café e pastagem, enquanto o componente de

sazonalidade não apresentou nenhuma alteração ao longo do tempo, para as três classes analisadas. Segundo Verbesselt et al. (2010), mudanças no componente de tendência indicam mudanças graduais ou abruptas, mas isso não necessariamente evidencia uma mudança no tipo de cobertura. Já as mudanças no componente sazonal indicam mudanças fenológicas (Verbesselt et al., 2010), o que pode evidenciar alteração no tipo de cobertura da terra. As classes de cobertura da terra foram bem separadas pelo algoritmo BFAST.

No componente sazonal, o intervalo entre as curvas geradas por cada um dos tipos de cobertura tem um limiar e é possível observar a diferença de amplitude entre as classes e a diferença entre vegetação nativa, café e pastagem. Para pastagem, a sazonalidade variou de -0,15 a 0,10; já para a vegetação nativa, este intervalo foi menor, variando de -0,04 a 0,08, enquanto, para o café, a variação foi de -0,05 a 0,06. A forma dos perfis de crescimento dos diferentes tipos de cobertura da terra, representada por análises de séries temporais, contém informações sobre o comportamento de cada tipo de vegetação analisada (Xiao et al., 2002; Brown et al., 2013).

O componente de tendência também mostra variações ao longo do tempo, mostrando *breakpoints* para pastagem, ocorrido em 2008 e para o café, ocorrido em 2003, enquanto a vegetação nativa não apresentou nenhum *breakpoint*. Uma possível explicação para este comportamento pode ser porque a vegetação nativa, muitas vezes, sofre uma menor intervenção antrópica que outros tipos de cobertura, não apresentando mudanças ao longo do tempo, ao contrário das áreas de café e da pastagem, que podem ser manejadas e sofrerem modificações antrópicas durante os anos. Galford et al. (2008) afirmam que áreas agrícolas têm uma fenologia mais complexa em relação à vegetação nativa, apresentando mais *breakpoints*. Já para a pastagem, os *breakpoints* no componente de tendência podem ter ocorrido devido à falta de gestão dessas áreas e por serem bastante degradadas (Nascimento Júnior et al., 1994).



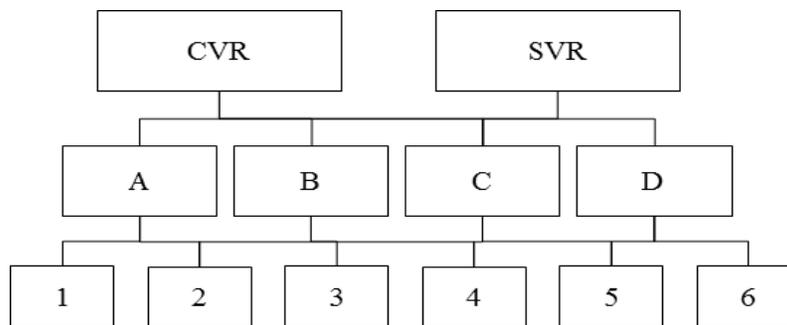
**Figura 3** Resultado do processamento do BFAST para vegetação natural, café e pastagem. a) dados originais da série temporal; b) sazonalidade dos três tipos de uso; c) tendência; d) comparação entre os diferentes tipos de uso da terra. Em azul: vegetação; vermelho: café; preto: pastagem

De acordo com Jensen (2006), as assinaturas temporais são muito importantes ao mapear diferentes tipos vegetacionais e extrair informações biofísicas (por exemplo, biomassa) a partir de dados de sensoriamento remoto. Estes resultados expressam a importância destas variáveis na diferenciação da cobertura da terra, pois o comportamento de cada tipo de cobertura é diferente ao longo do tempo. Sendo assim, usar estas variáveis no processo de classificação pode ser promissor no processo de separação entre classes.

### 3.3 Análise da classificação

No total, 48 classificações foram realizadas, usando quatro números diferentes de amostras de treinamento, seis conjuntos de variáveis distintas e usando com duas abordagens diferentes: uma abordagem com variáveis espectrais da imagem RapidEye e outra sem a utilização destas variáveis. A forma como foi tratada cada uma das variáveis e amostras de treinamento para melhor compreensão dos resultados é mostrada no Quadro 1.

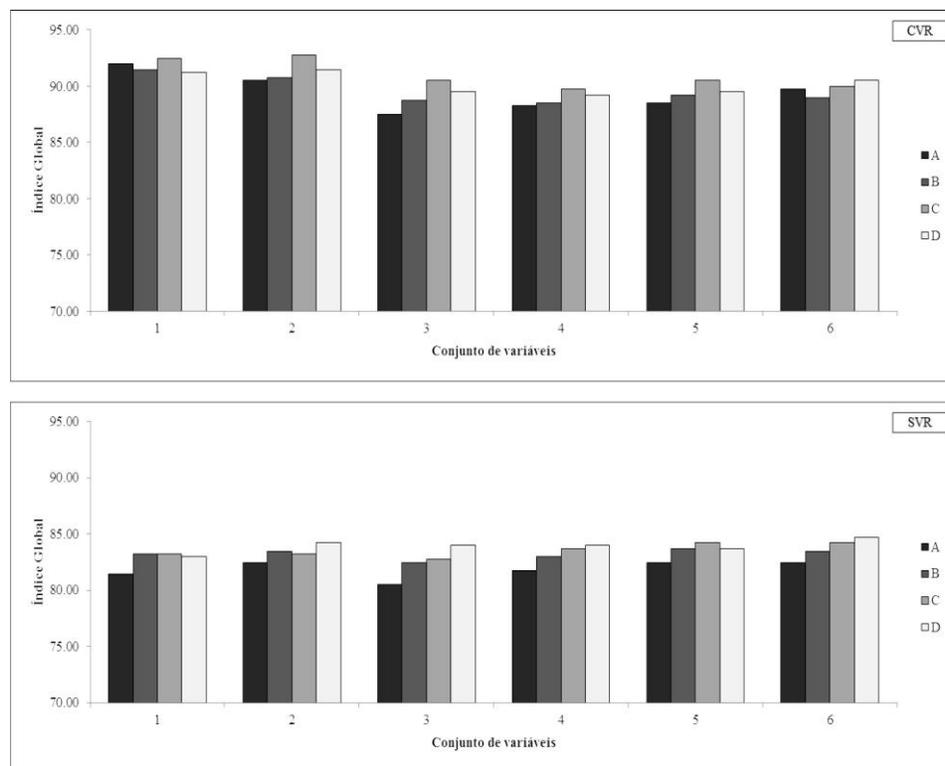
**Quadro 1** Parâmetros testados no processo de classificação. CVR = com variáveis espectrais Rapideye; SVR = sem variáveis espectrais Rapideye; A = 500 amostras; B = 700 amostras; C = 1.000 amostras; D = 1.500 amostras; 1 = GetStatistic; 2 = GetStatistic+AAT; 3 = GetStatistic+AAT+STM; 4 = GetStatistic+AAT+STM+ MAC; 5 = GetStatistic+AAT+STM+ MAC+SSA; 6 = Melhores variáveis



