



MINISTÉRIO DA CIÊNCIA, TECNOLOGIA, INOVAÇÕES E COMUNICAÇÕES
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

sid.inpe.br/mtc-m21b/2017/02.02.15.58-TDI

**MINERAÇÃO DE DADOS DE SENSORIAMENTO
REMOTO PARA DETECÇÃO E CLASSIFICAÇÃO DE
ÁREAS DE PASTAGEM NA AMAZÔNIA LEGAL**

Alana Kasahara Neves

Dissertação de Mestrado do
Curso de Pós-Graduação em
Sensoriamento Remoto, orientada
pelos Drs. Thales Sehn Körting,
e Leila Maria Garcia Fonseca,
aprovada em 15 de fevereiro de
2017.

URL do documento original:

<<http://urlib.net/8JMKD3MGP3W34P/3NAADAS>>

INPE
São José dos Campos
2017

PUBLICADO POR:

Instituto Nacional de Pesquisas Espaciais - INPE

Gabinete do Diretor (GB)

Serviço de Informação e Documentação (SID)

Caixa Postal 515 - CEP 12.245-970

São José dos Campos - SP - Brasil

Tel.:(012) 3208-6923/6921

Fax: (012) 3208-6919

E-mail: pubtc@inpe.br

COMISSÃO DO CONSELHO DE EDITORAÇÃO E PRESERVAÇÃO DA PRODUÇÃO INTELECTUAL DO INPE (DE/DIR-544):

Presidente:

Maria do Carmo de Andrade Nono - Conselho de Pós-Graduação (CPG)

Membros:

Dr. Plínio Carlos Alvalá - Centro de Ciência do Sistema Terrestre (CST)

Dr. André de Castro Milone - Coordenação de Ciências Espaciais e Atmosféricas (CEA)

Dra. Carina de Barros Melo - Coordenação de Laboratórios Associados (CTE)

Dr. Evandro Marconi Rocco - Coordenação de Engenharia e Tecnologia Espacial (ETE)

Dr. Hermann Johann Heinrich Kux - Coordenação de Observação da Terra (OBT)

Dr. Marley Cavalcante de Lima Moscati - Centro de Previsão de Tempo e Estudos Climáticos (CPT)

Silvia Castro Marcelino - Serviço de Informação e Documentação (SID)

BIBLIOTECA DIGITAL:

Dr. Gerald Jean Francis Banon

Clayton Martins Pereira - Serviço de Informação e Documentação (SID)

REVISÃO E NORMALIZAÇÃO DOCUMENTÁRIA:

Simone Angélica Del Ducca Barbedo - Serviço de Informação e Documentação (SID)

Yolanda Ribeiro da Silva Souza - Serviço de Informação e Documentação (SID)

EDITORAÇÃO ELETRÔNICA:

Marcelo de Castro Pazos - Serviço de Informação e Documentação (SID)

André Luis Dias Fernandes - Serviço de Informação e Documentação (SID)



MINISTÉRIO DA CIÊNCIA, TECNOLOGIA, INOVAÇÕES E COMUNICAÇÕES
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

sid.inpe.br/mtc-m21b/2017/02.02.15.58-TDI

**MINERAÇÃO DE DADOS DE SENSORIAMENTO
REMOTO PARA DETECÇÃO E CLASSIFICAÇÃO DE
ÁREAS DE PASTAGEM NA AMAZÔNIA LEGAL**

Alana Kasahara Neves

Dissertação de Mestrado do
Curso de Pós-Graduação em
Sensoriamento Remoto, orientada
pelos Drs. Thales Sehn Körting,
e Leila Maria Garcia Fonseca,
aprovada em 15 de fevereiro de
2017.

URL do documento original:

<<http://urlib.net/8JMKD3MGP3W34P/3NAADAS>>

INPE
São José dos Campos
2017

Dados Internacionais de Catalogação na Publicação (CIP)

Neves, Alana Kasahara.

N414m Mineração de dados de sensoriamento remoto para detecção e classificação de áreas de pastagem na Amazônia Legal / Alana Kasahara Neves. – São José dos Campos : INPE, 2017.
xxiv + 77 p. ; (sid.inpe.br/mtc-m21b/2017/02.02.15.58-TDI)

Dissertação (Mestrado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2017.

Orientadores : Drs. Thales Sehn Körting, e Leila Maria Garcia Fonseca.

1. Pasto limpo. 2. Pasto sujo. 3. Árvore de decisão. 4. Séries temporais. 5. TerraClass. I.Título.

CDU 528.8:633.2.033(811.3)



Esta obra foi licenciada sob uma Licença [Creative Commons Atribuição-NãoComercial 3.0 Não Adaptada](https://creativecommons.org/licenses/by-nc/3.0/).

This work is licensed under a [Creative Commons Attribution-NonCommercial 3.0 Unported License](https://creativecommons.org/licenses/by-nc/3.0/).

Aluno (a): *Alana Kasahara Neves*

Título: "MINERAÇÃO DE DADOS DE SENSORIAMENTO REMOTO PARA DETECÇÃO E CLASSIFICAÇÃO DE ÁREAS DE PASTAGEM NA AMAZÔNIA LEGAL"

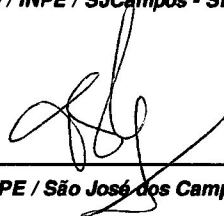
Aprovado (a) pela Banca Examinadora
em cumprimento ao requisito exigido para
obtenção do Título de **Mestre** em
Sensoriamento Remoto

Dra. Maria Isabel Sobral Escada



Presidente / INPE / SJC Campos - SP

Dr. Thales Sehn Körting



Orientador(a) / INPE / São José dos Campos - SP

Dra. Leila Maria Garcia Fonseca



Orientador(a) / INPE / SJC Campos - SP

Dr. Marcos Adami



Membro da Banca / INPE / São José dos Campos - SP

Dr. Júlio César Dalla Mora Esquerdo



Convidado(a) / EMBRAPA / Campinas - SP

Este trabalho foi aprovado por:

maioria simples

unanimidade

São José dos Campos, 15 de fevereiro de 2017

*“We keep moving forward,
opening new doors
and doing new things,
because we’re curious,
and curiosity keeps
leading us down new paths.”*

- Walt Disney

A meus pais, Silvia e David.

AGRADECIMENTOS

Agradeço a Deus, por ter me dado coragem e saúde para chegar até aqui.

Ao meu orientador, Dr. Thales Körting, por todos os ensinamentos e conversas e por sempre se mostrar solícito em ajudar, com muita bondade e paciência.

A minha co-orientadora, Dra. Leila Fonseca, pelas contribuições ao longo do trabalho e por sempre tentar oferecer o melhor aos seus alunos.

Aos doutores Marcos Adami, Isabel Escada e Julio Esquerdo, por fazerem parte da banca e pelo interesse em contribuir com o trabalho. É uma honra tê-los como avaliadores.

Ao Dr. Claudio Almeida, por ter sido o primeiro “inpeano” a abrir as portas da instituição pra mim, ainda em Belém.

Ao INPE e a todos os professores e colaboradores que fizeram parte dessa jornada nada fácil, mas cheia de aprendizados e crescimento.

Aos colegas de trabalho de campo: Taise Farias, Carlos da Costa (“Machu”) e Jefferson Souza. Vocês tornaram essa experiência muito mais leve e divertida.

À CAPES, pelo apoio financeiro.

A minha família: meus pais (Silvia e David), minha irmã (Celine) e minha avó (Agripina), meus grandes exemplos de força e determinação e meus maiores incentivadores.

Ao Cesare, meu namorado, que esteve comigo em todos os momentos, fazendo o que estivesse ao seu alcance pra me dar um pouco de paz e alegria.

Aos amigos de longa data e àqueles que conheci no INPE, obrigada por todas as conversas, apoio e momentos de descontração.

RESUMO

Aproximadamente 60% das áreas desflorestadas na Amazônia Legal são ocupadas por pastagens. A expansão das áreas de pastagem sobre áreas de floresta pode ser associada a fatores como o mercado de terras e a perda de produtividade da pastagem ao longo do tempo. De toda forma, essa expansão permanece como um obstáculo ao combate ao desflorestamento. A detecção e avaliação das condições das pastagens permitem melhores monitoramento e controle, assim como a identificação de áreas propícias para a recuperação. Dentro deste contexto, o objetivo deste trabalho é desenvolver uma metodologia de reconhecimento de áreas de pastagem na Amazônia, com base na detecção e classificação por meio de atributos de séries temporais e técnicas de mineração de dados, de acordo com as condições da cobertura vegetal. A área de estudo consiste em três órbitas/ponto do satélite Landsat 8 distribuídas em quatro estados (AC, MT, RO e AM): 001/067, 226/068 e 231/067. Foram utilizadas imagens de reflectância de superfície do sensor OLI do Landsat 8, referentes ao período entre abril de 2013 e dezembro de 2015. As nuvens e sombras de nuvens foram detectadas pelo algoritmo FMask e excluídas. A classificação foi feita em duas etapas: detecção de pastagem (diferenciá-las de outros alvos da cena) e, posteriormente, a classificação das pastagens em Pasto Limpo e Pasto Sujo. Para a caracterização destas classes, os seguintes atributos foram usados: índices de vegetação (NDVI, EVI, EVI2, SAVI e NDII), frações do Modelo Linear de Mistura Espectral (frações vegetação, NPV e solo), componentes da transformação *Tasselled Cap* (*greenness*, *brightness* e *wetness*), outros atributos espectrais e atributos texturais. As duas etapas da classificação foram realizadas utilizando três classificadores (árvore de decisão, *random forest* e rede neural) e duas abordagens: por *pixel* e baseada em objetos. A avaliação dos resultados de classificação baseou-se em trabalho de campo e interpretação visual de imagens do satélite RapidEye. Os resultados mostraram melhoras nas taxas de acerto quando houve a utilização de segmentos, uma vez que as pastagens possuem uma grande quantidade de mistura de elementos em sua composição. Os modelos criados e avaliados na mesma cena obtiveram altas taxas de acerto (próximas a 90%), entretanto não foram capazes de classificar outras cenas com a mesma eficiência. Quando amostras de duas cenas diferentes foram combinadas para a geração do modelo, as taxas de acerto ficaram parecidas entre as imagens, por volta de 80%. A maior dificuldade esteve na separação entre Pasto Limpo e Pasto Sujo, uma vez que as pastagens na Amazônia variam de acordo com muitos fatores: manejo adotado, tipo de solo, regime de chuvas, tipo de gramínea utilizada e outros.

Palavras-chave: Pasto Limpo. Pasto Sujo. Árvore de decisão. Séries temporais. TerraClass.

REMOTE SENSING DATA MINING TO DETECT AND CLASSIFY PASTURE LANDS IN THE LEGAL AMAZON

ABSTRACT

The highest percentage of deforested areas in the Legal Amazon is occupied by pastures. The expansion of pasture areas over the forest may be associated with factors such as land speculation and loss of productivity over time. In any case, this expansion remains an obstacle to fight against deforestation. The detection and evaluation of pasture conditions allow better monitoring and control it, as well as the identification of suitable areas for recovery. In this context, the objective of this work is to develop a methodology for the recognition of pasture areas in the Amazon, based on the detection and classification by time series attributes and data mining techniques, according to the vegetation conditions. The study area consists of three path/rows of Landsat 8 satellite distributed in four Brazilian states (AC, MT, RO and AM): 001/067, 226/068 and 231/067. Surface reflectance images from OLI sensor (Landsat 8 satellite) were used for the period between April 2013 and December 2015. Clouds and cloud shadows were detected by the FMask algorithm and excluded from the dataset. The classification was carried out in two steps: pasture detection (differentiate them from other targets in the scene) and, later, the classification of pastures between Herbaceous Pasture and Shrubby Pasture. For the characterization of these classes, the following attributes were used: vegetation indices (NDVI, EVI, EVI2, SAVI and NDII), fractions of the Linear Spectral Mixture Model (vegetation, NPV and soil), bands of the Tasseled Cap Transformation (Greenness, Brightness and Wetness), other spectral attributes and textured attributes. The two steps of classification were performed using three classifiers (decision tree, random forest and neural network) and two different approaches: per pixel and object-based. The evaluation of the classification results was based on fieldwork and visual interpretation of RapidEye satellite images. The results showed improvements in the accuracy when segments were used instead of pixels, since pastures have a large amount of mixture of elements in their composition. The models created and evaluated in the same scene obtained high accuracy (close to 90%), but they were not able to classify other scenes with the same efficiency. When samples from two different scenes were combined for model generation, the accuracy was similar between the images, around 80%. The greatest difficulty was in the separation between Herbaceous Pasture and Shrubby Pasture, since pastures in Amazon may vary according to some factors, such as: adopted management, soil type, rainfall regime and type of grass used.

Keywords: Herbaceous Pasture. Shrubby Pasture. Decision Tree. Time Series. TerraClass.

LISTA DE FIGURAS

	<u>Pág.</u>
Figura 2.1 – Dinâmica do rebanho bovino (milhões de cabeças) nas grandes regiões do Brasil em 1975 e 2006.....	7
Figura 2.2 – Dinâmica das áreas de pastagem (milhões de hectares) nas grandes regiões do Brasil em 1975 e 2006.....	7
Figura 2.3 – Padrões das classes de pastagem. A) Pasto limpo. B) Pasto sujo. C) Regeneração com pasto. D) Pasto com solo exposto.....	9
Figura 2.4 – Representação de um plano de estrutura de dados. a) Em um sensor de duas bandas. b) Em um sensor de três bandas, o plano de estrutura de dados seria diferente dependendo da perspectiva de visualização.	18
Figura 2.5 – Dados de um sensor hipotético de três bandas. A) O cubo é o espaço total disponível para os dados. A linha tracejada representa uma estrutura retangular (plana) dentro do cubo. B) A perspectiva mudou para a estrutura retangular dos dados ser vista diretamente. C) Direção da variação espectral associada com dois processos físicos distintos em classes presentes na cena. D) Ajustamento adicional na perspectiva de observação para alinhar as direções relacionadas a processos físicos com os novos eixos de coordenadas.....	19
Figura 2.6 – Construção da GLCM. a) Imagem em níveis de cinza. b) Valores dos níveis de cinza. c) Matriz de estrutura. d) GLCM.....	21
Figura 2.7 – Direções de deslocamento na GLCM..	22
Figura 2.8 – Composição de uma série temporal..	24
Figura 2.9 – Processo da descoberta de conhecimento em bases de dados.....	26
Figura 3.1 – Fluxograma descrevendo a metodologia do trabalho.....	31
Figura 3.2 - Localização da área de estudo.	32
Figura 3.3 – Diferentes padrões de pastagens mapeados pelo TerraClass em imagens do sensor OLI do Landsat 8. À esquerda, um recorte da cena 226/068, no Mato Grosso. À direita, um recorte da cena 231/067, em Rondônia.	32

Figura 3.4 – Disponibilidade de imagens do sensor OLI para as cenas e período estudado.....	35
Figura 3.5 – Exemplo de utilização de segmentos constantes na série temporal, feito a partir de recorte da cena 001/067... ..	36
Figura 3.6 – Composição colorida R(6)G(5)B(4) e frações do MLME de recorte da imagem 231/067 referente à 15/10/16.....	38
Figura 3.7 – Espectros dos <i>endmembers</i> das quatro frações do Modelo Linear de Mistura Espectral.....	38
Figura 3.8 – Classificação hierárquica de pastagens... ..	40
Figura 3.9 – Trajeto do trabalho de campo e pontos de coleta.....	44
Figura 3.10 – Identificação dos segmentos dos pontos de campo para extração de atributos.....	45
Figura 4.1 – Espacializações da a) Referência, b) Caso 1 e c) Caso 2.....	50
Figura 4.2 – Árvore de decisão para detecção de Pastagem do modelo criado com amostras das duas cenas, 001/067 e 226/068.... ..	55
Figura 4.3 – Árvore de decisão para a classificação de Pasto Limpo e Pasto Sujo do modelo criado com amostras das duas cenas, 001/067 e 226/068..... ..	56
Figura 4.4 – Matriz de correlação (Spearman) relacionando todos os atributos utilizados..... ..	56
Figura 4.5 – Espacialização das classificações da cena 001/067. a) Referência. b) Modelo criado/avaliado na 001/067. c) Modelo criado na combinação (001/067 e 226/068) e avaliado na 001/067. d) Modelo criado na 226/068 e avaliado na 001/067..... ..	57
Figura 4.6 – Espacialização das classificações da cena 226/068. a) Referência. b) Modelo criado/avaliado na 226/068. c) Modelo criado na combinação (001/067 e 226/068) e avaliado na 226/068. d) Modelo criado na 001/067 e avaliado na 226/068..... ..	58
Figura 4.7 – Espacialização da classificação da cena 231/067. a) Referência. b) Modelo criado na combinação (001/067 e 226/068) e avaliado na 231/067..... ..	59

Figura 4.8 – Comparação das referências e dos resultados gerados pelo modelo criado na combinação (001/067 e 226/068). a) 001/067. b) 226/068. c) 231/067.....	60
Figura 4.9 – Alguns pontos de pastagem observados em campo. a) Pasto Limpo na cena 001/067, b) Pasto Sujo na cena 231/067, c) Pasto Sujo na cena 226/068, d) Pasto Limpo na cena 231/067.....	62
Figura 4.10 – Alguns aspectos que contribuem para a heterogeneidade das pastagens. a) e b) Presença de afloramentos rochosos e variação de altitude do terreno, c) Pasto queimado.....	63

LISTA DE TABELAS

	<u>Pág.</u>
Tabela 2.1 – Coeficientes da transformação TC para dados do sensor OLI (Landsat 8).	20
Tabela 2.2 – Exemplo de uma matriz de confusão com valores verdadeiros positivos (VP), falsos negativos (FN), verdadeiros negativos (VN) e falsos positivos (FP).....	29
Tabela 3.1 – Órbitas/ponto e municípios e estados com áreas compreendidas nas respectivas cenas..	33
Tabela 3.2 – Informações sobre o sensor OLI do satélite Landsat 8.....	34
Tabela 3.3 – Informações das imagens do sensor OLI utilizadas na segmentação.....	36
Tabela 3.4 – Informações sobre os índices de vegetação.....	37
Tabela 3.5 – Alguns atributos espectrais e texturais para a abordagem baseada em objetos.....	39
Tabela 3.6 – Tipologia das classes de pastagem.....	42
Tabela 3.7 – Informações das imagens RapidEye utilizadas na validação.....	43
Tabela 4.1 – Taxas de acerto (%) na construção dos modelos.....	47
Tabela 4.2 – Taxas de acerto (%) na validação dos modelos gerados pelo classificador J48.....	48
Tabela 4.3 – Taxas de acerto (%) na geração e aplicação dos modelos nas três cenas... 52	52
Tabela 4.4 – Validação dos modelos de classificação por pontos aleatórios.....	61
Tabela 4.5 – Classificação dos pontos do trabalho de campo.....	62
Tabela A.1 – Matrizes de confusão, em valores absolutos, da validação dos modelos gerados pelo classificador J48 para a cena 001/067.....	75
Tabela A.2 – Matrizes de confusão, em valores absolutos, da validação dos modelos nas três cenas de estudo.....	76

LISTA DE SIGLAS E ABREVIATURAS

CFS	<i>Correlation Feature Selection</i>
EVI	<i>Enhanced Vegetation Index</i>
EVI2	<i>Enhanced Vegetation Index 2</i>
GLCM	<i>Gray Level Co-occurrence Matrix</i>
INPE	Instituto Nacional de Pesquisas Espaciais
IV	Índice de Vegetação
KDD	<i>Knowledge Discovery in Databases</i>
Landsat	<i>Land Remote Sensing Satellite</i>
MLME	Modelo Linear de Mistura Espectral
MLP	<i>Multilayer Perceptron</i>
MODIS	<i>Moderate Resolution Imaging Spectroradiometer</i>
NDII	<i>Normalized Difference Infrared Index</i>
NDVI	<i>Normalized Difference Vegetation Index</i>
NIR	<i>Near Infrared</i>
NPV	<i>Non Photosynthetic Vegetation</i>
OLI	<i>Operational Land Imager</i>
PV	<i>Photosynthetic Vegetation</i>
RF	<i>Random Forest</i>
SAVI	<i>Soil-Adjusted Vegetation Index</i>
SUDAM	Superintendência do Desenvolvimento da Amazônia
SWIR	<i>Short-Wave Infrared</i>
TC	<i>Tasselled Cap</i>
TCT	<i>Tasselled Cap Transformation</i>

SUMÁRIO

	<u>Pág.</u>
1 INTRODUÇÃO	1
1.1. Objetivos	4
1.1.1. Objetivo Geral	4
1.1.2. Objetivos Específicos	4
2 REFERENCIAL TEÓRICO	5
2.1. Histórico	5
2.2. As Pastagens na Amazônia	7
2.3. Classificação de Pastagens	9
2.4. Métodos de Processamento Digital de Imagens	14
2.4.1. Índices de Vegetação	14
2.4.2. Modelo Linear de Mistura Espectral	15
2.4.3. Transformação Tasseled Cap	17
2.4.4. Medidas de Textura	21
2.5. Séries Temporais de Imagens de Sensoriamento Remoto	23
2.6. Mineração de Dados	25
3 MATERIAIS E MÉTODOS	31
3.1. Área de Estudo	31
3.2. Preparação da Base de Dados	34
3.3. Extração de Atributos	36
3.4. Seleção de Atributos	39
3.5. Classificação	40
3.5.1. Detecção de Pastagens	41
3.5.2. Classificação de Pastagens	41
3.6. Dados de Referência e Validação	43
4. RESULTADOS E DISCUSSÃO	47
4.1. Abordagens de Classificação	47

4.2. Classificação das Cenas de Estudo.....	51
4.3. Validação e Trabalho de Campo.....	60
5. CONCLUSÕES.....	65
REFERÊNCIAS BIBLIOGRÁFICAS	67
APÊNDICE A – MATRIZES DE CONFUSÃO.....	75

1 INTRODUÇÃO

A Amazônia Legal (tratada a partir deste ponto apenas como Amazônia) abrange uma área de cerca de 5 milhões de km², englobando nove estados brasileiros (Amazonas, Pará, Acre, Amapá, Roraima, Rondônia, Tocantins, Mato Grosso e parte do Maranhão). Estima-se que, até o ano 2012, uma área de 751.340 km² foi desmatada, o que corresponde a 18,5% do total de floresta disponível inicialmente (INPE, 2013). O desflorestamento na Amazônia varia de acordo com as forças econômicas e os atores sociais a ele relacionados e pode estar associado a diversas causas, das quais podemos destacar: construção de estradas, extração de madeira, agricultura, pecuária e especulação de terra (FEARNSIDE, 2006; BANCO MUNDIAL, 2003).

As consequências relacionadas com o desflorestamento são diversas: erosão do solo, resultando no esgotamento de seus nutrientes; mudanças no regime hidrológico, uma vez que, sem a floresta, a água escoar mais rápido, podendo alterar e intensificar os períodos de secas e cheias nos rios de desaguamento; perda de biodiversidade, pois a floresta Amazônica possui alto grau de endemismo; e aumento das emissões de gases do efeito estufa através de queimadas e retirada da cobertura vegetal que serve como um estoque de gases, principalmente carbono (FEARNSIDE, 2005).

A conversão de floresta para outros usos da terra, como as pastagens, possui importante impacto no ciclo de carbono e na dinâmica climática da região, dentre outros impactos ambientais (ASNER *et al.*, 2004). Aproximadamente 60% das áreas desflorestadas na Amazônia são ocupadas por pastagens (INPE, 2014). Entre os anos de 1975 e 2006, as áreas de pastagem na região norte, onde predomina o bioma amazônico, aumentaram de 5,28 milhões de hectares para 32,63 milhões de hectares, uma variação de 517,9%. Em relação à dinâmica do rebanho bovino criado nessas pastagens, o aumento foi de 2,11 milhões de cabeças para 41,06 milhões de cabeças (IBGE, 2007).

No Brasil e na Amazônia, a maior parte do rebanho é criada a pasto, em sistemas extensivos, nos quais grandes áreas de pastagem são utilizadas e o animal obtém seu próprio alimento. Em países que praticam sistemas intensivos, de

confinamento (EUA e Austrália, por exemplo), a colheita e oferta do alimento aos animais requerem altos custos de mão de obra, máquinas, combustível, etc. (DIAS-FILHO, 2010). A abundância de terra como recurso na Amazônia, a degradação de pastagens, a falta de manejo adequado e o mercado de terras na região são fatores fundamentais na expansão da pastagem em áreas de vegetação natural (MUCHAGATA; BROWN, 2003; HOMMA, 1999). Assim, o manejo e monitoramento das áreas de pastagem são fundamentais no controle do desflorestamento e de suas consequências e, além disso, são capazes de fornecer um panorama da qualidade das pastagens para ajudar a pensar em políticas de incentivo para a melhoria dos sistemas produtivos.

A degradação da pastagem pode ser definida como a perda da produtividade de forrageiras e da capacidade de recuperação natural da pastagem, com qualidade insuficiente para a produção animal e superação dos efeitos causados por pragas, doenças e plantas invasoras (ZIMMER *et al.*, 2012). A degradação está relacionada com o empobrecimento de nutrientes do solo e o aumento da acidez e compactação (ASNER *et al.*, 2004).

Na região amazônica, a degradação pode estar associada ao rápido processo de sucessão secundária (regeneração da floresta), devido à competição com plantas invasoras (ASNER *et al.*, 2004; AGUIAR, 2013). Entretanto, é importante ressaltar que nem sempre a regeneração está associada exclusivamente à degradação, pois pode representar também o abandono da terra, problemas de manejo ou especulação fundiária.

A avaliação dos níveis de degradação, relacionados com as condições da cobertura vegetal, e produtividade das pastagens permitem a identificação de áreas propícias para a recuperação, seja com o objetivo de reincorporar essas áreas ao sistema de produção pecuária ou de convertê-las para outros usos, como o agrícola (VALENTIM, 2009).

As causas da degradação já são bem conhecidas, mas a sua dinâmica espaço-temporal ainda permanece desconhecida. As pastagens na Amazônia sofrem influência da sazonalidade climática, diferentes idades, tipos de solo, coberturas vegetais, estratégias de manejo e produtividades (ASNER *et al.*, 2004). Segundo

Müller *et al.* (2015), as técnicas de sensoriamento remoto e outras ferramentas associadas podem contribuir para o estudo e melhor compreensão da dinâmica das pastagens de maneira a compreender a sua ampla distribuição espacial na Amazônia.

Contribuindo para o conhecimento da dinâmica de pastagens, o projeto TerraClass (Mapeamento do Uso e Cobertura da terra nas Áreas Desflorestadas da Amazônia Legal) mapeia visualmente, a cada dois anos, as áreas de pastagens na Amazônia de acordo com quatro classes de interesse: pasto limpo, pasto sujo, regeneração com pasto e pasto com solo exposto (ALMEIDA *et al.*, 2016). Apesar de ser um mapeamento sistemático e de grande abrangência espacial, a classificação das pastagens é realizada por meio de interpretação visual, o que torna o processo bastante custoso em relação ao tempo e aos recursos humanos.

A coleta sistemática de imagens pelos sensores remotos permite a observação de fenômenos terrestres a partir de séries temporais de dados (HUETE *et al.*, 2012). Por exemplo, no mapeamento de uso e cobertura da terra, atributos como índices de vegetação, extraídos a partir de séries temporais, podem auxiliar nos modelos de classificação (GALFORD *et al.*, 2008).

