sid.inpe.br/mtc-m19/2012/02.23.17.05-TDI

DADOS ÓPTICOS E SAR PARA CLASSIFICAÇÃO DE COBERTURA DA TERRA POR ÁRVORE DE DECISÃO NA REGIÃO DA FLORESTA NACIONAL DE TAPAJÓS, PARÁ

Luciane Yumie Sato

Dissertação de Mestrado do Curso de Pós-Graduação em Sensoriamento Remoto, orientada pelos Drs. Tatiana Mora Kuplich, e Yosio Edemir Shimabukuro, aprovada em 02 de março de 2012.

URL do documento original:
<http://urlib.net/8JMKD3MGP7W/3BDM3CP>

INPE
São José dos Campos
2012

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@sid.inpe.br

CONSELHO DE EDITORAÇÃO E PRESERVAÇÃO DA PRODUÇÃO INTELECTUAL DO INPE (RE/DIR-204):

Presidente:

Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)

Membros:

Dr. Antonio Fernando Bertachini de Almeida Prado - Coordenação Engenharia e Tecnologia Espacial (ETE)
Drª Inez Staciarini Batista - Coordenação Ciências Espaciais e Atmosféricas (CEA)
Dr. Gerald Jean Francis Banon - Coordenação Observação da Terra (OBT)
Dr. Germano de Souza Kienbaum - Centro de Tecnologias Especiais (CTE)
Dr. Manoel Alonso Gan - Centro de Previsão de Tempo e Estudos Climáticos (CPT)
Drª Maria do Carmo de Andrade Nono - Conselho de Pós-Graduação
Dr. Plínio Carlos Alvalá - Centro de Ciência do Sistema Terrestre (CST)

BIBLIOTECA DIGITAL:

Dr. Gerald Jean Francis Banon - Coordenação de Observação da Terra (OBT)

REVISÃO E NORMALIZAÇÃO DOCUMENTÁRIA:

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

EDITORAÇÃO ELETRÔNICA:

Vivéca Sant´Ana Lemos - Serviço de Informação e Documentação (SID)

sid.inpe.br/mtc-m19/2012/02.23.17.05-TDI

DADOS ÓPTICOS E SAR PARA CLASSIFICAÇÃO DE COBERTURA DA TERRA POR ÁRVORE DE DECISÃO NA REGIÃO DA FLORESTA NACIONAL DE TAPAJÓS, PARÁ

Luciane Yumie Sato

Dissertação de Mestrado do Curso de Pós-Graduação em Sensoriamento Remoto, orientada pelos Drs. Tatiana Mora Kuplich, e Yosio Edemir Shimabukuro, aprovada em 02 de março de 2012.

URL do documento original:
<http://urlib.net/8JMKD3MGP7W/3BDM3CP>

INPE
São José dos Campos
2012

Dados Internacionais de Catalogação na Publicação (CIP)

Sato, Luciane Yumie.
Sa83d Dados ópticos e SAR para classificação de cobertura da terra por árvore de decisão na região da Floresta Nacional de Tapajós, Pará / Luciane Yumie Sato. – São José dos Campos : INPE, 2012.
xxii + 86 p. ; (sid.inpe.br/mtc-m19/2012/02.23.17.05-TDI)

Dissertação (Mestrado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2012.
Orientadores : Drs. Tatiana Mora Kuplich, e Yosio Edemir Shimabukuro.

1. multisensores. 2. SAR. 3. mineração de dados. 4. ávore de decisão. 5. floresta Tropical. I.Título.

CDU 528.811(811.5)

Copyright © 2012 do MCT/INPE. Nenhuma parte desta publicação pode ser reproduzida, armazenada em um sistema de recuperação, ou transmitida sob qualquer forma ou por qualquer meio, eletrônico, mecânico, fotográfico, reprográfico, de microfilmagem ou outros, sem a permissão escrita do INPE, com exceção de qualquer material fornecido especificamente com o propósito de ser entrado e executado num sistema computacional, para o uso exclusivo do leitor da obra.

Copyright © 2012 by MCT/INPE. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, microfilming, or otherwise, without written permission from INPE, with the exception of any material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use of the reader of the work.

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

Dr. Camilo Daleles Rennó

Presidente / INPE / SJCampos - SP

Dr. Yosio Edemir Shimabukuro

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

Dra. Tatiana Mora Kuplich

Orientador(a) / INPE / SJCampos - SP

Dr. Egídio Arai

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

Dr. Elódio Sebem

Convidado(a) / UFSM / Santa Maria - RS

Este trabalho foi aprovado por:

( ) maioria simples

(X) unanimidade

Aluno (a): Luciane Yumie Sato

São José dos Campos, 02 de março de 2012

"Escolha um trabalho que você ame e não terás que trabalhar um único dia em sua vida".

Confúcio

A meus pais,
a meus irmãos,
e ao meu companheiro e amigo Vitor.

AGRADECIMENTOS

Ao Instituto Nacional de Pesquisas Espaciais - INPE, pelo apoio, oportunidade e fornecimento dos recursos e instalações.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq, pela concessão da bolsa para o desenvolvimento e realização deste trabalho.

Aos meus orientadores, Dra. Tatiana Mora Kuplich e ao Dr. Yosio Edemir Shimabukuro, pelo apoio, dedicação, conhecimentos compartilhados, paciência e motivação para realização desta pesquisa.

Aos professores, pesquisadores e funcionários do INPE, pelos conselhos e apoio durante esses dois anos. Meus agradecimentos especiais, para o pesquisador Dr. Egídio Arai pela ajuda e incentivo. Á Vera, pela terapia e ajuda.

Aos colegas e amigos da turma PG-SER 2010, companheiros de estudos e comemorações. Em especial, as minhas amigas Tathiane, pela compreensão, apoio, comemorações e amizade e a Ana Elisa, por estar todos os dias ao meu lado e por cuidar de mim. E ao amigo Rafael, pela companhia nos estudos, risadas e cafés.

Aos meus colegas de campo Ramon e Adami, pelos ensinamentos, conselhos e risadas. Aos funcionários da FLONA do Tapajós pela ótima estadia e comida nas atividades de campo.

Ao meu companheiro e amigo Vitor, pelo apoio e dedicação, sem ele esse sonho não seria realizado.

A minha família, pai, mãe e irmãos, pelo amor, por me ajudarem, incentivarem e entenderem a minha ausência em alguns momentos ao longo desses meses.

A todas as pessoas que torceram por mim, muito obrigada!

RESUMO

A floresta Amazônica ocupa uma área de aproximadamente 5 milhões de km^2, sendo responsável por concentrar grande parte da biodiversidade do planeta. Apesar de sua importância, esta região sofre constantemente com o processo de desmatamento e por isso é fonte de estudos e centro de atenção da comunidade científica mundial. A FLONA do Tapajós é uma importante referência de unidade de conservação dos recursos florestais tropicais, e é frequentemente alvo de diversos estudos. Entretanto, ainda são poucos os trabalhos que integram informações de duas ou mais fontes de dados coletados remotamente para o mapeamento de cobertura da terra dessa área. Nesse contexto, o principal objetivo deste trabalho é realizar o mapeamento da cobertura da terra da região da FLONA do Tapajós, incluindo classes de degradação florestal e de regeneração. Para isso, foi utilizada a técnica de árvore de decisão aplicada às informações provenientes de múltiplos sensores de observação terrestre para a classificação da área de estudo. Foram testadas diferentes combinações de atributos como dados de entrada no treinamento da árvore de decisão, e, com isso, foi possível avaliar a contribuição de cada informação nas classificações. Os atributos utilizados a partir de imagens ópticas foram: seis bandas provenientes do sensor Thematic Mapper (TM) do satélite Landsat-5, imagens-fração do MLME, índices de vegetação NDVI, NDWI e SAVI. Das imagens RADARSAT-2 foram utilizadas as imagens intensidade, as diferenças de fase, as componentes de espalhamento volumétrico, *double-bounce* e superficial obtidas da decomposição de alvos. Além disso, foram utilizados atributos de textura extraídos de dados ópticos e SAR. Os resultados quando comparados aos resultados que utilizaram apenas os dados TM, avaliados através do índice Kappa, mostraram que a integração de dados ópticos e SAR não apresentaram uma melhora significativa. Entretanto, a utilização de todos os atributos resultou em uma classificação com índice Kappa considerado Excelente. A utilização da técnica de árvore de decisão permitiu a integração de dados de diferentes sensores e foi computacionalmente eficiente nas classificações da cobertura florestal.

SAR AND OPTICAL DATA FOR LAND COVER CLASSIFICATION BY DECISION TREE IN THE REGION OF THE TAPAJÓS NATIONAL FOREST, PARÁ

ABSTRACT

The Amazon rainforest covers an area of approximately 5 million km^2, and is responsible for concentrating much of the planet's biodiversity. Despite its importance, this region is constantly in the process of deforestation and therefore is a source of study and the center of attention of the scientific community. The Tapajós National Forest is an important reference unit for conservation of tropical forest resources, and is often the target of several studies. However, there are few studies that integrate information from two or more sources of remotely sensed data for mapping land cover in this area. In this context, the main objective of this study is to map the land cover in the region of the Tapajós National Forest, including classes of forest degradation and regeneration. For this, we used the technique of decision tree applied to information from multiple Earth observation sensors for the classification of the study area. We tested different combinations of attributes as input data to train the decision tree, and thus it was possible to evaluate the contribution of each source of information in the classification. The attributes used from optical images were six Thematic Mapper (TM)/Landsat-5 bands, MLME-fraction images, and NDVI, NDWI and SAVI indices. We also used the RADARSAT-2 intensity images, the phase difference and volume, double-bounce and surface scattering, obtained from target decomposition technique. Furthermore, we used texture attributes extracted from optical and SAR data. The integration of optical and SAR data showed no significant improvement when compared to the results when only TM data were used as measured by the Kappa index. However, the proposed use of all attributes resulted in a Kappa Index classification considered excellent. The use of decision tree technique enabled the integration of data from different sensors and was computationally efficient in the classification of forest cover.

LISTA DE FIGURAS

Pág.

Figura 2.1 - Metodologia do trabalho “Monitoramento de cobertura da terra no estado do Acre, no oeste da Amazônia Brasileira, usando dados multitemporais de sensoriamento remoto”. 6
Figura 2.2 - A) mapa das idades da regeneração, (B) bandas do SAR SIR-C filtradas – polarização VV (vermelho – banda L), polarização HV (verde - banda C) e polarização HV (azul – banda L), (C) bandas do Landsat/TM – banda 5 (vermelho), banda 4 (verde) e banda 3 (azul) e (D) combinação das bandas do SAR SIR-C e Landsat/TM – banda 5 do TM (vermelho), banda polarização HV filtrada (verde – banda C) e banda polarização HV filtrada (azul – banda L). 9
Figura 2.3 - Estrutura de uma árvore de decisão genérica. 11
Figura 3.4 - Região onde está localizada a Flona do Tapajós. A Flona do Tapajós está delimitada pelo traço vermelho. 15
Figura 3.5 - Etapas da metodologia do trabalho. 17
Figura 3.6 - Etapas de pré-processamento dos dados. 18
Figura 3.7 - Etapas do processamento dos dados. 19
Figura 3.8 - Composição colorida em vermelho, verde e azul das imagens-fração solo, vegetação e sombra do MLME dos anos de 2008 (a) e 2009 (b). 22
Figura 3.9 - Imagens NDVI de 2008 (a) e 2009 (b). 23
Figura 3.10 - Imagens NDWI de 2008 (a) e 2009 (b). 24
Figura 3.11 - Imagens SAVI de 2008 (a) e 2009 (b). 25
Figura 3.12 - Composição colorida R (HH), G (VV) e B (HV) das imagens intensidade do ano de 2008 ascendente 27
Figura 3.13: Composição colorida R (HH), G (VV) e B (HV) das imagens intensidade do ano de 2008 descendente. 27
Figura 3.14: Composição colorida R (HH), G (VV) e B (HV) das imagens intensidade do ano de 2009 ascendente. 28
Figura 3.15: Composição colorida R (Pd), G (Pv) e B (Ps) das componentes de espalhamento da decomposição de Freeman-Durden do ano de 2008 ascendente. 29
Figura 3.16: Composição colorida R (Pd), G (Pv) e B (Ps) das componentes de espalhamento da decomposição de Freeman-Durden do ano de 2008 descendente. 29
Figura 3.17: Composição colorida R (Pd), G (Pv) e B (Ps) das componentes de espalhamento da decomposição de Freeman-Durden do ano de 2009 ascendente. 30
Figura 3.18 - Método Holdout. 34
Figura 3.19 - Distribuição dos pontos amostrados em 2008. 37
Figura 3.20 - Distribuição dos pontos amostrados em 2009. 38

Figura 3.21: Classes e amostras de uso e cobertura da terra, utilizados nas classificações das imagens. Os retângulos vermelhos destacam as áreas onde foram coletados os pixels para formar o conjunto de amostra para treinamento e validação das classificações. 39
Figura 4.22 - Primeiro atributo usado na árvore de decisão da classificação 2009a. 44
Figura 4.23 - Resultado da classificação 2009a. 44
Figura 4.24 - Detalhes da classificação 2009a da Comunidade São Jorge. 45
Figura 4.25 - Primeiro atributo usado na árvore de decisão da classificação 2009b. 47
Figura 4.26 - Resultados da classificação 2009b. 48
Figura 4.27 - Detalhes da classificação 2009b da Comunidade São Jorge. Os círculos na cor preta mostram algumas áreas que perderam informações de borda. 49
Figura 4.28 - Os dois primeiros atributos usados na árvore de decisão da classificação 2009c. 52
Figura 4.29 - Resultado da classificação 2009c. 53
Figura 4.30 - Detalhes dos erros de classificação (círculos pretos) das áreas encobertas por nuvens da classificação 2009c. 53
Figura 4.31 - Atributos usados para a separação da classe Água da classificação 2008a. 57
Figura 4.32 - Resultado da classificação 2008a. 58
Figura 4.33 - Detalhes dos erros (círculo preto) de classificação das áreas encobertas por nuvens na classificação 2008a. 59
Figura 4.34 - Atributos usados para a separação da classe Água da classificação 2008b. 62
Figura 4.35 - Resultado da classificação 2008b. 63
Figura 4.36 - Detalhes dos erros (círculo preto) da classificação de uma área encoberta por nuvens. 63
Figura 4.37 - Atributos usados para a separação da classe Água da classificação 2008c. 67
Figura 4.38 - Atributos usados para a classificação 2009d. 70

LISTA DE TABELAS

Pág.