#### 3.3.1 Importância das variáveis temporais na classificação

Algumas variáveis temporais foram testadas no intuito de tentar separar áreas cafeeiras de outros tipos de cobertura da terra, principalmente como vegetação nativa e pastagem. As variáveis temporais obtiveram um efeito significativo na classificação da cobertura da terra (Figura 4). Os resultados

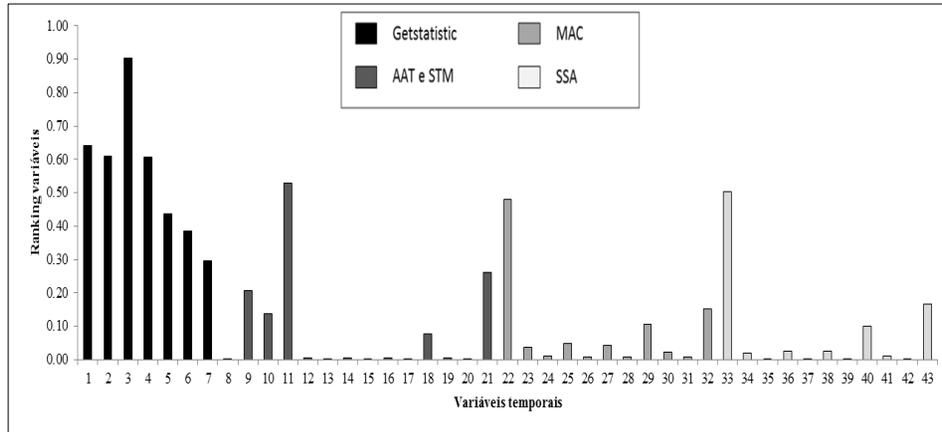
obtidos foram eficazes, mostrando índices altos de acurácia nas classificações para todos os conjuntos de variáveis usados, com acurácia mínima de 80,50% de acerto (SVR, A, 3), mesmo utilizando imagens preditas para compor as séries temporais. As imagens sintéticas se mostraram favoráveis para derivar os parâmetros temporais utilizados na classificação. Jia et al. (2014) e Watts et al. (2011) também mostraram a eficácia destas imagens, para classificação usando dados temporais.



**Figura 4** Acurácia das classificações para os diferentes conjuntos de variáveis temporais usadas. CVR = com variáveis RapidEye; SVR = sem variáveis RapidEye; A = 500 amostras; B = 700 amostras; C = 1.000 amostras; D = 1.500 amostras; 1 = GetStatistic; 2 = GetStatistic+AAT; 3 = GetStatistic+AAT+STM; 4 = GetStatistic+AAT+STM+ MAC; 5 = GetStatistic+AAT+STM+ MAC+SSA; 6 = melhores variáveis

Os conjuntos de variáveis 1 e 2 mostraram os melhores desempenhos nas classificações automáticas, com acurácia máxima de 92,75% de acerto, usando também variáveis RapidEye (CVR). É importante salientar que todas as classificações usando estes conjuntos de variáveis (CVR) obtiveram resultados superiores a 90,00% de acerto, o que, segundo Landis & Koch (1977), são resultados excelentes. Os demais conjuntos de variáveis usados (3, 4, 5 e 6) mostraram desempenhos alternados, porém, todos eles com bons índices de acerto. O conjunto de variáveis 1 usa o GetStatistic, que é composto por diferentes variáveis temporais, tais como sazonalidade, tendência, diferença intra- e interanual, diferentemente das demais variáveis que trabalham mais com aspectos de tendência (Forkel et al., 2013). Por ser um conjunto que tem uma quantidade de parâmetros mais heterogêneos, com mais análises temporais, é possível que isto explique o melhor desempenho das classificações. Além disso, Forkel et al (2013) afirmam que o *greenbrown* tem a capacidade de estimar mudanças ocorridas e também caracterizar as diferentes coberturas da terra existentes nos ecossistemas utilizando uma série temporal. Dessa forma, é possível afirmar que os dados oriundos das suas análises foram capazes de compor as variáveis de entrada no processo de classificação da cobertura da terra e das áreas cafeeiras.

É importante salientar que o conjunto de variáveis 6 é composto pelas melhores variáveis de separabilidade na classificação (Figura 5), porém, mesmo usando este conjunto de dados, as classificações não obtiveram resultados superiores. Isso mostra que as variáveis mais significativas na separabilidade das classes foram as do conjunto 1.



**Figura 5** Pontuação das variáveis usadas no processo de classificação, obtidas com os algoritmos *Random Forest* e *Decison Tree*

É necessário um tempo de processamento extenso para obter todas as análises disponíveis no pacote *greenbrown*, usando imagens Landsat (Tabela 3). O conjunto que usa como variáveis temporais só o GetStatistic mostrou o menor tempo de processamento, e foi este conjunto de dados que apresentou os melhores resultados nas classificações. Isto mostra que as outras variáveis não precisam ser colocadas no processo de classificação, o que diminuirá o tempo de processamento e esta metodologia poderá ser utilizada em áreas maiores.

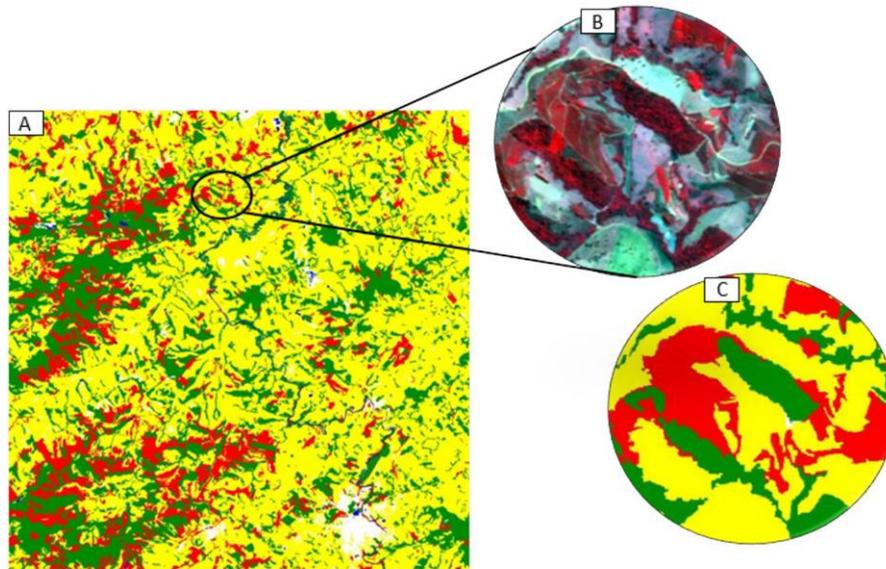
**Tabela 3** Tempos de processamento das variáveis temporais, usando o pacote *greenbrown*, no software RStudio