Uma das técnicas utilizadas para manipular grandes volumes de dados e observações presentes em séries temporais é a mineração de dados. A ideia central das técnicas de mineração de dados consiste em encontrar correlações, padrões e tendências nos dados e, para isto, utiliza combinações de técnicas matemáticas, estatísticas e de reconhecimento de padrões (LAROSE; LAROSE, 2014).

Dessa forma, entende-se que técnicas de mineração de dados, aplicadas a imagens de sensoriamento remoto (de períodos específicos ou de séries temporais), podem ser utilizadas para o reconhecimento de áreas de pastagens na Amazônia, de acordo com as condições de sua cobertura vegetal. O processo de reconhecimento foi dividido em duas etapas: a detecção das áreas de pastagens (diferenciá-las de outros alvos) e posterior classificação das pastagens em Pasto Limpo e Pasto Sujo.

1.1. Objetivos

1.1.1. Objetivo Geral

Desenvolver uma metodologia de reconhecimento de áreas de pastagem, com base na detecção e classificação, por meio de atributos de séries temporais e técnicas de mineração de dados, de acordo com as condições da cobertura vegetal.

1.1.2. Objetivos Específicos

- Identificar os atributos mais relevantes na detecção e classificação das áreas de pastagem.
- Avaliar o desempenho de diferentes algoritmos de classificação nas áreas de pastagem.
- Avaliar diferentes modelos de classificação em diferentes localidades da Amazônia para detectar a diferença espacial nos padrões de pastagens.
- Validar os experimentos de classificação por meio de dados de campo.

2 REFERENCIAL TEÓRICO

2.1. Histórico

A intensificação da ocupação na Amazônia teve início a partir da década de 1960, induzida pelo governo por meio de diversas ações, como a abertura de estradas e o financiamento de grandes projetos agropecuários, que geraram grandes fluxos migratórios ocorridos durante as décadas de 70 e 80 (BECKER, 1990, 2009). A iniciativa do governo federal, que tinha como objetivo integrar a região norte ao restante do país, criou em 1966 a SUDAM (Superintendência do Desenvolvimento da Amazônia) (RIBEIRO, 2005). A SUDAM concedia diversos incentivos fiscais, como dedução de até 50% do imposto de renda para aqueles que investissem em projetos pecuários, agrícolas, industriais ou de serviços na região. Lotes de terra também eram dados ou vendidos a preços muito baixos para aqueles que se dispusessem a ocupar a região (DIAS-FILHO, 2015; VEIGA *et al.*, 2002).

A infraestrutura necessária para o processo de expansão foi construída no mesmo período, com a abertura de importantes estradas que serviam de escoamento da produção e de acesso para regiões remotas da Amazônia, como a Belém-Brasília (BR-010), a Pará-Maranhão (BR-316), a Cuiabá-Santarém (BR-163), a Cuiabá-Rio Branco (BR-364), a Manaus-Boa Vista (BR-174) e a Transamazônica (BR-230) (VEIGA *et al.*, 2002). Além disso, a década de 1970 foi marcada por uma crise no abastecimento de carne bovina na região sul e sudeste do Brasil. Enquanto isso, a região norte possuía atrativos climáticos, como temperaturas com pouca variação ao longo do ano e estações secas menos severas (DIAS-FILHO, 2010). Na região de Paragominas – PA, já eram noticiados casos de sucesso na criação de gado. Por conta desses vários fatores, a expansão da pecuária para a Amazônia foi vista por um longo período como uma atividade promissora (DIAS-FILHO, 2015).

A contínua e rápida expansão das pastagens pode ser atribuída à utilização das áreas por pouco tempo devido ao rápido decaimento da fertilidade do solo ao longo do tempo, devido à falta de manejo adequado. A fertilidade varia de acordo com a idade da pastagem, o tipo de solo e sistema de manejo (NUMATA, 2000).

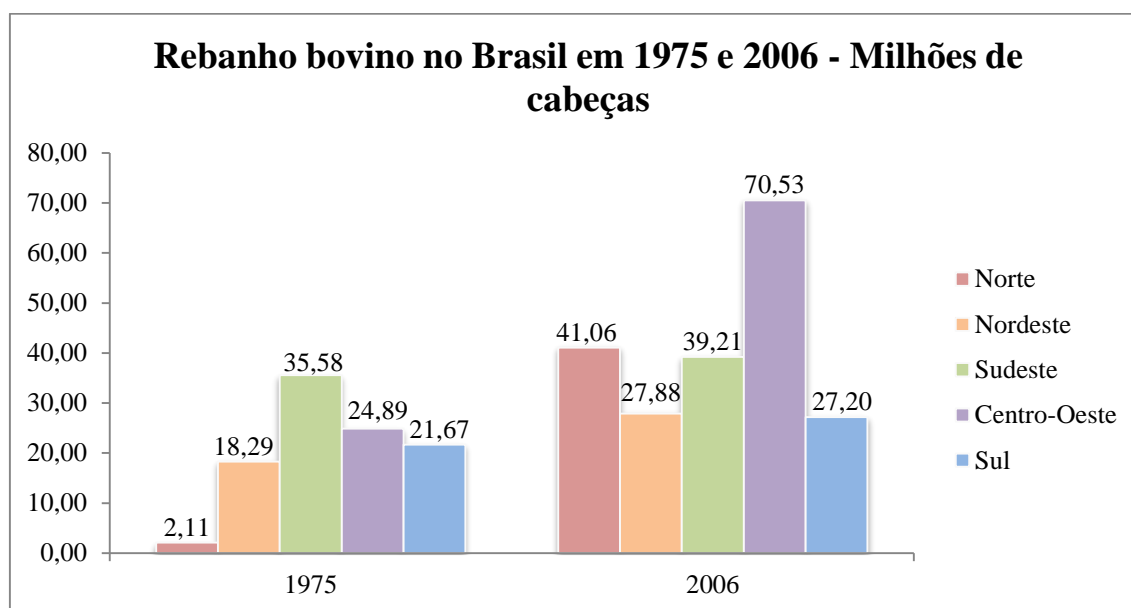
Áreas improdutivas eram constantemente abandonadas e as pastagens passavam a ocupar novas áreas de floresta. Essas práticas contribuíram para o aumento do desflorestamento e de áreas de pastagens degradadas na região (DIAS-FILHO, 2015).

Entretanto, outros estudos apontam que muitas terras foram mais bem sucedidas do que o esperado e a migração e expansão contínua das pastagens em áreas de floresta não é resultado somente da degradação. Pode ser considerada como resultado do mercado de terras na região e outros incentivos econômicos (MUCHAGATA; BROWN, 2003).

Aliado ao crescimento das áreas de pastagem, um aumento considerável do número de animais também ocorreu. No Brasil, o rebanho bovino cresceu de 102 milhões de cabeças em 1975 para 205 milhões de cabeças em 2006, enquanto as áreas de pastagem passaram de 165 milhões de hectares em 1975 para 172 milhões de hectares em 2006 (IBGE, 2007). Apesar de o aumento de área ter sido de apenas 4%, quando a análise é feita para cada grande região, nota-se que as pastagens na região norte tiveram um crescimento notável (IBGE, 2007). Nas Figuras 2.1 e 2.2, podem ser observadas as dinâmicas do rebanho bovino e das áreas de pastagem nas grandes regiões do Brasil em 1975 e 2006, contabilizando pastagens naturais e plantadas. Na região norte, onde predomina o bioma amazônico, as áreas de pastagem em 1975 eram de 5,28 milhões de hectares e em 2006 esse número subiu para 32,63 milhões de hectares (IBGE, 2007).

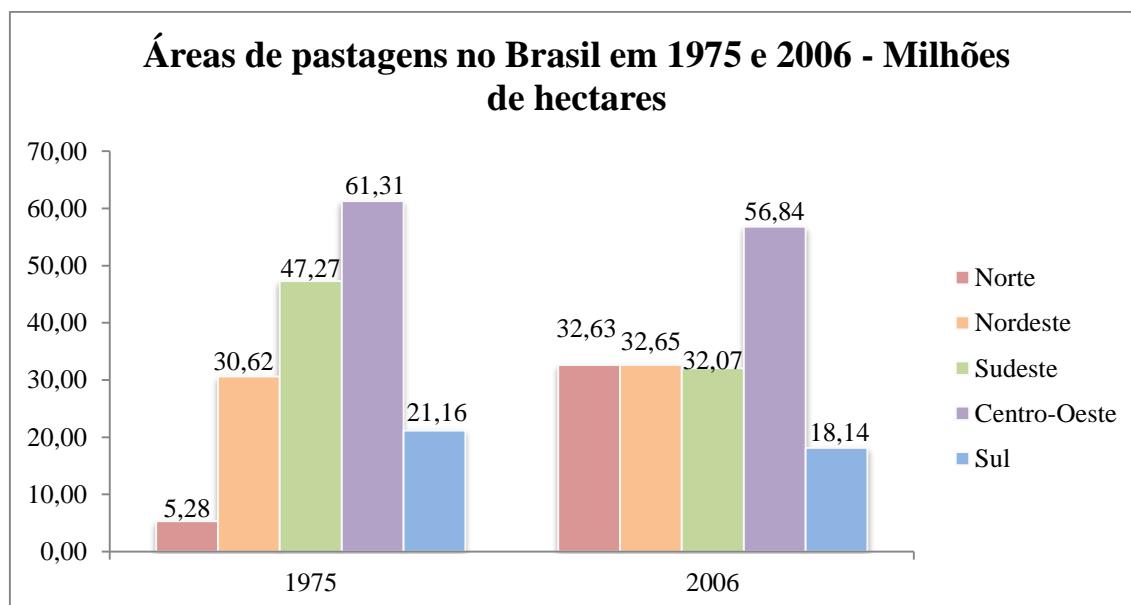
De acordo com o mapeamento do PRODES (Projeto de Monitoramento do Desflorestamento na Amazônia Legal) (INPE, 2013), estima-se que, até o ano de 2012, 18,5% da área da floresta Amazônica (aproximadamente 751.340 km²) encontrava-se desmatada. Desse total de área desmatada, no mesmo ano, aproximadamente 60% estavam cobertas por pastagem, sendo esta a maior responsável pela ocupação das áreas desflorestadas na Amazônia (INPE, 2014).

Figura 2.1 - Dinâmica do rebanho bovino (milhões de cabeças) nas grandes regiões do Brasil em 1975 e 2006.



Fonte: IBGE (2007).

Figura 2.2 - Dinâmica das áreas de pastagem (milhões de hectares) nas grandes regiões do Brasil em 1975 e 2006.



Fonte: IBGE (2007).

2.2. As Pastagens na Amazônia

Considerando a vasta extensão territorial ocupada por pastagens na Amazônia, é importante frisar que são áreas bastante heterogêneas, possuindo diferentes condições de manejo, produtividade e degradação (DAVIDSON *et al.*, 2008; AGUIAR, 2013). O aumento da quantidade de vegetação invasora em pastagens

diminui a produtividade, devido ao menor acesso às forrageiras e à maior competição por nutrientes. A intensidade de uso da terra nas pastagens é dada por diversos fatores, como a utilização de herbicidas e fertilizantes, taxa de lotação, queimadas, retirada de vegetação arbóreo-arbustiva. Normalmente, é comum o abandono de áreas que tiveram a sua capacidade produtiva esgotada e o investimento de capital para a recuperação de pastagens degradadas ainda é uma exceção (RUFIN *et al.*, 2015).

O monitoramento da dinâmica e qualidade das pastagens, em boas condições ou degradadas, auxilia no entendimento de suas tendências espaciais, temporais e de produtividade e também no combate ao desflorestamento (RUFIN *et al.*, 2015).

O projeto TerraClass (ALMEIDA *et al.*, 2016), que realiza o mapeamento sistemático do uso e cobertura da terra na Amazônia desde 2008, define quatro classes de pastagens, considerando exclusivamente áreas de criação de gado em regime extensivo. O mapeamento é feito baseado em imagens da classe Landsat, imagens de aproximadamente 30 m de resolução espacial. As descrições das classes, de acordo com Almeida *et al.* (2016) são:

Pasto limpo: Áreas de pastagem em processo produtivo, com predomínio de vegetação herbácea. Refere-se às áreas recém-implantadas ou com baixa infestação por invasoras herbáceas e arbustivas. Ausência de indivíduos arbóreos;

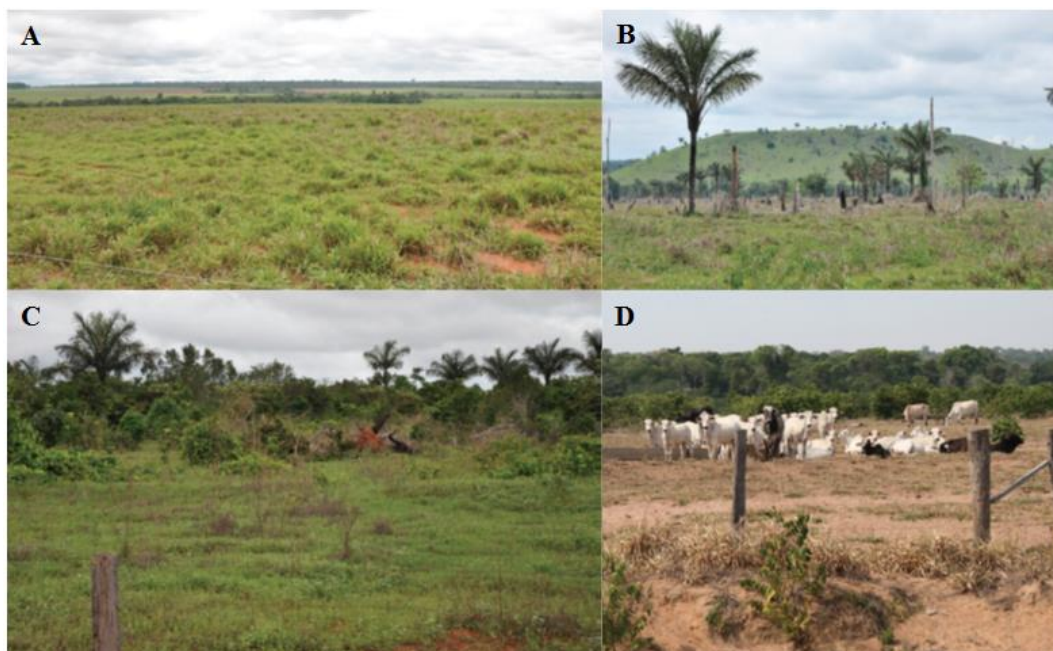
Pasto sujo: Áreas de pastagem em processo produtivo, com predomínio de vegetação herbácea e com a presença de vegetação arbustiva esparsa. Presença significativa de invasoras arbustivas e presença de alguns indivíduos arbóreos;

Regeneração com pasto: Áreas que, após o corte raso da vegetação natural e o desenvolvimento de alguma atividade agropastoril, encontram-se no início do processo de regeneração da vegetação nativa, sendo ocupadas predominantemente por espécies de hábitos arbustivo e arbóreo. Áreas caracterizadas pela grande presença de plantas invasoras, em diversos níveis de desenvolvimento, dificultando muito sua utilização para o pastejo de animais;

Pasto com solo exposto: Áreas que, após o corte raso da vegetação natural e o desenvolvimento de alguma atividade agropastoril, apresentam o solo exposto ou com baixíssima cobertura vegetal;

Os padrões, em campo, apresentados pelas classes descritas podem ser visto na Figura 2.3.

Figura 2.3 – Padrões das classes de pastagem. A) Pasto limpo. B) Pasto sujo. C) Regeneração com pasto. D) Pasto com solo exposto.



Fonte: Coutinho *et al.* (2013).

De acordo com os dados do TerraClass (INPE, 2014), em 2012, aproximadamente 60% do desflorestamento na Amazônia estavam cobertos por pastagem, discriminadas nas seguintes porcentagens: Pasto limpo (45,97%), Pasto sujo (6,72%), Regeneração com pasto (6,18%) e Pasto com solo exposto (0,01%). A classe “Pasto com solo exposto” possui abrangência minoritária se comparada às outras que envolvem pastagens.

2.3. Classificação de Pastagens

Desde 2008, o TerraClass produz um mapa temático, a cada dois anos, contendo o uso e cobertura da terra na Amazônia. Esta classificação se divide em 12 padrões, dentre os quais quatro envolvem pastagens: pasto limpo, pasto sujo, regeneração com pasto e pasto com solo exposto (ALMEIDA *et al.*, 2016). Apesar de ser um

mapeamento sistemático e de grande abrangência espacial, a classificação das pastagens é realizada por meio de interpretação visual, o que torna o processo bastante custoso em relação ao tempo e aos recursos humanos.

Outra iniciativa relacionada com o tema é o projeto GeoDegrade – Desenvolvimento de geotecnologias para identificação e monitoramento de níveis de degradação em pastagens (SILVA *et al.*, 2012). O GeoDegrade teve início em 2011 e término em 2014. O projeto fez a identificação de pastagens degradadas e não degradadas em sete propriedades utilizadas como áreas amostrais, distribuídas nos estados do Pará, São Paulo e Mato Grosso do Sul. Envolveu, portanto, a Amazônia, a Mata Atlântica e o Pantanal. Foram utilizadas séries temporais de índices de vegetação do sensor MODIS (250m de resolução espacial), imagens do satélite Landsat 5 e dados de campo, como amostras de biomassa vegetal. A dependência espacial dos dados foi analisada e foram geradas interpolações e análises geoestatísticas (GREGO *et al.*, 2015). Apesar de o projeto utilizar imagens de sensoriamento remoto e índices de vegetação para a diferenciação de pastagens, sua análise está associada com variáveis coletadas em campo, dificultando a expansão da abrangência espacial para o restante dos biomas analisados.

Numata *et al.* (2000) classificaram a cronossequência de pastagens formadas a partir da conversão floresta-pastagem, na Amazônia, para análise de fertilidade do solo. Para separar desmatamento de floresta, os autores utilizaram, primeiramente, apenas a fração sombra do modelo linear de mistura espectral – MLME (SHIMABUKURO; SMITH, 1991). Trabalho de campo foi realizado para diferenciar pastagens de outros usos da terra na área de estudo e fazer amostragem do solo para análise de parâmetros de fertilidade. As imagens fração sombra foram dispostas em composições falsa-cor e posteriormente classificadas pelo classificador MAXVER para que as informações das idades da pastagem pudessem estar contidas em uma única imagem. A avaliação dos resultados indicou que várias áreas de confusão ocorreram devido a características físicas das pastagens, como a presença de diferentes espécies vegetais e tipos de solo. Isto demonstrou a dificuldade na classificação automática de pastagens que, além de possuírem comportamento espectral semelhante ao de outras classes, possuem

padrões heterogêneos dentro da mesma classe. Sugeriu-se a utilização da fração solo e da fração de NPV (*Non Photosynthetic Vegetation*) para melhorar a caracterização espectral das pastagens.

O monitoramento da conversão floresta-pastagem e das pastagens em si é relevante devido à sua contribuição às dinâmicas climáticas e de carbono na região. Asner *et al.* (2004) focaram seu estudo em mudanças a longo prazo das propriedades biofísicas do dossel e nas propriedades biogeoquímicas de pastagens de dois tipos de solo na Amazônia central. O objetivo foi entender a relação entre as medidas das condições de dossel estimadas por imagens e as dinâmicas biogeoquímicas medidas em campo. Os resultados indicaram que as frações NPV e vegetação (PV – *photosynthetic vegetation*) foram altamente correlacionadas com as medidas em campo de biomassa, carbono orgânico, nitrogênio, fósforo e cálcio disponíveis no solo. O aumento de NPV correlacionou-se com a diminuição de PV e da disponibilidade de nutrientes no solo e na folhagem viva. Dessa forma, Asner *et al.* (2004) defendem o uso das frações como indicadores das condições da pastagem e de estoque de carbono. Entretanto, a presença de espécies invasoras nas áreas de estudo não foi analisada no trabalho.

A utilização das frações PV e NPV como indicadores das condições de pastagens pode ser útil no estudo de sua degradação. Davidson *et al.* (2008) utilizaram as duas frações derivadas de imagens do satélite Landsat 5 para caracterizar a degradação de pastagens na Amazônia. Além disso, eles também utilizaram a fração solo, uma vez que o pastejo em excesso resulta na exposição do solo, que é um indicador de degradação possível de ser determinado por sensoriamento remoto. Para relacionar as medidas das frações com a degradação, foram utilizados registros locais sobre o manejo, a intensidade de pastejo e as condições do solo.

Davidson *et al.* (2008) observaram que as pastagens mais degradadas possuíam maiores frações solo e NPV e menor fração PV. Os valores de PV foram maiores na estação chuvosa em comparação com a estação seca, demonstrando a importância de considerar a sazonalidade climática nos estudos de degradação. A quantidade de tratamentos (gradagem, calagem e aplicação de herbicidas) recebida pelas pastagens foi baixa, porém significativa e positivamente

correlacionada com o NPV e negativamente correlacionada com o PV. Essas relações podem ser justificadas devido: (1) aos tratamentos terem levado a uma redução na cobertura e biomassa de certa vegetação que contribuiria para baixo valor de PV e (2) considerando que os pastos que receberam maior quantidade de tratamentos foram os mais utilizados (sofreram maior pressão), essas pastagens possuíam maior quantidade de vegetação senescente em relação à vegetação verde (DAVIDSON *et al.*, 2008).

A relação positiva entre os valores de NPV e a maior quantidade de tratamentos foi observada na pastagem com maior lotação (1,5 – 1,8 cabeças/ha), enquanto as pastagens com menores valores de NPV apresentaram lotação entre 0,25 e 1,0 cabeças/ha. De acordo com Davidson *et al.* (2008), o manejo pode superar a tendência para a degradação e manter a produtividade suficiente para a operação da pecuária. Entretanto, as diferentes idades e tipos de solo das pastagens estudadas não foram levados em consideração.

Aliada às imagens fração, é comum a utilização de índices de vegetação (IV) na caracterização das pastagens. Numata *et al.* (2007a) avaliaram a relação entre medidas biofísicas obtidas em campo e imagens fração e IVs obtidos a partir de imagens do satélite Landsat para investigar o impacto da intensidade de pastejo (animais/ha) nas medidas biofísicas das pastagens em Rondônia, Amazônia. As medidas obtidas em campo foram biomassa (t/ha), teor de água (t/ha) e altura do dossel (cm). As medidas obtidas a partir da transformação de imagens de satélite foram frações NPV, PV e sombra e os IVs: NDVI (*Normalized Difference Vegetation Index*), SAVI (*Soil-Adjusted Vegetation Index*), NDII5 (*Normalized Difference Infrared Index – banda 5 – Short-Wave Infrared – SWIR1*) e NDII7 (*banda 7 – SWIR2*).

Todos os IVs apresentaram correlações positivas com as medidas em campo. Os baixos valores de coeficiente de correlação indicaram que pastos com intensa criação de gado possuem menor quantidade de folhagem verde e os materiais não fotossintéticos dominam a biomassa. Numata *et al.* (2007a) também verificaram a possibilidade de estimar as propriedades biofísicas das pastagens a partir dos produtos de sensoriamento remoto utilizados, como a estimativa da altura do dossel usando a fração sombra.

Além dos métodos que utilizam técnicas de transformação de imagens para extrair informações sobre áreas de pastagens, a inclusão do aspecto temporal utilizando séries temporais na análise pode auxiliar na separação de alvos espectralmente similares. Com esse objetivo, Müller *et al.* (2015) utilizaram a mineração de métricas de variabilidade espectro-temporal para discriminar as seguintes classes: floresta, pastagem artificial, agricultura, vegetação natural de formações savânicas (campo cerrado, cerrado *sensu stricto* e cerradão), corpos d'água e área sem vegetação (área construída, solo exposto e formações rochosas). As métricas estatísticas (média, mediana, desvio padrão, quartil 75% e intervalo interquartil 25-75%) e o índice SWIR (*Short-Wave Infrared*) foram extraídos a partir de séries temporais de valores de reflectância das imagens do satélite Landsat, referentes à bacia do Rio das Mortes - MT entre os anos de 2009 e 2012. Utilizaram o classificador *random forest* (RF) e obtiveram uma taxa de acerto de 93%, com um intervalo de confiança de 95%.

Rufin *et al.* (2015) exploraram o potencial de séries temporais de imagens TM do satélite Landsat para caracterizar a qualidade de pastagens na Amazônia, em Novo Progresso – PA. Para isso, as áreas de floresta primária mapeadas pelo PRODES foram mascaradas. Nas áreas desflorestadas, os autores identificaram como métricas da estação seca os valores médios, máximos, mínimos e de desvio padrão da componente *wetness* (TCW) da transformação *tasseled cap* em uma série temporal de imagens de 1984 a 2012. A análise das métricas permitiu a diferenciação das pastagens pelo algoritmo RF em dois estágios de sucessão diferentes, sendo um dominado por gramíneas (pasto limpo, em boas condições) e outro dominado por plantas invasoras. Nesse estudo, os dados do TerraClass 2008 foram utilizados como referência.

Na análise da trajetória, a alta variabilidade do TCW nos dois estágios de sucessão indica a existência de heterogeneidade na sucessão, sugerindo que existem transições graduais entre os diferentes estágios (RUFIN *et al.*, 2015). Os autores não utilizaram métricas relacionadas com a textura e com os formatos dos polígonos. Além disso, a aplicação da metodologia em outras regiões do bioma amazônico pode trazer novas informações acerca da heterogeneidade no comportamento das pastagens.

Neves *et al.* (2016) exploraram técnicas de mineração de dados para classificar séries temporais de EVI2, produto do sensor MODIS (250m de resolução espacial), em uma área no norte do Mato Grosso. Os anos de 2008 e 2010 foram classificados em quatro classes: Agricultura, Pastagem, Floresta e Outros. Além da própria série temporal com os valores de EVI2, outros atributos foram extraídos a partir da série: atributos básicos (média, desvio padrão, amplitude, etc.) e atributos polares (área, ângulo de orientação, excentricidade, etc.). As classificações foram feitas por *pixel* e também baseada em objetos. Os resultados utilizando somente os valores de EVI2 ou os valores de EVI2 junto com os atributos básicos e polares foram bem próximos, alcançando taxas de acerto de até 91,70%.

Apesar de as taxas de acerto global para as classificações por *pixel* terem dado superiores àquelas baseadas em objetos, Neves *et al.* (2016) observaram que a utilização de objetos aumentou consideravelmente as taxas de acerto para a classe de Pastagem. Para esta, os acertos passaram de 75%, em média, para 90%, em média, quando explorada a abordagem baseada em objetos. Os resultados demonstram a importância do aspecto espacial para a identificação desta classe. A inclusão de novos atributos e a utilização de uma melhor resolução espacial dos dados poderia melhorar o conhecimento acerca da separação das classes estudadas. Nesse sentido, Neves *et al.* (2016) pode ser considerado como ponto de partida para o presente trabalho.

2.4. Métodos de Processamento Digital de Imagens

Esta seção apresenta uma descrição resumida das técnicas de processamento digital de imagens que foram utilizadas no desenvolvimento deste trabalho.