Tabela 3.1 - Descrição dos PI`s utilizados neste trabalho..... 18
Tabela 3.2: Parâmetros das imagens RADARSAT-2. 26
Tabela 3.3 - Exemplo de um arquivo no formato ARFF do WEKA. 33
Tabela 3.4 - Exemplo do arquivo no formato CSV. 33
Tabela 3.5 - Matriz de confusão adotada para análise da classificação. 36
Tabela 4.6 - Classificações geradas para o ano de 2009. 41
Tabela 4.7 - Atributos de entrada na árvore de decisão da classificação 2009a. 42
Tabela 4.8 - Matriz de confusão da classificação 2009a. 43
Tabela 4.9 - Atributos de entrada na árvore de decisão da classificação 2009b. 46
Tabela 4.10 - Matriz de confusão da segunda classificação. 47
Tabela 4.11 - Atributos de entrada na árvore de decisão da classificação 2009c. 50
Tabela 4.11 - Conclusão 51
Tabela 4.12 - Matriz de confusão da classificação 2009c. 51
Tabela 4.13 - Classificações geradas para o ano de 2008. 54
Tabela 4.14 - Atributos de entrada no treinamento da árvore de decisão da classificação 2008a..... 55
Tabela 4.14 – Conclusão 56
Tabela 4.15 - Matriz de confusão da classificação 2008a. 56
Tabela 4.17 - Matriz de confusão da classificação 2008b. 61
Tabela 4.18 - Classificações geradas para o ano de 2008 e 2009. 64
Tabela 4.19 - Atributos de entrada no treinamento da arvore de decisão da classificação 2008c 65
Tabela 4.19 - Conclusão 66
Tabela 4.20 - Matriz de confusão da classificação 2008c 66
Tabela 4.21 - Atributos de entrada no treinamento da arvore de decisão da classificação 2009d..... 68
Tabela 4.21 - Conclusão 69
Tabela 4.22 - Matriz de confusão da classificação 2008c. 69
Tabela 4.23 - Resumo dos resultados das classificações. 71

LISTA DE SIGLAS E ABREVIATURAS

6S	*Second Simulation of the Satellite Signal in the Solar Spectrum*
CART	*Classification and regression tree*
CRA	Centro Regional da Amazônia
CSV	Comma-separated values
CCD	*Charge-Coupled Device*
ENVI	*Environment for Visualizing Images*
ETM+	*Enhanced Thematic Mapper Plus*
FLONA	Floresta Nacional
GLCF	*Global Land Cover Facility*
ICMBio	Instituto Chico Mendes para Conservação da Biodiversidade
INPA	Instituto Nacional de Pesquisas da Amazônia
INPE	Instituto Nacional de Pesquisas Espaciais
LBA	Experimento de Grande Escala da Biosfera-Atmosfera na Amazônia
LEGAL	Linguagem Espacial para Geoprocessamento Algébrico
MLME	Modelo Linear de Mistura Espectral
MODIS	*MODerate-Resolution Imaging Spectroradiometer*
MSS	*Multispectral Scanner Subsystem*
NASA	*National Aeronautics and Space Administration*
ND	Número digital
NDVI	*Normalized Difference Vegetation Index*
NDWI	*Normalized Difference Water Index*
P_d	*Double-bounce*
PI's	Planos de informações
PRODES	Monitoramento da Floresta Amazônica Brasileira por Satélite
P_s	Espalhamento superficial
P_v	Espalhamento volumétrico
REDD	*Reduction of Emissions from Deforestation and Degradation*
SAR	*Synthetic Aperture Radar*
SAVI	*Soil Adjusted Vegetation Index*

SIG	Sistema de Informações Geográficas
SPOT	*Satellite pour L'observation de la Terre*
SRTM	*Space Shuttle Radar and Topography Mission*
DEM	*Digital Elevation Model*
TM	*Thematic Mapper*
WEKA	*Waikato Environment for Knowledge Analysis*

SUMÁRIO

		Pág.
1	INTRODUÇÃO	1
2	REVISÃO BIBLIOGRÁFICA	5
2.1.	Dados multisensores	5
2.2.	Classificação	7
2.2.1.	Atributos de Textura	8
2.3.	Árvore de decisão	9
3	MATERIAIS E MÉTODOS	15
3.1.	Área de estudo	15
3.1.1.	Clima	16
3.1.2.	Vegetação	16
3.2.	Procedimentos metodológicos	17
3.2.1.	Construção e manipulação do banco de dados	17
3.2.2.	Pré-processamento dos dados	18
3.2.3.	Processamento dos dados	19
3.2.3.1.	Correção Atmosférica	19
3.2.3.2.	Normalização Radiométrica	20
3.2.3.3.	Modelo Linear de Mistura Espectral – MLME	21
3.2.3.4.	Índices de Vegetação	22
3.2.3.5.	Dados SAR	25
3.2.3.6.	Árvore de decisão e classificação	30
3.2.3.6.1.	Preparação dos dados para o sistema WEKA	32
3.2.3.6.2.	Validação cruzada pelo Método Holdout	33
3.2.3.6.3.	Geração de árvore de decisão para o ENVI usando o WEKA	35
3.2.4.	Análise e validação dos resultados	35
3.2.5.	Dados de campo	36
4	RESULTADOS E DISCUSSÕES	41
4.1.	Classificações	41
4.1.1.	Classificações 2009	41
4.1.1.1.	Classificação da imagem Landsat-5 de 2009	41
4.1.1.2.	Classificação da imagem Landsat-5 e dos atributos de textura de 2009	45

4.1.1.3. Classificação das imagens Landsat-5 e RADARSAT-2 de 2009... 49
4.2. Classificação com dados de 2008 54
4.2.1. Classificação da imagem Landsat-5 e RADARSAT-2/Descendente de 2008 54
4.2.2. Classificação da imagem Landsat-5 e RADARSAT-2/Ascendente do ano de 2008 59
4.3. Classificação dos anos de 2008 e 2009 com todos os atributos ... 64
4.3.1. Classificação da imagem do ano de 2008 com todos os atributos 64
4.3.2. Classificação da imagem do ano de 2009 com todos os atributos 67
5 CONCLUSÕES E RECOMENDAÇÕES 73
REFERÊNCIAS BIBLIOGRÁFICAS 75

1 INTRODUÇÃO

Nas últimas décadas, o sensoriamento remoto vem se destacando pela possibilidade de fornecer resultados rápidos, pela facilidade de acesso às informações e por contribuir na avaliação de diferentes fatores ligados ao ambiente e às atividades humanas. Além disso, esta tecnologia é um dos meios mais eficientes para o monitoramento e mapeamento de grandes extensões da superfície terrestre, devido à obtenção de dados de diversos instrumentos de observação da Terra em diferentes níveis de aquisição (terrestre, aéreo e orbital) e resoluções (espacial, temporal, espectral e radiométrica) (SHIMABUKURO et al., 2009; ESPÍRITO-SANTO, 2003).

As atividades humanas causam mudanças na cobertura da terra, que podem acarretar em alterações nos ciclos hidrogeoquímicos e climáticos do sistema terrestre (ANDERSON et al., 2005). Por estes motivos, o mapeamento e monitoramento da cobertura florestal e do uso da terra são de extrema importância, pois permitem analisar as alterações da superfície terrestre e auxiliam na elaboração e na tomada de decisões para a conservação e preservação dos ecossistemas (SHIMABUKURO et al., 2003).

O bioma Amazônia é objeto de estudos e centro de atenção da comunidade científica mundial, pois as florestas tropicais são responsáveis por abrigar uma das maiores biodiversidades do mundo. Além disso, apresenta um papel fundamental na ciclagem da água, armazenamento de carbono e na regulação do clima local e regional (CLEMENTE; HIGUCHI, 2006; NOBRE et al., 2007).

A Floresta Nacional (FLONA) do Tapajós é uma importante referência de unidade de conservação dos recursos florestais tropicais e é frequentemente alvo de diversos estudos. Destes, o projeto LBA (Experimento de Grande Escala da Biosfera-Atmosfera na Amazônia), uma parceria do Instituto Nacional de Pesquisas Espaciais (INPE), Instituto Nacional de Pesquisas da Amazônia (INPA), a National Aeronautics and Space Administration (NASA) e outras instituições dos Estados Unidos e Europa, destaca-se na produção de novos conhecimentos quanto ao funcionamento climatológico, ecológico,

biogeoquímico e hidrológico, bem como a compreensão dos impactos das mudanças do uso da terra (LBA, 1995; IBAMA, 2004).

Mesmo que o LBA utilize dados coletados remotamente para várias aplicações, ainda são poucos os trabalhos que integram informações de duas ou mais fontes de dados para o mapeamento temático da cobertura vegetal da FLONA do Tapajós (SHIMABUKURO, 2003; ANGELIS, 2000; ESPÍRITO-SANTO, 2003; IBAMA, 2004; LBA, 1995).

Neste contexto, uma nova metodologia que utilize informações multitemporais de multisensores associadas a uma classificação de imagens baseada em regras, como a árvore de decisão, pode contribuir para uma eficiente classificação das diferentes tipologias florestais e de cobertura da terra na região da FLONA do Tapajós.

Para a elaboração desta dissertação foram levantadas as seguintes hipóteses:

1) A utilização de informações temporais, espectrais e espaciais provenientes de multisensores permitirá uma classificação mais precisa da cobertura e uso da terra da região da FLONA do Tapajós.

2) Uma abordagem de mineração de dados, como a árvore de decisão, contribuirá para a caracterização com alto grau de detalhes e elevada acurácia das diferentes tipologias florestais presentes na FLONA do Tapajós, assim como seus estados de conservação.

Dadas estas hipóteses e com intuito de inovar metodologicamente, este trabalho tem como principal objetivo realizar o mapeamento detalhado da cobertura da terra da região da FLONA do Tapajós, incluindo classes de floresta degradada e de regeneração. Neste mapeamento será utilizada a técnica de árvore de decisão aplicada às informações provenientes de múltiplos sensores de observação terrestre para a classificação da área de estudo.

Para atender o objetivo geral deste trabalho, foram definidos os seguintes objetivos específicos:

- Reunir informações de diferentes fontes de dados de sensoriamento remoto (Landsat-5/TM e RADARSAT-2/SAR) e dados de apoio ou verdade terrestre (dados de campo) em um sistema de informações geográficas (SIG).

- Desenvolver, baseada em árvore de decisão, uma classificação semi-automática para a separação das diferentes tipologias florestais da FLONA do Tapajós, sendo incluídas as áreas que estão sobre o processo de degradação florestal e de regeneração.

- Realizar o mapeamento temático da cobertura florestal através da integração de dados multitemporais de multisensores.

2 REVISÃO BIBLIOGRÁFICA

2.1. Dados multisensores

As informações coletadas por um único sensor podem apresentar algumas limitações quando se deseja atingir um resultado de boa qualidade para uma dada aplicação. Na maior parte das vezes, uma maior acurácia pode ser alcançada através da combinação de dados provenientes de diferentes sensores. Esta integração apresenta um grande potencial para detectar mudanças do uso da terra e antecipar os impactos de forma mais consistente em relação à utilização de apenas uma fonte de dados (HUANG et al., 2007).