Variável temporal	Tempo de processamento
GetStatistic	4 horas
AAT	30 horas
STM	30 horas
MAC	3 semanas
SSA	3 semanas

O aumento da quantidade de variáveis temporais utilizadas mostrou que os desempenhos das classificações diminuíram um pouco, usando os dados CVR, o que não ocorreu com os dados SVR cujas acurácias apresentaram resultados alternados para os diferentes conjuntos de variáveis usadas. Porém, os melhores resultados foram obtidos usando o conjunto de variáveis 6, que contém um número de variáveis menores.

O uso de variáveis temporais é uma alternativa para melhorar a precisão das classificações e foi bastante importante para distinguir de maneira mais eficaz as áreas cafeeiras dos demais tipos de uso, como é expresso na Figura 6. As áreas cafeeiras foram bem distinguidas de áreas de vegetação nativa. Dados temporais têm um grande potencial para melhorar a classificação da cobertura da terra (Jia et al., 2014), principalmente se as características de imagens com resoluções melhores estiveram envolvidos no processo de classificação (Jia et al., 2014). Estes resultados corroboram os encontrados por Jia et al. (2014), uma vez que os melhores resultados foram obtidos utilizando o CVR.

Zhu & Woodcock (2014) também utilizaram algumas variáveis (média, diferença intra e interanual, sazonalidade) derivadas de séries temporais para classificação de uso da terra e obtiveram bons resultados na acurácia, com índice de acerto de 90,48%. Vieira et al.(2012) também mostraram bons resultados no mapeamento de cana, usando dados de séries temporais, juntamente com classificação baseada em objeto, obtendo excelentes resultados de acurácia com 94% de acerto. Estes resultados corroboram os encontrados neste estudo, o que mostra a importância destas variáveis para a classificação de áreas com grande similaridade espectral.



**Figura 6** Classificação CVR-A1 (A), detalhe em uma cena RapidEye de uma área ampliada (B) e detalhe correspondente no mapa da classificação (C)

Na Tabela 4 são mostradas as comparações entre a diferença dos conjuntos de variáveis usadas. Observa-se que não houve diferenças significativas entre as classificações, quando comparados os conjuntos de variáveis usados. Entretanto, as maiores diferenças foram para os resultados obtidos em CVR, em que as classificações mostraram uma diferença maior entre os índices de acerto, mas estes conjuntos de dados apresentaram melhores índices de acerto, quando comparadas as classificações usando os dados SVR.

**Tabela 4** Matrizes de comparações das classificações, mostrando a diferença entre os conjuntos de variáveis, realizadas pelo teste McNemar com chi-quadrado tabulado  $\chi^2=3,81$ . CVR = com variáveis RapidEye; SVR = sem variáveis RapidEye; A = 500 amostras; B = 700 amostras; C = 1000 amostras; A = 1500 amostras; 1 = GetStatistic; 2 = GetStatistic+AAT; 3 = GetStatistic+AAT+STM; 4 = GetStatistic+AAT+STM+ MAC; 5 = GetStatistic+AAT+STM+ MAC+SSA; 6 = Melhores variáveis

							CVR						
A	1	2	3	4	5	6	C	1	2	3	4	5	6
1	-						1	-					
2	0,18	-					2	0,00	-				
3	1,62	0,72	-				3	0,32	0,40	-			
4	1,12	0,40	0,04	-			4	0,60	0,72	0,04	-		
5	0,98	0,32	0,08	0,00	-		5	0,32	0,40	0,00	0,04	-	
6	0,40	0,04	0,40	0,18	0,12	-	6	0,50	0,60	0,02	0,00	0,02	-
B	1	2	3	4	5	6	D	1	2	3	4	5	6
1	-						1	-					
2	0,04	-					2	0,00	-				
3	0,60	0,32	-				3	0,24	0,32	-			
4	0,72	0,40	0,00	-			4	0,32	0,40	0,00	-		
5	0,40	0,18	0,02	0,04	-		5	0,24	0,32	0,00	0,00	-	
6	0,50	0,24	0,00	0,02	0,00	-	6	0,08	0,12	0,04	0,08	0,04	-
A	1	2	3	4	5	6	C	1	2	3	4	5	6
1	-						1	-					
2	0,04	-					2	0,00	-				
3	0,08	0,24	-				3	0,02	0,02	-			
4	0,00	0,02	0,12	-			4	0,02	0,02	0,08	-		
5	0,08	0,00	0,32	0,04	-		5	0,08	0,08	0,18	0,02	-	
6	0,08	0,00	0,32	0,04	0,00	-	6	0,08	0,08	0,18	0,02	0,00	-
B	1	2	3	4	5	6	D	1	2	3	4	5	6
1	-						1	-					
2	0,00	-					2	0,12	-				
3	0,04	0,08	-				3	0,08	0,00	-			
4	0,00	0,02	0,02	-			4	0,08	0,00	0,00	-		
5	0,02	0,00	0,12	0,04	-		5	0,04	0,02	0,00	0,00	-	
6	0,00	0,00	0,08	0,02	0,00	-	6	0,24	0,02	0,04	0,04	0,08	-

O uso de variáveis temporais leva em consideração as características das dinâmicas temporais (Vieira et al., 2012). No caso deste estudo, a área apresenta culturas com diversas variações temporais e, nesse sentido, tais parâmetros foram relevantes para potencializar melhores classificações e separabilidade entre as classes. Os resultados obtidos mostraram a importância destas variáveis para uma boa classificação do café, com índices superiores aos de outros

trabalhos realizados (Sarmiento et al., 2014; Andrade et al., 2013; Martínez-Verduzco et al., 2012; Santos et al., 2012; Vieira et al., 2007), até mesmo com metodologias que têm melhor acurácia. Entretanto, levam um maior tempo para o mapeamento (Cordero-Sancho & Sader, 2007).

Mello et al. (2012) mostraram, em um estudo realizado na região centro-oeste do Brasil, que variáveis derivadas de dados temporais mostraram resultados significativos no mapeamento de áreas agrícolas. Variáveis resultantes de dados temporais têm um potencial considerável para representar as mudanças de características espectrais ao longo do tempo na superfície da terra, o que as tornam úteis não só para fins de classificação, mas também para outras aplicações, como a compreensão de alterações na cobertura da terra (Mello et al., 2012).