2.4.1. Índices de Vegetação

Os índices de vegetação (IV) são obtidos por técnicas de transformação de imagens, que geram novos produtos digitais a partir de dados originais de satélite, reduzindo a dimensionalidade dos dados sem perda de informação espectral (ADAMI, 2010; MOREIRA; SHIMABUKURO, 2004). Os índices são utilizados para realçar a resposta espectral da vegetação e resultam da combinação dessas respostas em duas ou mais bandas espectrais do espectro eletromagnético

(HUETE *et al.*, 2002). A proposição dessa combinação fundamenta-se nas respostas espectrais distintas que a vegetação possui nas regiões do visível e do infravermelho próximo (NIR – *Near Infrared*). No visível, a baixa reflectância está relacionada à absorção da REM (Radiação Eletromagnética) pela ação de pigmentos fotossintetizantes, enquanto na região do infravermelho próximo, a alta reflectância ocorre devido ao espalhamento da radiação no interior das folhas pela sua estrutura celular (PONZONI; SHIMABUKURO, 2009).

Os IVs podem ser relacionados com uma série de variáveis biofísicas da vegetação, como a biomassa, o índice de área foliar (IAF) e a evapotranspiração (DORIGO *et al.*, 2007). Além disso, minimizam os efeitos de iluminação da cena, de declividade da superfície e de geometria de aquisição, que influenciam nos valores de reflectância da vegetação (PONZONI; SHIMABUKURO, 2009).

A utilização dos IVs permite o monitoramento sazonal, interanual e a longo prazo de parâmetros estruturais, fenológicos e biofísicos da vegetação (HUETE *et al.*, 2002). Entretanto, a utilização deles deve ser feita com certo cuidado, uma vez que atendem a necessidades particulares sob condições específicas. Se essas condições (por exemplo, as geometrias de iluminação e visada) são modificadas, os resultados dos índices se alteram, sem necessariamente terem ocorrido mudanças nas características da vegetação (ADAMI, 2010).

O índice de vegetação por diferença normalizada (*Normalized Difference Vegetation Index* – NDVI) é amplamente utilizado (MOREIRA; SHIMABUKURO, 2004). Além deste, muitos outros índices vêm sendo utilizados nos estudos envolvendo pastagens, como o *Enhanced Vegetation Index* (EVI), o *Soil-Adjusted Vegetation Index* (SAVI) e o *Normalized Difference Infrared Index* (NDII) (NUMATA *et al.*, 2003; NUMATA *et al.*, 2007a).

2.4.2. Modelo Linear de Mistura Espectral

A radiância registrada no campo de visada instantâneo (*Instantaneous Field of View* – IFOV) de um sensor corresponde à soma das radiâncias de vários alvos (folhas, galhos, sombra, solo, etc.) contidos nele. A mistura está relacionada com a resolução espacial do sensor, uma vez que sensores com resolução espacial

maior, e IFOV menor, terão menos componentes em um *pixel* (SHIMABUKURO; SMITH, 1991).

Os modelos de mistura são utilizados para separar as contribuições dos componentes, ou *endmembers*, em um *pixel*, gerando imagens a partir dos espectros de “componentes puros” (ROBERTS *et al.*, 1998). Diversos modelos já foram propostos para identificar a proporção de misturas e alguns trabalhos indicam os principais componentes como solo, vegetação e água ou sombra (SHIMABUKURO; SMITH, 1991; TEIXEIRA, 2004). O Modelo Linear de Mistura Espectral (MLME), conforme Shimabukuro e Smith (1991), propõe uma relação linear entre os componentes de um mesmo *pixel*, de acordo com o seguinte modelo:

$$r_i = \sum_{j=1}^N (a_{ij}x_j) + e_i \quad (2.1)$$

Onde:

- $j = 1, 2, \dots, N$ = número de componentes assumidos no problema;
- $i = 1, 2, \dots, M$ = número de bandas espectrais do sensor utilizado;
- r_i = média da reflectância espectral para a i -ésima banda espectral de um *pixel* contendo N componentes;
- a_{ij} = reflectância espectral do j -ésimo componente do *pixel* para a i -ésima banda espectral;
- x_j = valor proporcional do j -ésimo componente do *pixel*;
- e_i = erro para a i -ésima banda espectral.

Além dos componentes já citados, pode ser feita a inclusão de um quarto importante componente: vegetação não fotossinteticamente ativa (NPV) (ADAMS *et al.*, 1995). O NPV, como já diz o nome, consiste no material vegetal que não possui função de fotossíntese, incluindo troncos, vegetação seca e serrapilheira. Consiste, portanto, em um importante indicador da produtividade em pastagens, assim como em outros ecossistemas. Ecologicamente, o NPV possui papel importante no controle de carbono e de água e na captação de nutrientes (LI; GUO, 2015).

O aumento da idade da pastagem relaciona-se com o crescimento da fração NPV. O aumento dessas duas variáveis pode indicar uma redução na disponibilidade de

nutrientes no solo e também de nutrientes disponíveis na folhagem, de forma que o NPV pode ser um indicador viável na análise da degradação de pastagens (DAVIDSON *et al.*, 2008; NUMATA *et al.*, 2007a).

A presença de NPV contribui significativamente na reflectância dos alvos de vegetação. Portanto, as medidas da fração NPV estimadas pelo MLME podem ser relacionadas com a diferenciação entre pastagens e vegetação natural (VICENTE *et al.*, 2011), com parâmetro de fertilidade do solo em pastagens, com suas idades (NUMATA *et al.*, 2003) e com processos e níveis de degradação (NUMATA *et al.*, 2007a, 2007b).

2.4.3. Transformação Tasselled Cap

A transformação de imagens *tasselled cap* (TC), ou transformação KT, nome dado em homenagem aos seus criadores Kauth e Thomas (KAUTH; THOMAS, 1976), consiste em uma transformação linear, capaz de reduzir o volume e a complexidade dos dados, realçando determinadas características das cenas de estudo.

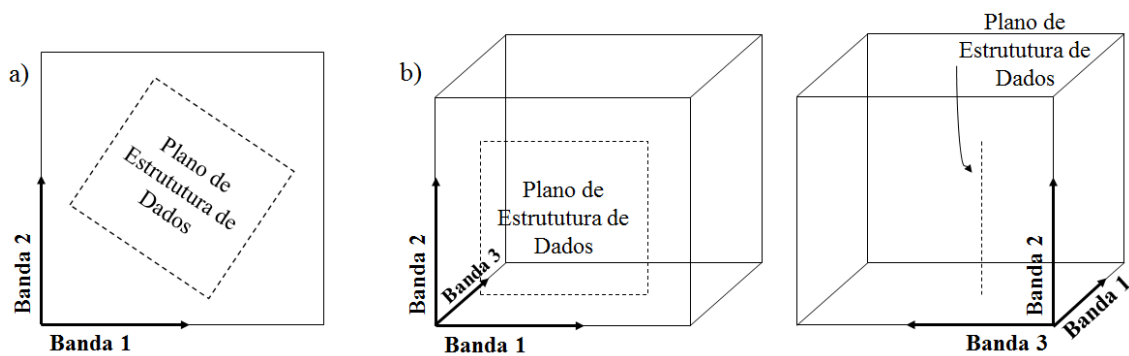
Os dados de um determinado sensor podem ser representados por um espaço multi-dimensional, no qual cada dimensão corresponde a uma banda do sensor. Em um sensor hipotético de três bandas, o espaço resultante seria um cubo, como na Figura 2.4b. Esse raciocínio pode ser aplicado para sensores com mais bandas, apesar de a visualização dos dados tornar-se mais complexa (CRIST; KAUTH, 1986).

Alvos como vegetação e solo possuem curvas de reflectância com formatos característicos, pois seus indivíduos compartilham propriedades físicas (como a estrutura celular, no caso da vegetação) que possuem respostas com padrões previsíveis. Então, os dados não possuem igual probabilidade de ocorrência no espaço multidimensional, mas se concentram em certas porções, formando o que se pode chamar de “estrutura de dados” (CRIST; KAUTH, 1986).

Se cada banda do sensor fosse relacionada com as características de uma classe de estudo e não fosse correlacionada com outras bandas, então modificações nessas características iriam causar alterações somente nessa banda. Porém, é mais

comum que essas correlações sejam imperfeitas. As alterações de um alvo causam variações em diversas bandas e a informação apenas será captada por meio da combinação de bandas. A estrutura desses dados em um sensor de duas bandas apenas ficaria desalinhada dos eixos das bandas (Figura 2.4a). Em um sensor de três bandas, a estrutura fica um pouco mais complexa e, dependendo da perspectiva de visualização, a informação mais distorcida (Figura 2.4b) e assim por diante (CRIST; KAUTH, 1986).

Figura 2.4 – Representação de um plano de estrutura de dados. a) Em um sensor de duas bandas. b) Em um sensor de três bandas, o plano de estrutura de dados seria diferente dependendo da perspectiva de visualização.



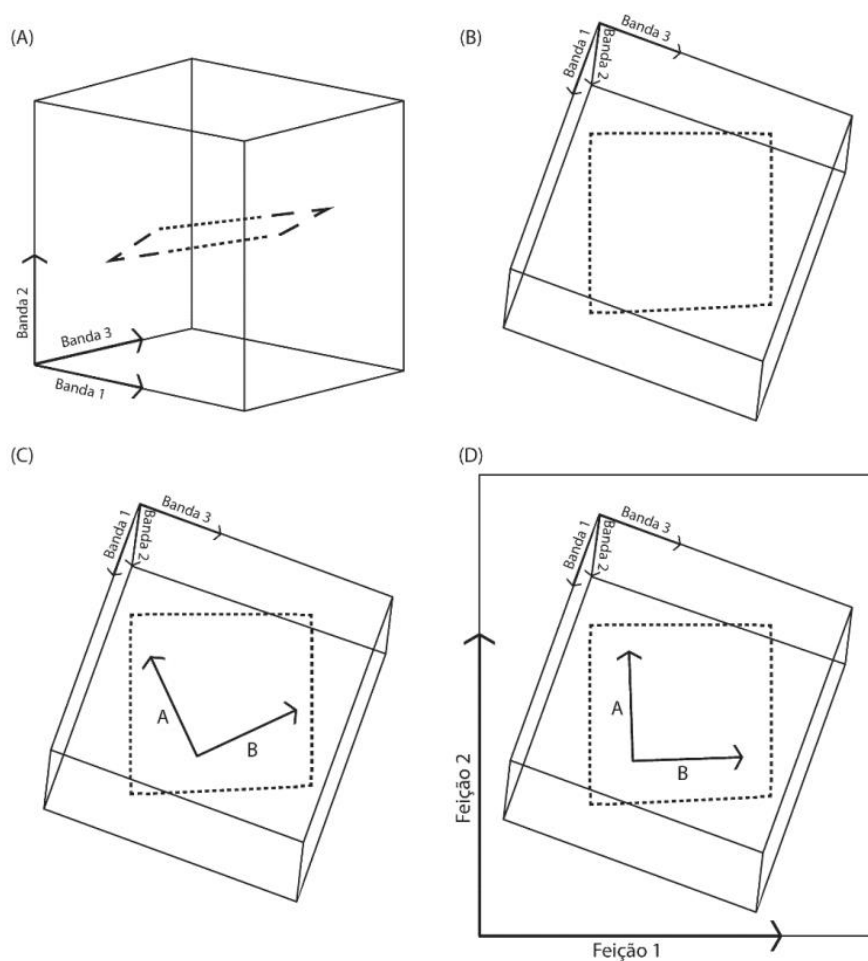
Fonte: Adaptado de Crist e Kauth (1986).

Já que a estrutura dos dados existente em cada sensor está relacionada com características das classes de estudo, um determinado alvo seria melhor interpretado se a informação fosse vista de maneira mais direta possível e separadas de outras estruturas. A transformação *Tasselled Cap* (*Tasselled Cap Transformation* – TCT) é capaz de identificar a estrutura de dados para determinados conjuntos de sensor/aplicação e, rotacionando os eixos das bandas, definir novas direções (eixos) associadas exclusivamente às características físicas de um alvo da cena (Figura 2.5). Esses novos eixos serão as componentes da transformação TC. Não há alteração de dados após a transformação, apenas a perspectiva de visualização muda (CRIST; KAUTH, 1986).

Suponhamos que os dados de um sensor de três bandas estão concentrados em uma estrutura plana (Figura 2.5a). A mudança da perspectiva dos dados permite a visualização da totalidade da variação dos dados no plano (Figura 2.5b). A e B são classes de uma cena que variam primeiramente nas direções indicadas na Figura 2.5c e podem ser rotacionadas para que as direções de variação fiquem alinhadas

com os novos eixos, que correspondem às feições (componentes) da TCT (Figura 2.5d). Os novos eixos são combinações das bandas originais e correspondem mais fielmente às variações das classes da cena e, assim, poderão ser interpretadas de maneira mais correta e sem ambiguidade (CRIST; KAUTH, 1986).

Figura 2.5 – Dados de um sensor hipotético de três bandas. A) O cubo é o espaço total disponível para os dados. A linha tracejada representa uma estrutura retangular (plana) dentro do cubo. B) A perspectiva mudou para a estrutura retangular dos dados ser vista diretamente. C) Direção da variação espectral associada com dois processos físicos distintos em classes presentes na cena. D) Ajustamento adicional na perspectiva de observação para alinhar as direções relacionadas a processos físicos com os novos eixos de coordenadas.



Fonte: Pessoa (2016) – Adaptado de Crist e Kauth (1986).

Os sensores possuem bandas espectrais cujas informações são altamente correlacionadas. TC não só diminui a quantidade de bandas (componentes), como também retira a correlação entre as bandas, transformando-as em eixos relacionados com características físicas (BAIG *et al.*, 2014).

Para cada diferente sensor, deve ser proposta uma diferente transformação adequada às suas bandas espectrais. Porém, a mudança do conjunto de dados não requer uma nova transformação, permitindo a comparação de produtos de muitas datas de aquisição e diferentes sensores (CRIST; KAUTH, 1986; RUFIN *et al.*, 2015).

Em trabalhos anteriores, foram criados para diferentes sensores três componentes: verdor (“*greenness*”), brilho (“*brightness*”) e umidade (“*wetness*”) (KAUTH; THOMAS, 1976; CRIST; CICONE, 1984; ZHANG *et al.*, 2002). A componente *greenness* está relacionada com a reflectância característica da vegetação verde. Perpendicularmente, a componente *brightness* foi definida na direção de variação da reflectância do solo e pode estar relacionada com solos descobertos ou parcialmente cobertos, feições naturais ou antrópicas e variações na topografia. Por último, a componente *wetness* corresponde à direção observada pela variação da umidade do solo no plano ocupado por solos descobertos, também podendo estar relacionada com a água (CRIST; KAUTH, 1986; BAIG *et al.*, 2014).

Baig *et al.* (2014) propuseram os coeficientes (Tabela 2.1) utilizados para a TCT em dados do sensor *Operational Land Imager* (OLI) do Landsat 8.

Tabela 2.1 – Coeficientes da transformação TC para dados do sensor OLI (Landsat 8).

Landsat 8						
TC	Banda 2 (azul)	Banda 3 (verde)	Banda 4 (vermelho)	Banda 5 (NIR)	Banda 6 (SWIR1)	Banda 7 (SWIR2)
Brightness	0,3029	0,2786	0,4733	0,5599	0,508	0,1872
Greenness	-0,2941	-0,243	-0,5424	0,7276	0,0713	0,1608
Wetness	0,1511	0,1973	0,3283	0,3407	-0,7117	-0,4559
TC4	-0,8239	0,0849	0,4396	-0,058	0,2013	0,2773
TC5	-0,3294	0,0557	0,1056	0,1855	-0,4349	0,8085
TC6	0,1079	-0,9023	0,4119	0,0575	-0,0259	0,0252

Fonte: Baig *et al.* (2014).

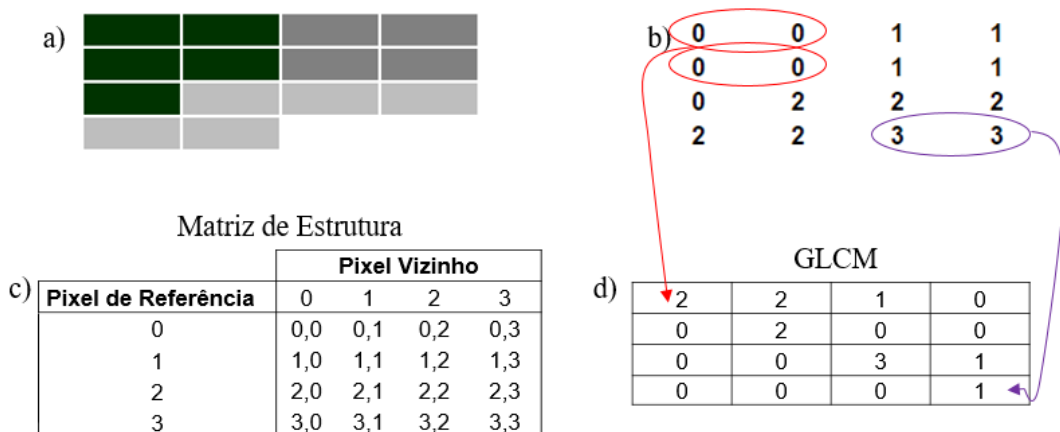
Devido à sua forte relação com a estrutura da vegetação e baixa sensibilidade aos tipos de solo se comparada com os outros índices de vegetação, Rufin *et al.* (2015) utilizaram componentes da transformação TC para o monitoramento das sucessões secundárias em pastagens.

2.4.4. Medidas de Textura

Os alvos em uma imagem de sensoriamento remoto também podem ser caracterizados por meio de suas texturas. Uma textura lisa está associada a pequenas diferenças entre os maiores e menores valores de níveis de cinza de uma imagem, ou de cada segmento da imagem. A textura rugosa refere-se a grandes diferenças entre os menores e maiores valores. Tais valores podem representar diferentes grandezas, como radiância, número digital, reflectância de superfície e outros.

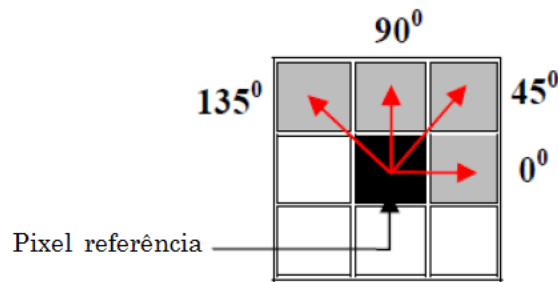
Para a representação da textura e de atributos relacionados a ela, Haralick *et al.* (1973) propôs a matriz de coocorrência de níveis de cinza (*Grey Level Co-occurrence Matrix – GLCM*), uma tabulação de todas as diferentes combinações de brilho de *pixel* (níveis de cinza) que ocorrem em uma imagem (ou região). Dado que temos uma imagem (Figura 2.6a) com seus respectivos níveis de cinza (Figura 2.6b), a estrutura de sua GLCM será feita de acordo com as combinações possíveis entre os *pixels* de referência e os *pixels* vizinhos (Figura 2.6c). A escolha de quais serão considerados vizinhos depende da direção de deslocamento escolhida (0°, 45°, 90°, 135° ou todas as direções). Neste exemplo, a direção 0° (Figura 2.7) foi utilizada. O preenchimento da matriz será de acordo com a quantidade de ocorrência de cada combinação (Figura 2.6d).

Figura 2.6 – Construção da GLCM. a) Imagem em níveis de cinza. b) Valores dos níveis de cinza. c) Matriz de estrutura. d) GLCM.



Fonte: Adaptado de Hall-Beyer (2007).

Figura 2.7 – Direções de deslocamento na GLCM.



Fonte: Adaptado de Pathak e Barooah (2013).

A matriz deve ser transposta para ser simétrica em relação à diagonal e também deve ser normalizada para valores de probabilidade de ocorrência das combinações. A matriz resultante é quadrada e possui o mesmo número de linhas e colunas que o nível de quantização da imagem. Assim, uma imagem com 256 níveis de cinza formará uma matriz quadrada de 256 linhas por 256 colunas, totalizando 65.536 células. A partir da GLCM, medidas de textura podem ser criadas e estas são, geralmente, ponderações derivadas da matriz. Estão divididas em três grupos: de contraste, de ordem e estatístico (HALL-BEYER, 2007).

No grupo de contraste, as medidas são ponderadas de acordo com a distância em relação à diagonal da GLCM, o peso aumenta com o aumento da distância. Estão nesse grupo: Contraste, Dissimilaridade e Homogeneidade. O grupo de ordem reflete quão regularmente ordenados os valores dos *pixels* ocorrem na imagem. São deste grupo o 2º Momento Angular e a Entropia. O terceiro grupo, como já diz o nome, representa medidas estatísticas extraídas da matriz: Média, Desvio Padrão ou Correlação (HALL-BEYER, 2007). Dada uma GLCM i,j com N células de valor P , as fórmulas das medidas de textura estão a seguir:

$$\text{Contraste} = \sum_{i,j=0}^{N-1} P_{i,j}(i-j)^2 \quad (2.2)$$

$$\text{Dissimilaridade} = \sum_{i,j=0}^{N-1} P_{i,j}|i-j| \quad (2.3)$$

$$\text{Homogeneidade} = \sum_{i,j=0}^{N-1} \frac{P_{i,j}}{1+(i-j)^2} \quad (2.4)$$

$$2^{\text{o}} \text{ Mom. Ang.} = \sum_{i,j=0}^{N-1} P_{i,j}^2 \quad (2.5)$$

$$\text{Entropia} = \sum_{i,j=0}^{N-1} P_{i,j} (-\ln P_{i,j}) \quad (2.6)$$

$$\text{Média} = \mu_i = \sum_{i,j=0}^{N-1} i(P_{i,j}) \quad (2.7)$$

$$\text{Desvio Padrão} = \sigma_i = \sqrt{\sum_{i,j=0}^{N-1} P_{i,j} (i - \mu_i)^2} \quad (2.8)$$

$$\text{Correlação} = \sum_{i,j=0}^{N-1} P_{i,j} \left[\frac{(i - \mu_i)(i - \mu_j)}{\sqrt{(\sigma_i^2)(\sigma_j^2)}} \right] \quad (2.9)$$

2.5. Séries Temporais de Imagens de Sensoriamento Remoto

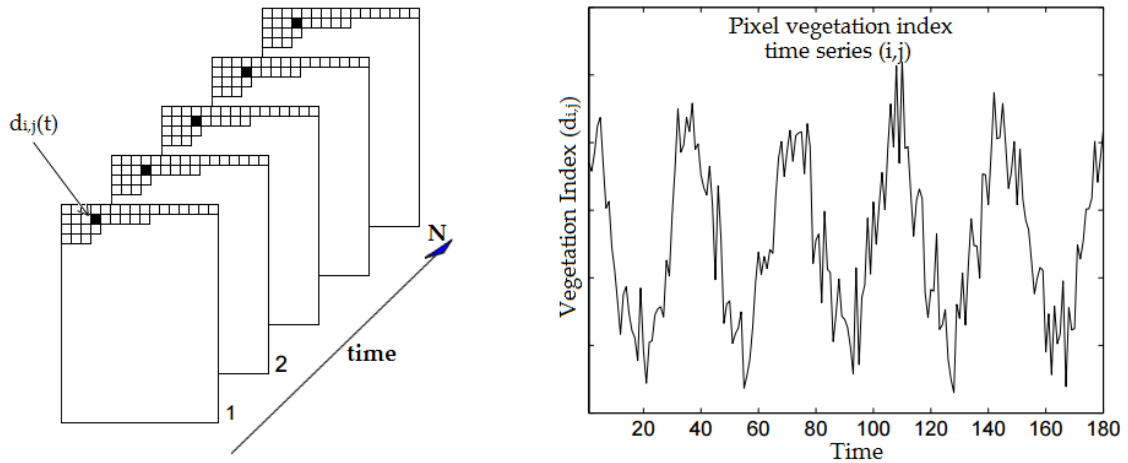
Satélites orbitais geram imagens das mesmas regiões do planeta várias vezes ao ano. O empilhamento destas imagens, levando em consideração as datas de aquisição, dá origem ao que chamamos de série temporal de imagens de sensoriamento remoto.

Séries temporais consistem em um conjunto de observações $d_{i,j}(t)$, em que cada uma delas foi capturada em um momento de tempo t e encontra-se em uma coordenada específica i,j (Figura 2.8). Podem ser utilizadas para representar a dinâmica (ou perfil) temporal de fenômenos naturais ou artificiais (BROCKWELL, 2002).

Sabe-se que as florestas podem ser caracterizadas no seu estado atual, por meio de atributos relacionados à sua estrutura e à sua composição, ou na sua dinâmica temporal. No estudo desta última, as séries temporais são imprescindíveis. As mudanças na dinâmica temporal podem estar relacionadas com a sazonalidade natural ou com outros eventos resultantes de atividades humanas. Entender essa

dinâmica e suas alterações pode auxiliar no manejo das florestas e em políticas de desenvolvimento (BANSKOTA *et al.*, 2014).

Figura 2.8 – Composição de uma série temporal.



Fonte: Adaptado de Eklundh e Jönsson (2012).

Existem diversas técnicas para extrair informações das séries temporais. Primeiramente pode-se estabelecer um modelo probabilístico adequado para representar os dados. Depois, é possível estimar parâmetros e utilizar o modelo para auxiliar o nosso entendimento acerca do fenômeno estudado. A partir do desenvolvimento de um modelo adequado, ele pode ser utilizado de diversas maneiras, como numa descrição mais compacta dos dados, na separação dos efeitos de sazonalidade e tendência, na filtragem de ruídos nos sinais, ou ainda na predição de valores futuros (BROCKWELL, 2002).

Séries temporais de índices de vegetação podem ser usadas para obter informações acerca do desenvolvimento sazonal da vegetação, auxiliando no entendimento das características estruturais e funcionais da cobertura terrestre. As longas séries temporais de IVs podem, então, auxiliar no monitoramento de variações nos padrões de uso e cobertura da terra (EKLUNDH; JÖNSSON, 2012). O uso de séries temporais em estudos que envolvem vegetação possui diversas aplicações, como a caracterização e mapeamento fenológico e ambiental (EKLUNDH; JÖNSSON, 2003; SILVA *et al.*, 2013), o melhoramento de base de dados em classificações ambientais (NUMATA *et al.*, 2003; NEVES *et al.*, 2016) e o monitoramento de mudanças ambientais naturais ou antrópicas (VERBESSELT *et al.*, 2010; DEVRIES, *et al.*, 2015).

Diversas fontes de dados são utilizadas para a composição de séries temporais. Dados do MODIS (*Moderate Resolution Imaging Spectroradiometer*) e do AVHRR (*Advanced Very High Resolution Radiometer*), que possuem revisita diária, são adequados para a análise de séries temporais em escalas regionais ou globais. Entretanto, devido às suas resoluções espaciais (250 m e 1,1 km, respectivamente), são menos apropriadas para a captura e análise de pequenas alterações ambientais relevantes para o manejo e monitoramento ambiental (WALKER *et al.*, 2012). Apesar da melhor resolução temporal do MODIS e do AVHRR, a vantagem da utilização do satélite Landsat é a sua resolução espacial (30 metros) mais adequada na captura de impactos antropogênicos (BANSKOTA *et al.*, 2014).

A utilização de séries temporais de IVs auxilia na diferenciação de classes de uso da terra que possuem assinaturas espectrais parecidas no período seco, mas apresentam comportamentos diferentes ao longo de seu perfil temporal, como a pastagem e a agricultura anual (ALMEIDA *et al.*, 2014). O monitoramento detalhado do processo de degradação de pastagens requer uma frequência de observações intensa, especialmente no bioma amazônico, onde ocorre o rápido crescimento da vegetação. A fenologia da vegetação durante a estação seca é fundamental para a diferenciação das condições da pastagem (RUFIN *et al.*, 2015).