Com a grande variedade de sensores de coleta de dados da superfície terrestre com características espaciais, espectrais, temporais e radiométricas diferentes, existem, atualmente, desafios para a integração destes dados. Muitos cientistas consideram a combinação de dados de múltiplos sensores uma estratégia eficiente para a discriminação e estudo dos alvos que compõem a superfície terrestre (SHIMABUKURO et al., 1991; KUPLICH et al., 2000; ZURITA-MILLA et al., 2009).

Cada sensor possui diferentes especificações que garantem vantagens ou desvantagens dependendo da aplicação de interesse. Por exemplo, o sensor MODerate-Resolution Imaging Spectroradiometer (MODIS), do satélite Terra, obtém informações da superfície terrestre com resolução espacial de 250m, sendo considerada baixa em relação ao sensor Thematic Mapper (TM) que possui resolução espacial de 30m. Entretanto, o sensor MODIS fornece dados quase diários em nível global, permitindo acompanhar a dinâmica de ocupação da terra. Por outro lado o sensor TM, com uma resolução espacial mais refinada, realiza a coleta de dados em menor frequência (SHIMABUKURO et al., 1991).

Shimabukuro et al. (2010) utilizaram com sucesso dados provenientes de multisensores para o para o monitoramento das mudanças da cobertura vegetal do Estado do Acre. Neste estudo foram utilizados dados do sensor Multispectral Scanner Subsystem (MSS), TM, Enhanced Thematic Mapper Plus

(ETM+) do satélite Landsat e dados do sensor MODIS. A Figura 2.1 apresenta a metodologia utilizada neste estudo:

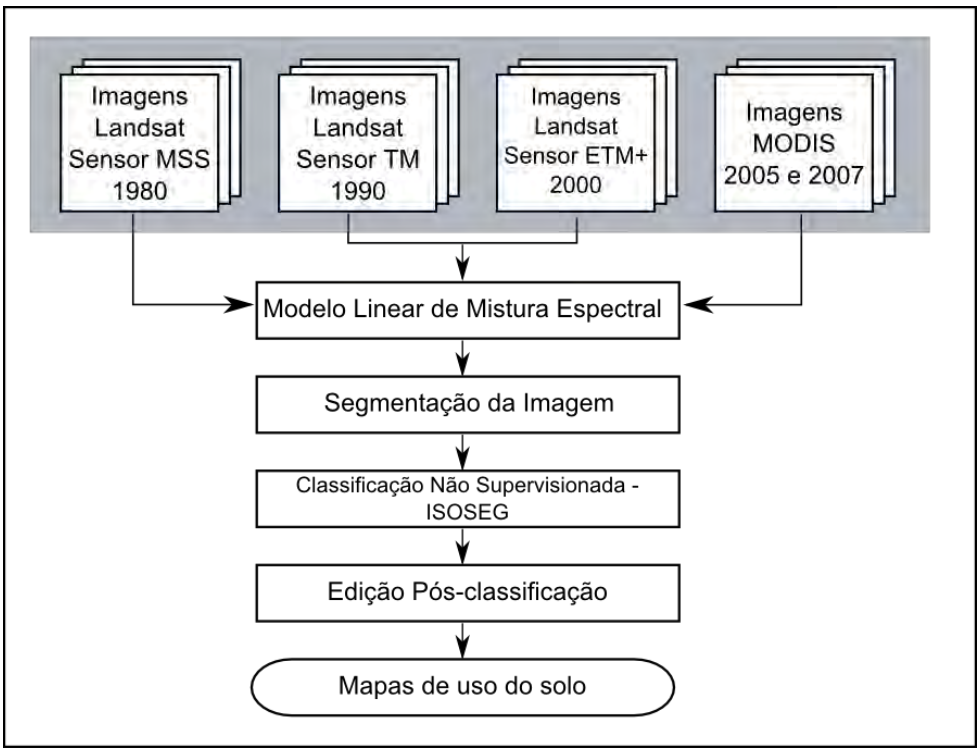

Figura 2.1 - Metodologia do trabalho "Monitoramento de cobertura da terra no estado do Acre, no oeste da Amazônia Brasileira, usando dados multitemporais de sensoriamento remoto".
Fonte: Adaptado de Shimabukuro et al. (2010).

Ainda dentro da temática de dados de multisensores, outros trabalhos mostram que a combinação de dados contribui para uma melhor classificação da cobertura vegetal. Roberts et al. (2002) utilizaram uma série temporal de 33 imagens do sensor MSS e TM, para gerar um mapa com as classes de uso da terra e os estágios de regeneração da cobertura vegetal, de uma região central do Estado de Rondônia. Além do uso de dados multisensores, os autores utilizaram a técnica baseada em regras por árvore de decisão para a classificação das imagens.

Wolter e Townsend (2011) utilizaram dados SAR (Synthetic Aperture Radar) e dados ópticos do Landsat e do SPOT (Satellite pour L'observation de la Terre) para analisar a distribuição de dez espécies florestais e dois gêneros dominantes em uma floresta da província de Ontário no Canadá. Estes autores concluíram que o uso de dados de multisensores é uma eficiente ferramenta

para mapear e analisar a quantidade de indivíduos de algumas espécies florestais da região em estudo.

Outra importante aplicação de dados ópticos associados a dados SAR aparece em estudos voltados para florestas tropicais. Este conjunto de dados permite explorar ao máximo as informações dos alvos presentes na superfície terrestre, pois os dados ópticos ficam muitas vezes limitados com a quase constante presença de nuvens sobre estas áreas (ANGELIS, 2000; GUERRA et al., 2010; KUPLICH, 2003).

2.2. Classificação

O mapeamento acurado da superfície terrestre tem grande importância para a compreensão das interações e das relações entre os fenômenos naturais e as atividades humanas. Além disso, auxilia na tomada de decisões para a conservação e utilização adequada dos recursos naturais. O mapeamento do uso e cobertura da terra pode ser realizado através da classificação de imagens, pela identificação e distinção dos objetos que compõem a superfície terrestre (WANG, 2002).

Lu et al. (2004) realizaram uma revisão bibliográfica, comparando diversas técnicas de classificação de imagens para a detecção e o acompanhamento das mudanças no uso e cobertura da terra. Para estes autores, determinar qual a melhor técnica ou o melhor algoritmo para a análise das alterações dos objetos ao longo do tempo não é uma tarefa trivial. Por isso, é importante conhecer as características da área de estudo e quais informações que serão utilizadas para auxiliar na escolha da técnica adequada e na obtenção de resultados de boa qualidade.

Atualmente, criar metodologias para obter mapas de uso e cobertura da terra com bons resultados de maneira rápida e fácil é um desafio. Estes mapas, em especial os de florestas tropicais, são importantes para auxiliar na avaliação da extensão das mudanças na cobertura da terra e na estimativa da contribuição destas florestas no sequestro de carbono regional e global, que reduzem os gases do efeito estufa (HOUGHTON, 2003; POWELL et al., 2004). O mecanismo REDD (Redução de Emissão por Desmatamento e Degradação), é

uma iniciativa neste sentido e que pode ser beneficiado pelo aumento da capacidade de mapeamento da cobertura de florestas tropicais (ANGELSEN et al., 2009; PACHECO et al., 2011).

2.2.1. Atributos de Textura

As medidas de textura são importantes características utilizadas nas classificações de imagens, pois aumentam o desempenho dos classificadores digitais (ULABY et al., 1986; MARCEAU et al., 1989). A textura é um atributo amplamente utilizado em classificações que apresentam classes com comportamento espectral muito próximo, como por exemplo, para separar as classes de floresta nativa e floresta plantada (CLARO, 2010).

O conceito de textura de uma imagem não é um consenso, por isso, sua quantificação é dificultada (ULABY et al., 1986; MARCEAU et al., 1989; LOCK; KIRCHNER, 1997). Segundo Haralick et al. (1973), textura de uma imagem representa a informação da distribuição espacial das variações de tonalidades de um objeto.

Rennó (1995) avaliou o uso de medidas de textura para a classificação de áreas agrícolas. Nesse trabalho foi utilizada uma imagem SIR-C/SAR e a área de estudo fica localizada em uma região semi-árida do município de Petrolina, no estado de Pernambuco. Os resultados obtidos pelo autor comprovam que a utilização de medidas de textura melhora a classificação de imagens SAR, pois com o uso de 15 atributos de textura o índice Kappa foi maior do que 0,90.

No trabalho desenvolvido por Kuplich (2006), foram classificadas através de redes neurais áreas de regeneração florestal e para isso, foram utilizados dados SAR (JERS-1, SIR-C e XSAR), TM e atributos de textura. Nesse estudo, os principais resultados das classificações mostraram que o uso de dados SAR filtrados apresentou resultados melhores na discriminação da cobertura da terra, quando comparados com a classificação que utilizou bandas SAR não filtradas. Além disso, a classificação obtida através do uso dos dados TM, dados SAR e bandas de textura da imagem SAR, permitiu discriminar as área de regeneração florestal com até cerca de 18 anos de idade. A Figura 2.2 apresenta os resultados obtidos neste estudo.

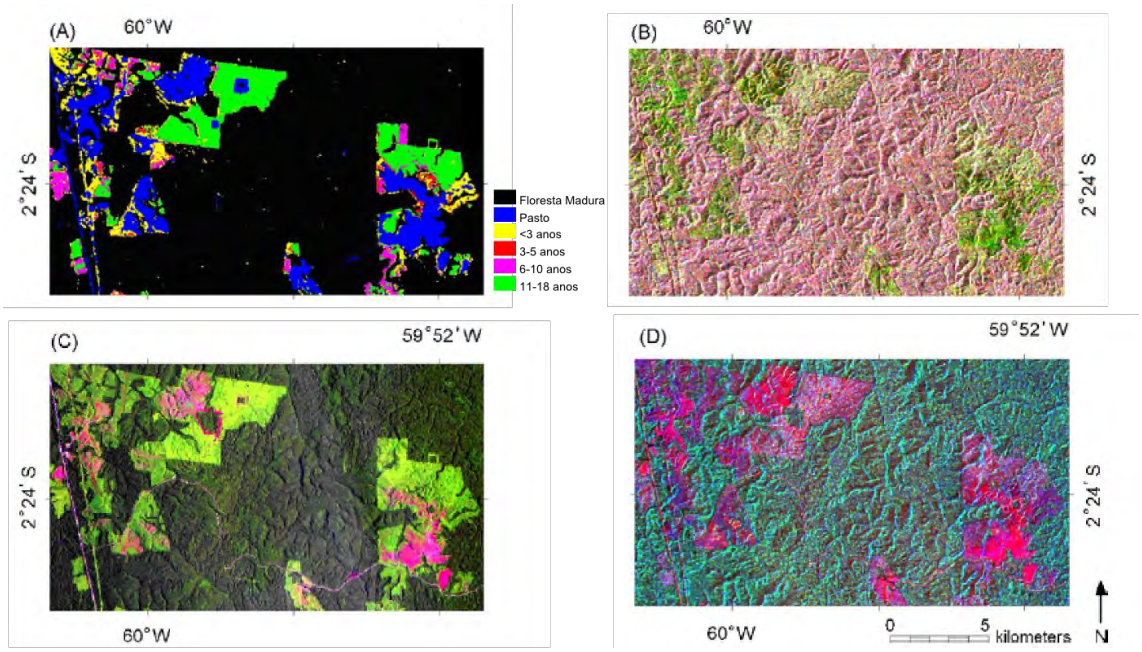

Figura 2.2 - A) mapa das idades da regeneração, (B) bandas do SAR SIR-C filtradas – polarização VV (vermelho – banda L), polarização HV (verde - banda C) e polarização HV (azul – banda L), (C) bandas do Landsat/TM – banda 5 (vermelho), banda 4 (verde) e banda 3 (azul) e (D) combinação das bandas do SAR SIR-C e Landsat/TM – banda 5 do TM (vermelho), banda polarização HV filtrada (verde – banda C) e banda polarização HV filtrada (azul – banda L).
Fonte: Adaptado de Kuplich (2006).

Murray et al. (2010), em seu estudo realizaram a classificação com medidas de textura de uma imagem IKONOS, para a determinação das comunidades vegetais de uma ilha localizada no oceano Índico. Os autores realizaram três combinações de classificações. A primeira classificação foi realizada com o uso das bandas da imagem IKONOS, na segunda classificação, foram utilizadas as informações de textura e na terceira classificação, foram combinadas as medidas de textura e as bandas espectrais da imagem. O melhor resultado foi obtido através da combinação das bandas da imagem IKONOS mais os atributos de textura, com uma melhora de 6% na classificação e com uma precisão total na classificação de 84,7%.