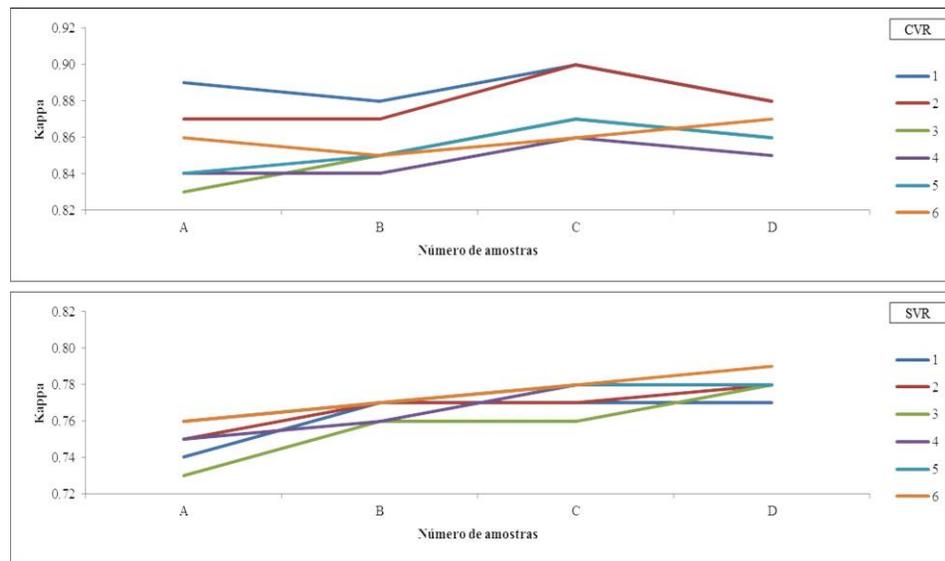
Áreas muito heterogêneas, como é o caso da região de estudo, são bem complexas de serem mapeadas (Li et al., 2014), porém, este novo método, utilizando variáveis temporais, se mostrou um procedimento viável e de fácil implementação, diminuindo a confusão entre áreas de café e vegetação nativa. Este método é recomendado, inclusive, para ser aplicado no mapeamento de outras áreas e apresenta grande potencial no mapeamento de áreas cafeeiras, e na separação entre classes de uso e cobertura da terra.

### **3.3.2 Uso de diferentes números de amostras de treinamento**

Diferentes conjuntos de treinamento foram utilizados para avaliar como o número de amostras de treinamento pode interferir na acurácia das classificações. Os resultados obtidos pelas classificações apresentaram performances distintas (Figura 7). Os melhores resultados encontrados foram usando o conjunto de 1.000 amostras de treinamento (C), usando CVR, que corresponde a 20 vezes o número de variáveis utilizadas na classificação. Usando as classificações SVR, as performances foram diferenciadas e a melhor classificação foi obtida pelo

conjunto D (1.500 amostras de treinamento), 30 vezes o número de variáveis usadas na classificação, usando o conjunto de variáveis 6.

Os piores resultados foram encontrados usando CVR para as classificações que utilizaram o conjunto de treinamento B (700 amostras), com exceção do conjunto de variáveis 3 e 5, cujos piores resultados foram encontrados pelo conjunto de treinamento A (500 amostras). Li et al. (2014) afirmam que quando o número de amostras é pequeno, nenhum algoritmo de classificação funciona de forma satisfatória, o que mostra a importância de um número mínimo de amostras de treinamento. No entanto, alguns algoritmos podem produzir bons resultados com número de amostras pequenos, como é o caso do SVM (Li et al., 2014).



**Figura 7** Acurácia das classificações para os diferentes conjuntos de amostras de treinamento usadas. CVR = com variáveis RapidEye; SVR = sem variáveis RapidEye; A = 500 amostras; B = 700 amostras; C = 1.000 amostras; D = 1.500 amostras; 1 = GetStatistic; 2 = GetStatistic+AAT; 3 = GetStatistic+AAT+STM; 4 = GetStatistic+AAT+STM+ MAC; 5 = GetStatistic+AAT+STM+ MAC+SSA; 6 = melhores variáveis

As análises estatísticas mostraram que não houve diferença significativa utilizando um número diferenciado de amostras de treinamento (Tabela 5). As maiores diferenças foram encontradas para os conjuntos C, em relação aos outros conjuntos, uma vez que foram o que apresentaram os maiores índices de acurácia para a maioria das classificações.

**Tabela 5** Matrizes de comparações das classificações, mostrando a diferença entre os conjuntos de amostras de treinamento, realizadas pelo teste McNemar com chi quadrado tabulado  $X^2=3.81$ . CVR = com variáveis RapidEye; SVR = sem variáveis RapidEye; A = 500 amostras; B = 700 amostras; C = 1000 amostras; A = 1500 amostras; 1 = GetStatistic; 2 = GetStatistic+AAT; 3 = GetStatistic+AAT+STM; 4 = GetStatistic+AAT+STM+ MAC; 5 = GetStatistic+AAT+STM+ MAC+SSA; 6 = Melhores variáveis

CVR											
A-1				A-2				A-3			
A-1	B-1	C-1	D-1	A-2	B-2	C-2	D-2	A-3	B-3	C-3	D-3
-	-	-	-	-	-	-	-	-	-	-	-
0,02	-	-	-	0,00	-	-	-	0,12	-	-	-
0,02	0,08	-	-	0,40	0,32	-	-	0,72	0,24	-	-
0,04	0,00	0,12	-	0,08	0,04	0,12	-	0,32	0,04	0,08	-
A-4				A-5				A-6			
A-4	B-4	C-4	D-4	A-5	B-5	C-5	D-5	A-6	B-6	C-6	D-6
-	-	-	-	-	-	-	-	-	-	-	-
0,00	-	-	-	0,04	-	-	-	0,04	-	-	-
0,18	0,12	-	-	0,00	0,08	-	-	0,00	0,08	-	-
0,08	0,04	0,02	-	0,02	0,02	0,12	0,00	0,02	0,12	0,00	-
SVR											
A-1				A-2				A-3			
A-1	B-1	C-1	D-1	A-2	B-2	C-2	D-2	A-3	B-3	C-3	D-3
-	-	-	-	-	-	-	-	-	-	-	-
0,24	-	-	-	0,12	-	-	-	0,32	-	-	-
0,24	0,00	-	-	0,08	0,00	-	-	0,08	0,04	0,08	-
0,18	0,00	0,00	-	0,32	0,04	0,08	-	0,12	-	-	-
A-4				A-5				A-6			
A-4	B-4	C-4	D-4	A-5	B-5	C-5	D-5	A-6	B-6	C-6	D-6
-	-	-	-	-	-	-	-	-	-	-	-
0,32	-	-	-	0,12	-	-	-	0,12	-	-	-