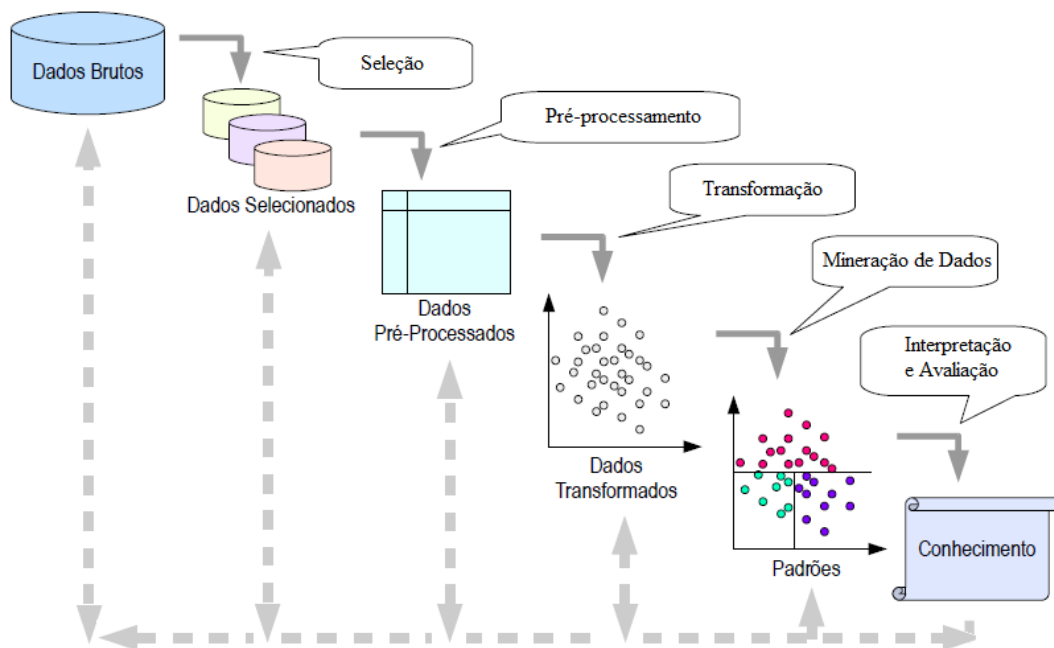
2.6. Mineração de Dados

Com o avanço das tecnologias de aquisição de dados nas últimas décadas e, como consequência, o grande aumento da quantidade de dados, tornou-se inviável a utilização desses dados para a produção de informações de forma manual (muitas vezes lenta, custosa e subjetiva). A partir da necessidade de transformar o grande volume de dados em informação útil e compacta, surgiu o campo da descoberta de conhecimento em bancos de dados (KDD – *Knowledge Discovery in Databases*). Nesse sentido, surgiram diversas técnicas computacionais capazes de automatizar o processo de reconhecimento de padrões e tendências nos dados. O processo de descoberta (Figura 2.9) consiste em várias etapas: seleção de dados, pré-processamento, transformação, mineração de dados e interpretação e avaliação (FAYYAD *et al.*, 1996).

A mineração de dados consiste em uma das mais importantes etapas do KDD, é responsável pela aplicação de algoritmos baseados em técnicas estatísticas, de reconhecimento de padrões e de aprendizado de máquinas para extrair padrões, correlações e tendências dos dados (FAYYAD *et al.*, 1996; LAROSE; LAROSE, 2014). A mineração de dados pode executar as seguintes tarefas: descrição, estimativa, predição, classificação, agrupamento e associação (LAROSE; LAROSE, 2014).

No âmbito do sensoriamento remoto, a classificação consiste em uma técnica de extração de informação de imagens digitais, responsável pela atribuição de significado ou rótulo a um *pixel* (ou a um conjunto deles) de acordo com suas propriedades, sejam elas espectrais, espaciais, etc. (NOVO, 2010).

Figura 2.9 – Processo da descoberta de conhecimento em bases de dados.



Fonte: Santos (2009) – Adaptado de Fayyad *et al.* (1996).

Em um conjunto de dados, diversos atributos estão relacionados com cada registro ou instância. Por exemplo, em uma imagem digital, cada *pixel* pode ser um registro ou instância e os valores de reflectância de cada banda espectral podem ser os atributos dos *pixels*. Entretanto, alguns desses atributos podem ser parcial ou completamente irrelevantes (não afetam o conjunto de dados) ou redundantes (não adicionam nenhuma nova informação) na geração do modelo. Existem

técnicas da mineração de dados para realizar a seleção de atributos, ou seja, selecionar os atributos que melhor contribuem para a classificação. A exclusão dos atributos desnecessários é capaz de reduzir o tempo de geração do modelo e aumentar a taxa de acerto da classificação (DASH; LIU, 1997).

Os métodos de reconhecimento de padrões nos dados podem ser de dois tipos: não supervisionados ou supervisionados. Os não supervisionados não identificam classes especificadas pelo usuário, apenas procuram padrões e estruturas capazes de agrupar os dados semelhantes. Enquanto isso, os métodos supervisionados indicam a existência de classes preestabelecidas e, além disso, os seus algoritmos recebem “treinamento”, no qual amostras das classes escolhidas são apresentadas para que os algoritmos entendam qual o padrão das classes procuradas (LAROSE; LAROSE, 2014).

O método não supervisionado mais comum consiste no agrupamento (*clustering*), no qual os grupos, ou *clusters*, representam uma coleção de registros similares entre si e dissimilares dos registros dos outros grupos. Podem ser utilizadas diversas distâncias de dissimilaridade e algoritmos, como o k-médias, o ISODATA e os mapas auto-organizáveis (SOM – *Self Organizing Maps*) (TAN *et al.*, 2001).

Dentre os métodos de mineração de dados supervisionados, existem os métodos de regressão, que encontram funções para predizer valores numéricos. Para a classificação, existem as árvores de decisão, as florestas aleatórias, redes neurais, dentre outros (LAROSE; LAROSE, 2014). As árvores de decisão consistem em uma coleção de regras de decisões (testes sobre os atributos). A partir do conjunto de dados de entrada, cria-se um nó raiz, que possui um teste lógico capaz de subdividir o conjunto em ramos. Cada ramo possuirá novos testes lógicos que irão subdividir o conjunto de dados até atingir as folhas, que representam as classes (QUINLAN, 2014). As árvores de decisão possuem representações visuais simples do conhecimento e permitem a visualização da interação dos atributos na classificação (APTE; WEISS, 1997).

Classificadores compostos (os que usam vários classificadores), como as florestas aleatórias (*random forest* – RF), são mais robustos e lidam melhor com ruídos e

outliers. O RF cria um conjunto de árvores de decisão sobre subconjuntos dos dados de treinamento, estando cada uma delas responsável por um voto que irá decidir qual a classe mais “popular” (BREIMAN, 2001). O classificador RF possui as seguintes vantagens: funcionamento eficiente quando se usam grandes conjuntos de dados, capacidade de estimar os atributos importantes na classificação e menor custo computacional se comparado com outros classificadores compostos. O RF consiste em um conjunto de árvores de decisão, cada uma capaz de gerar um voto; o *pixel*/objeto recebe a classe que obtiver a maior quantidade de votos. As diferentes árvores de decisão do RF são criadas a partir de subconjuntos de dados, formados por *bootstrap*, uma técnica de reamostragem aleatória com reposição (RODRIGUEZ-GALIANO *et al.*, 2012).

As redes neurais (RN) artificiais apresentam um modelo matemático baseado na estrutura neural de organismos inteligentes e adquirem conhecimento por meio de experiência. As RN são formadas por camadas (de entradas, intermediárias e de saída) interligadas por conexões. As camadas de entrada recebem sinais de entrada ou funcionais (estímulos), que se propagam pelas conexões (neurônio a neurônio) através da rede e chegam às camadas de saída como um sinal de saída. Se o valor do sinal de saída não atende ao sinal desejado pelo usuário, um sinal de “erro” é calculado e propagado no sentido inverso da rede (*error backpropagation*), de forma que as conexões sofrem ajustes (HAYKIN, 2009).

Após a construção de modelos preditivos, existem algumas medidas de avaliação e desempenho dos modelos utilizadas para analisar a capacidade de predição do modelo criado. Algumas dessas técnicas são: matriz de confusão, taxa de acerto (ou erro), falsos positivos, falsos negativos e índice Kappa (LAROSE; LAROSE, 2014; HAN *et al.*, 2012).

A forma geral de uma matriz de confusão, contendo os erros e acertos de uma classificação, com duas classes de estudo (SIM e NÃO), pode ser vista na Tabela 2.2. As colunas representam os valores preditos na classificação, as linhas representam as classes verdadeiras e as células contêm os conceitos de verdadeiros positivos (VP), falsos negativos (FN), verdadeiros negativos (VN) e falsos positivos (FP) (LAROSE; LAROSE, 2014).

Tabela 2.2 – Exemplo de uma matriz de confusão com valores verdadeiros positivos (VP), falsos negativos (FN), verdadeiros negativos (VN) e falsos positivos (FP).

		Predição	
		NÃO	SIM
Classe real	NÃO	VN: Predito NÃO Real NÃO	FP: Predito SIM Real NÃO
	SIM	FN: Predito NÃO Real SIM	VP: Predito SIM Real SIM

A partir dos valores definidos de VP, FN, VN e FP, podem ser definidas as seguintes medidas de avaliação:

$$Taxa\ de\ erro = \frac{FN+FP}{VN+FN+VP+FP} \quad (2.10)$$

$$Taxa\ de\ acerto = 1 - taxa\ de\ erro = \frac{VN+VP}{VN+FN+VP+FP} \quad (2.11)$$

$$Sensitividade = \frac{VP}{VP+FN} \quad (2.12)$$

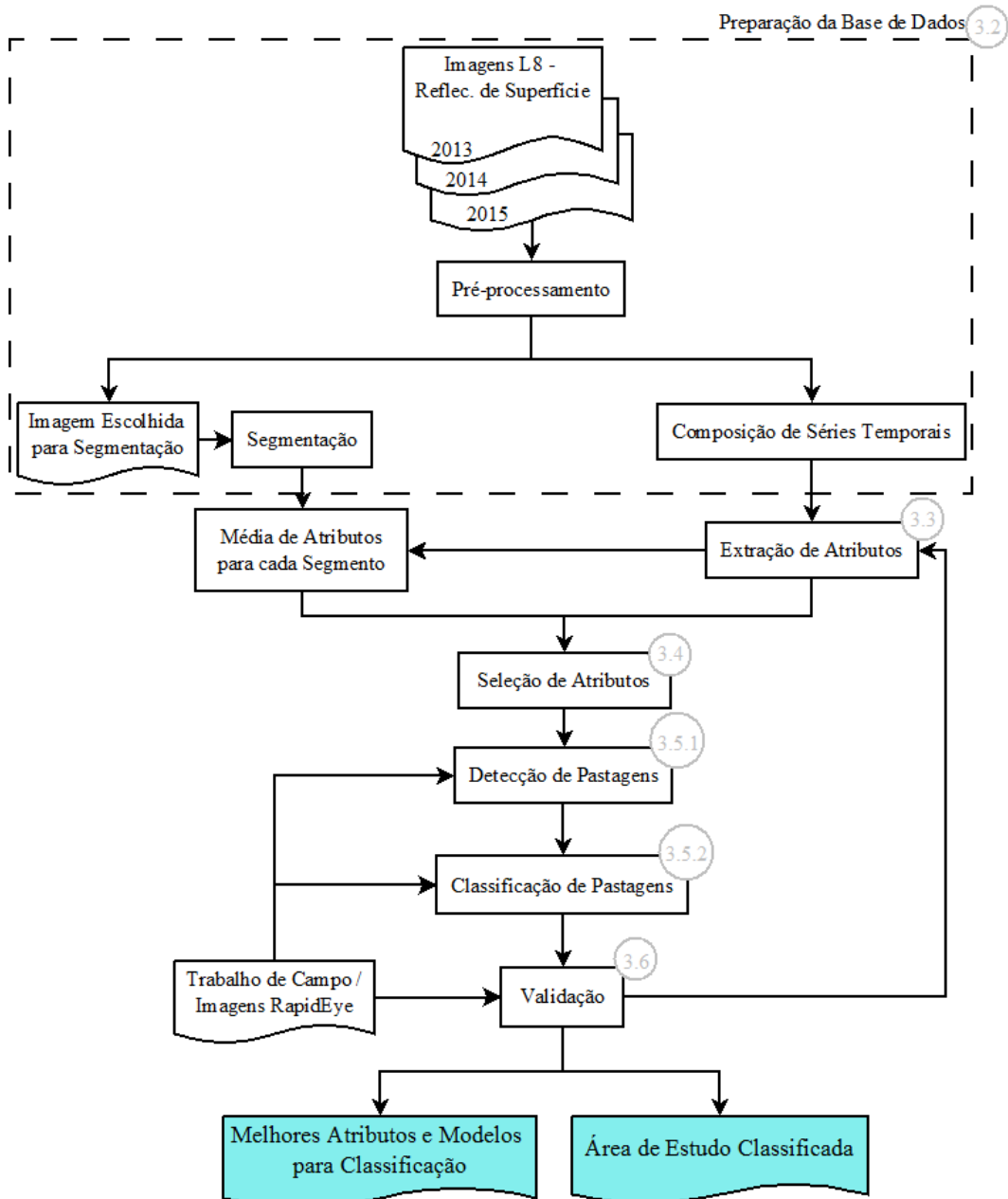
$$Especificidade = \frac{VN}{VN+FP} \quad (2.13)$$

Um bom modelo de classificação precisa ser sensível (Equação 2.12), ou seja, deve identificar uma alta proporção de registros verdadeiros (positivos) quando o registro for realmente verdadeiro. Além disso, um bom modelo deve ser específico (Equação 2.13), ou seja, deve ser capaz de identificar uma alta taxa de registros falsos (negativos) quando estes realmente o forem. Quanto maiores forem os valores dessas medidas, melhor será a classificação. A utilização dessas medidas permite a comparação de diferentes modelos criados por diferentes algoritmos e diferentes abordagens (baseada em *pixels* ou em objetos) (LAROSE; LAROSE, 2014).

3 MATERIAIS E MÉTODOS

A Figura 3.1 apresenta o fluxograma proposto para a metodologia deste trabalho. Cada etapa será detalhada no decorrer deste capítulo.

Figura 3.1 – Fluxograma descrevendo a metodologia do trabalho.



Os números inseridos nos círculos representam a seção correspondente à etapa do fluxograma.

3.1. Área de Estudo

As áreas de estudo (Figura 3.2) escolhidas para o trabalho consistem em três órbitas/ponto do satélite Landsat 8 distribuídas em quatro estados da Amazônia:

Mato Grosso, Rondônia, Amazonas e Acre. Procurou-se escolher cenas espacialmente bem distribuídas, com diferentes padrões de pastagem (Figura 3.3) e abrangendo todas as classes de interesse (as quais serão detalhadas na Seção 3.5.2). Cada cena abrange áreas de diversos municípios pertencentes, em alguns casos, a mais de um estado (Tabela 3.1).

Figura 3.2 – Localização da área de estudo.

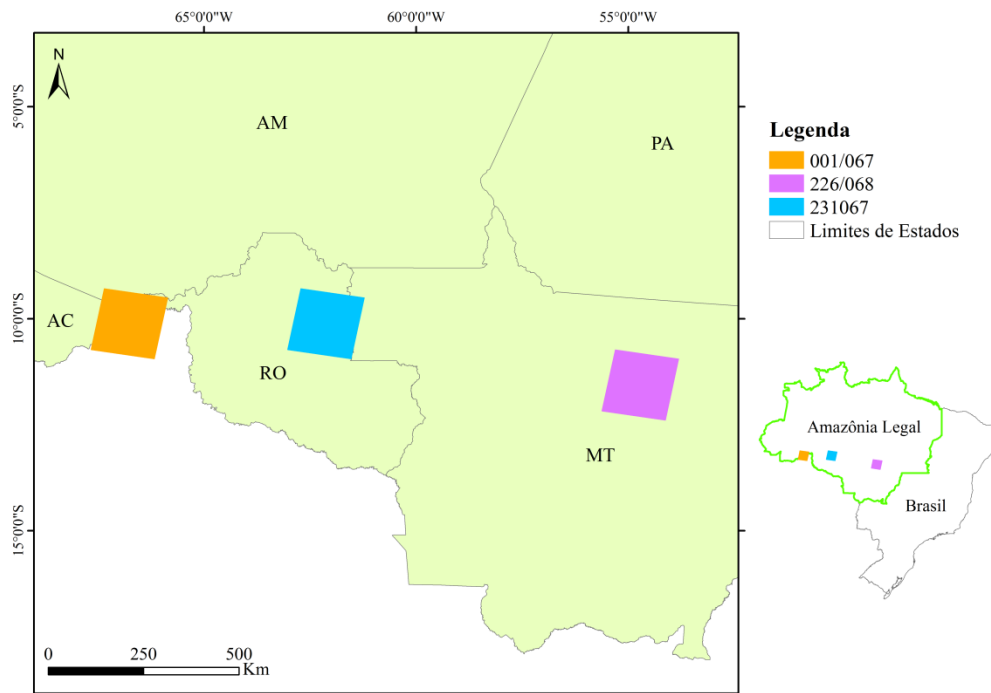


Figura 3.3 – Diferentes padrões espaciais de pastagens mapeados pelo TerraClass em imagens do sensor OLI do Landsat 8. À esquerda, um recorte da cena 226/068, no Mato Grosso. À direita, um recorte da cena 231/067, em Rondônia.

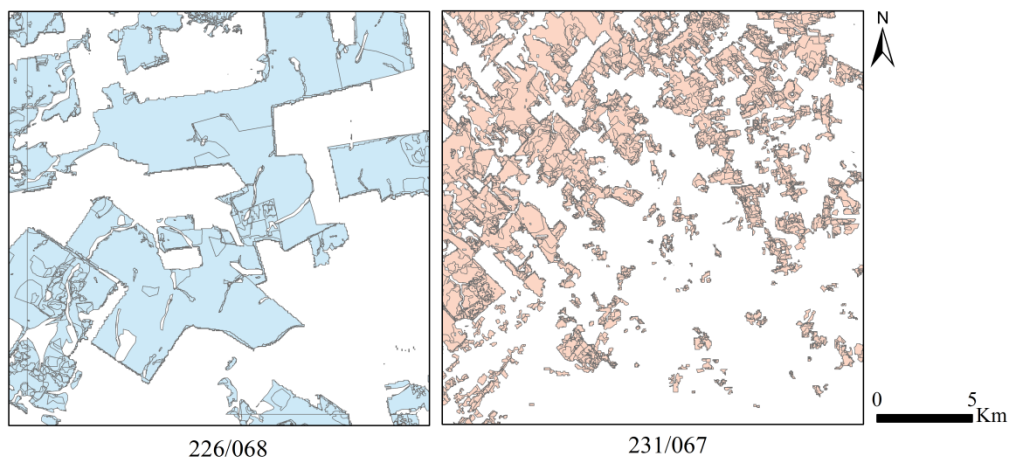


Tabela 3.1 - Órbitas/ponto e municípios e estados com áreas compreendidas nas respectivas cenas.

Órbita/Ponto	Municípios	Estado
226/68	Marcelândia, Nova Santa Helena, Colíder, Itaúba, Cláudia, Sinop, Vera, Santa Carmem, Feliz Natal, Sorriso, União do Sul e Nova Ubiratá.	Mato Grosso
231/67	Ji-Paraná, Colniza, Rondolândia, Vale do Anari, Machadinho D'Oeste, Cujubim, Rio Crespo, Ariquemes, Theobroma, Vale do Paraíso, Ouro Preto do Oeste, Jaru, Governador Jorge Teixeira e Cacaulândia.	Rondônia
001/067	Porto Velho.	Rondônia
	Lábrea, Boca do Acre.	Amazonas
	Rio Branco, Porto Acre, Senador Guimard, Plácido de Castro, Acrelândia.	Acre

A Amazônia possui clima equatorial quente e úmido. A temperatura média da região está entre 24°C e 26°C, com uma amplitude térmica sazonal de 1 a 2°C e umidade relativa em torno de 80%. A precipitação média da Amazônia é de 2300 mm/ano, entretanto existem algumas áreas em que o total anual pode chegar a 3500 mm/ano (FISCH *et al.*, 2014).

O clima, no geral, possui dois períodos bem definidos: seco (maio a setembro) e úmido (novembro a março), ficando os meses de abril e outubro como transição entre os períodos. Porém, os comportamentos da temperatura e da precipitação na Amazônia variam espacial e temporalmente. As temperaturas apresentam pouca variação ao longo do ano, mas regiões meridionais da Amazônia (Acre, Rondônia e Mato Grosso) podem receber sistemas frontais, denominados localmente de “friagens”, capazes de reduzir a temperatura em até 8°C (FISCH *et al.*, 2014). As regiões de fronteira entre Brasil, Colômbia e Venezuela, o litoral do Pará ao Amapá e a região central da Amazônia (próximo à latitude 5°S) possuem alta atividade convectiva e não possuem período de seca definido (FIGUEROA; NOBRE, 1990).

A vegetação da Amazônia é composta predominantemente por floresta ombrófila densa, típica floresta tropical com alto grau de umidade. Pode ser dividida em matas de planícies de inundação (mata de várzea e mata de igapó) e matas de terra firme. As matas de igapó estão presentes em locais constantemente inundados, enquanto as matas de várzea ocorrem em áreas de inundação intermitente. As matas de terra firme, onde é comum a instalação de pastagens, possuem maior

diversidade e complexidade de espécies. Além da floresta ombrófila densa, existem outras seis formações de menor extensão na Amazônia: campinaranas, florestas estacionais decíduais ou semidecíduais, florestas ombrófilas abertas, formações pioneiras com influência fluvial e/ou marinha, refúgios montanos e savanas amazônicas (IBGE, 2012).

3.2. Preparação da Base de Dados

Neste trabalho, foram utilizadas imagens do sensor OLI do satélite Landsat 8 convertidas para reflectância de superfície, disponibilizadas para *download* no *website* <http://earthexplorer.usgs.gov/>. As imagens estudadas correspondem ao período entre abril de 2013 e dezembro de 2015. Aquelas que possuíam cobertura de nuvem superior a 30% foram excluídas. Foram usadas as bandas 2 a 7 das nove bandas disponíveis para cada imagem do sensor OLI (Tabela 3.2). De acordo com esses critérios, 28, 26 e 25 imagens compuseram as séries temporais das cenas 001/067, 226/068 e 231/067, respectivamente.

Tabela 3.2 – Informações sobre o sensor OLI do satélite Landsat 8.

Sensor OLI – Landsat 8 – Bandas	Comprimento de Onda (μm)	Resolução espacial (m)
Banda 1 – <i>Coastal aerosol</i>	0,43 – 0,45	30
Banda 2 – <i>Blue</i>	0,45 – 0,51	30
Banda 3 – <i>Green</i>	0,53 – 0,59	30
Banda 4 – <i>Red</i>	0,64 – 0,67	30
Banda 5 – NIR*	0,85 – 0,88	30
Banda 6 – SWIR* 1	1,57 – 1,65	30
Banda 7 – SWIR* 2	2,11 – 2,29	30
Banda 8 – <i>Panchromatic</i>	0,50 – 0,68	15
Banda 9 - <i>Cirrus</i>	1,36 – 1,38	30

*NIR = *Near Infrared*; SWIR = *Short-Wave Infrared*.

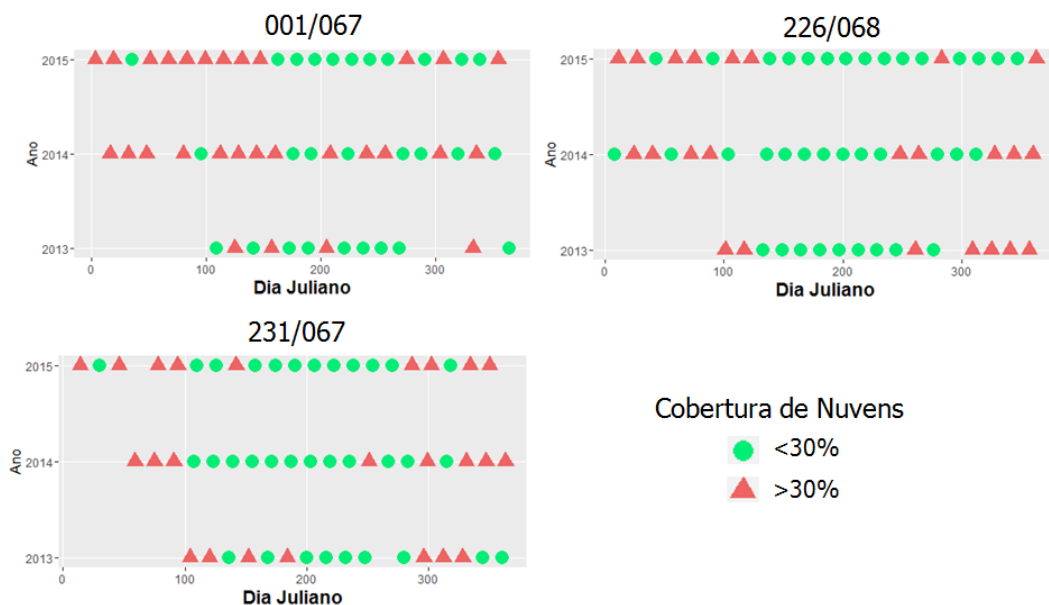
Fonte: <http://landsat.usgs.gov/>

O satélite Landsat 8 possui um período de revisita (resolução temporal) de 16 dias. Entretanto, como qualquer outro sensor óptico, possui interferência de nuvens e sombras de nuvens em suas imagens, comuns principalmente em regiões tropicais úmidas, como é o caso da Amazônia (ASNER, 2001). Na Figura 3.4, pode ser observada a disponibilidade de imagens para as cenas e período estudado, de acordo com o limiar de cobertura de nuvens adotado. A grande

quantidade de imagens com cobertura de nuvens superior a 30% (representadas por triângulos na Figura 3.4) dificulta a construção de séries temporais com intervalos igualmente espaçados no tempo.

Para impedir a influência de nuvens e sombras de nuvens nos padrões que foram utilizados para a classificação, o pré-processamento das imagens incluiu a detecção desses alvos por meio do algoritmo Fmask. Esse algoritmo utiliza informações espectrais e texturais para a detecção de nuvens e sombras de nuvens com bases probabilísticas (ZHU *et al.*, 2015). As regiões identificadas foram excluídas das imagens e, conseqüentemente, das séries temporais.

Figura 3.4 – Disponibilidade de imagens do sensor OLI para as cenas e período estudado. Cada círculo ou triângulo representa a existência de uma imagem e a cor dos símbolos está associada à porcentagem de cobertura de nuvens.



O software R (IHAKA; GENTLEMAN, 1996) foi usado para a composição das séries temporais. Duas abordagens de análise de imagens foram testadas no trabalho: por *pixel* e baseada em objetos. Na primeira, as séries temporais foram compostas para cada *pixel*, utilizando as várias imagens das datas disponíveis, semelhante à Figura 2.8.