2.3. Árvore de decisão

Nos últimos anos, técnicas de mineração de dados estão sendo cada vez mais empregadas para o mapeamento da cobertura florestal e de uso da terra em escalas locais, regionais e globais. Esta tecnologia consiste em extrair conhecimentos de grandes volumes de dados, através de técnicas inteligentes

para extração de padrões de interesse (RUSSELL; NORVING, 2009; HANSEN et al., 1996; IM; JENSEN, 2005; HANSEN et al., 2008; GONG et al., 2011).

A árvore de decisão é uma ferramenta que se destaca no contexto de classificação de imagens, fornecendo aumento da acurácia na classificação de imagens e a possibilidade de classificar conjuntos distintos de dados (HANSEN et al., 1996; IM; JENSEN, 2005; HANSEN et al., 2008; GONG et al., 2011).

A árvore de decisão, também denominada como árvore binária, pode ser explicada como uma função que apresenta como dado de entrada um vetor de valores de atributos e uma decisão como valor de saída: sim ou não (BREIMAN et al., 1984).

Os valores de entrada da árvore de decisão podem ser discretos ou contínuos. As decisões são tomadas através de uma sequência de testes, onde os resultados serão verdadeiros ou falsos. Sendo assim, a árvore de decisão é considerada com uma estrutura de classificação booleana (RUSSELL; NORVIG, 2009).

O funcionamento básico de uma árvore de decisão ocorre através do particionamento de um conjunto de dados de treinamento em subconjuntos de forma recursiva. A separação dos dados de treinamento ocorre até que cada subconjunto esteja homogêneo, isto é com casos de uma única classe (WITTEN et al., 2011)

A fácil interpretação dos resultados obtidos na árvore de decisão está entre as principais vantagens na utilização desta técnica para a classificação de imagens, quando comparada aos métodos tradicionais de classificação. Devido ao fato da classificação ocorrer de forma explícita, sua interpretação torna-se facilitada. Além destas vantagens, os resultados são gerados rapidamente, pois as árvores de decisão possuem eficiência computacional (HANSEN et al., 1996; FRIELD; BRODLEY, 1997).

Uma árvore de decisão é estruturada de forma hierárquica de cima para baixo, sendo formada basicamente por nós internos e folhas que são conectados por ramos. Os nós são as regiões onde ocorrem as tomadas de decisões, cuja

função é determinar, através de um teste lógico, qual será o próximo nó descendente (nó filho). O primeiro nó da árvore é chamado de nó raiz e os nós que não apresentam nós descendentes e estão associados a um rótulo ou a um valor são chamados nós folhas (FRIELD; BRODLEY, 1997; SOBRAL, 2005; POZZER, 2006; AITKENHEAD, 2008).

A formação ideal de uma árvore de decisão é manter as tomadas de decisões mais importantes próximas ao nó raiz, pois desta forma busca-se resolver o problema com o menor número possível de regras (POZZER, 2006). A estrutura típica de uma árvore de decisão é apresentada na Figura 2.3.

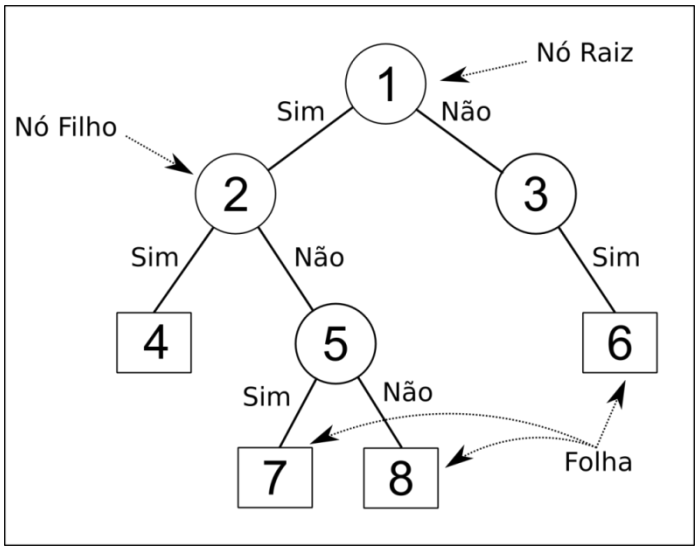

Figura 2.3 - Estrutura de uma árvore de decisão genérica.

A simplicidade de sua estrutura e os resultados satisfatórios têm motivado a comunidade científica no uso de árvores de decisão para o mapeamento e monitoramento da cobertura florestal e de uso da terra através de imagens obtidas por satélites.

Sesnie et al. (2010) utilizaram árvore de decisão aplicada a uma série temporal para acompanhar as alterações do uso e cobertura da terra de uma floresta tropical úmida da Costa Rica. Neste estudo, foram utilizados dados do sensor TM e dados SRTM-DEM (*Space Shuttle Radar and Topography Mission - Digital Elevation Model)* para separar as diferentes tipologias florestais. Os

autores concluíram que com o uso da árvore de decisão foi possível classificar o complexo ambiente de uma floresta tropical úmida com boa acurácia.

Pouliot et al. (2009) utilizaram a técnica de classificação por árvore de decisão em um estudo multitemporal com imagens MODIS de 250m de uma floresta no Canadá. O objetivo deste estudo foi monitorar em tempo quase real as mudanças nas florestas desta região. Na primeira classificação, com dados de 2005 a 2006, a precisão do mapa de classificação atingiu um Kappa de 0,91. Na segunda classificação, com dados de 2001 a 2005, o índice Kappa reduziu para 0,80. Entretanto, quando os autores adicionaram a esta segunda classificação dados contextuais e realizaram a normalização radiométrica das imagens, o Kappa atingiu o valor de 0,90.

Davrache et al. (2010) utilizaram árvore de decisão para monitorar áreas úmidas próximas ao Mar Mediterrâneo no sul da França. Neste estudo, foram utilizados dados de campo, imagens SPOT-5 com 10m de resolução espacial e alguns índices de vegetação. Este estudo utilizou uma série temporal de 2005 a 2006 e obteve uma boa classificação da região de estudo.

Latorre et al. (1998) desenvolveram uma metodologia de integração de multisensores para um sistema de monitoramento da Amazônia e propuseram um sistema baseado na técnica de árvore de decisão. Neste estudo, os autores utilizaram dados obtidos com diferentes resoluções espaciais, temporais e radiométricas. Para isso, foram utilizados dados do sensor TM do Landsat-5, do sensor ETM+ do Landsat-7, do sensor MODIS do Terra e da câmera C*harge-Coupled Device* (CCD) do CBERS-2.

Ainda no trabalho desenvolvido por Latorre et al. (1998), foram utilizados como dados de entrada para a árvore de decisão um conjunto de métricas multitemporais e de NDVI do sensor MODIS (variáveis independentes), dados de treinamento obtidos da classificação das imagens dos sensores TM, ETM+ e CCD (variáveis dependentes). Foram determinadas três classes de cobertura da terra em uma área de Floresta Amazônica: floresta; não floresta; e áreas inundáveis. Os resultados obtidos neste estudo foram comparados com os

dados provenientes do projeto PRODES (Monitoramento da Floresta Amazônica Brasileira por Satélite), sendo considerados satisfatórios.

3 MATERIAIS E MÉTODOS

3.1. Área de estudo

A FLONA do Tapajós está localizada no oeste do estado do Pará entre os paralelos 2º 45’ e 4º 10’ de latitude sul e entre os meridianos de 54º 45’ e 55º 30’ de longitude oeste (Figura 3.4). Esta FLONA foi criada pelo decreto nº. 73.684 de 19 de fevereiro de 1974, sendo seus limites estabelecidos ao norte com o paralelo que cruza o km 50 da rodovia Cuiabá-Santarém (BR-163), ao sul com a rodovia Transamazônica e os rios Cupari e Cuparitinga (ou Santa Cruz), a leste com a rodovia Cuiabá-Santarém (BR-163) e a oeste com o rio Tapajós.

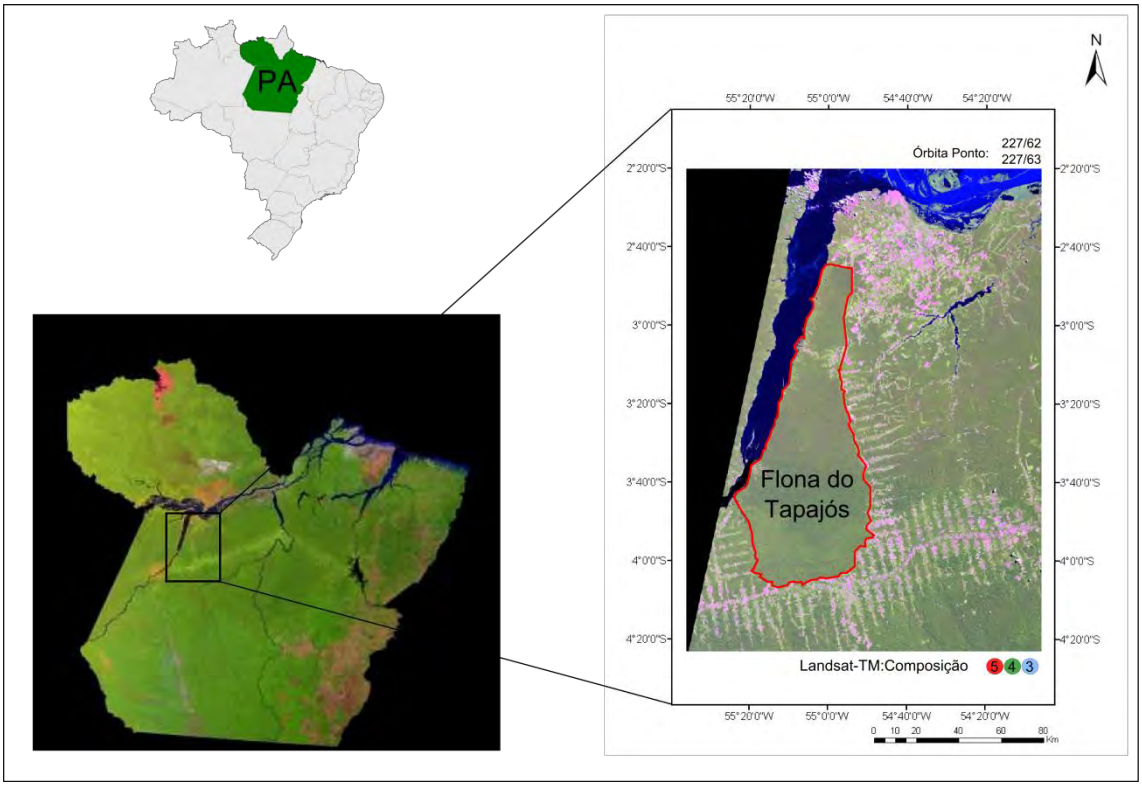

Figura 3.4 - Região onde está localizada a Flona do Tapajós. A Flona do Tapajós está delimitada pelo traço vermelho.

A FLONA do Tapajós é uma Unidade de Conservação com aproximadamente 600.000 ha, administrada pelo Instituto Chico Mendes para Conservação da Biodiversidade (ICMBio) conforme a lei nº. 11.516 de 28 de agosto de 2007.

Das 66 Florestas Nacionais existentes no país, a FLONA do Tapajós foi a 13ª criada e é definida

> Como uma área com cobertura florestal de espécies predominantemente nativas e tem como objetivo básico o uso múltiplo sustentável de recursos florestais e a pesquisa científica, com ênfase em métodos para exploração sustentável de florestas nativas (SNUC, 2004).

A escolha desta área é motivada pela disponibilidade de uma ampla base de dados de inventário florestal adquirida por diversos estudos científicos. A dinâmica da área inclui atividades de desmatamento que ocorrem no entorno da FLONA do Tapajós, embora seja restrito o uso dos recursos florestais em Unidades de Conservação. Os desmatamentos das florestas primárias desta área ocorrem para a ocupação humana ao longo da BR-163, seguindo o padrão conhecido como “espinha de peixe”. Este padrão é caracterizado pela abertura de estradas vicinais para o estabelecimento de pequenas propriedades rurais, que após alguns anos são abandonadas quando se tornam improdutivas. Estas ações resultam na formação de capoeiras, criando paisagens compostas por mosaicos de vegetação secundária (NEPSTAD, 2001; ALVES, 2002; ESPÍRITO-SANTO, 2003; ARAGÃO, 2004).

3.1.1. Clima

O clima da região é classificado no sistema de Köppen como AmW- tropical chuvoso, apresentando temperatura média anual de 26ºC e variação térmica anual inferior a 5ºC. As chuvas são distribuídas ao longo dos anos com regime pluviométrico variando de 1.700 a 2.000 mm. No bioma Amazônia, o clima divide-se em duas estações bem características: um período chuvoso e outro seco. O predomínio de chuvas ocorre durante os meses de dezembro a maio (ESPÍRITO-SANTO; 2003; ARAGÃO, 2004; SERVELLO, 2010).