**Tabela 5, conclusão**

	A-3	B-3	C-3	D-3		A-4	B-4	C-4	D-4
C-3	0,40	0,00	-		C-4	0,32	0,04	-	
D-3	0,98	0,18	0,12	-	D-4	0,40	0,08	0,00	-
	A-5	B-5	C-5	D-5		A-6	B-6	C-6	D-6
A-5	-				A-6	-			
B-5	0,12	-			B-6	0,08	-		
C-5	0,24	0,02	-		C-6	0,24	0,04	-	
D-5	0,12	0,00	0,02	-	D-6	0,40	0,12	0,02	-

A dimensionalidade do espaço de características influencia o tamanho dos dados de treinamento que necessitam ser capturados e o número de amostras de treinamento deve ser significativamente maior do que o número de variáveis (Wieland & Pittore, 2014). Uma razão de pelo menos 10 vezes mais do que o número de variáveis é recomendada na literatura (Koprinska, 2010).

Os dados obtidos neste estudo evidenciaram que este número de 500 amostras (10 vezes mais o número de variáveis usadas) já seria suficiente para um bom desempenho nas classificações, entretanto, quando foram adicionadas mais amostras de treinamento, a acurácia aumentou, mostrando que o número ideal foi de 20 vezes o número de variáveis utilizadas (conjunto de dados C). Quando o número de amostras foi maior, com 30 vezes o número de variáveis, os resultados começaram a decrescer, usando CVR. Porém, usando os dados SVR, o conjunto de amostras D apresentou os melhores resultados, o que corrobora o recomendado por Van-Nil et al. (2005), que afirmam que este número pode ser de até 30 vezes a quantidade de características. Sendo assim, pode-se afirmar, para este estudo, que os conjuntos de amostras de treinamento aqui utilizados foram eficientes para os bons resultados no processo de classificação.

### **3.3.3 Eficácia de variáveis RapidEye para a classificação**

Para avaliar o desempenho das variáveis espectrais da imagem RapidEye, foram feitos testes retirando-se estes dados das variáveis de análise. Os resultados mostraram que as classificações que não usaram estes dados obtiveram desempenhos inferiores, quando comparadas com as classificações que usaram os dados espectrais. Porém, é importante salientar que, mesmo não usando estes dados, os resultados foram superiores ao de pesquisas anteriores mapeando café, como é o caso de Andrade et al. (2013), Martínez-Verduzco et al. (2012) e Cordero-Sancho & Sander (2007).

Na Tabela 6 apresenta-se a diferença entre os dois grupos de classificações, CVR e SVR. Uma análise de similaridade também foi realizada e os resultados mostraram que os dois grupos são diferentes e que há diferença significativa entre eles ( $p < 0,05$  e  $\chi^2 > 3,81$ ), portanto, as classificações podem ser consideradas diferentes.

**Tabela 6** Comparações das classificações, mostrando a diferença entre os conjuntos usando variáveis RapidEye (CVR) e sem variáveis RapidEye (SVR), realizadas pelo teste McNemar com Z tabelado e chi-quadrado tabulado  $X^2=3,81$ . Em negrito, os resultados significativos

<b>CVR x SVR</b>	<b>Z</b>	<b>X<sup>2</sup></b>	<b>Significativo</b>
<b>CVR-A1 x SVR-A1</b>	<b>2,96</b>	<b>8,82</b>	Sim, <0,05
<b>CVR-A2 x SVR-A2</b>	<b>2,33</b>	<b>5,44</b>	Sim, <0,05
<b>CVR-A3 x SVR-A3</b>	<b>2,64</b>	<b>7,00</b>	Sim, <0,05
CVR-A4 x SVR-A4	1,83	3,38	Não, > 0,05
CVR-A5 x SVR-A5	1,69	2,88	Não, > 0,06
<b>CVR-A6 x SVR-A6</b>	<b>2,05</b>	<b>4,20</b>	Sim, <0,05
<b>CVR-B1 x SVR-B1</b>	<b>2,33</b>	<b>5,44</b>	Sim, <0,05
<b>CVR-B2x SVR-B2</b>	<b>2,05</b>	<b>4,20</b>	Sim, <0,05
<b>CVR-B3x SVR-B3</b>	<b>2,17</b>	<b>4,71</b>	Sim, <0,05
CVR-B4x SVR-B4	1,55	2,42	Não, > 0,05
CVR-B5 x SVR-B5	1,55	2,42	Não, > 0,05
CVR-B6 x SVR-B6	1,55	2,42	Não, > 0,05
<b>CVR-C1 x SVR-C1</b>	<b>2,61</b>	<b>6,84</b>	Sim, <0,05
<b>CVR-C2 x SVR-C2</b>	<b>2,68</b>	<b>7,22</b>	Sim, <0,05
<b>CVR-C3 x SVR-C3</b>	<b>2,48</b>	<b>6,18</b>	Sim, <0,05
CVR-C4 x SVR-C4	1,69	2,88	Não, > 0,05
CVR-C5 x SVR-C5	1,76	3,12	Não, > 0,05
CVR-C6 x SVR-C6	2,64	2,64	Não, > 0,05
<b>CVR-D1 x SVR-D1</b>	<b>2,33</b>	<b>5,44</b>	Sim, <0,05
<b>CVR-D2 x SVR-D2</b>	<b>2,05</b>	<b>4,20</b>	Sim, <0,05
CVR-D3 x SVR-D3	1,81	3,27	Não, > 0,05
CVR-D4 x SVR-D4	1,48	2,20	Não, > 0,05
CVR-D5 x SVR-D5	1,62	2,64	Não, > 0,05
CVR-D6 x SVR-D6	1,55	2,42	Não, > 0,05

Os dados temporais, provenientes do pacote *greenbrown*, se mostraram significativos para uma boa classificação da cobertura da terra. Resultados semelhantes foram encontrados por Santos (2014), nos quais os dados temporais foram significativos no mapeamento de remanescentes florestais. Os resultados abordados no capítulo anterior, mapeando áreas cafeeiras na mesma região deste estudo, mostraram resultados inferiores usando somente dados espectrais para o

mapeamento, com índices de acerto de 82,00%. Isso mostra que, mesmo sem o uso das variáveis RapidEye, empregando somente variáveis temporais, os resultados obtidos neste estudo foram superiores. Entretanto, quando inseridas as variáveis RapidEye, a melhoria dos índices de acerto foi significativa. Isso demonstra a importância da combinação entre os dados espectrais e temporais para um bom desempenho no processo de classificação.