No caso da abordagem baseada em objetos, para cada cena de estudo, uma imagem da série temporal foi escolhida como entrada para o algoritmo de segmentação (Tabela 3.3). A segmentação implica na partição da imagem em regiões homogêneas, de forma que características espaciais possam ser incluídas

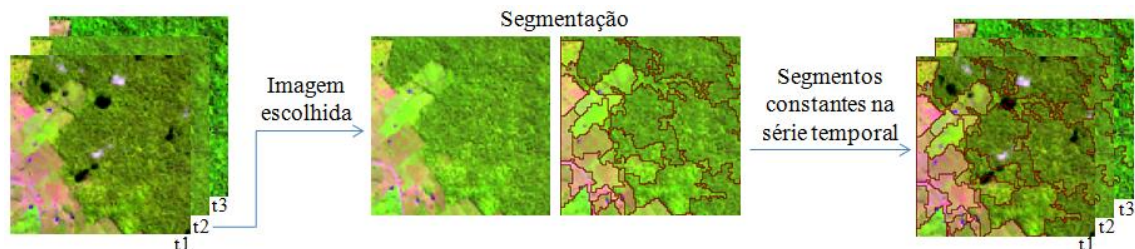
na análise a ser feita (HAY; CASTILLA, 2006). O método utilizado foi a segmentação multirresolução, baseado no método de crescimento de regiões, implementado no *software* eCognition Developer 9.0 e baseado em três parâmetros: escala, forma e compacidade (BAATZ; SCHAPE, 2000). Os parâmetros da segmentação (Tabela 3.3) foram definidos de maneira empírica.

Tabela 3.3 – Informações das imagens do sensor OLI utilizadas na segmentação.

Cena	Data da Imagem	Parâmetros da Segmentação		
		Escala	Forma	Compacidade
001/067	11/jul/14	60	0,5	0,8
226/068	22/jul/2015	60	0,5	0,8
231/067	10/out/2014	60	0,5	0,8

A escolha das imagens foi baseada na proximidade ao tempo médio da série temporal, baixa cobertura de nuvens e proximidade às datas das imagens RapidEye da validação (Seção 3.6). Para cada cena, os segmentos resultantes permaneceram constantes ao longo da série temporal (Figura 3.5).

Figura 3.5 – Exemplo de utilização de segmentos constantes na série temporal, feito a partir de recorte da cena 001/067.



3.3. Extração de Atributos

Nesta fase, vários atributos foram extraídos: índices de vegetação, frações do Modelo Linear de Mistura Espectral, componentes resultantes da transformação *tasselled cap*, atributos espectrais e atributos de textura.

A Tabela 3.4 apresenta uma descrição matemática dos métodos de índices de vegetação.

Tabela 3.4 – Informações sobre os índices de vegetação.

Índice de Vegetação	Fórmula*	Autor
NDVI (<i>Normalized Difference Vegetation Index</i>)	$NDVI = \frac{\rho_{NIR} - \rho_V}{\rho_{NIR} + \rho_V}$	Rouse Jr. <i>et al.</i> (1973)
EVI (<i>Enhanced Vegetation Index</i>)	$EVI = G \cdot \frac{(\rho_{NIR} - \rho_V)}{(\rho_{NIR} + C_1 \cdot \rho_V - C_2 \cdot \rho_A + L)}$	Huete <i>et al.</i> (1997)
EVI2 (<i>Two Band Enhanced Vegetation Index</i>)	$EVI2 = G \cdot \frac{(\rho_{NIR} - \rho_V)}{(\rho_{NIR} + 2,4 \cdot \rho_V + 1)}$	Jiang <i>et al.</i> (2008)
SAVI (<i>Soil-Adjusted Vegetation Index</i>)	$SAVI = \left(\frac{\rho_{NIR} - \rho_V}{\rho_{NIR} + \rho_V + L_1} \right) * (1 + L_1)$	Huete (1988)
NDII (<i>Normalized Difference Infrared Index</i>)	$NDII = \frac{\rho_{NIR} - \rho_{SWIR}}{\rho_{NIR} + \rho_{SWIR}}$	Hardisky <i>et al.</i> (1983)

* ρ_{NIR} = reflectância no infravermelho próximo; ρ_{SWIR} = reflectância no infravermelho de ondas curtas; ρ_V = reflectância no vermelho; ρ_A = reflectância no azul; L = Ajuste da linha do solo (1); G = fator de ganho (2,5); C1 e C2 = coeficientes de resistência de aerossóis (6 e 7,5); L_1 = fator de correção dependente da cobertura vegetal.

Para cada imagem, as proporções dos principais componentes existentes na mistura de cada *pixel* foram estimadas. Para isso, foi aplicado o Modelo Linear de Mistura Espectral – MLME (SHIMABUKURO; SMITH, 1991). O MLME foi gerado no *software* ENVI 4.8, utilizando espectros de referência e as bandas de cada imagem para formar um sistema de equações (Equação 2.1 da Seção 2.4.2).

Para a criação dos espectros de referência, a escolha dos *endmembers*, *pixels* “puros” que representem as frações desejadas, foi feita em cada imagem analisada. Após a aplicação do MLME, quatro imagens sintéticas, para cada cena e cada data, foram geradas: fração vegetação, fração NPV, fração solo e fração água/sombra (Figura 3.6). Na figura 3.7, podem-se ver os espectros de cada fração gerados pela média dos *endmembers* das cenas.

Figura 3.6 – Composição colorida R(6)G(5)B(4) e frações do MLME de recorte da imagem 231/067 referente à 15/10/16.

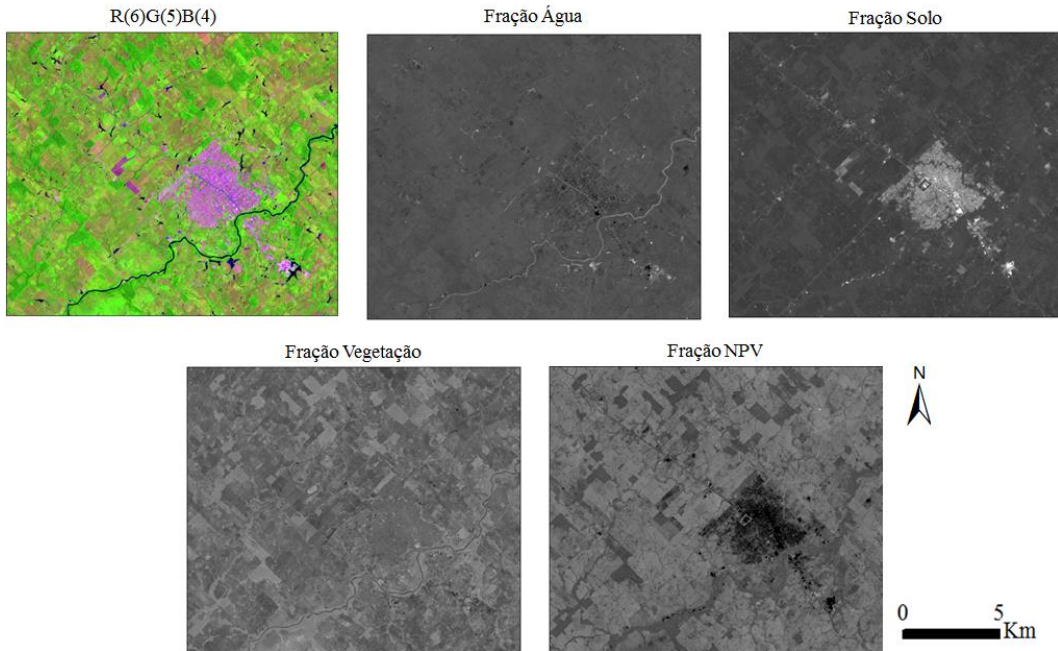
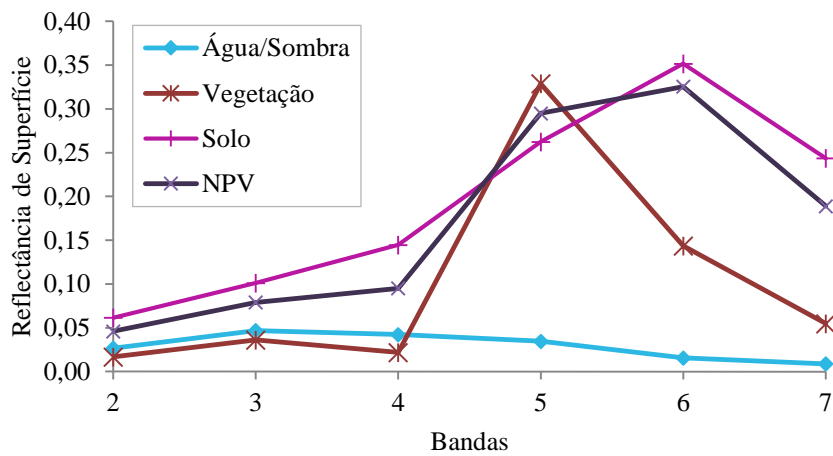


Figura 3.7 – Espectros dos *endmembers* das quatro frações do Modelo Linear de Mistura Espectral.



A transformação *tasseled cap* foi realizada para agregar as informações das seis bandas do sensor OLI em componentes que se relacionam fortemente com as características físicas das imagens. A transformação foi executada de acordo com os coeficientes apresentados na Tabela 2.1 (Seção 2.4.3). As componentes resultantes foram *Greenness*, *Brightness* e *Wetness*.

Os cálculos dos valores de IVs, das frações do MLME e dos componentes da transformação *tasselled cap* foram realizados para todos os dados disponíveis nas datas definidas neste estudo, gerando assim séries temporais.

A extração dos atributos foi feita primeiramente para a abordagem baseada na classificação por *pixels*, na qual cada *pixel* possui um valor correspondente para um determinado atributo. Para a abordagem baseada em objetos, os atributos de cada segmento foram representados pela média dos valores do atributo para todos os *pixels* contidos naquele segmento. Alguns atributos espectrais e os atributos de textura foram utilizados somente na abordagem baseada em objetos (Tabela 3.5).

Tabela 3.5 – Alguns atributos espectrais e texturais para a abordagem baseada em objetos.

Atributo	Descrição	Tipo
Brilho	Média da soma das médias das bandas	Espectral
Média – Banda	Média do valor de reflectância da banda	Espectral
Desvio Padrão	Desvio padrão das médias das bandas	Espectral
Máxima Diferença	Amplitude média das bandas	Espectral
GLCM <i>Entropy</i>	Entropia da matriz de níveis de cinza	Textural
GLCM <i>Standard Deviation</i>	Desvio padrão da matriz de níveis de cinza	Textural
GLCM <i>Dissimilarity</i>	Dissimilaridade da matriz de níveis de cinza	Textural
GLCM <i>Ang. 2nd Mom.</i>	Segundo momento angular da matriz de níveis de cinza	Textural
GLCM <i>Contrast</i>	Contraste da matriz de níveis de cinza	Textural
GLCM <i>Homogeneity</i>	Homogeneidade da matriz de níveis de cinza	Textural
GLCM <i>Correlation</i>	Correlação da matriz de níveis de cinza	Textural

3.4. Seleção de Atributos

Após a extração dos atributos, a seleção de atributos foi aplicada para reduzir o conjunto de dados e selecionar os atributos mais relevantes para a detecção e classificação de pastagens. Esta etapa foi realizada no *software* Weka (HALL *et al.*, 2009). O método adotado foi a Seleção Baseada em Correlação (CFS – *Correlation Feature Selection*).

O método CFS baseia-se na correlação dos atributos com as classes e é independente do algoritmo de classificação. Além de considerar as altas correlações dos atributos com as classes, também considera relevante a não

correlação entre os atributos. O CFS não gera um *ranking* de atributos, mas escolhe um conjunto S (*Best First*) de k atributos com maior mérito (equação 3.1). O numerador da equação indica o poder preditivo do conjunto, enquanto o denominador indica o grau de redundância existente entre os atributos.

$$\text{Mérito}(S) = \frac{k \cdot \overline{r_{ac}}}{\sqrt{k + k(k-1)\overline{r_{aa}}}} \quad (3.1)$$

Onde:

k = número de atributos;

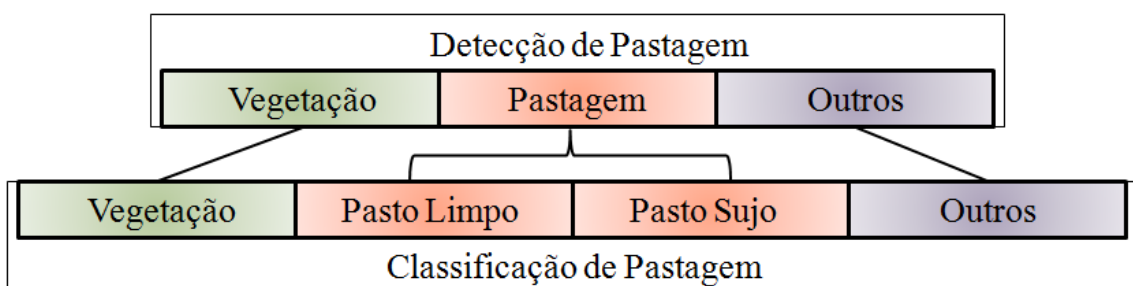
$\overline{r_{ac}}$ = média da correlação entre atributo-classe;

$\overline{r_{aa}}$ = média da correlação entre atributo-atributo.

3.5. Classificação

A fase de classificação foi subdividida em duas etapas hierárquicas: a detecção de pastagens e, posteriormente, a classificação das pastagens em Pasto Limpo e Pasto Sujo (Figura 3.8), conforme será descrito nas seções seguintes. A divisão da classificação em etapas segue a estratégia “dividir para conquistar”, bem conhecida em diversas áreas do conhecimento. A estratégia consiste em dividir um problema maior em subproblemas menos complexos. Depois de resolvidos os subproblemas, as soluções são combinadas para solucionar o problema maior (DASGUTPA *et al.*, 2006).

Figura 3.8 – Classificação hierárquica de pastagens.



Dois experimentos foram realizados. O primeiro consistiu na investigação da melhor abordagem para a classificação. Foram testadas as abordagens por *pixel* e baseada em objetos, com e sem seleção de atributos, apenas para a primeira imagem (001/067), utilizando três algoritmos diferentes (Seção 3.5.1). O segundo experimento baseou-se na avaliação da melhor abordagem nas três cenas, ou seja,

a abordagem que obteve o melhor desempenho no primeiro experimento foi avaliada na cena em que foi gerada e também nas outras cenas.

3.5.1. Detecção de Pastagens

A detecção de pastagens foi a primeira etapa de classificação, na qual a classe de Pastagem foi diferenciada de outros alvos (áreas agrícolas, áreas urbanas, mineração, vegetação secundária, etc.), agregados em duas classes: Vegetação e Outros. A classe Vegetação contém feições de floresta e de vegetação secundária e a classe Outros abrange o restante, inclusive áreas agrícolas. Foram utilizados algoritmos supervisionados de classificação, nos quais as amostras das classes de interesse são fornecidas. As amostras para o treinamento dos modelos foram coletadas dos dados de referência gerados em imagens RapidEye (detalhados na Seção 3.6). Foram coletadas 500 amostras aleatórias (200 de Vegetação, 200 de Pastagem e 100 de Outros) em cada uma das três cenas de estudo.

Os algoritmos de classificação usados no sistema Weka, foram os seguintes: árvores de decisão (J48), RF e redes neurais artificiais (*Multilayer Perceptron - MLP*). Os parâmetros utilizados nos algoritmos foram escolhidos empiricamente: número mínimo de objetos por folhas = 12, número de árvores = 100 e número de camadas = 50 para o J48, o RF e o MLP, respectivamente.

3.5.2. Classificação de Pastagens

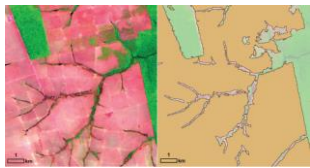

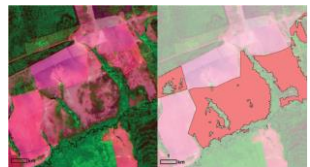

Os *pixels* e objetos identificados como pastagem na etapa anterior passaram a ser o conjunto de dados de entrada para a classificação das pastagens em duas classes (Tabela 3.6): Pasto Limpo e Pasto Sujo. A escolha das classes foi baseada na legenda adotada pelo projeto TerraClass (ALMEIDA *et al.*, 2016) cujas descrições foram detalhadas na Seção 2.2. Das duas outras classes que abrangem áreas de pastagem no TerraClass (Regeneração com Pasto e Pasto com Solo Exposto), a primeira foi incluída na classe de pasto sujo e a outra, como possui área minoritária, não foi analisada neste trabalho. O Pasto Sujo não é necessariamente um pasto degradado, pois está mais associado com a regeneração em áreas de pastagem. A regeneração, como um problema de manejo, pode estar associada à degradação, mas também ao abandono ou especulação fundiária. As duas classes identificadas por Rufin *et al.* (2015) em pastagens também foram

consideradas: *woody pasture* (semelhante ao Pasto Sujo) e *grass-dominated pasture* (semelhante ao Pasto Limpo).

As amostras utilizadas no treinamento dos modelos de classificação foram extraídas do mesmo conjunto de dados de referência utilizados para a etapa de detecção. Esta etapa foi realizada no Weka e foram utilizados os mesmos classificadores descritos na Seção 3.5.1 (árvores de decisão, RF e redes neurais).

Esta segunda etapa de classificação foi feita, em cada cena, para apenas um momento da série temporal, usando as mesmas datas da segmentação (Tabela 3.3), pois existe uma considerável variação da condição da pastagem ao longo de um ano, conforme observado por Rufin *et al.* 2015.

Tabela 3.6 – Tipologia das classes de pastagem.

Classe	Tipologia – Landsat*	Tipologia - Fotografia	Padrão
Pasto Limpo			Cor: magenta; Textura: lisa; Correspondem às áreas (quase) totalmente cobertas por gramíneas. Baixa infestação de invasoras herbáceas e arbustivas e sem a presença de arbóreas.
Pasto Sujo			Cor: magenta e verde claro; Textura: lisa ou levemente rugosa; Correspondem às pastagens com grande concentração de espécies invasoras, arbustos e árvores.

*Utilizou-se a composição falsa-cor R(SWIR)G(NIR)B(Vermelho). Fonte: Adaptada de Coutinho *et al.* (2013).

Além do método de detecção (3 classes) seguido da classificação (2 classes), na abordagem baseada em objetos também foi realizada uma classificação em uma única etapa. Essa etapa foi responsável pela separação das duas classes de interesse (Pasto Limpo e Pasto Sujo) das demais classes, resultando em uma classificação com 5 classes: *Pasto Limpo*, *Pasto Sujo*, *Floresta*, *Vegetação Secundária* e *Outros*.

3.6. Dados de Referência e Validação

Dois tipos de dados de referência foram utilizados: trabalho de campo e interpretação visual de imagem de alta resolução espacial. A interpretação visual foi realizada em imagens do satélite RapidEye, de 5 m de resolução espacial, 12 *bits* de resolução radiométrica e *swath* de 77 km. As imagens contém 5 bandas: 3 nas faixas espectrais do visível; uma *Red Edge* de 690 a 730 nm, utilizada para avaliar a atividade fotossintética da vegetação; e uma no infravermelho próximo. As imagens foram obtidas no geocatálogo do Ministério do Meio Ambiente (MMA) (<http://geocatalogo.mma.gov.br/>) e correspondem às seguintes datas:

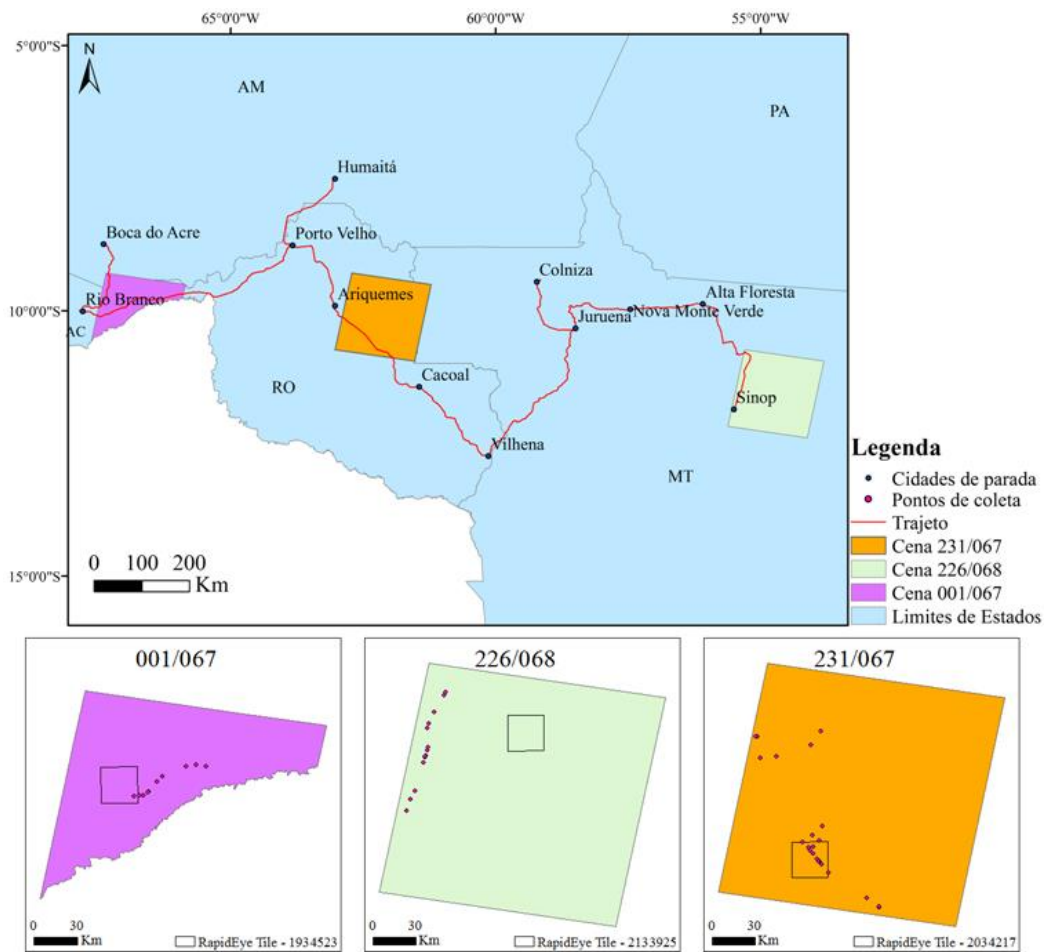
Tabela 3.7 – Informações das imagens RapidEye utilizadas na validação.

Cena correspondente	RapidEye - Tile	Data
001/067	1934523	11/jul/2014
226/068	2133925	16/jul/2015
231/067	2034217	09/Ago/2014

As classificações visuais foram comparadas com as classificações geradas semi-automaticamente, resultando em taxas de acerto global e matrizes de confusão. Além disso, foram gerados mil pontos aleatórios em cada classificação e novas taxas de acerto globais foram geradas.

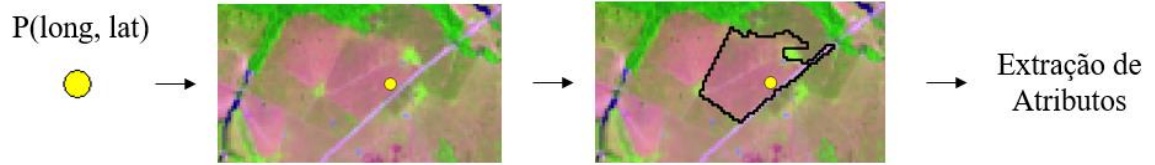
O trabalho de campo foi feito em colaboração com o Centro Regional da Amazônia (INPE/CRA), no âmbito do projeto TerraClass, no período entre 16 e 30 de outubro de 2016, realizando o trajeto destacado na Figura 3.9. Nos 51 pontos de coleta indicados, as classes a que pertenciam o ponto foram identificadas, as coordenadas foram coletadas e fotos e vídeos das áreas foram registrados.

Figura 3.9 – Trajeto do trabalho de campo e pontos de coleta.



Para cada ponto, identificou-se o segmento correspondente para extração dos atributos necessários (Figura 3.10). Como o trabalho de campo foi realizado em 2016 e os dados do trabalho correspondem ao período de 2013 a 2015, algumas modificações foram necessárias para a validação. O modelo que obteve melhor desempenho nos experimentos de classificação foi adaptado para o triênio que inclui 2016 (2014, 2015 e 2016) e os atributos necessários foram extraídos para esse último ano. Por exemplo, se o modelo originalmente utiliza o atributo NDVI de julho de 2013, então, para o dado de campo foi usado o NDVI de julho de 2014 e assim por diante.

Figura 3.10 – Identificação dos segmentos dos pontos de campo para extração de atributos.



4 RESULTADOS E DISCUSSÃO

Este capítulo apresenta os resultados dos três experimentos realizados neste trabalho. O primeiro experimento consiste na investigação da melhor abordagem de classificação (por *pixel* ou por segmento e utilizando três ou cinco classes), utilizando apenas uma cena (001/067). O segundo experimento utiliza a abordagem de melhor desempenho para realizar a classificação nas demais cenas de estudo (226/068 e 231/067). Por fim, o terceiro experimento valida os modelos de classificação com pontos aleatórios e, além disso, os aplica nos dados coletados em trabalho de campo.

4.1. Abordagens de Classificação

A Tabela 4.1 apresenta as taxas de acerto na construção dos modelos em três casos de classificação. Os dados utilizados foram as 500 amostras aleatórias obtidas na classificação de referência, 66% delas foram usadas na construção do modelo e os outros 33% restantes, no teste do modelo, resultando nos dados da Tabela 4.1. O Caso 1 consiste na classificação por *pixel* para discriminar Vegetação, Pastagem e Outros (o que chamamos de *3 Classes*), seguida da diferenciação da Pastagem em Pasto Limpo e Pasto Sujo (*Tipos de Pastagem*). O Caso 2 consiste na classificação baseada em objetos para discriminar as mesmas três classes, seguida da diferenciação das pastagens. Finalmente, o Caso 3 classifica a área de estudo nas classes Floresta, Vegetação Secundária, Pasto Limpo, Pasto Sujo e Outros em uma só etapa (*5 Classes*), utilizando segmentos.

Tabela 4.1 – Taxas de acerto (%) na construção dos modelos.

		Classificador		
		J48	RF	MLP (Rede Neural)
(Caso 1) <i>Pixels</i>	3 Classes	82,94	83,53	82,94
	Tipos de Pastagem	72,44	75	78,21
(Caso 2) <i>Objetos</i>	3 Classes	84,24	85,2	83,99
	Tipos de Pastagem	78,25	79,15	76,46
(Caso 3) <i>Objetos</i>	5 Classes	65,1	68,59	65,46

Para os Casos 1 e 2, os três classificadores obtiveram os seguintes desempenhos: na detecção de Pastagem, os classificadores J48, RF e MLP apresentaram desempenho médio de 83,59%, 84,36% e 83,46%, respectivamente. Já na etapa de

classificação de Pasto Limpo e Pasto Sujo, os desempenhos médios foram de 75,34%, 77,07% e 77,33%, respectivamente. Houve uma pequena superioridade da taxa de acerto na classificação pelo algoritmo RF (1,83% maior, em média, considerando todos os casos). Entretanto, o J48 gera uma única árvore de decisão para realizar a classificação (QUINLAN, 2014), facilitando a reprodutibilidade do método e a visualização da contribuição de cada atributo. Por esse motivo, as árvores de decisão geradas na construção dos modelos foram implementadas no *software* R para serem aplicadas nas imagens completas, obtendo as taxas de acerto apresentadas na Tabela 4.2. Além disso, para esse estudo, a seleção de atributos não apresentou vantagens, uma vez que os conjuntos obtidos na fase de seleção de atributos obtiveram desempenho inferior aos conjuntos completos.