3.1.2. Vegetação

A maior parte da área da FLONA do Tapajós é constituída por Floresta Tropical Densa. Espírito-Santo (2003), a partir de dados multitemporais de multisensores e com dados de inventários florestais, identificou três padrões fitogeográficos predominantes na FLONA do Tapajós. No primeiro padrão, na

porção norte, ocorre o predomínio do babaçu (*Orbignya phalerata* Mart.). O segundo padrão, na porção central, é caracterizado pela presença de uma floresta densa, semelhante às espécies entre o alto e baixo platô. Por fim, o terceiro padrão, na porção sul, possui alta diversidade florística e a ocorrência frequente de palmeiras como mombaca (*Astrocaryum mumbaca* Mart.), tucumã (*Astrocaryumvulgare* Mart.), inajá (*Maximiliana martiana* Karst.), entre outras (exceto o babaçu).

3.2. Procedimentos metodológicos

As atividades deste trabalho foram divididas em quatro etapas: (1) construção e manipulação de um banco de dados da FLONA do Tapajós em um SIG; (2) pré-processamento dos dados; (3) processamento e tratamento dos dados; e (4) análise das informações e resultados. A Figura 3.5 sintetiza as etapas deste trabalho.

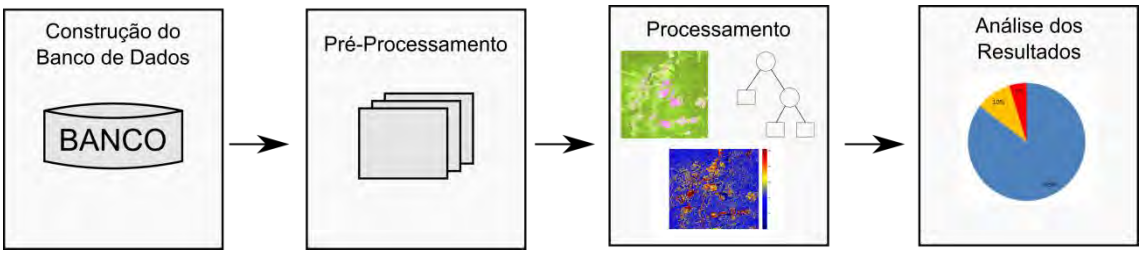

Figura 3.5 - Etapas da metodologia do trabalho.

Para todas as atividades de processamento de imagens, foi utilizada uma máquina equipada com processador Core i7, 4 Gigabytes de memória RAM e 1Terabytes de disco rígido. As diferentes atividades foram cronometradas para avaliar o desempenho do sistema utilizado na geração dos resultados.

3.2.1. Construção e manipulação do banco de dados

Um conjunto de imagens, pontos de GPS e levantamentos de campo da área de estudo foram adicionadas a um banco de dados geográficos. Para isso, foi utilizado o aplicativo SPRING 4.3.3, onde foi criado um projeto para a integração de todos os planos de informações (PI's) do banco de dados. A Tabela 3.1 lista os PI's utilizados neste trabalho.

Tabela 3.1 - Descrição dos PI`s utilizados neste trabalho

PI's	Categoria	Descrição	Aquisições
Imagens TM (2000, 2008 e 2009)	Imagem	Imagem registrada, com correção atmosférica e normalizada radiometricamente	05/09/2000 23/06/2008 28/07/2009
Imagens RADARSAT-2 (2008 e 2009)	Imagem	Modo ScanSAR e polarimétrica	24/09/2008 12/10/2009 18/10/2008
Pontos de GPS	Temático	Pontos adquiridos em trabalhos de campo	2008 e 2009
Levantamento de Campo	Cadastral	Pontos de verdade terrestre coletados no campo	2008 e 2009

3.2.2. Pré-processamento dos dados

Após a criação do banco de dados, foi feito o registro das imagens para garantir que os pixels das imagens sejam referentes às mesmas áreas. Para essa atividade, foi utilizada como referência uma imagem Geocover de mesma órbita/ponto, disponível na base de dados da *Global Land Cover Facility* (GLCF) da NASA.

Após o registro das imagens, foram elaborados mosaicos de imagens TM Landsat-5 que cobrem toda a área da FLONA do Tapajós. Sobre esses mosaicos, foram georreferenciados os dados de campo. O fluxograma mostrado na Figura 3.6 apresenta as etapas de pré-processamento dos dados.

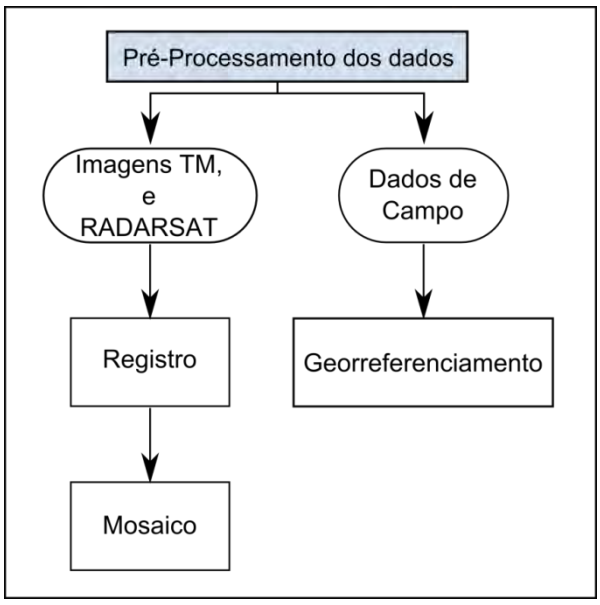

Figura 3.6 - Etapas de pré-processamento dos dados.

3.2.3. Processamento dos dados

Os dados utilizados para o treinamento das árvores de decisão e posterior classificação das imagens, foram obtidos a partir de processamentos realizados sobre as imagens Landsat-5/TM e SAR. Estes processamentos foram divididos em sete etapas: (1) correção atmosférica; (2) normalização radiométrica; (3) aplicação do modelo linear de mistura espectral (MLME); (4) cálculo dos índices de vegetação; (5) processamento das imagens SAR; (6) árvore de decisão; e (7) classificação das imagens. As etapas do processamento dos dados são apresentadas na Figura 3.7.

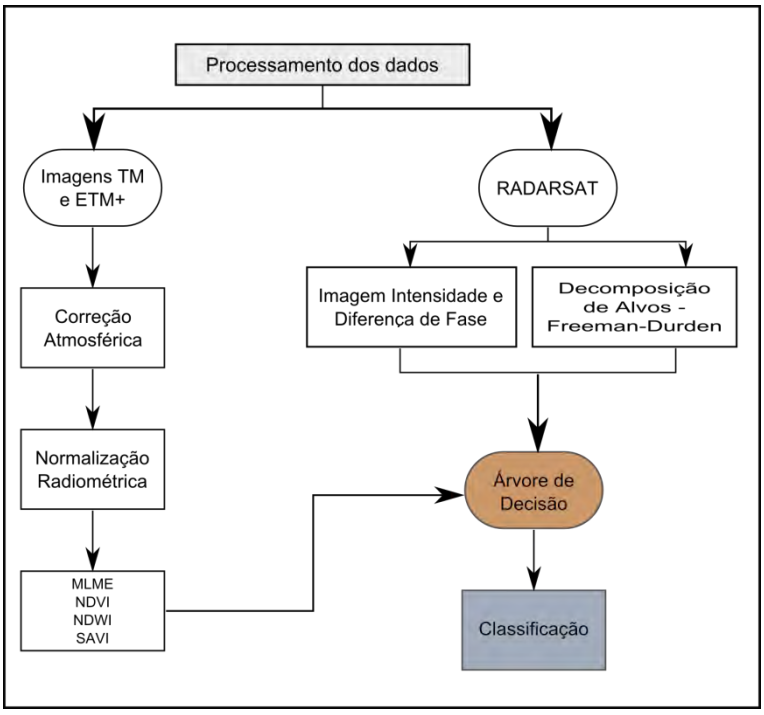

Figura 3.7 - Etapas do processamento dos dados.

3.2.3.1. Correção Atmosférica

A correção atmosférica é uma importante etapa para a caracterização espectral dos objetos que compõem a superfície terrestre e para a realização de operações com o uso de imagens de diferentes bandas espectrais ou de diferentes sensores. Esta correção busca eliminar os efeitos causados pelos gases presentes na atmosfera, que podem absorver ou espalhar a radiação eletromagnética (PONZONI; SHIMABUKURO, 2010).

Neste trabalho, são utilizadas imagens multitemporais TM do satélite Landsat-5 que passaram pelo processo de correção atmosférica. Para isso, estas imagens foram convertidas de números digitais (ND) para valores de reflectância aparente. Para esta transformação, foi utilizada a equação da radiância no topo da atmosfera ($\boldsymbol{L_\lambda}$) e a equação da reflectância aparente ($\boldsymbol{\rho_\lambda}$), com o uso da ferramenta Linguagem Espacial para Geoprocessamento Algébrico (LEGAL) do programa SPRING 4.3.3.

Para o cálculo da $\boldsymbol{L_\lambda}$foram utilizados os parâmetros extraídos do trabalho de Chander et al. (2009) e a equação

$$\boldsymbol{L_\lambda = \left(\frac{Lmax_\lambda - Lmin_\lambda}{QCALMAX}\right) QCAL} \quad (3.1)$$

onde $\boldsymbol{Lmax_\lambda}$ é a radiância máxima quando QCAL = QCALMAX; $\boldsymbol{Lmin_\lambda}$ é a radiância mínima quando QCAL = 0; o $\boldsymbol{QCAL}$ são os valores de números digitais referentes e proporcionais a $\boldsymbol{L_\lambda}$; e $\boldsymbol{\lambda}$ é a faixa espectral.

A $\boldsymbol{\rho_\lambda}$ é obtida pela equação

$$\boldsymbol{\rho_\lambda = \frac{\pi * L_\lambda * d^2}{Esun_\lambda * cos(\theta_s)}} \quad (3.2)$$

onde $\boldsymbol{d}$ é a distância Terra-Sol em unidades astronômicas;$\boldsymbol{Esun_\lambda}$ é a irradiância solar no topo da atmosfera; o $\boldsymbol{\theta_s}$ é o ângulo zenital solar no horário de aquisição da imagem; e $\boldsymbol{\lambda}$ é a faixa espectral.

Após esta etapa, foi aplicado o modelo 6S (Second Simulation of the Satellite Signal in the Solar Spectrum) de transferência radiativa (VERMOTE et al., 1997) para a realização da correção atmosférica das imagens. Finalizado este procedimento, os valores de reflectância aparente foram convertidos para valores de reflectância de superfície.

3.2.3.2. Normalização Radiométrica

Após a etapa de correção atmosférica, a imagem TM de 2000 foi usada como referência para a normalização radiométrica das demais imagens Landsat-5. Este procedimento foi realizado pelo método proposto por Hall et al. (1991), com o uso da ferramenta LEGAL do aplicativo SPRING 4.3.3.

O processo de normalização radiométrica consiste basicamente em compatibilizar a iluminação e as condições atmosféricas dos dados adquiridos. Em estudos multitemporais de multisensores este procedimento deve ser realizado com intuito de tornar os dados compatíveis entre si.

3.2.3.3. Modelo Linear de Mistura Espectral – MLME

A fim de realçar as áreas de uso e cobertura da terra da área de estudo, foi aplicado o MLME proposto por Shimabukuro e Smith (1991), nas bandas 1, 2, 3, 4, 5 e 7 das imagens TM. Este modelo tem como objetivo estimar a proporção dos componentes solo, sombra e vegetação de cada pixel, através das respostas espectrais nas diversas bandas.

A aplicação deste modelo é feita através da definição dos componentes puros - solo, sombra e vegetação - nas imagens, resultando em três imagens-fração correspondentes. A formulação do MLME é dada por

$$\boldsymbol{r_i = a * \mathbf{vege_i} + b * \mathbf{solo_i} + c * \mathbf{agua_i} + \mathbf{e_i}} \quad (3.3)$$

onde r_i é a resposta do pixel na banda $\boldsymbol{i}$; $\boldsymbol{a}$ é a proporção de vegetação; $\boldsymbol{b}$ é a proporção de solo; $\boldsymbol{c}$ é a proporção de sombra ou água; $vege_i$ é a resposta espectral do componente vegetação na banda $\boldsymbol{i}$; $solo_i$ é a resposta espectral do componente solo na banda $\boldsymbol{i}$; $agua_i$ é a resposta espectral do componente sombra ou água na banda $\boldsymbol{i}$; e_i é o erro na banda $\boldsymbol{i}$; sendo que $\boldsymbol{i}$ varia entre as bandas 1, 2, 3, 4, 5 e 7.