Santos (2014) mostrou a eficácia de uma série temporal Landsat TM para mapear dados provenientes de imagens RapidEye, como também visto neste estudo. Diante disso, os resultados deste estudo evidenciaram que os dados derivados de uma série temporal Landsat TM, composta por imagens originais e sintéticas, foram eficazes para caracterizar objetos derivados de uma imagem RapidEye, com alta resolução espacial, obtendo excelentes índices de acurácia e conseguindo separar, de forma satisfatória, as áreas cafeeiras dos demais tipos de cobertura da terra.

#### 4 CONCLUSÕES

No intuito de propor uma nova metodologia para mapeamento de áreas cafeeiras, este estudo mostrou que:

1) dados temporais, juntamente com uma abordagem orientada a objeto, foram eficientes na classificação de áreas cafeeiras e na separabilidade entre as classes de cobertura da terra, diminuindo a confusão existente entre os alvos. As variáveis estatísticas (Getstatistic) se mostraram melhores para o mapeamento, aumentando o índice de acerto das classificações;

2) o aumento do número de variáveis diminuiu o percentual de acerto das classificações, mostrando que os melhores resultados foram obtidos pelo Getstatistic, usando também variáveis espectrais;

3) o número de amostras de treinamento não influenciou o índice de acerto. Porém, os melhores resultados foram obtidos usando o conjunto C, com 1.000 amostras de treinamento, totalizando 20 vezes o número de variáveis utilizadas;

4) dados derivados de uma série temporal Landsat TM foram eficazes, caracterizando e mapeando corretamente os objetos derivados de uma imagem RapidEye, com alta resolução espacial.

Mais estudos devem ser realizados, utilizando esta metodologia, em áreas distintas. Outras variáveis também podem ser incorporadas ao processo de classificação, como, por exemplo, dados físicos (dados de relevo, altitude e entre outros).

## REFERÊNCIAS

- ADAMI, M.; MOREIRA, M. A.; BARROS, M. A.; Avaliação da exatidão do mapeamento da cultura do café no estado de Minas Gerais. In: XIV Simpósio Brasileiro de Sensoriamento Remoto, 2009, Natal. **Anais**. INPE, 2009, p. 1–8.
- 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.
- ANDRADE, L. N.; VIEIRA, T. G. C.; LACERDA, W. S.; VOLPATO, M. M. L.; DAVIS-JUNIOR, C. A. Aplicação de redes neurais artificiais na classificação de áreas cafeeiras em Machado - MG. **Coffee Science**, v. 8, n. 1, p. 78–90, 2013.
- ARANTES, T. B. **Análise de séries temporais compostas por imagens sintetizadas a partir da fusão de dados MODIS - TM**. 2014. 90 p. Dissertação (Mestrado em Engenharia Florestal) - Universidade Federal de Lavras, Lavras - MG.
- BAI, J.; PERRON, P. Computation and analysis of multiple structural change models. **Journal of Applied Econometrics**, v. 18, n. 1, p. 1–22, 2003.
- BHANDARI, S.; PHINN, S.; GILL, T. Preparing Landsat Image Time Series (LITS) for Monitoring Changes in Vegetation Phenology in Queensland, Australia. **Remote Sensing**, v. 4, n. 12, p. 1856–1886, 2012.
- BROWN, J. C.; KASTENS, J. H.; COUTINHO, A. C.; VICTORIA, D. D.; BISHOP, C. R. Classifying multiyear agricultural land use data from Mato Grosso using time-series MODIS vegetation index data. **Remote Sensing Environment**, 130, 39–50.2013.
- BRUZZONE, L.; SMITS, P. C.; TILTON, J. C. Special Issue on Analysis of Multitemporal Remote Sensing Images. **IEEE Transactions on Geoscience and Remote Sensing**, v. 41, n. 11, p. 2419–2422, 2003.
- CARVALHO, L. M. T. DE. Detecção de modificações na cobertura do solo. In: CARVALHO, L. M. T.; SCOLFORO, J. R. S. **Inventário Florestal de Minas Gerais: Monitoramento da Flora Nativa 2005 – 2007**: Monitoramento da Flora Nativa. Lavras: UFLA, 2008. p. 2–24.

CONAB. **Acompanhamento de Safra Brasileira**, 2014.

CORDERO-SANCHO, S.; SADER, S. A. Spectral analysis and classification accuracy of coffee crops using Landsat and a topographic-environmental model. **International Journal of Remote Sensing**, v. 28, n. 7, p. 1577–1593, 2007.

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**, v. 118, p. 259–272, 2012.

FOODY, G. M. Thematic Map Comparison : Evaluating the Statistical Significance of Differences in Classification Accuracy. **Photogrammetric Engineering & Remote Sensing**, v. 70, n. 5, p. 627–633, 2004.

FORKEL, M. CARVALHAIS, N.; VERBESSELT, J.; MAHECHA, M. D.; NEIGH, C. S. R.; REICHSTEIN, M. Trend Change Detection in NDVI Time Series: Effects of Inter-Annual Variability and Methodology. **Remote Sensing**, v. 5, n. 5, p. 2113–2144, 2013.

GALFORD, G. L.; MUSTARD, J. F.; MELILLO, J.; GENDRIN, A.; CERRI, C. C.; CERRI, C. E. P. Wavelet analysis of MODIS time series to detect expansion and intensification of row-crop agriculture in Brazil. **Remote Sensing of Environment**, v. 112, n. 2, p. 576–587, 2008.

GAO, F.; MASEK, J.; SCHWALLER, M.; HALL, F. On the blending of the Landsat and MODIS surface reflectance: predicting daily Landsat surface reflectance. **IEEE Transactions on Geoscience and Remote Sensing**, v. 44, n. 8, p. 2207–2218, 2006.

HILKER, T.; WULDER, M. A.; COOPS, N. C.; SEITZ, N.; WHITE, J. C.; GAO, F.; MASEK, J. G.; STENHOUSE, G. 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**, v. 113, n. 9, p. 1988–1999, 2009.

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**, v. 12, p. S3–S10, 2010.

JENSEN, J. R. **Remote sensing of the environment: an Earth Resource perspective**. 2 ed. Upper Saddle River, NJ, USA: Pearson Prentice Hall, 2006. 608 p. (Prentice Hall series in geographic information science).

JIA, K.; LIANG, S.; ZHANG, N.; WEI, X.; GU, X.; ZHAO, X.; YAO, Y.; XIE, X. 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**, v. 93, p. 49–55, 2014.

JIA, K.; WU, B.; LI, Q. Crop classification using HJ satellite multispectral data in the North China Plain. **Journal of Applied Remote Sensing**, v. 7, n. 1, p. 1–12, 2013.