O Caso 3 obteve as piores taxas de acerto (65,76%, em média), confirmando uma melhora de desempenho quando é possível detectar o padrão das áreas de pastagem e posteriormente separá-las de acordo com outro critério como, por exemplo, a condição da cobertura vegetal. Isso ocorre devido à consideração da relação de dependência entre as classes (só será Pasto Limpo ou Sujo um objeto já classificado como Pastagem). De acordo com Silla Jr. e Freitas (2011), a representação dessa dependência em uma classificação hierárquica (“dividir para conquistar”) é capaz de melhorar o poder de predição dos classificadores.

Tabela 4.2 – Taxas de acerto (%) na validação dos modelos gerados pelo classificador J48.

		Caso 1 (pixels):					
		3 Classes - 81,75%			Tipos de Pastagem - 64,77%		
		Classificado como (%):			Classificado como (%):		
		Vegetação	Pastagem	Outros	Pasto Limpo	Pasto Sujo	
	Vegetação	92,11	16,55	11,52	Pasto Limpo	61,50	12,82
Ref.:	Pastagem	6,51	73,81	45,71	Pasto Sujo	30,75	71,25
	Outros	1,38	9,64	42,77	Outros	7,76	15,93

Caso 2 (objetos):							
3 Classes - 90,09%				Tipos de Pastagem - 76,43%			
Classificado como (%):				Classificado como (%):			
	Vegetação	Pastagem	Outros		Pasto Limpo	Pasto Sujo	
	Vegetação	94,07	5,51	7,12	Pasto Limpo	84,03	18,49
Ref.:	Pastagem	5,26	91,49	21,35	Pasto Sujo	10,32	70,95
	Outros	0,67	3,00	71,54	Outros	5,66	10,56

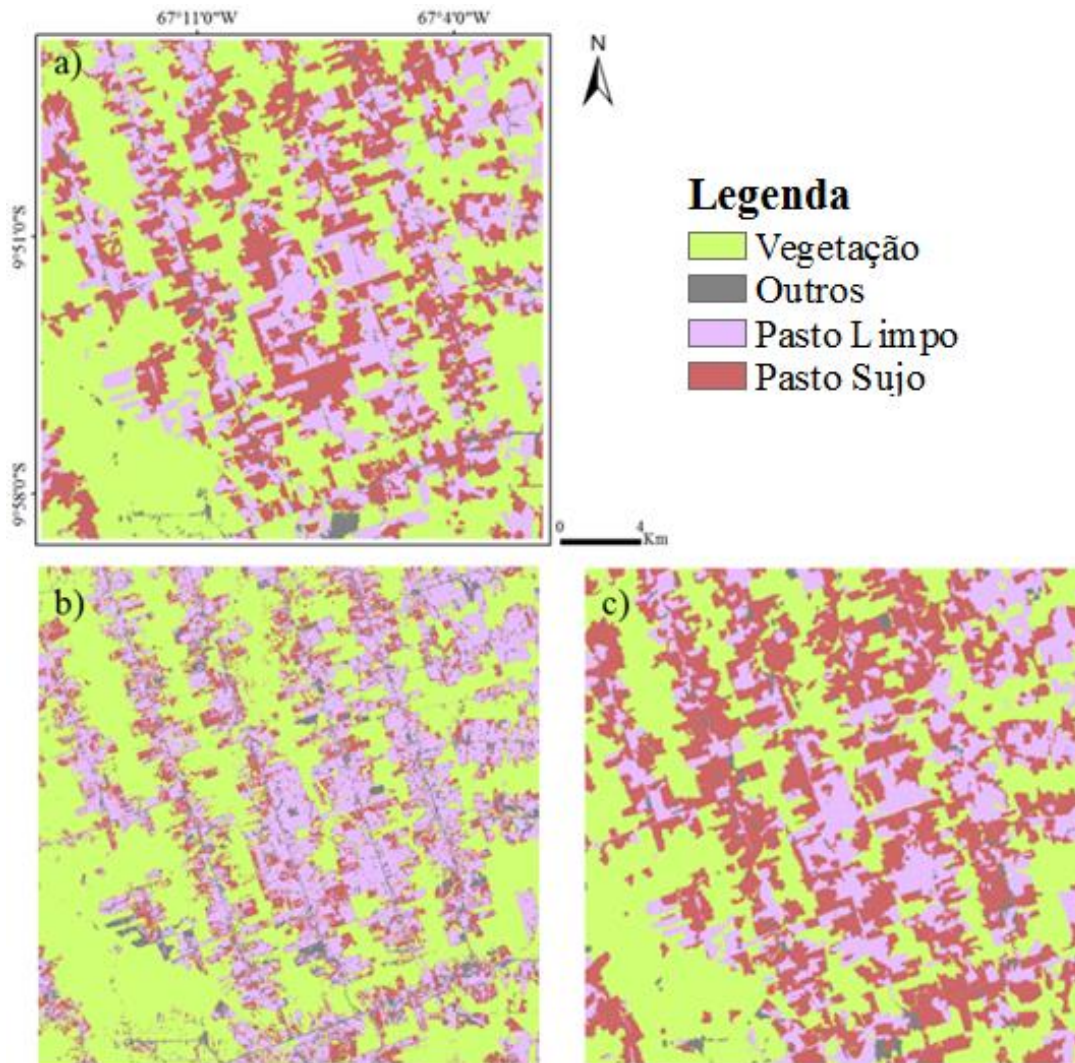
Caso 3 (objetos):						
5 Classes - 32,22%						
Classificado como:						
	Floresta	Pasto Limpo	Pasto Sujo	Vegetação Secundária	Outros	
	Floresta	35,30	20,08	18,84	18,81	12,50
	Pasto Limpo	20,70	35,67	26,44	21,29	30,36
Ref.:	Pasto Sujo	25,88	31,04	32,83	22,28	29,64
	Veg. Sec.	11,28	5,06	11,85	27,72	8,21
	Outros	6,84	8,15	10,03	9,90	19,29

Ainda na Tabela 4.1, observa-se que as taxas de acerto da classificação hierárquica baseada em objetos (Caso 2) foram superiores daquelas por *pixel* (Caso 1). Na validação dos modelos (Tabela 4.2), a superioridade da classificação baseada em objetos para o estudo tornou-se mais explícita. No Caso 1, a taxa de acerto na detecção de pastagem foi de 81,75% e na classificação do Pasto Limpo e Sujo foi de 64,77%, enquanto no Caso 2 as taxas foram de 90,09% e 76,43%, respectivamente. Na Figura 4.1, as espacializações dos dados de referência, nas áreas referentes às cenas RapidEye, e dos casos 1 e 2 de classificação podem ser vistas. Na espacialização do Caso 1 (Figura 4.1b) é possível notar a menor capacidade na diferenciação entre Pasto Limpo e Sujo em comparação com a classificação gerada no Caso 2 (Figura 4.1c).

Pode-se observar que a utilização da segmentação e a inclusão de outros parâmetros espectrais e parâmetros texturais na abordagem baseada em objetos foi fundamental para a melhoria da diferenciação entre Pastagem e Outros e entre Pasto Limpo e Pasto Sujo, principalmente. A melhor identificação dos alvos de Pastagem quando se utiliza a abordagem baseada em objetos também foi observada em Neves *et al.* (2016). De acordo com Seyler (2002), as pastagens são

difíceis de serem identificadas somente por meio de dados de sensoriamento remoto. Devido a grande quantidade de mistura de elementos em sua composição, como grama, árvore, arbustos e outros, as pastagens possuem padrões mais difíceis de serem caracterizados em séries temporais se comparadas com outras classes, como floresta e agricultura, por exemplo.

Figura 4.1 – Espacializações da a) Referência, b) Caso 1 e c) Caso 2.



Nas árvores de decisão aplicadas, todos os cinco IVs, as três frações MLME e os três componentes da TCT aparecem relacionados a alguma imagem da série temporal (ex.: NDVI de julho de 2014, *Wetness* de setembro de 2014, etc.). Isto mostra a importância de cada um dos onze atributos escolhidos na identificação dos padrões de interesse.

4.2. Classificação das cenas de estudo

O Caso 2 da seção anterior foi a melhor abordagem para este trabalho: classificação baseada em objetos para a diferenciação de três classes (Floresta, Pastagem e Outros) e posterior separação da Pastagem em Pasto Limpo e Pasto Sujo. Por isso, esta abordagem foi aplicada na classificação das demais cenas de estudo, utilizando o algoritmo J48.

Um modelo criado com amostras apenas de uma cena pode ser específico demais e não ser capaz de detectar os padrões desejados em outras cenas. O modelo criado e testado na mesma cena 001/067 (Tabela 4.3a), utilizando o J48, obteve altas taxas de acerto global (90,09% para a detecção de Pastagem e 76,43% na classificação de Pasto Limpo e Sujo). Entretanto, as taxas de acerto caíram mais de 20% quando o mesmo modelo foi testado na cena 226/068 (67,84% de acerto na detecção de Pastagem e 47,54% na classificação de Pasto Limpo e Sujo – Tabela 4.3b). As Tabelas 4.2 e 4.3 estão representadas em valores absolutos (quantidade de *pixels* e quantidade de segmentos) nos Apêndices A.1 e A.2, respectivamente.

O contrário também obteve resultados semelhantes: o modelo criado e testado na cena 226/068 obteve taxas de acerto superiores a 80% (90,90% na detecção de Pastagem e 80,03% na classificação do Pasto Limpo e Sujo – Tabela 4.3c), mas quando testado na primeira cena (Tabela 4.3d) sofreu uma queda expressiva: 74,99% na detecção de Pastagem e 43,68% na separação de Pasto Limpo e Sujo.

Para aumentar a capacidade de interpretação dos padrões desejados, um novo modelo foi construído combinando as amostras das duas imagens, 001/067 e 226/068. As taxas de acerto foram ligeiramente menores daquelas dos modelos específicos de uma imagem testada nela mesma (Tabela 4.3e e Tabela 4.3f), porém nesse caso as mesmas árvores de decisão (Figura 4.2 e 4.3) foram capazes de classificar as duas imagens. Para testá-la, a cena 231/067 foi classificada e teve um acerto de 87,19% para a detecção de Pastagem e 52,76% para a diferenciação de Pasto Limpo e Sujo (Tabela 4.3g). Em todas as sete classificações compreendidas na Tabela 4.3, cada segmento classificado corretamente

correspondeu a 1 (um) acerto, independente do seu tamanho, e posteriormente os acertos foram convertidos para valores em porcentagem.

Tabela 4.3 – Taxas de acerto (%) na geração e aplicação dos modelos nas três cenas.

a) Modelo criado/avaliado na 001/067:

3 Classes - 90,09%				Tipos de Pastagem - 76,43%			
Classificado como (%):				Classificado como (%):			
	Vegetação	Pastagem	Outros		Pasto Limpo	Pasto Sujo	
	Vegetação	94,07	5,51	7,12	Pasto Limpo	84,03	18,49
Ref.:	Pastagem	5,26	91,49	21,35	Pasto Sujo	10,32	70,95
	Outros	0,67	3	71,54	Outros	5,66	10,56

b) Modelo criado na 001/067 e avaliado na 226/068:

3 Classes - 67,84%				Tipos de Pastagem - 47,54%			
Classificado como (%):				Classificado como (%):			
	Vegetação	Pastagem	Outros		Pasto Limpo	Pasto Sujo	
	Vegetação	99,11	19,71	25	Pasto Limpo	53,66	3,13
Ref.:	Pastagem	0,3	58,35	40	Pasto Sujo	12,2	13,54
	Outros	0,59	21,94	35	Outros	34,15	83,33

c) Modelo criado/avaliado na 226/068:

3 Classes - 90,90%				Tipos de Pastagem - 80,03%			
Classificado como (%):				Classificado como (%):			
	Vegetação	Pastagem	Outros		Pasto Limpo	Pasto Sujo	
	Vegetação	94,19	4,91	4,55	Pasto Limpo	82,89	15,09
Ref.:	Pastagem	2,74	87,99	5,45	Pasto Sujo	6,39	66,98
	Outros	3,08	7,11	90	Outros	10,72	17,92

d) Modelo criado na 226/068 e avaliado na 001/067:

3 Classes - 74,99%				Tipos de Pastagem - 43,68%			
Classificado como (%):				Classificado como (%):			
	Vegetação	Pastagem	Outros		Pasto Limpo	Pasto Sujo	
	Vegetação	68,95	6,36	22,29	Pasto Limpo	46,06	43,38
Ref.:	Pastagem	20,43	86,5	53,71	Pasto Sujo	42,54	39,04
	Outros	10,62	7,14	24	Outros	11,4	17,58

e) Modelo criado na combinação (001/067 e 226/068) e avaliado na 001/067:

3 Classes - 87,76%				Tipos de Pastagem - 73,15%			
Classificado como (%):				Classificado como (%):			
	Vegetação	Pastagem	Outros		Pasto Limpo	Pasto Sujo	
	Vegetação	90,93	6,71	2,95	Pasto Limpo	76,32	16,43
Ref.:	Pastagem	7,75	89	27	Ref.: Pasto Sujo	15,2	70,3
	Outros	1,31	4,29	70,04	Outros	8,48	13,27

f) Modelo criado na combinação (001/067 e 226/068) e avaliado na 226/068:

3 Classes - 86,25%				Tipos de Pastagem - 72,08%			
Classificado como (%):				Classificado como (%):			
	Vegetação	Pastagem	Outros		Pasto Limpo	Pasto Sujo	
	Vegetação	93,24	5,58	8,33	Pasto Limpo	81,29	25,35
Ref.:	Pastagem	4,33	82,91	14,47	Ref.: Pasto Sujo	6,24	42,96
	Outros	2,43	11,51	77,19	Outros	12,47	31,69

g) Modelo criado na combinação (001/067 e 226/068) e avaliado na 231/067:

3 Classes - 87,19%				Tipos de Pastagem - 52,76%			
Classificado como (%):				Classificado como (%):			
	Vegetação	Pastagem	Outros		Pasto Limpo	Pasto Sujo	
	Vegetação	76,68	4,29	3,41	Pasto Limpo	53,03	8,33
Ref.:	Pastagem	21,93	91,21	19,51	Ref.: Pasto Sujo	39,18	41,67
	Outros	1,38	4,49	77,07	Outros	7,79	50

As Figuras 4.2 e 4.3 mostram as árvores de decisão do modelo gerado com amostras da combinação das cenas 001/067 e 226/068. Em ambas, todos os atributos selecionados automaticamente pelo J48 foram referentes ao período seco. Neste período, as classes de estudo diferenciam-se melhor. Em pastagens tropicais, a sazonalidade é realçada devido à diferença significativa nos regimes de chuva dos períodos seco e chuvoso (AGUIAR, 2010). Os maiores valores de NDVI no período seco para o Pasto Sujo (Figura 4.3) em relação ao pasto limpo podem ser explicados pela capacidade das plantas invasoras em permanecerem verdes devido à sua maior capacidade de captar água mesmo na época seca (AGUIAR, 2013).

Os atributos espectrais ocuparam as posições superiores na árvore (Figuras 4.2 e 4.3), demonstrando maior capacidade de classificação, uma vez que influenciam uma maior quantidade de amostras (NEVILLE, 1999). Entretanto, o atributo de

textura *GLCM Homogeneity* teve um papel importante. Mais de 50% das amostras de pastagem na geração do modelo seguiram pelo ramo indicando valores maiores de *GLCM Homogeneity*. A homogeneidade da matriz GLCM está inversamente relacionada com as diferenças de contraste existentes dentro de cada polígono. Os valores menores desse atributo para a classe Outros estão relacionados com a presença de polígonos de área urbana, que possuem um padrão de textura mais rugosa do que as Pastagens.

Além de índices de vegetação, a separação entre Pasto Limpo e Sujo (Figura 4.3) baseou-se na média e desvio padrão da banda 5 (NIR). A interação da pastagem com a região do NIR pode ser relacionada com a presença de indicadores utilizados para avaliar a qualidade da pastagem como alimento para os animais (LUGASSI *et al.*, 2015). Mudanças ocorridas nessa região do espectro indicam mudanças na composição da vegetação (LUGASSI *et al.*, 2015), de forma que um menor desvio padrão desses valores associou-se com o Pasto Limpo.

Todos os atributos escolhidos para este trabalho foram comparados entre si por meio de correlação de Spearman (Figura 4.4). Todos obtiveram correlações menores (em módulo) que 0,5 com mais da metade dos demais atributos, indicando que cada um deles é capaz de fornecer novas informações sobre o conjunto de dados trabalhado. Para a construção da Figura 4.4, foram utilizados os dados das mesmas duas cenas, porém somente do período de julho de 2014.

Figura 4.2 – Árvore de decisão para detecção de Pastagem do modelo criado com amostras das duas cenas, 001/067 e 226/068.

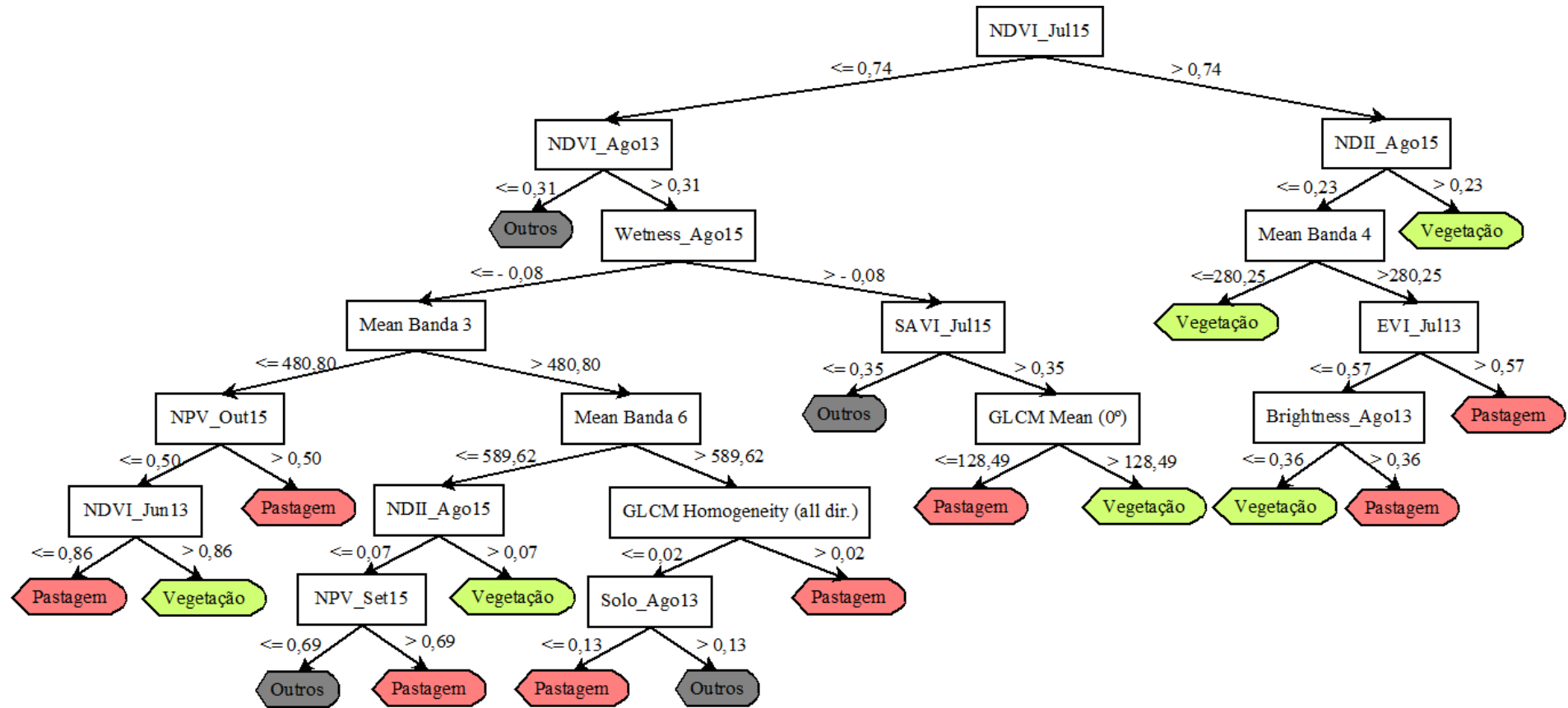


Figura 4.3 – Árvore de decisão para a classificação de Pasto Limpo e Pasto Sujo do modelo criado com amostras das duas cenas, 001/067 e 226/068.

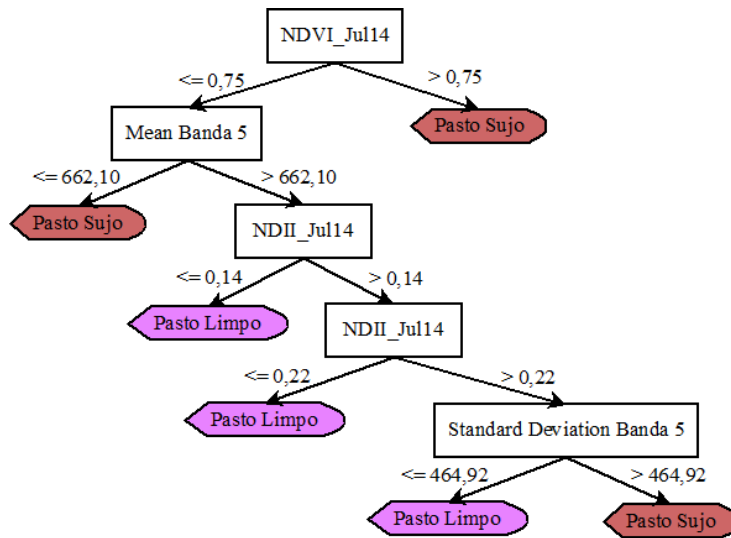
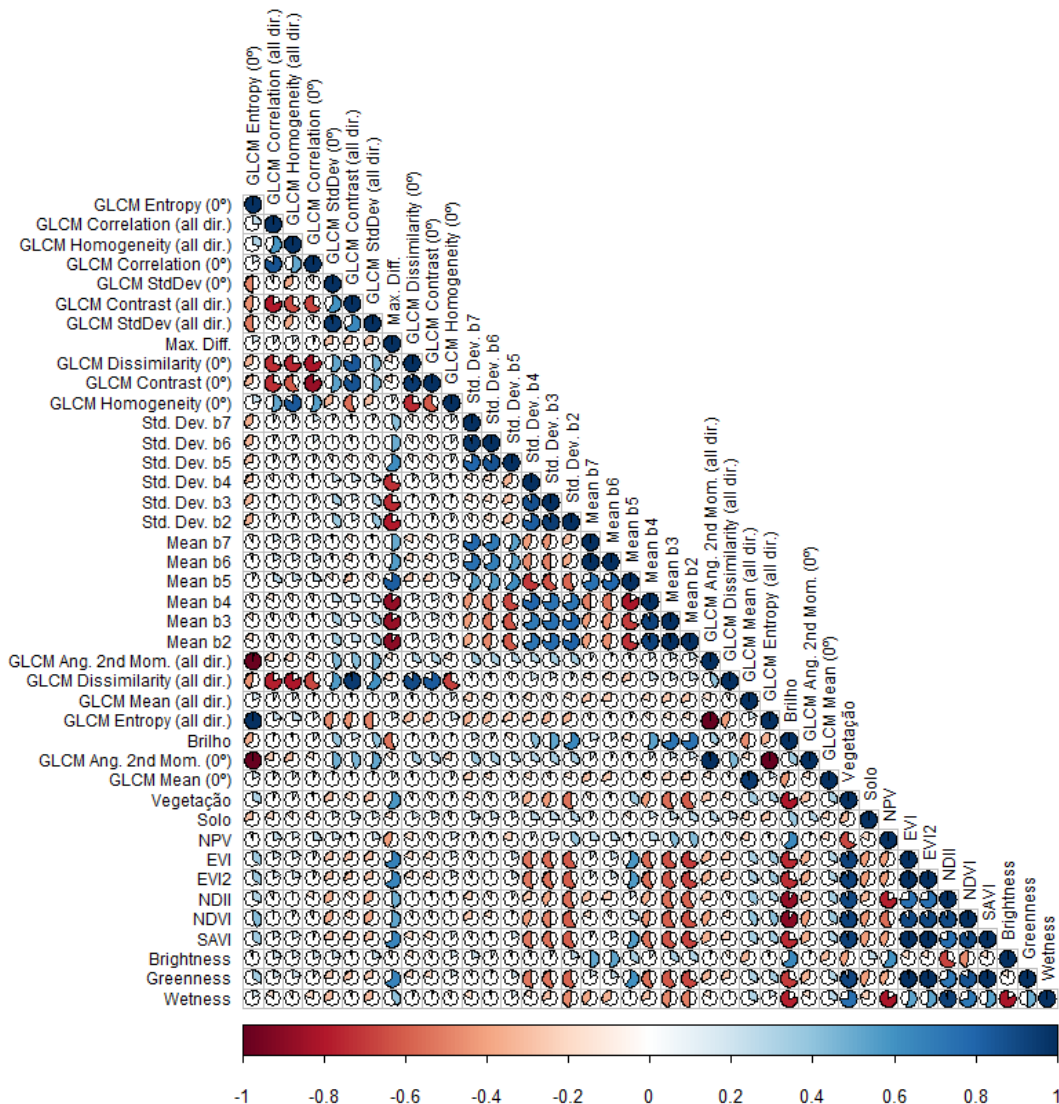
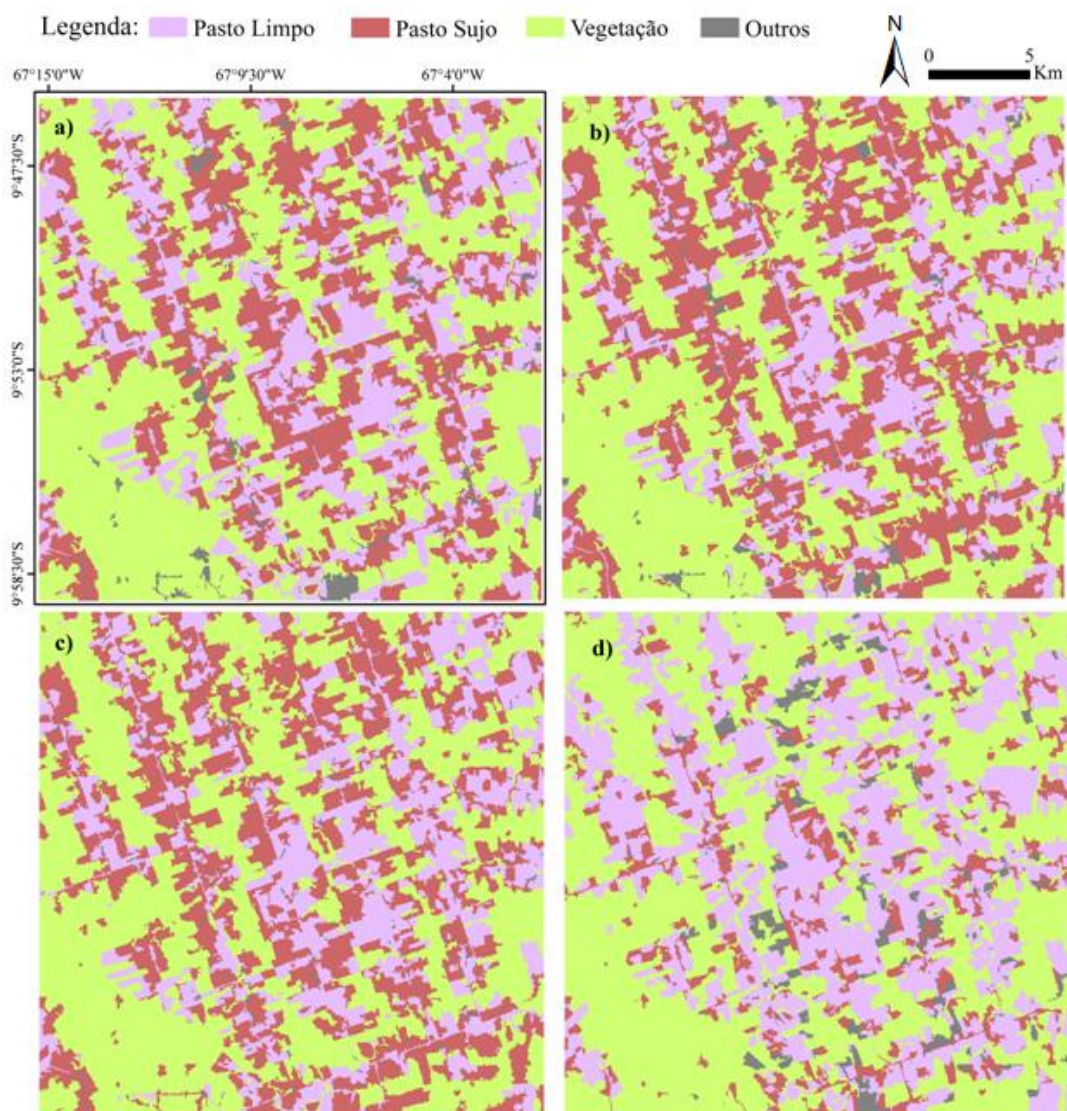


Figura 4.4 – Matriz de correlação (Spearman) relacionando todos os atributos utilizados.