Nesse estudo para a determinação dos valores espectrais de cada um dos componentes puros do modelo, foram selecionadas amostras de solo exposto para a componente solo, áreas de floresta secundária para a componente vegetação e por fim, água limpa para a componente água. A qualidade das amostras coletadas para posterior execução do MLME sobre as imagens foi realizada através de análise do gráfico com as curvas espectrais de cada componente.

As composições coloridas das imagens-fração adquiridas a partir das imagens Landsat-5 de 2008 e 2009 são apresentadas na Figura 3.9. A imagem-fração

solo encontra-se em vermelho, a imagem-fração vegetação em verde e imagem-fração sombra em azul.

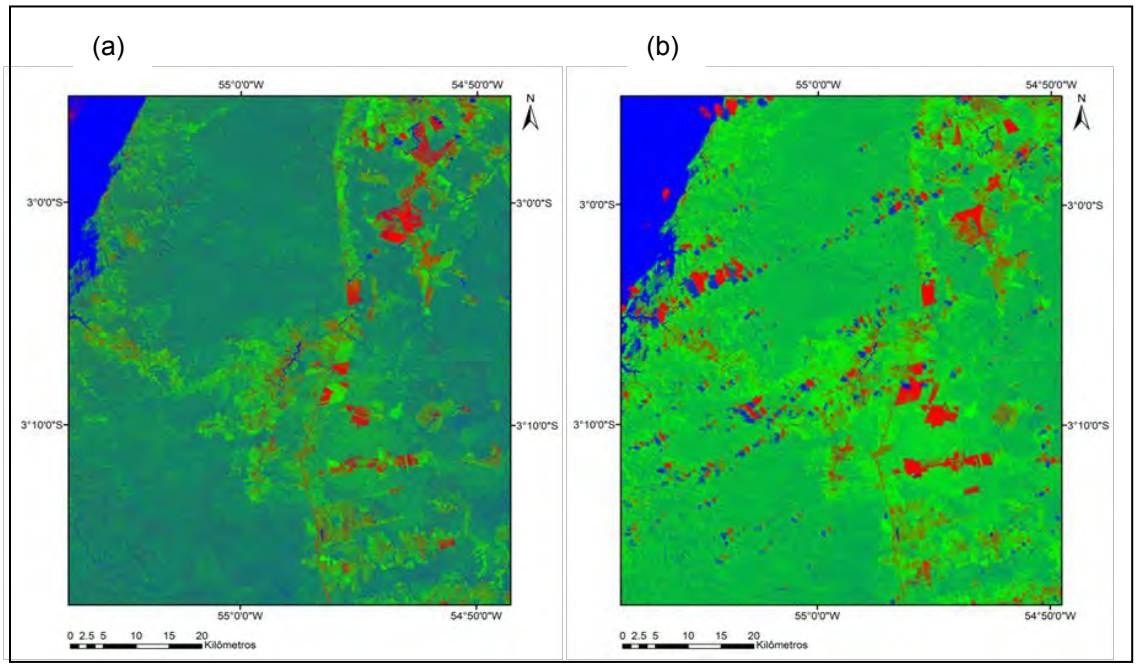

Figura 3.8 - Composição colorida em vermelho, verde e azul das imagens-fração solo, vegetação e sombra do MLME dos anos de 2008 (a) e 2009 (b).

3.2.3.4. Índices de Vegetação

Diversos índices para o estudo da vegetação foram propostos pela comunidade científica. O principal objetivo desses índices é permitir a exploração das propriedades espectrais da vegetação, tomando como base o comportamento espectral antagônico da vegetação nas regiões espectrais do visível e infravermelho próximo (PONZONI; SHIMABUKURO, 2010).

O índice de vegetação da diferença normalizada (NDVI), proposto por Rouse et al. (1973), é baseado na normalização da razão simples para o intervalo de -1 a +1. Esse índice é amplamente utilizado como ferramenta para o monitoramento da vegetação. Além disso, o NDVI permite o acompanhamento das mudanças sazonais e interanuais da atividade e desenvolvimento da vegetação, das mudanças fisiológicas das folhas, dos períodos de senescência, etc (PONZONI; SHIMABUKURO, 2010; JENSEN, 2009).

A normalização é realizada através da equação

$$\boldsymbol{NDVI} = \frac{(\boldsymbol{\rho_{IVP}} - \boldsymbol{\rho_V})}{(\boldsymbol{\rho_{IVP}} + \boldsymbol{\rho_V})} \quad (3.4)$$

onde $\boldsymbol{\rho_{IVP}}$ é o valor de reflectância no infravermelho próximo e $\boldsymbol{\rho_V}$ é o valor de reflectância no vermelho.

A Figura 3.9 apresenta os índices de vegetação NDVI obtidos a partir das imagens Landsat-5 dos anos 2008 e 2009.

Gao (1996) desenvolveu o índice de umidade por diferença normalizada (NDWI), também denominado de índice de água, o qual está diretamente relacionado com a quantidade de água das folhas da vegetação. Este índice é sensível às mudanças que ocorrem no conteúdo hídrico destas estruturas.

O NDWI é formulado como

$$\boldsymbol{NDWI} = \frac{(\boldsymbol{\rho_{NIR}} - \boldsymbol{\rho_{MIDIR}})}{(\boldsymbol{\rho_{NIR}} + \boldsymbol{\rho_{MIDIR}})} \quad (3.5)$$

onde $\boldsymbol{\rho_{NIR}}$ representa os valores de reflectância no infravermelho próximo e o $\boldsymbol{\rho_{MIDIR}}$ representa os valores de reflectância no infravermelho médio.

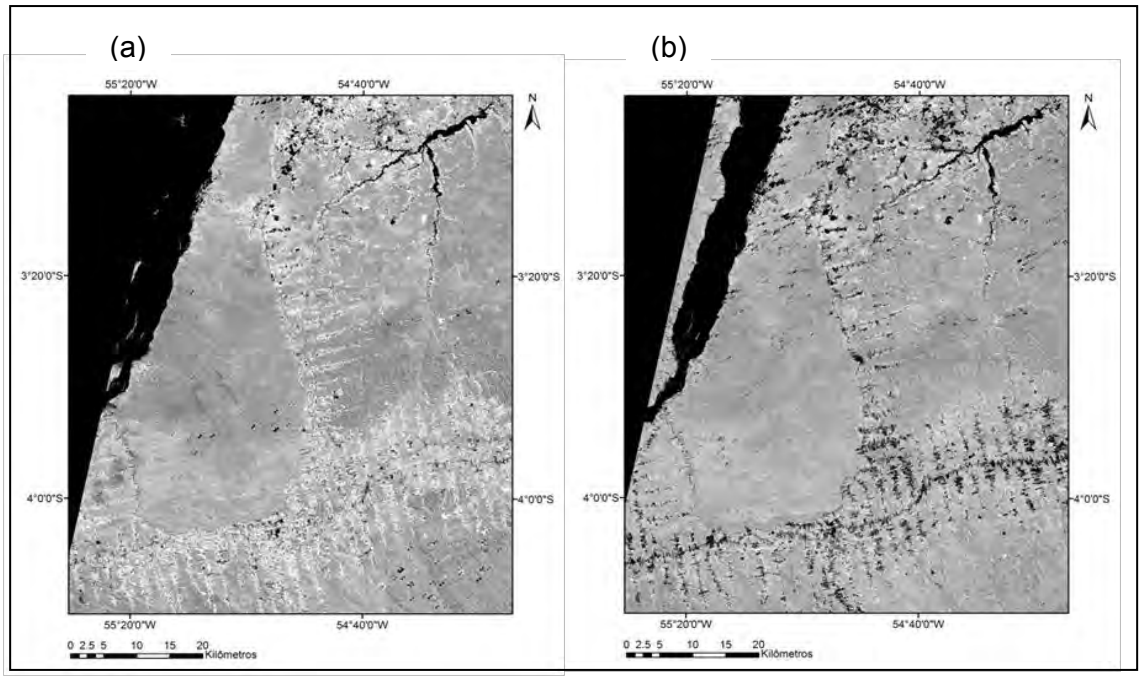

Figura 3.9 - Imagens NDVI de 2008 (a) e 2009 (b).

A Figura 3.10 apresenta os índices NDWI obtidos a partir das imagens Landsat-5 dos anos de 2008 e 2009.

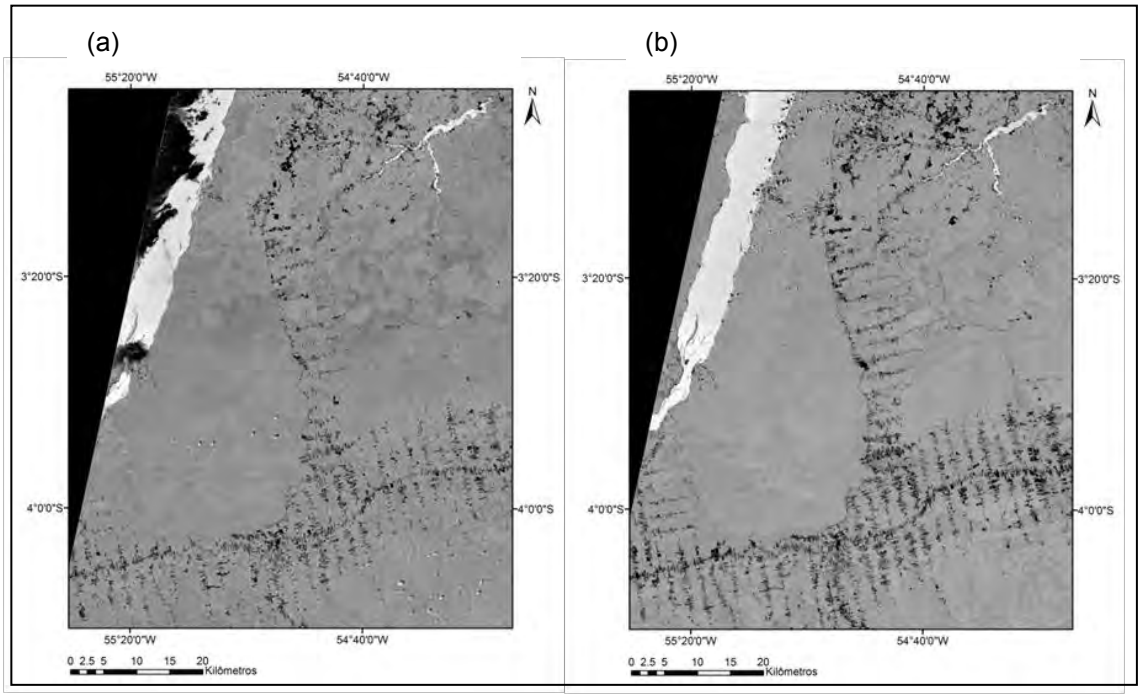

Figura 3.10 - Imagens NDWI de 2008 (a) e 2009 (b).

Devido à influência do solo no espectro da radiação da vegetação, Huete (1988) sugeriram o índice de vegetação ajustado para o solo (SAVI). Este índice permite minimizar os efeitos do solo no resultado do índice da vegetação, sendo definido por

$$\boldsymbol{SAVI} = \frac{(\boldsymbol{\rho_{NIR}} - \boldsymbol{\rho_R})}{(\boldsymbol{\rho_{NIR}} - \boldsymbol{\rho_R} + \boldsymbol{L})} * (\mathbf{1} + \boldsymbol{L}) \qquad (3.6)$$

onde $\boldsymbol{\rho_{NIR}}$ é o valor de reflectância no infravermelho próximo; $\boldsymbol{\rho_R}$ é o valor de reflectância no vermelho; e $\boldsymbol{L}$ é uma constante que minimiza o efeito do solo e pode variar de 0 a 1.

Os valores ótimos de L, segundo os autores são:

L = 1, para densidades baixas de vegetação

L= 0,5, para densidades médias de vegetação

L = 0,25, para densidades altas de vegetação

Para esse estudo optou-se utilizar o valor 0,5 para a constante L, pois como nem todas as áreas são de vegetação com alta densidade, como as áreas que estão no interior da Flona do Tapajós, o uso de um valor intermediário é mais adequado para toda a região de estudo.

A Figura 3.11 apresenta os índices SAVI obtidos a partir das imagens Landsat-5 dos anos de 2008 e 2009.