KOPRINSKA, I. Feature Selection for Brain-Computer Interfaces. PAKDD Workshops. **Anais...**, Berlin: LNAI 5669, p. 100-111, 2010.

LANDIS, J. R.; KOCH, G. G. The Measurement of Observer Agreement for Categorical Data Data for Categorical of Observer Agreement The Measurement. **Biometrics**, v. 33, n. 1, p. 159–174, 1977.

LI, C.; WANG, J.; WANG, L.; HU, L.; GONG, P. Comparison of classification algorithms and training sample sizes in urban land classification with Landsat Thematic Mapper imagery. **Remote Sensing**, v. 6, n. 2, p. 964–983, 2014.

MACHADO, M. L.; ALVES, H. M. R.; VIEIRA, T. G. C.; FERNANDES-FILHO, E. I.; LAERDA, M, P. C. Mapeamento de áreas cafeeiras (Coffea arabica L.) da Zona da Mata Mineira usando sensoriamento remoto. **Coffee Science**, v. 5, n. 2, p. 113–122, 2010.

MANN, H. B. Nonparametric tests against trend. **Econometrica**, v. 13, n. 3, p. 245–259, 1945.

MARTÍNEZ-VERDUZCO, G. C.; GALEANA-PIZAÑA, J. M.; CRUZ-BELLO, G. M. Coupling community Mapping and supervised classification to discriminate Shade coffee from Natural vegetation. **Applied Geography**, v. 34, p. 1–9, 2012.

MARUJO, R. F. B.; VOLPATO, M. M. L.; VIEIRA, T. G. C.; ALVES, H. M. R.; RIBEIRO, M. B. P. Classificação orientada a objetos aplicada a cultivos cafeeiros em Três Pontas - MG.. In: XVI Simpósio Brasileiro de Sensoriamento Remoto, 2013, Foz do Iguaçu. **Anais**. INPE, 2013, p. 1338–1345.

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

GAO, F.; MASEK, J. WOLFE, R. E. Automated registration and orthorectification package for Landsat and Landsat-like data processing. **Journal of Applied Remote Sensing**, v. 3, n. 1, p. 033515, 2009.

MELLO, M. P.; VIEIRA, C. A. O.; RUDORFF, B. F. T.; APLIN, P. STARS : A New Method for Multitemporal Remote Sensing. **IEEE Transactions on Geoscience and Remote Sensing**, p. 1–17, 2012.

METZGER, J. P. Editorial conservation issues in the Brazilian Atlantic Forest. **Biological Conservation**, v. 142, n. 6, p. 1138-1140, 2009.

MOREIRA, M. A.; BARROS, M. A.; FARIA, V. G. C.; ADAMI, M. Tecnologia de informação: imagens de satélite para o mapeamento de áreas de café de Minas Gerais 1. **Informe Agropecuário**, v. 28, n. 241, p. 27–37, 2007.

MOREIRA, M. A.; ADAMI, M.; THEODOR, F. Análise espectral e temporal da cultura do café em imagens Landsat Spectral and temporal behavior analysis of coffee crop in Landsat images. **Pesquisa Agropecuária Brasileira**, v. 39, n. 3, p. 223–231, 2004.

NIEL, T. VAN; MCVICAR, T.; DATT, B. On the relationship between training sample size and data dimensionality: Monte Carlo analysis of broadband multi-temporal classification. **Remote Sensing of Environment**, v. 98, n. 4, p. 468–480, 2005.

PAPE, A. D.; FRANKLIN, S. E. MODIS-based Change Detection for Grizzly Bear Habitat Mapping in Alberta. **Photogrammetric Engineering & Remote Sensing**, v. 74, n. 8, p. 973–985, 2008.

R Core Team. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. 2014. URL <http://www.R-project.org/>.

ROUSE, J. W. MONITORING THE VERNAL ADVACEMENT AND RETROGRADATION OF NATURAL VEGETATION. **Remote Sensing Center**, p. 1–8, 1974.

SÁ JÚNIOR, A. **Aplicação da classificação de Koppen para o zoneamento climático do estado de Minas Gerais**. 2009. 101 p. Dissertação (Mestrado em Engenharia Agrícola, Engenharia de Água e Solo) – Universidade Federal de Lavras, Lavras, MG.

SANTOS, J. A.; GOSSELIN, P. H.; PHILIPP-FOLIGUET, S.; TORRES, R. S.; FALCÃO, A. X. Multiscale Classification of Remote Sensing Images. **IEE Transactions on Geoscience and Remote Sensing**, v. 50, n. 10, p. 3764–3775, 2012.

SANTOS, P. A. **Avaliação do uso de variáveis temporais na classificação da cobertura da terra**. 2014. 45 p. Dissertação (Mestrado em Ecologia Aplicada) - Universidade Federal de Lavras, Lavras - MG.

SARMIENTO, C. M.; RAMIREZ, G. M.; COLTRI, P. P.; SILVA, L. F. L.; NASSUR, A. C.; SOARES, J. F. Comparação de classificadores supervisionados na discriminação de áreas cafeeiras em Campos Gerais – Minas Gerais. **Coffee Science**, v. 9, n. 4, p. 546-557, 2014.

SONNENSCHNEIN, R.; KUEMMERLE, T.; UDELHOVEN, T.; STELLMES, M.; HOSTERT, P. Differences in Landsat-based trend analyses in drylands due to the choice of vegetation estimate. **Remote Sensing of Environment**, v. 115, n. 6, p. 1408–1420, 2011.

SOUZA, V. C. O.; VIEIRA, T. G. C.; ALVES, H. M. R.; VOLPATO, M. M. L. Análise e classificação de áreas de mata e café na região de Machado - MG. XIV Simpósio Brasileiro de Sensoriamento Remoto. **Anais...**, Natal: INPE, p. 7925-7932, 2009.

TRABAQUINI, K.; MIGLIORANZA, E.; FRANÇA, V. PEREIRA NETO, O. C. Caracterização de lavouras cafeeiras, utilizando técnicas de geoprocessamento e sensoriamento remoto, no município de Umuarama – PR. **Ciência e Agrotecnologia**, v. 35, n. 1, p. 35–44, 2011.

VELOSO, H. P.; RANGEL FILHO, A. L. R.; LIMA, J. C. A. **Classificação da vegetação brasileira, adaptada a um sistema universal**. Rio de Janeiro: IBGE, 1991. 123 p.

VERBESSELT, J.; HYNDMAN, R.; NEWNHAM, G.; CULVENOR, D. Detecting trend and seasonal changes in satellite image time series. **Remote Sensing of Environment**, v. 114, n. 1, p. 106–115, 2010.