As espacializações das classificações geradas podem ser vistas nas Figuras 4.5, 4.6 e 4.7. Em todos os casos, as taxas de acerto para a detecção de Pastagem foram maiores do que as taxas para a classificação entre Pasto Limpo e Sujo. Apesar de se tratar de três cenas pertencentes ao mesmo bioma, Amazônia, região sul, elas possuem diferenças entre si e as pastagens mudam consideravelmente de uma região para outra. Esse aspecto será mais detalhado na próxima seção, referente ao trabalho de campo.

Figura 4.5 – Espacialização das classificações da cena 001/067. a) Referência. b) Modelo criado/avaliado na 001/067. c) Modelo criado na combinação (001/067 e 226/068) e avaliado na 001/067. d) Modelo criado na 226/068 e avaliado na 001/067.



A região da cena 226/068 é a única que possui grandes áreas de agricultura, incluídas na classe Outros, e as suas pastagens são maiores e possuem fronteiras mais bem definidas se comparadas às das outras cenas. As outras duas cenas são

predominantemente cobertas por pastagem, mas a 001/067 ainda possui aproximadamente metade de sua área coberta por floresta ou vegetação secundária. A cena 231/067 possui poucos polígonos de vegetação (floresta e vegetação secundária) e suas pastagens representam predominantemente polígonos menores, com formatos irregulares e fronteiras indefinidas. Na Figura 4.7 podemos ver a grande quantidade de áreas de Pastagem nessa cena. Apesar de a classificação (Figura 4.7b) parecer bastante diferente de sua referência (Figura 4.7a), a detecção de Pastagem teve uma taxa de acerto de 87,19% (Tabela 4.3). A maior quantidade de erros, que sobressai na espacialização da classificação, foi na segunda etapa, de diferenciação entre Pasto Limpo e Sujo, cuja taxa de acerto foi de 52,76%.

Figura 4.6 – Espacialização das classificações da cena 226/068. a) Referência. b) Modelo criado/avaliado na 226/068. c) Modelo criado na combinação (001/067 e 226/068) e avaliado na 226/068. d) Modelo criado na 001/067 e avaliado na 226/068.

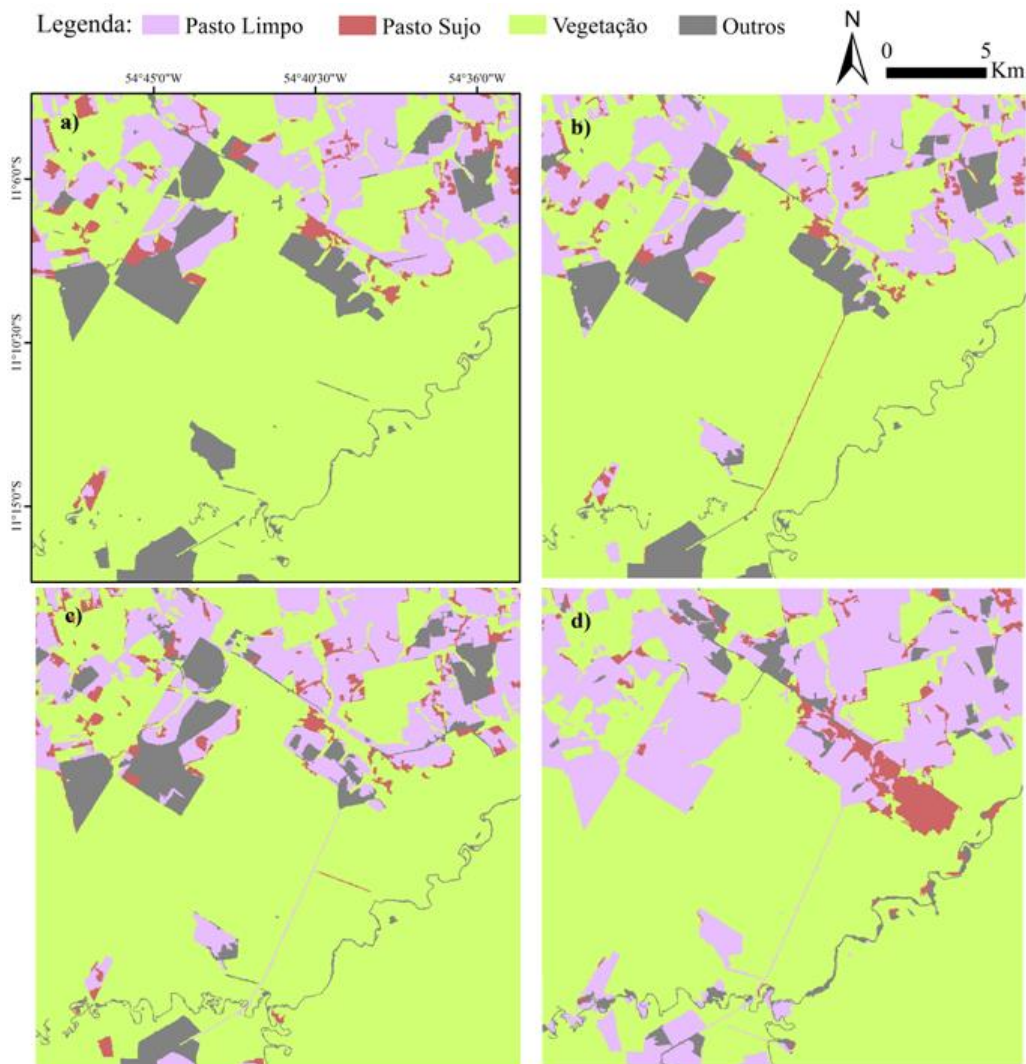
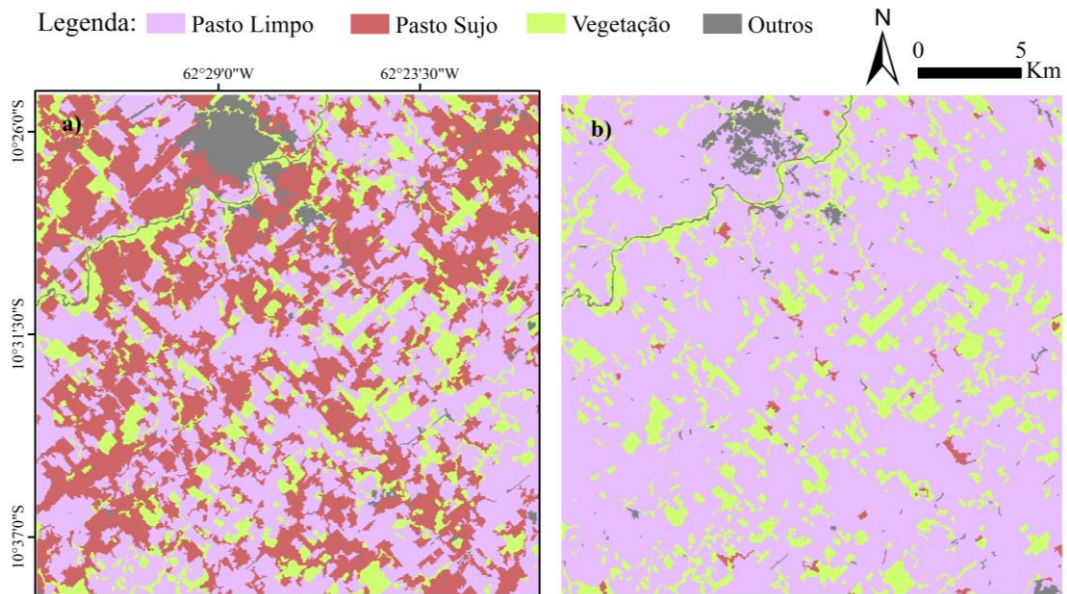
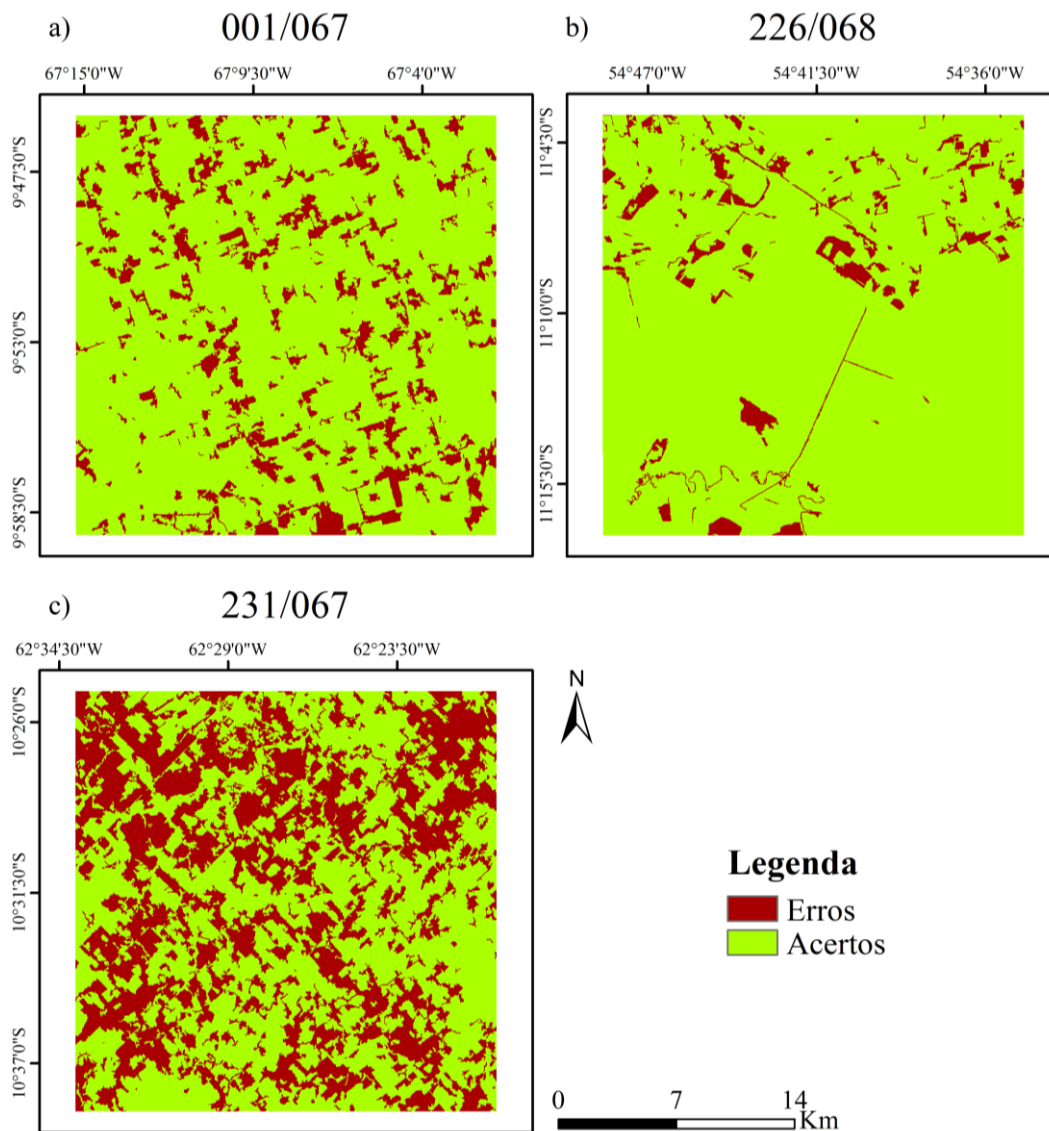


Figura 4.7 – Espacialização da classificação da cena 231/067. a) Referência. b) Modelo criado na combinação (001/067 e 226/068) e avaliado na 231/067.



Para melhorar a visualização dos acertos e erros, os três resultados das classificações gerados a partir do modelo criado na combinação 001/067 e 226/068 foram comparados com suas respectivas referências (Figura 4.8). Os erros são, em sua maioria, referentes às áreas de confusão entre Pasto Limpo e Pasto Sujo. Entre as duas classes, o modelo foi menos sensível para a detecção de Pasto Sujo devido a maior quantidade de “desordem” nessa classe (várias espécies diferentes de invasoras e diferentes práticas de manejo ou ausência dele).

Figura 4.8 – Comparação das referências e dos resultados gerados pelo modelo criado na combinação (001/067 e 226/068). a) 001/067. b) 226/068. c) 231/067.



4.3. Validação e Trabalho de Campo

Para a construção da Tabela 4.3, as classificações geradas já foram comparadas com suas respectivas referências mapeadas em imagens RapidEye. Porém, independente das áreas dos polígonos, ao serem classificados corretamente, contaram como 1 acerto cada. Para evitar possíveis tendências causadas por polígonos de tamanhos muito diferentes, a validação foi feita por meio da geração de 1000 pontos aleatórios em cada classificação (já representando quatro classes: Vegetação, Pasto Limpo, Pasto Sujo e Outros) e sua referência (Tabela 4.4).

Tabela 4.4 – Validação dos modelos de classificação por pontos aleatórios.

Modelo	Taxa de acerto global (%)
Criado/avaliado na 001/067	86,9
Criado na 001/067 e avaliado na 226/068	85,93
Criado/avaliado na 226/068	96,1
Criado na 226/068 e avaliado na 001/067	65,9
Criado na combinação (001/067 e 226/068) e avaliado na 001/067	86
Criado na combinação (001/067 e 226/068) e avaliado na 226/068	94,4
Criado na combinação (001/067 e 226/068) e avaliado na 231/067	57,1

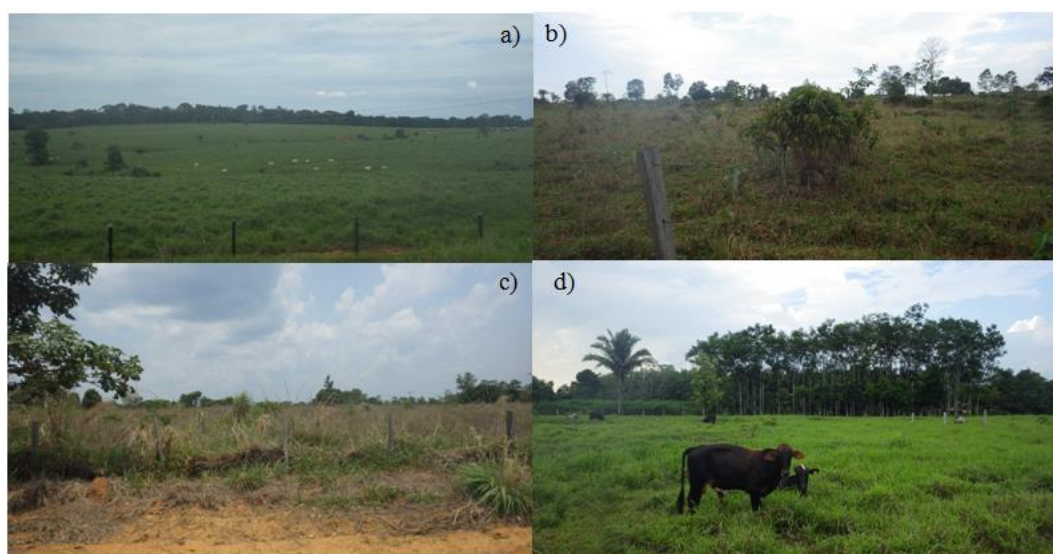
As taxas de acerto global seguiram padrão semelhante às aquelas encontradas na Tabela 4.3. As taxas de acerto também caíram na validação dos modelos criados em uma cena e testados em outra. A classificação da cena 231/067 aparentou uma queda: ocorreu um acerto de 57,10% na validação, enquanto na Tabela 4.3 os valores eram de 87,19% para a detecção de Pastagem e 52,76% na classificação de Pasto Limpo e Sujo. Esses valores são coerentes se for observada (Figura 4.7a) a quantidade elevada de pastagens nessa área. Como a maior dificuldade do modelo esteve na separação de Pasto Limpo e Pasto Sujo, a validação por pontos permitiu que a confusão entre os dois tipos de pastagem sobressaísse na taxa de acerto global.

Dos 51 pontos identificados em campo, 12 pontos foram na cena 001/067, 17 na 226/068 e 22 na 231/067 (Figura 4.9). O modelo criado com amostras das duas primeiras cenas foi utilizado para classificar os dados de campo e resultou em taxas de acerto maiores para a fase de detecção de pastagem em comparação à fase de classificação de Pasto Limpo e Sujo (Tabela 4.5). A redução das taxas de acerto pode ter sido causada pelo deslocamento do modelo na janela temporal (2013 a 2015 para 2014 a 2016) para compreender o ano do campo (2016), uma vez que as pastagens são afetadas por fatores ambientais, como a variabilidade climática.

Tabela 4.5 – Classificação dos pontos do trabalho de campo.

Imagem	Pontos em campo	Fase de classificação	Taxa de acerto (%)
001/067	12	Detecção de Pastagem	66,66
		Pasto Limpo x Pasto Sujo	50
226/068	17	Detecção de Pastagem	82,35
		Pasto Limpo x Pasto Sujo	52,94
231/067	22	Detecção de Pastagem	77,27
		Pasto Limpo x Pasto Sujo	63,63

Figura 4.9 – Alguns pontos de pastagem observados em campo. a) Pasto Limpo na cena 001/067, b) Pasto Sujo na cena 231/067, c) Pasto Sujo na cena 226/068, d) Pasto Limpo na cena 231/067.

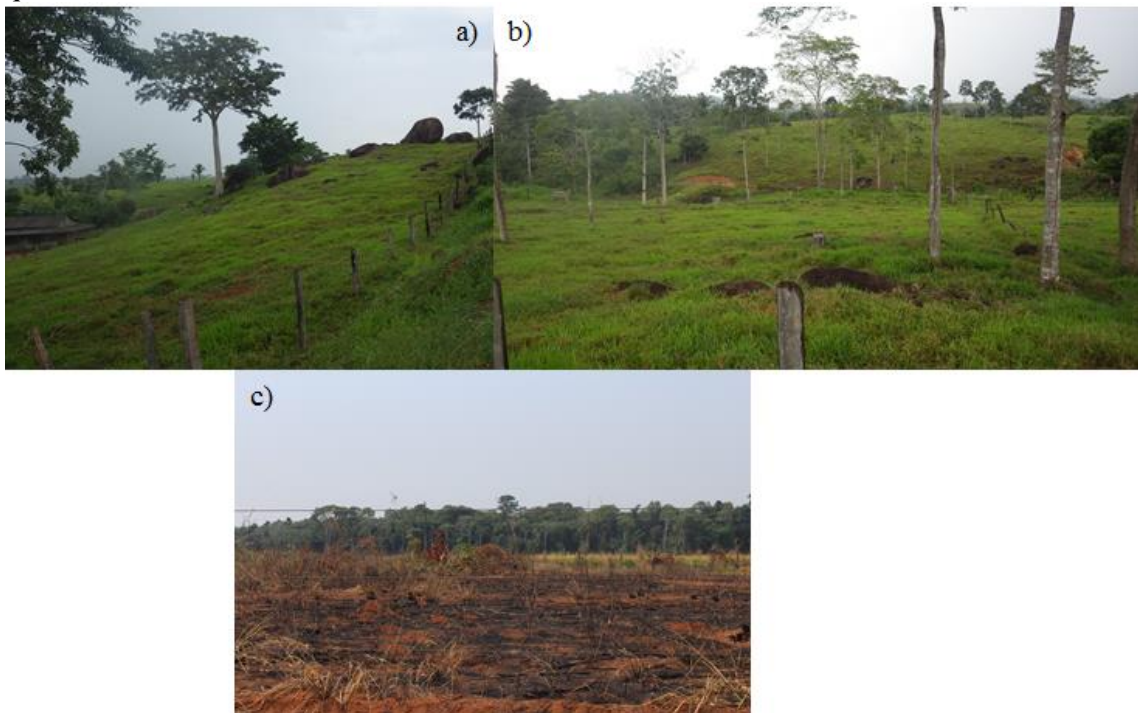


Durante o trabalho de campo, foi possível constatar a diversidade dos padrões de pastagem ao longo da área de estudo. No estado de Rondônia, existem diversos materiais rochosos e depósitos minerais originados pela evolução geológica. A maior parte das pastagens nessa região encontra-se em terrenos de relevo dissecados em colinas e morros, que de acordo com CPRM (2010), podem variar de 20 a 500m de altitude (Figura 4.10a e b).

No estado do MT, as pastagens observadas ocorrem em grandes áreas planas ou de baixa ondulação. No estado do Acre, a região percorrida no trabalho de campo, a leste da cidade de Rio Branco, possui grandes áreas de pastagens contínuas em terrenos planos e de baixa amplitude. De acordo com CPRM (2015), essa área pertence ao domínio geomorfológico Tabuleiros da Amazônia Centro-oriental, dominada por extensos tabuleiros com amplitude de relevo entre 20 e 50m apenas.

Os aspectos descritos contribuem na dificuldade de identificar padrões homogêneos em grandes áreas. Além disso, as diferenças observadas ocorrem não só espacialmente, em alguns pontos também se observou uma dinâmica temporal no uso da terra. Alguns exemplos de áreas que não se mantiveram constantes ao longo do tempo: área que em 2014 possuía padrão de pastagem na imagem OLI, mas em 2016 mudou para um início de reflorestamento; área de pastagem que no período do campo foi observado como área queimada (Figura 4.10c); área de agricultura abandonada; e área de pastagem que estava sendo limpa para dar lugar à agricultura.

Figura 4.10 – Alguns aspectos que contribuem para a heterogeneidade das pastagens. a) e b) Presença de afloramentos rochosos e variação de altitude do terreno, c) Pasto queimado.



5 CONCLUSÕES

O presente trabalho apresentou um conjunto de experimentos para o reconhecimento de pastagens na Amazônia, por meio da detecção e classificação de dois padrões de pastagem: Pasto Limpo e Pasto Sujo. O manejo e monitoramento das áreas de pastagem são capazes de fornecer um panorama da qualidade das pastagens para ajudar a pensar em políticas de incentivo para a melhoria de manejo nessas áreas ou para a conversão para outros sistemas produtivos.

Comparando as classificações por *pixel*, baseada em objetos (ambas hierárquicas) e a de 5 Classes, o Caso 2 (baseado em objetos, utilizando o J48) atingiu 90,09% de acurácia na classificação e obteve as maiores taxas de acerto dentre os demais casos. Ao comparar com os outros casos, pode-se observar que três aspectos foram fundamentais para este resultado: o uso da segmentação, a inclusão dos atributos texturais e a realização da classificação hierárquica.

As pastagens na Amazônia variam de acordo com diversos parâmetros ambientais (tipo de solo, relevo, sazonalidade, etc.) e de manejo (como espécie de gramínea, taxa de lotação e idade da pastagem). Por esse motivo, os modelos que combinam amostras de cenas diferentes possuem uma capacidade maior de classificação, se escalas maiores forem consideradas. Os resultados dos modelos criados e avaliados na mesma cena obtiveram as maiores taxas de acerto, porém foram mais restritivos. O modelo criado com a combinação de amostras das cenas 001/067 e 226/068 obteve taxas de acerto menores do que o caso anterior, mas com valores superiores a 80%, demonstrando a importância de tentar abranger o máximo de padrões possíveis no momento da coleta de amostras.

Para a fase de detecção de Pastagem, no modelo criado com amostras das duas cenas (001/067 e 226/068), os atributos mais relevantes foram: NDVI, NDII, SAVI, fração NPV, fração Solo, Brightness, Wetness, média da banda *Green*, média da banda *Red*, média da banda SWIR 1, média da GLCM e homogeneidade da GLCM. No mesmo modelo, os atributos mais relevantes para a classificação de Pasto Limpo e Pasto Sujo foram: NDVI, NDII, média da banda NIR e desvio padrão da banda NIR.

A observação dos diferentes padrões de pastagem em campo foi fundamental para reconhecer o que estava sendo observado em imagens Landsat. Pode-se compreender a complexidade de tipos de pastagem existentes no mesmo bioma. A diferença temporal do modelo gerado para o intervalo 2013-2015 e aplicado em 2014-2016 pode ter sido responsável por uma diminuição na taxa de acerto, classificando de maneira correta aproximadamente 75% dos pontos amostrados.

Para dar continuidade a esta pesquisa, pode-se testar a metodologia em outras regiões mais ao norte da Amazônia Legal, uma vez que as particularidades de cada região e, nesse caso, a diferença dos regimes de chuva, podem modificar a dinâmica que as pastagens apresentam no tempo e espaço. Na etapa de segmentação, algoritmos que incluam a variabilidade dos segmentos no tempo (segmentação “multitemporal”) também devem ser testados. Além disso, pode-se investigar o tempo de permanência de cada tipo de pastagem ao longo da série temporal e como elas estão integradas na dinâmica de mudanças de uso da terra na Amazônia.

REFERÊNCIAS BIBLIOGRÁFICAS

ADAMS, J. B.; SABOL, D. E.; KAPOV, V.; ALMEIDA FILHO, R.; ROBERTS, D. A.; SMITH, M. O.; GILLESPIE, A. R. Classification of multispectral image based on fractions of endmembers: application to land-cover change in the Brazilian Amazon. **Remote Sensing of Environment**, v. 52, p.137-154, 1995.

ADAMI, M. **Estimativa da data de plantio da soja por meio de séries temporais de imagens MODIS**. 2010. 163p. Tese (Doutorado) do Curso de Pós-Graduação em Sensoriamento Remoto - INPE, 2010. Disponível em: <<http://urlib.net/sid.inpe.br/mtc-m19/2010/09.15.21.47>>.