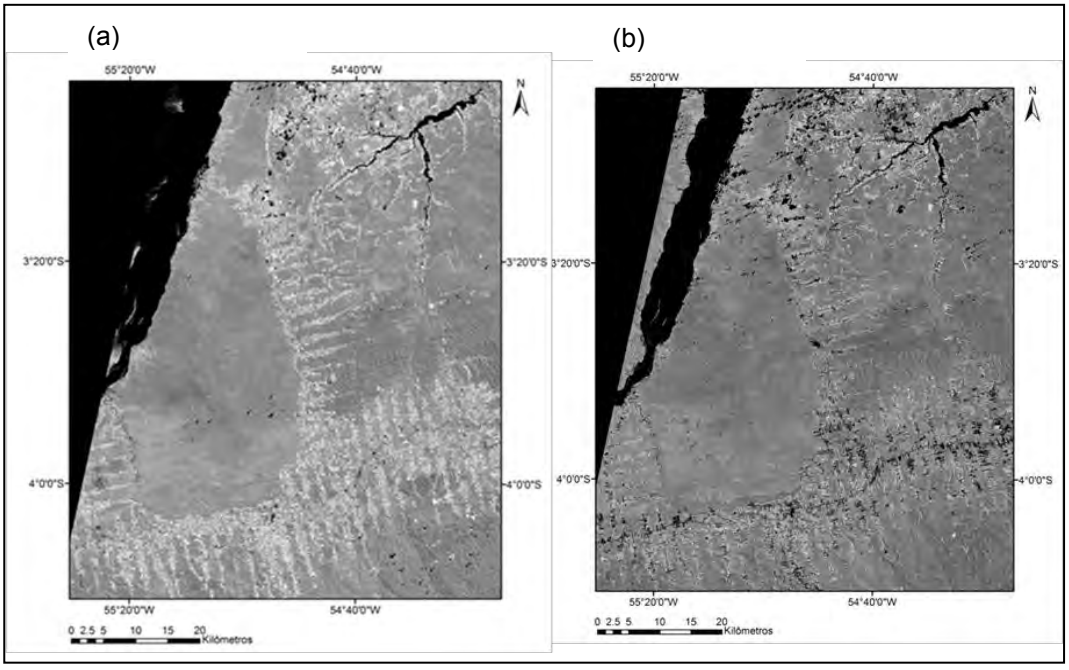

Figura 3.11 - Imagens SAVI de 2008 (a) e 2009 (b).

3.2.3.5. Dados SAR

O uso de dados ópticos para o monitoramento e mapeamento de áreas de florestas tropicais é muitas vezes limitado, devido à constante presença de nuvens sobre essas regiões. Neste sentido, imagens SAR aliadas aos dados ópticos podem ser utilizadas como uma alternativa para o estudo dessas áreas, pois o sinal de microondas é capaz de penetrar as nuvens e obter informações da superfície terrestre independente das condições atmosféricas (SANTOS; GONÇALVES, 2009; SAATCHI et al., 1997; ANGELIS, 2003).

Para este estudo foram utilizadas três imagens na banda C adquiridas pelo RADARSAT-2. Duas destas imagens foram adquiridas no modo Standard Quad-polarization nos dias 24 de setembro de 2008 e 13 de outubro de 2009. A

terceira imagem RADARSAT-2 utilizada foi adquirida no modo Fine resolution Quad-polarization do dia 18 de outubro de 2008. A Tabela 3.2 lista os parâmetros das imagens RADARSAT-2 utilizadas.

Tabela 3.2: Parâmetros das imagens RADARSAT-2.

Características	**2008**	**2008**	**2009**
Banda/freqüência	C - 5,3cm/5,6 Hz	C - 5,3cm/5,6 Hz	C - 5,3 cm/5,6 Hz
Largura da banda	100 MHz	100 MHz	100 MHz
Polarização	Full (HH, HV, VH, VV e fase)	Full (HH, HV, VH, VV e fase)	Full (HH-HV-VH-VV e fase)
Resolução	25 x 28 m	11 x 9 m	25 x 28 m
Largura da cena	25 x 25 km	25 x 25 Km	25 x 25 km
Ângulo de incidência	27,07°	32,56º	33,18°
Nº de looks	4	1	4
Órbita	Descendente	Ascendente	Ascendente

Estas imagens foram processadas no software Radar Tools (RAT) para a extração das imagens intensidade, das diferenças de fases e dos atributos provenientes do teorema da decomposição de alvos proposta por Freeman e Durden (1998).

As imagens intensidades e as diferenças de fase foram obtidas a partir da matriz de covariância complexa [C],definida por

$$[C] = \begin{bmatrix} C11 & C12 & C13 & C14 \\ C21 & C22 & C23 & C24 \\ C31 & C32 & C33 & C34 \\ C41 & C42 & C43 & C44 \end{bmatrix} = \begin{bmatrix} \langle |S_{hh}|^2 \rangle & \langle S_{hh}S_{hv}^* \rangle & \langle S_{hh}S_{vh}^* \rangle & \langle S_{hh}S_{vv}^* \rangle \\ \langle S_{hv}S_{hh}^* \rangle & \langle |S_{hv}|^2 \rangle & \langle S_{hv}S_{vh}^* \rangle & \langle S_{hv}S_{vv}^* \rangle \\ \langle S_{vh}S_{hh}^* \rangle & \langle S_{vh}S_{hv}^* \rangle & \langle |S_{vh}|^2 \rangle & \langle S_{vh}S_{vv}^* \rangle \\ \langle S_{vv}S_{hh}^* \rangle & \langle S_{vv}S_{hv}^* \rangle & \langle S_{vv}S_{vh}^* \rangle & \langle |S_{vv}|^2 \rangle \end{bmatrix} \quad (3.7)$$

onde * é o conjugado e ⟨ ⟩ indica a média espacial de células vizinhas. A matriz de covariância complexa [C] contém na sua diagonal principal as informações de intensidade e nas demais posições, a informações das diferenças de fase. (ULABY; ELACHI, 1990; GUERRA, 2008). As Figura 3.12, Figura 3.13 e Figura 3.14 apresentam uma composição colorida em vermelho as imagens intensidade de polarização HH, em verde as imagens intensidade de polarização HV e em azul as imagens intensidade de polarização VV para cada ano e órbita de aquisição.

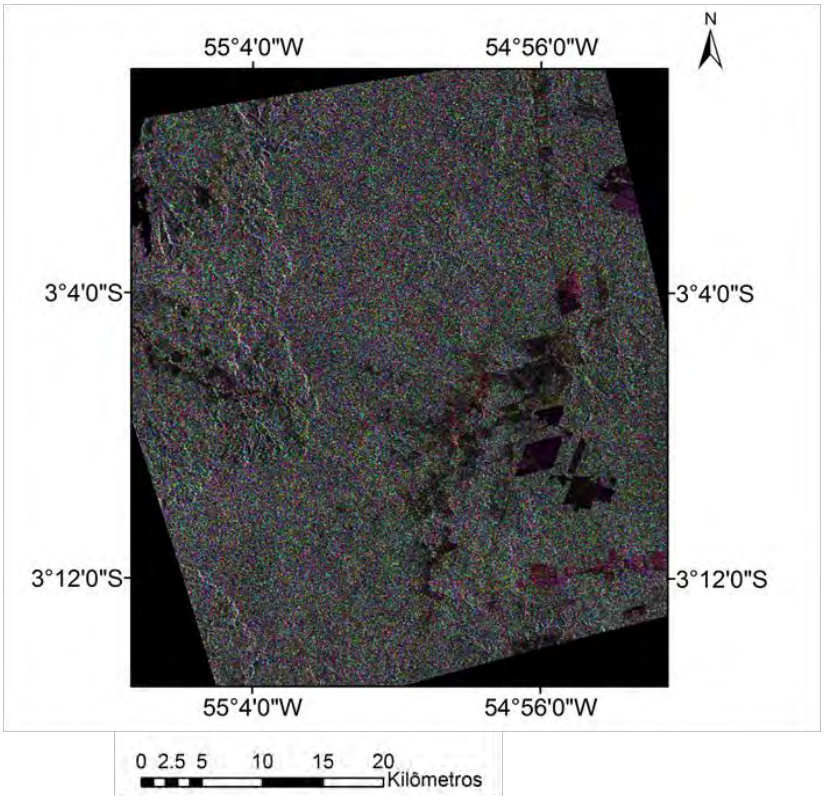

Figura 3.12 - Composição colorida R (HH), G (VV) e B (HV) das imagens intensidade do ano de 2008 ascendente

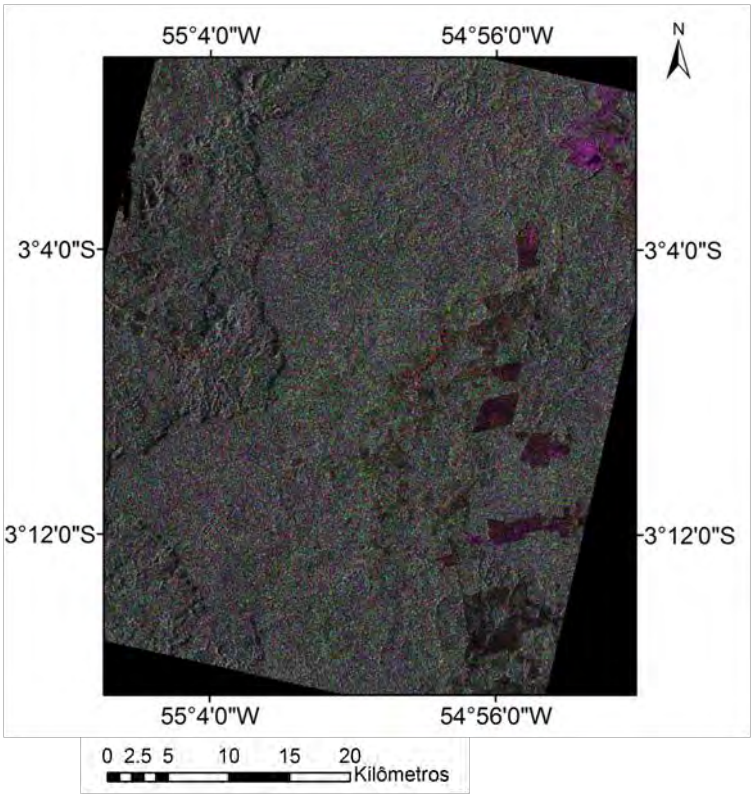

Figura 3.13: Composição colorida R (HH), G (VV) e B (HV) das imagens intensidade do ano de 2008 descendente.

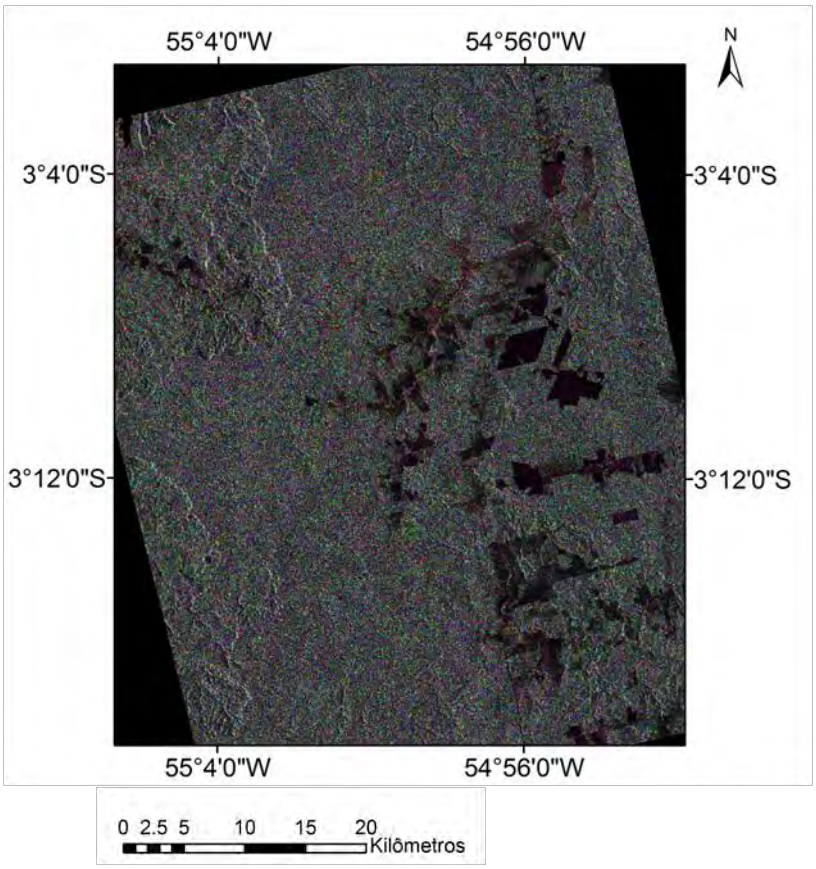

Figura 3.14: Composição colorida R (HH), G (VV) e B (HV) das imagens intensidade do ano de 2009 ascendente.