VIEIRA, M. A.; FORMAGGIO, A. R.; RENNÓ, C. D.; ATZBERGER, C.; AGUIAR, D. A.; MELLO, M. P. 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**, v. 123, p. 553–562, 2012.

VIEIRA, T. G. C.; ALVES, H. M. R.; LACERDA, M. P. C.; VEIGA, R. D.; EPIPHANIO, J. C. N. Crop parameters and spectral response of coffee (*Coffea arabica* L.) areas within the state of Minas Gerais, Brazil. **Coffee Science**, v. 1, n. 2, p. 111–118, 2006.

VIEIRA, T. G. C.; ALVES, H. M. R.; BERTOLDO, M. A.; SOUZA, V. C. O. Geotechnologies in the assessment of land use changes. **Coffee Science**, v. 2, n. 2, p. 142–149, 2007.

WALKER, J. J.; BEURS, K. M.; WYNNE, R. H.; GAO, F. Evaluation of Landsat and MODIS data fusion products for analysis of dryland forest phenology. **Remote Sensing of Environment**, v. 117, p. 381–393, 2012.

WATTS, J. D.; POWELL, S. L.; LAWRENCE, R. L.; HILKER, T. Improved classification of conservation tillage adoption using high temporal and synthetic satellite imagery. **Remote Sensing of Environment**, v. 115, n. 1, p. 66–75, 2011.

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

XIAO, X.; BOLES, S.; LIU, J.; ZHUANG, D.; LIU, M. Characterization of forest types in Northeastern China, using multi-temporal SPOT-4 VEGETATION sensor data. **Remote Sensing of Environment**, v. 82, n. 2-3, p. 335–348, 2002.

ZEILEIS, A.; KLEIBER, C.; KRÄMER, W.; HORNIK, K. Testing and dating of structural changes in practice. **Computational Statistics & Data Analysis**, v. 44, n. 1-2, p. 109–123, 2003.

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

## CONSIDERAÇÕES FINAIS

### 1 CONSIDERAÇÕES GERAIS

De acordo com os objetivos gerais deste trabalho, uma nova estratégia de mapeamento da cafeicultura, baseada em dados espectrais e dados temporais, oriundos de dados de sensoriamento remoto, foi desenvolvida, para a região sul do estado de Minas Gerais. Um procedimento viável e de fácil implementação foi proposto, inclusive para ser aplicado no mapeamento de outras áreas. Este método apresenta um grande potencial no mapeamento de áreas cafeeiras e na separação entre classes de uso e cobertura da terra.

As áreas cafeeiras são extremamente difíceis de serem identificadas corretamente. No entanto, neste estudo, mostrou-se que as novas abordagens utilizando variáveis temporais e espectrais, juntamente com uma classificação baseada em objeto, para o mapeamento desta cultura, se mostraram promissoras, com índices de acurácia muito superiores aos encontrados na literatura.

#### 1.1 Principais conclusões

As principais conclusões deste estudo estão listadas a seguir, considerando os principais objetivos desta tese.

No capítulo 3, o principal objetivo foi comparar algoritmos de AM usando diferentes conjuntos de variáveis derivadas de dados de sensoriamento remoto e identificar as melhores combinações algoritmos-variáveis para o mapeamento de áreas cafeeiras. O estudo mostrou que as variáveis espectrais e de textura, quando usadas conjuntamente, mostraram bom desempenho na classificação destas áreas, utilizando os algoritmos de AM, tendo o *Support Vector Machine* e o *Random Forest* mostrado os melhores desempenhos nas classificações, das três áreas estudadas. Quando usadas separadamente, somente as variáveis espectrais mostraram bons desempenhos no processo de

classificação. Porém, as classificações ainda mostraram muita confusão entre alvos.

No Capítulo 4, o objetivo foi desenvolver um padrão metodológico para o mapeamento da cafeicultura por meio de dados multitemporais, juntamente com dados espectrais, aliados a uma abordagem baseada em objeto. O estudo evidenciou que as variáveis derivadas do pacote *greenbrown*, utilizando uma série temporal densa, foram eficientes para o mapeamento de áreas cafeeiras, conseguindo identificar corretamente estas áreas. Além disso, foi possível observar que os níveis de confusão entre áreas de vegetação nativa e café, bem como pastagem e café, diminuíram, o que mostra o quanto esta metodologia favorece o mapeamento dessas áreas com maior precisão.

## **1.2 Recomendações e perspectivas**

Esta metodologia foi desenvolvida no intuito de solucionar problemas referentes à confusão entre alvos, principalmente relacionados a áreas cafeeiras e vegetação nativa, ocasionada pela semelhança espectral encontrada nas imagens de sensoriamento remoto. No entanto, os resultados obtidos neste estudo mostraram o quanto esta técnica foi eficiente no mapeamento dessas áreas.

Os métodos descritos neste trabalho podem ser aplicados em diversas áreas para o mapeamento não só do café, mas de diferentes tipos de uso e cobertura da terra. O algoritmo para extrair as variáveis temporais, que foi usado neste estudo, é robusto e pode ser aplicado a qualquer tipo de sensor. Além disso, este algoritmo disponibiliza dados matriciais de diversos componentes de sazonalidade e tendência, que podem ser usados como variáveis no processo de classificação.

Recomenda-se o uso desta metodologia para o mapeamento de outras áreas cafeeiras e, assim, obter informações mais rápidas e precisas sobre o seu sistema produtivo para melhorar a precisão de estimativas de área plantada, área

colhida, produção, distribuição espacial e o ambiente em que é cultivada, principalmente devido à sua grande importância econômica e social. Dessa forma, tem-se a oportunidade melhorar as estratégias de planejamento ambiental e as práticas de manejo, auxiliando o planejamento dos governos municipais, estaduais e nacionais.

### **1.3 Sugestão de trabalhos futuros**

No intuito de diminuir a quantidade de processamentos e otimizar o trabalho, além de tentar sanar as lacunas, recomendam-se, como trabalhos futuros:

- testar o uso de séries temporais derivadas do pacote *greenbrown* em outras áreas de estudo. Além disso, aliar dados físicos, como exposição de vertentes e declividade das áreas a serem mapeadas, juntamente com estas variáveis temporais;

- testar a efetividade de séries temporais com frequências temporais diferentes para o mapeamento desta cultura, no intuito de diminuir o tempo de processamento da série temporal utilizando o pacote *greenbrown*.

Além disso, é importante realizar uma caracterização mais detalhada dos perfis temporais dos diversos tipos de cobertura da terra, caracterizando e comparando assinaturas temporais de café, vegetação nativa e pastagem, no intuito de mostrar a diferença no comportamento temporal destas feições.