AGUIAR, D. A. **Séries temporais de imagens MODIS para avaliação de pastagens tropicais**. 2013. 167 p. (sid.inpe.br/mtc-m19/2013/08.30.14.33-TDI). Tese (Doutorado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2013. Disponível em: <<http://urlib.net/8JMKD3MGP7W/3ENND62>>. Acesso em: 03 fev. 2017.
AGUIAR, Daniel Alves et al. MODIS time series to assess pasture land. In: **Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International**. IEEE, 2010. p. 2123-2126.

ALMEIDA, C. A.; COUTINHO, A. C.; ESQUERDO, J. C. D. M.; ADAMI, M.; VENTURIERI, A.; DINIZ, C. G.; DESSAY, N.; DURIEUX, L.; GOMES, A. R. High spatial resolution land use and land cover mapping of the Brazilian Legal Amazon in 2008 using Landsat-5/TM and MODIS data. **Acta Amazonica**, v. 46. n. 3, p. 291-302, 2016.

ALMEIDA, C.; SILVA, M; LOBO, F. L.; FARIAS, T. P.; GOMES, A.; COSTA, L. C.; ESCADA, M. I. S. TerraClass: classificação dos padrões de uso e cobertura da terra da Amazônia Legal. In: EMÍLIO, T.; LUIZÃO, F. (Org.). **Cenários para a Amazônia: clima, biodiversidade e uso da terra**. Manaus: INPA, 2014. p. 138-147.

APTE, C.; WEISS, S. Data mining with decision trees and decision rules. **Future Generation Computer Systems**. Amsterdam, v. 13, n.2-3, p.197-210, nov., 1997.

ASNER, G. P.. Cloud cover in Landsat observation of the Brazilian Amazon. **International Journal of Remote Sensing**, v. 22, n. 18, p. 3855–3862, 2001.

ASNER, G. P.; TOWNSEND, A. R.; BUSTAMANTES, M. M. C.; NARDOTO, G. B.; OLANDER, L. P. Pasture degradation in the central Amazon: linking changes in carbon and nutrient cycling with remote sensing. **Global Change Biology**, v. 10, p. 844-862, 2004.

BAATZ, M.; SCHAPE, A. Multiresolution segmentation: an optimization approach for high quality multi-scale image segmentation. In: ANGEWANDTE GEOGRAPHISCHE INFORMATIONSVERARBEITUNG, 12, 2000, Wichmann-Verlag. **Proceedings...** WichmannVerlag: Heidelberg, 2000. p. 12-23.

- BAIG, M. H. A.; ZHANG, L.; SHUAI, T.; TONG, Q. Derivation of a tasseled cap transformation based on Landsat 8 at satellite reflectance. **Remote Sensing Letters**, v.5, n.5, p. 423-431, 2014.
- BANCO MUNDIAL. **Causas do desmatamento da Amazônia Brasileira**. 1. ed. Brasília: Banco Mundial, 2003.
- BANSKOTA, A., N. KAYASTHA, M. FALKOWSKI, M.A. WULDER, R. FROESE, J.C. WHITE. Forest monitoring using Landsat time-series data- A review. **Canadian Journal of Remote Sensing**, v. 40, n. 5, p. 362-384, 2014..
- BECKER, B. K. Amazônia. **Série Princípios**. São Paulo: Ática, 1990. 92p.
- BECKER, B. K. **Amazônia: geopolítica na virada do III milênio**. Rio de Janeiro: Garamond, 2009. 172p.
- BROCKWELL, P. J.; DAVIS, R. A. **Introduction to time series and forecasting**. 2. ed. New York: Springer-Verlag. 2002. 434 p.
- BREIMAN, L. Random forests. **Machine learning**, v. 45, n. 1, p. 5-32, 2001.
- COMPANHIA DE PESQUISA DE RECURSOS MINERAIS (CPRM). **Geodiversidade do Estado do Acre**. Programa Geologia do Brasil. Porto Velho: CPRM, 2015. 321 p.
- COMPANHIA DE PESQUISA DE RECURSOS MINERAIS (CPRM). **Geodiversidade do Estado de Rondônia**. Programa Geologia do Brasil. Porto Velho: CPRM, 2010. 337 p.
- COUTINHO, A. C.; ALMEIDA, C.; VENTURIERI, A.; ESQUERDO, J. C. D. M.; SILVA, M. Projeto TerraClass: Uso e cobertura da terra nas áreas desflorestadas na Amazônia Legal. Brasília, DF: Embrapa; Belém: INPE, 2013.
- CRIST, E. P.; CICONE, R. C. A Physically-Based Transformation of Thematic Mapper Data-the TM Tasseled Cap. **IEEE Transactions on Geoscience and Remote Sensing**, GE-22, n. 3, p. 256-263, 1984.
- CRIST, E. P.; KAUTH, R. J. The Tasseled Cap De-Mystified. **Photogrammetric Engineering and Remote Sensing**, v. 52, n.1, p. 81-86, 1986.
- DASGUPTA, S. PAPANIMITRIOU, C. H. VAZIRAN, U. V. **Algorithms**. McGraw-Hill, 2006.
- DASH, M.; LIU, H. Feature selection for classification. **Intelligent Data Analysis**, v. 1, n.3, p. 131-156, 1997.
- DAVIDSON, E.A.; ASNER, G.P.; STONE, T.A.; NEILL, C.; FIGUEIREDO, R.O. Objective indicators of pasture degradation from spectral mixture analysis of Landsat imagery. **Journal of Geophysical Research**, v. 113, n. G1, 2008.
- DEVRIES, B.; DECUYPER, M.; VERBESSELT, J.; ZEILEIS, A.; HEROLD, M.; JOSEPH, S. Tracking disturbance-regrowth dynamics in tropical forests using

structural change detection and Landsat time series. **Remote Sensing of Environment**, n. 169, p. 320-334, 2015.

DIAS-FILHO, M. B. Produção de bovinos a pasto na fronteira agrícola. In: RODRIGUES, K.F.; FERREIRA, W.M.; MACEDO JR., G. L. (Orgs.). **Zootec na amazônia legal**. Palmas: UFT: ABZ, 2010. p. 131-145. CONGRESSO BRASILEIRO DE ZOOTECNIA, 20., 2010.

DIAS-FILHO, M. B. **Desafios e perspectivas na recuperação de pastagens degradadas na Amazônia**. Belém: Embrapa Amazônia Oriental, 2015. 38 f.

DORIGO, W. A.; ZURITA-MILLA, R.; DE WIT, A. J. W.; BRAZILE, J.; SINGH, R.; SCHAEPMAN, M. E. A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. **International Journal of Applied Earth Observation and Geoinformation**, n. 9, p. 165-193, 2007.

EKLUNDH, L.; JONSSON, P. Extracting information about vegetation seasons in Africa from Pathfinder AVHRR NDVI imagery using temporal filtering and least-squares fits to asymmetric Gaussian functions. In: INTERNATIONAL SYMPOSIUM ON REMOTE SENSING, 2003, Honolulu, Hawaii. **Proceedings...** Honolulu, Hawaii: International Society for Optics and Photonics, 2003. p. 215-225.

EKLUNDH, L.; AND JÖNSSON, P. **TIMESAT 3.2 with parallel processing - Software Manual**. 2012. Lund University, 88 p.

FAYYAD, U.; PIATETSKY-SHAPIO, G.; SMYTH, P. From data mining to knowledge discovery in databases. **AI Magazine**, v.17, n.3, p.37-54, jul., 1996.

FEARNSIDE, P. M. Desmatamento na Amazônia brasileira: história, índices e consequências. **Megadiversidade**, v. 1, n 1, julho de 2005.

FEARNSIDE, P. M. Desmatamento na Amazônia: dinâmica, impactos e controle. **Acta Amazonica**, v. 36, n. 3, p. 395-400, 2006.

FIGUEROA, S.N.; NOBRE, C.A. Precipitations distribution over Central and Western Tropical South America. **Climanálise - Boletim de Monitoramento e Análise Climática**, v. 5, n. 6, p. 36 - 45, 1990.

FISCH, G.; MARENGO, J. A.; NOBRE, C. A. Clima da Amazônia. **Boletim Climanalise**, v. 29, n. 2, 2014. Disponível em: <<http://climanalise.cptec.inpe.br/~rclimanl/boletim/>>.

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, p. 576-587, 2008.

GREGO, C. R.; GALDINO, S.; BOLFE, E. L.; NOGUEIRA, S. F.; BAYMA-SILVA, G.; ARAUJO, L. S.; POCCARD-CHAPUIS, R.; THALÊS, M. C. Distribuição espacial das classes de degradação de pastagem em Marabá, PA e

sua correlação com as classes de solo e relevo. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 17, 2015, João Pessoa. **Anais...** João Pessoa: INPE, 2015. p. 6196-6202.

HALL-BEYER, M. **The GLCM tutorial home page**: <www.fp.ucalgary.ca/mhallbey/tutorial.htm> . 2007

HALL M.; FRANK, E., HOLMES, G.; PFAHRINGER, B.; REUTEMANN, P.; WITTEN, I. H. The WEKA data mining software: an update. **SIGKDD Explorations**, v. 11, n. 1, 2009.

HAN, J.; KAMBER, M.; PEI, J. **Data mining concepts and techniques**. 3. ed. Elsevier, 2012.

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

HARDISKY, M.A.; KLEMAS, V.; SMART, R.M. The influences of soil salinity, growth form, and leaf moisture on the spectral reflectance of *Spartina alterniflora* canopies. **Photogrammetric Engineering and Remote Sensing**, n. 49, p. 77–83, 1983.

HAY, G. J.; CASTILLA, G. Object-based image analysis: strengths, weaknesses, opportunities and threats (swot). **The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences**. OBIA, 2006.

HAYKIN, S. **Neural Networks and Learning Machines**. 3. ed. Englewood Cliffs: Prentice- Hall, 2009.

HOMMA, A. K. O. As questões emergentes e a agricultura na Amazônia. In: CONGRESSO BRASILEIRO DE ECONOMIA E SOCIOLOGIA RURAL, 37., 1999, Foz do Iguaçu, 1999. **Anais...** Brasília: SOBER, 1999.

HUETE, A. R. A soil adjusted vegetation index SAVI. **Remote Sensing of Environment**. v. 25, p. 295–309, 1988.

HUETE, A. R.; DIDAN, K.; MIURA, T.; RODRIGUEZ, E.P.; GAO, X.; FERREIRA, L.G. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. **Remote Sensing of Environment**, v.83, n.1-2, p.195-213, 2002.

HUETE, A. R.; LIU, Q.Q.; BATCHILY, K.; VAN LEEUWEN, W. A comparison of vegetation indices over a global set of TM images for EOS-MODIS. **Remote Sensing of Environment**, v.59, n.3, p.440-451, 1997.

IHAKA, R.; GENTLEMAN, R. R: A language for data analysis and graphics. **Journal of Computational and Graphical Statistics**, v. 5, n. 3, p. 299-314, 1996.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Censo Agropecuário 1920/2006**. Até 1996, dados extraídos de: Estatísticas do Século XX. Rio de Janeiro: IBGE, 2007.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Manual Técnico da Vegetação Brasileira**. 2. ed revisada e ampliada. 2012. 271p.

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS (INPE). **Projeto de Monitoramento do Desflorestamento na Amazônia Legal – PRODES**. São José dos Campos, 2013. Disponível em: <<http://www.obt.inpe.br/prodes/index.php>>.

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS (INPE). **Projeto TerraClass 2012** – Mapeamento do uso e cobertura da terra na Amazônia Legal. São José dos Campos, 2014. Disponível em: <http://www.inpe.br/cra/projetos_pesquisas/TerraClass_2012_26nov2014.pdf>.

JIANG, Z.; HUETE, A.R.; DIDAN, K.; MIURA, T. Development of a two-band enhanced vegetation index without a blue band. **Remote Sensing of Environment**, v.112, n.10, p.3833-3845, 2008.

KAUTH, R. J.; THOMAS, G. S. The tasseled cap – a graphic description of the spectral-temporal development of agricultural crops as seen in Landsat. In: SYMPOSIUM ON MACHINE PROCESSING OF REMOTELY SENSED DATA, 1976, West Lafayette, Indiana. **Proceedings...** West Lafayette, Indiana: IEEE, 1976. p.41-51.

LAROSE, D. T.; LAROSE, C. D. **Discovering knowledge in data: an introduction to data mining**. 2. ed. John Wiley and Sons, Inc, 2014.

LI, Z.; GUO, X. Remote sensing of terrestrial non-photosynthetic vegetation using hyperspectral, multispectral, SAR, and LiDAR data. **Progress in Physical Geography**, v. 40, n. 2, May 2015. DOI: 10.1177/0309133315582005 .

LUGASSI, R.; CHUDNOVSKY, A.; ZAADY, E.; DVASH, L.; GOLDSHLEGER, N. Spectral slope as an indicator of pasture quality. **Remote Sensing**, v. 7, p. 256-274, 2015.

MOREIRA, M. A.; SHIMABUKURO, Y.E. Cálculo do índice de vegetação a partir do sensor AVHRR. In: FERREIRA, N. J. (Ed.). **Aplicações ambientais brasileiras dos satélites NOAA e TIROS-N**. São Paulo: Oficina de textos, 2004. p.79-101.

MUCHAGATA, M.; BROWN, K. Cows, colonists and trees: rethinking cattle and environmental degradation in Brazilian Amazon. **Agricultural Systems**, n. 76, p. 797-816, 2003.

MÜLLER, H.; RUFIN, P.; GRIFFITHS, P.; SIQUEIRA, A. J. B.; HOSTERT, P. Mining dense Landsat time series for separating cropland and pasture in a heterogeneous Brazilian savanna landscape. **Remote Sensing of Environment**, n. 156, p. 490-499, 2015.

- NEVES, A. K.; BENDINI, H. N.; KORTING, T. S.; FONSECA, L. M. G. Combining time series features and data mining to detect land cover patterns: a case study in northern Mato Grosso state, Brazil. **Brazilian Journal of Cartography**, n. 68, n. 6, p. 1133-1142, 2016.
- NEVILLE, P. G. **Decision trees for predictive modeling**. SAS Institute Inc, 1999. v. 4.
- NOVO, E. M. L. M. **Sensoriamento remoto** – princípios e aplicações. 4 ed. São Paulo: BLUCHER, 2010. 288 p.
- NUMATA, I.; SOARES, J. V.; BATISTA, G. T.; ROBERTS, D. A.; CHADWICK, O. A. Evaluation of the conversion from forest to pasture using remote sensing for soil fertility analysis. In: INTERNATIONAL CONGRESS FOR PHOTOGRAMMETRY AND REMOTE SENSING, 19., 2000, Amsterdam, Netherlands. **Proceedings...** 2000. p. 1033-1037. (INPE-8027-PRE/3845).
- NUMATA, I.; SOARES, J. V.; ROBERTS, D. A.; LEONIDAS, F. C.; CHADWICK, O. A.; BATISTA, G. T. Relationships among soil fertility dynamics and remotely sensed measures across pasture chronosequences in Rondônia, Brazil. **Remote Sensing of environment**, v. 87, p. 446-455, 2003.
- NUMATA, I.; ROBERTS, D. A.; CHADWICK, O. A.; SCHIMEL, J.; SAMPAIO, F. R.; LEONIDAS, F. C.; SOARES, J. V. Characterization of pasture biophysical properties and the impact of grazing intensity using remotely sensed data. **Remote Sensing of Environment**, v. 109, p. 314-327, 2007a.
- NUMATA, I.; ROBERTS, D. A.; SAWADA, Y.; CHADWICK, O. A.; SCHIMEL, J.; SOARES, J. V. Regional characterization of pasture changes through time and space in Rondônia, Brasil. **Earth Interactions**, v. 11, n. 14, 2007b.
- PATHAK, B. AND BAROOAH, D. Texture analysis based on the gray-level co-occurrence matrix considering possible orientations. **International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering**, 2(9), 2013. p.4206-4212.
- PONZONI, F. J.; SHIMABUKURO, Y. E. **Sensoriamento remoto no estudo da vegetação**. São José dos Campos: PARÊNTESES, 2009. 127p.
- QUINLAN, J. R. **C4. 5: programs for machine learning**. Elsevier. 2014
- RIBEIRO, N. de F. **A questão geopolítica da Amazônia**: da soberania difusa à soberania restrita. Brasília, DF: Senado Federal, 2005. 537 p.
- ROBERTS, D. A.; BATISTA, G. T.; PEREIRA, J. L. G.; WALLER, E. K.; NELSON, B.W. Change identification using multitemporal spectral mixture analysis: applications in eastern Amazonia. In: LUNETTA, R. S. ; ELVIDGE, C. D.(ed.). **Remote sensing change detection**: environmental monitoring methods and applications. Michigan: Ann Arbor Press, 1998. v.1, cap. 9, p. 137 - 161.

RODRIGUEZ-GALIANO, V. F., GHIMIRE, B., ROGAN, J., CHICA-OLMO, M., & RIGOL-SANCHEZ, J. P. An assessment of the effectiveness of a random forest classifier for land-cover classification. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 67, p. 93-104, 2012.

ROUSE, J.W.; HAAS, R.H.; SCHELL, J.A.; DEERING, D.W. Monitoring vegetation systems in the Great Plains with ERTS. In: ERTS SYMPOSIUM, 3., 1973, Greenbelt. Proceedings... Greenbelt: NASA, 1973. p. 309–317.

RUFIN, P; MULLER, H.; PFLUGMACHER, D.; HOSTERT, P. Land use intensity trajectories on Amazonian pastures derived from Landsat time series. **International Journal of Applied Earth Observation and Geoinformation**, v. 41, p. 1-10, (2015).

SEYLER, F.; CHAPLOT, V.; MULLER, F.; CERRI, C. E. P.; BERNOUX, M.; BALLESTER, V.; FELLER, C.; CERRI, C. C. C. Pasture mapping by classification of Landsat TM images. Analysis of the spectral behavior of the pasture class in a real medium-scale environment: the case of the Piracicaba Catchment (12 400 km², Brazil). **International Journal of Remote Sensing**, v. 23, n. 23, p. 4985-5004, 2002.

SILLA JR., C. N.; FREITAS, A. A. A survey of hierarchical classification across different application domains. **Data Mining and Knowledge Discovery**, v. 22, p. 31-72, 2011.

SILVA, G. B. S.; FERNANDES, R. M.; DE OLIVEIRA, B. P.; RÔVERE, S. D.; OSHIRO, O. T.; NOGUEIRA, S. F. **Sistema de informação geográfica das áreas amostrais do projeto GeoDegradate**. Campinas: Embrapa Monitoramento por Satélite, 2012. Disponível em: <<http://www.geodegrade.cnpem.embrapa.br/circular-tecnica-embrapa/>>.

SILVA, F. B.; SHIMABUKURO, Y. E.; ARAGÃO, L. E. O. C.; ANDERSON, L. O.; PEREIRA, G.; CARDOZO, F.; ARAI, E. Large-scale heterogeneity of Amazonian phenology revealed from 26-year long AVHRR/NDVI time-series. **Environmental research letter**, v. 8, n. 2, p. 024011-024011, 2013. DOI: <10.1088/1748-9326/8/2/024011>. Erratum: Large-scale heterogeneity of Amazonian phenology revealed from 26-year long AVHRR/NDVI time-series (Environmental Research Letters (2013) 8 (024011)). Disponível em: <<http://dx.doi.org/10.1088/1748-9326/8/2/024011>>.

SHIMABUKURO, Y. E.; SMITH, J. A. The least-squares mixing models to generate fractions images derived from remote sensing multispectral data. **IEEE Transactions on Geoscience and Remote Sensing**, v. 29, 1, p. 16-20, 1991.

TAN, P.; STEINBACH, M.; KUMAR, V.; POTTER, C.; KLOOSTER, S.; TORREGROSA, A. Clustering earth science data: Goals, issues and results. In: FOURTH KDD WORKSHOP ON MINING SCIENTIFIC DATASETS, 2001, San Francisco. **Proceedings...** San Francisco: U S Department of Energy, 2001. v. 25.

- TEIXEIRA, C. G. **Validação do modelo linear de mistura espectral em imagens Aster/Terra a partir de dados Ikonos**. 2004. 127 p. (INPE-13183-TDI/1029). Dissertação (Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2004. Disponível em: <<http://urlib.net/sid.inpe.br/jeferson/2005/02.15.15.35>>. Acesso em: 06 fev. 2017.
- VALENTIM, J.; ANDRADE, C. DE. Tendências e perspectivas da pecuária bovina na Amazônia. **Amazônia: Ci. & Desenv.**, v. 4, n.8, p. 9–32, 2009.
- VEIGA, J. B.; TOURRAND, J. F.; POCCARD-CHAPUIS, R.; PIKETTY, M. G. Cattle ranching in the Amazon rainforest. **Anim. Prod. Aust.**, v. 24, p. 253-256, 2002.
- VERBESSELT, J.; HYNDMAN, R.; NEWNHAM, G.; CULVENOR, D. Detecting trend and seasonal changes in satellite image time series. **Remote Sensing of Environment**, n. 114, p. 106-115, 2010.
- VICENTE, L. E.; VICTORIA, D. C.; BOLFE, E. L.; ANDRADE, R. G. Estimativa de propriedades biofísicas no mapeamento de pastagens utilizando espectroscopia de imageamento e dados do sensor EO1 – Hyperion. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 15., 2011, Curitiba. **Anais...** Curitiba: INPE, 2011. p. 8575-8582. ISBN 978-85-17-00056-0 (Internet), 978-85-17-00057-7 (DVD). IBI: <3ERPFQRTRW/3A6HA8E>. Disponível em: <<http://urlib.net/3ERPFQRTRW/3A6HA8E>>.
- WALKER, J. J., DE BEURS, K. M., WYNNE, R. H., AND 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.
- ZHANG, X. Y., C. B. SCHAAF, M. A. FRIEDL, A. H. STRAHLER, F. GAO, AND J. C. F. HODGES. MODIS Tasseled Cap Transformation and its Utility. In: IEEE INTERNATIONAL GEOSCIENCE AND REMOTE SENSING SYMPOSIUM (IGARSS'02), 2002, Toronto. **Proceedings...** Piscataway, NJ: IEEE, 2002. p. 1063–1065.
- ZIMMER, A. H.; MACEDO, M. C. M.; KICHEL, A. N.; DE ALMEIDA, R. G. **Degradação, recuperação e renovação de pastagens**. Campo Grande, MS: Embrapa gado de corte, 2012.
- ZHU, Z.; WANG, S.; WOODCOCK, C. E. Improvement and expansion of the Fmask algorithm: cloud, cloud shadow, and snow detection for Landsats 4-7, 8, and Sentinel 2 images. **Remote Sensing of Environment**, v. 159, p.269-277, 2015.

APÊNDICE A – MATRIZES DE CONFUSÃO

Tabela A.1 – Matrizes de confusão, em valores absolutos, da validação dos modelos gerados pelo classificador J48 para a cena 001/067.

Caso 1 (*pixels*):

		3 Classes - 81,75%			Tipos de Pastagem - 64,77%		
		Classificado como:			Classificado como:		
		Vegetação	Pastagem	Outros	Pasto Limpo	Pasto Sujo	
Ref.:	Vegetação	305300	56978	2077	Pasto Limpo	116070	12215
	Pastagem	21582	254201	8237	Pasto Sujo	58031	67885
	Outros	4575	33201	7708	Outros	14646	15173

Caso 2 (*objetos*):

		3 Classes - 90,09%			Tipos de Pastagem - 76,43%		
		Classificado como:			Classificado como:		
		Vegetação	Pastagem	Outros	Pasto Limpo	Pasto Sujo	
Ref.:	Vegetação	698	79	19	Pasto Limpo	505	154
	Pastagem	39	1312	57	Pasto Sujo	62	591
	Outros	5	43	191	Outros	34	88

Caso 3 (*objetos*):

		5 Classes - 32,22%				
		Classificado como:				
		Floresta	Pasto Limpo	Pasto Sujo	Vegetação Secundária	Outros
Ref.:	Floresta	191	143	124	38	35
	Pasto Limpo	112	254	174	43	85
	Pasto Sujo	140	221	216	45	83
	Vegetação Secundária	61	36	78	56	23
	Outros	37	58	66	20	54

Tabela A.2 – Matrizes de confusão, em valores absolutos, da validação dos modelos nas três cenas de estudo.

a) Modelo criado/avaliado na 001/067:

		3 Classes - 90,09%			Tipos de Pastagem - 76,43%		
		Classificado como:			Classificado como:		
		Vegetação	Pastagem	Outros	Pasto Limpo	Pasto Sujo	
	Vegetação	698	79	19	Pasto Limpo	505	154
Ref.:	Pastagem	39	1312	57	Pasto Sujo	62	591
	Outros	5	43	191	Outros	34	88

b) Modelo criado na 001/067 e avaliado na 226/068:

		3 Classes - 67,84%			Tipos de Pastagem - 47,54%		
		Classificado como:			Classificado como:		
		Vegetação	Pastagem	Outros	Pasto Limpo	Pasto Sujo	
	Vegetação	335	124	35	Pasto Limpo	286	3
Ref.:	Pastagem	1	367	56	Pasto Sujo	65	13
	Outros	2	138	49	Outros	182	80

c) Modelo criado/avaliado na 226/068:

		3 Classes - 90,90%			Tipos de Pastagem - 80,03%		
		Classificado como:			Classificado como:		
		Vegetação	Pastagem	Outros	Pasto Limpo	Pasto Sujo	
	Vegetação	551	29	10	Pasto Limpo	402	16
Ref.:	Pastagem	16	520	12	Pasto Sujo	31	71
	Outros	18	42	198	Outros	52	19

d) Modelo criado na 226/068 e avaliado na 001/067:

		3 Classes - 74,99%			Tipos de Pastagem - 43,68%		
		Classificado como:			Classificado como:		
		Vegetação	Pastagem	Outros	Pasto Limpo	Pasto Sujo	
	Vegetação	675	82	39	Pasto Limpo	392	190
Ref.:	Pastagem	200	1115	94	Pasto Sujo	362	171
	Outros	104	92	42	Outros	97	77

e) Modelo criado na combinação (001/067 e 226/068) e avaliado na 001/067:

3 Classes - 87,76%				Tipos de Pastagem - 73,15%			
Classificado como:				Classificado como:			
	Vegetação	Pastagem	Outros	Pasto Limpo	Pasto Sujo		
	Vegetação	692	97	7	Pasto Limpo	522	125
Ref.:	Pastagem	59	1286	64	Pasto Sujo	104	535
	Outros	10	62	166	Outros	58	101

f) Modelo criado na combinação (001/067 e 226/068) e avaliado na 226/068:

3 Classes - 86,25%				Tipos de Pastagem - 72,08%			
Classificado como:				Classificado como			
	Vegetação	Pastagem	Outros	Pasto Limpo	Pasto Sujo		
	Vegetação	538	33	19	Pasto Limpo	365	36
Ref.:	Pastagem	25	490	33	Pasto Sujo	28	61
	Outros	14	68	176	Outros	56	45

g) Modelo criado na combinação (001/067 e 226/068) e avaliado na 231/067:

3 Classes - 87,19%				Tipos de Pastagem - 52,76%			
Classificado como:				Classificado como:			
	Vegetação	Pastagem	Outros	Pasto Limpo	Pasto Sujo		
	Vegetação	444	87	7	Pasto Limpo	1049	4
Ref.:	Pastagem	127	1848	40	Pasto Sujo	775	20
	Outros	8	91	158	Outros	154	24