A partir do teorema da decomposição de alvos, de Freeman-Durden (1998), foram extraídas as componentes de espalhamento volumétrico (P_v), double-bounce (P_d) e superficial (P_s), resultante da matriz de covariância. O espalhamento volumétrico é modelado como o espalhamento aleatório que ocorre no interior dos dosséis florestais, enquanto que o espalhamento double-bounce é modelado como o espalhamento de refletores de canto diédricos, como por exemplo, resultante da interação tronco-solo que ocorre nas florestas. O espalhamento superficial é resultado de interações na superfície dos objetos, como na porção superior de dosséis e solos. As Figura 3.15, Figura 3.16 e Figura 3.17 apresentam uma composição colorida em vermelho as componentes de espalhamento P_v, em verde as componentes de espalhamento P_d e em azul as componentes de espalhamento P_s para cada ano e órbita de aquisição.

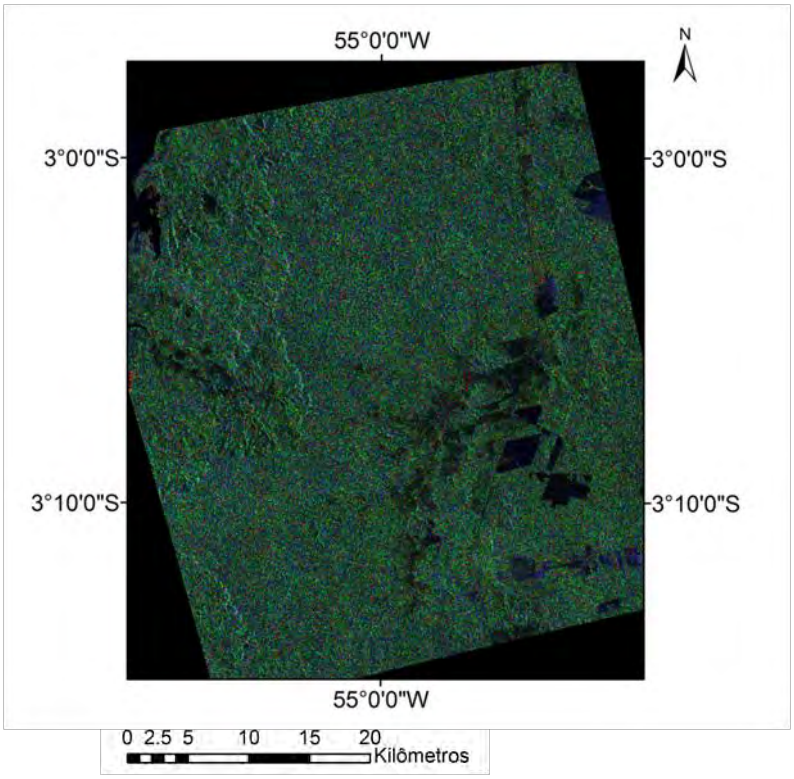

Figura 3.15: Composição colorida R (P_d), G (P_v) e B (P_s) das componentes de espalhamento da decomposição de Freeman-Durden do ano de 2008 ascendente.

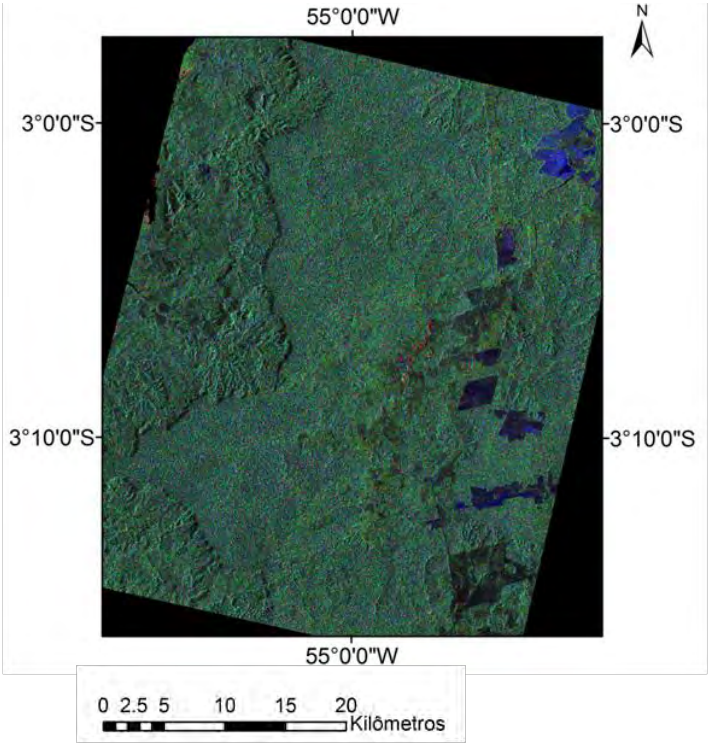

Figura 3.16: Composição colorida R (P_d), G (P_v) e B (P_s) das componentes de espalhamento da decomposição de Freeman-Durden do ano de 2008 descendente.

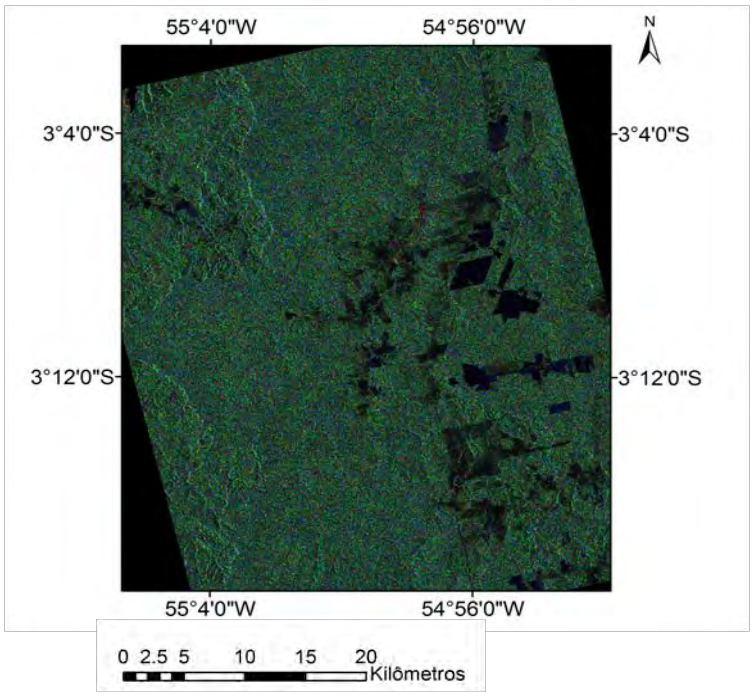

Figura 3.17: Composição colorida R (P_d), G (P_v) e B (P_s) das componentes de espalhamento da decomposição de Freeman-Durden do ano de 2009 ascendente.

Após o processamento das imagens SAR aplicou-se um filtro Gamma com uma janela de tamanho 7x7 para reduzir os efeitos do ruído speckle.

3.2.3.6. Árvore de decisão e classificação

As classificações das imagens de 2008 e 2009 foram realizadas através de uma técnica de mineração de dados baseadas em regras, denominada de árvore de decisão. As seis bandas das imagens Landsat-5, as três imagens-fração do MLME, os índices de vegetação, as informações extraídas das imagens SAR e os dados de textura formaram o conjunto de dados utilizados como entrada na construção das árvores de decisão.

Para a criação das árvores de decisão, foi utilizado o sistema livre para mineração de dados Waikato Environment for Knowledge Analysis (WEKA) na versão 3.6.4. Este aplicativo, disponível gratuitamente para download em http://www.cs.waikato.ac.nz/ml/weka/, fornece um conjunto de algoritmos, ferramentas de aprendizagem e a possibilidade de criação de novos algoritmos

em seu ambiente. Entre os algoritmos disponíveis estão: regressão, classificação, mineração de regras de associação, seleção de atributos, etc.

Este aplicativo possui interface amigável e rápido processamento dos dados. Estas características permitiram testar e analisar os doze algoritmos de árvore de decisão presentes no sistema. Com isso, foi possível identificar o melhor algoritmo para a classificação das imagens de acordo com os atributos utilizados como dados de entrada para a construção da árvore de decisão.

Os doze algoritmos de construção de árvore de decisão implementados no WEKA e testados foram: BFTree, DecisionStump, FT, J48, J48graft, LADTree, LMT, NBTree, RandomForest, RandomTree, REPTree e SimpleCart. A classificação mais precisa foi gerada pelo algoritmo SimpleCart, de acordo com índice Kappa, menor valor de pixels classificados incorretamente e acurácia geral das classificações. Esta análise foi realizada sobre o conjunto de dados formado pelas seis bandas do Landsat-5/TM, as três imagens-fração do MLME, os índices de vegetação NDVI, NDWI e SAVI.

O SimpleCart é uma implementação do algoritmo Classification and Regression Trees (CART) proposto por Breiman et al. (1984). Este algoritmo é considerado como não-paramétrico e permite o uso de diferentes tipos de dados, como dados ordinais, nominais e contínuos. Este algoritmo baseia-se em uma técnica exaustiva que define os melhores limiares que serão utilizados nos nós da árvore de decisão para a separação das classes. A árvore de decisão do algoritmo CART é uma árvore binária que testa apenas um atributo por nó (BREIMAN et al., 1984; VASCONCELLOS, 2011; BITTENCOURT; CLARKE, 2003; ELHADI; ZOMRAWI, 2010)

Entre as vantagens do uso do CART, destaca-se o fato dele não requerer nenhuma variável previamente identificada, pois o próprio algoritmo seleciona as variáveis mais significativas e elimina as demais. Outra vantagem, é que o algoritmo não é afetado por dados ruidosos, sendo estes isolados em nós separados. Diferentemente da maioria dos algoritmos, o CART não utiliza a técnica de pré-poda. Ao invés disso, utiliza um método que expande ao máximo o tamanho da árvore de decisão e realiza a pós-poda, tornando o

método mais eficiente e gerando árvores de decisão mais simples e precisas (FONCESA, 1994; TIMOFEEV, 2004).

A separação dos dados ao longo dos nós da árvore de decisão ocorre através do cálculo da impureza ou entropia entre os valores de entrada (VASCONCELLOS, 2011; GHOSE et al., 2010). No sistema WEKA, a entropia é denominada Gini ($i(t)$) e calculada através da equação

$$i(t) = -\sum_{i,j} C\,(i|j)\; p(i|t)p(j|t) \tag{3.8}$$

onde $C\,(i|j)$ é a probabilidade de um objeto da classe j ser classificado incorretamente como sendo da classe i, $p(i|t)$ é a probabilidade de um objeto da classe i ser classificado como um objeto da classe t, $p(j|t)$ é a probabilidade de um objeto que pertence a classe j ser classificado como um objeto da classe t.

A melhor separação entre os dados é aquela que tem a maior diferença obtida pela equação da variação do valor Gini ($\Delta I(s,t)$)

$$\Delta I(s,t) = I(t) - p_L I(t_L) - p_R I(t_R) \tag{3.9}$$

em que p_L e p_R são as probabilidades dos objetos serem passados respectivamente para o nó à esquerda t_L ou para o nó à direita t_R. Através destas equações, sucessivas subdivisões são determinadas até que ocorra uma diminuição significativa na medida da impureza. Quando este estágio é atingido, o nó não é mais subdividido e torna-se automaticamente em uma folha.

3.2.3.6.1. Preparação dos dados para o sistema WEKA

O aplicativo WEKA possui um formato próprio denominado ARFF para leitura do arquivo de dados. Esse formato consiste em uma lista com as instâncias e os valores dos atributos de cada classe. O arquivo é dividido em três partes, onde na primeira parte é definida a relação ou tarefa que será analisada. Na segunda parte, são declarados os atributos em linhas sequênciais e, na última parte, são listados os dados de treinamento que serão utilizados para a classificação. A Tabela 3.3 apresenta um exemplo genérico de um arquivo no

formato ARFF (VASCONCELLOS, 2011, WITTEN et al., 2011, BOUCKAERT et al., 2010).

Tabela 3.3 - Exemplo de um arquivo no formato ARFF do WEKA.

```
@Relation tempo

@Attribute Tempo sol,nublado chuva
@Attribute Temperatura Real
@Attribute Umidade Real
@Attribute Vento sim,não
@Attribute Decisão sim,não

@Data
Sol, 23.9,70,sim,sair
Sol,26.7,90,sim,não sair
Sol,29.4,85,não,não sair
Sol,22.2,95,não,não sair
Sol,20.6,70,não,sair
Nublado,22.2,87,sim,sair
```

Fonte: Vasconcellos (2011).

Além do formato ARFF, o WEKA permite a leitura de arquivos no formato Comma-separated values (CSV), tornando ágil o uso deste software, pois diversos aplicativos permitem a exportação de dados neste formato. Em um arquivo CSV cada linha representa um novo registro de dados, os quais são separados por vírgulas (WITTEN et al., 2011). A Tabela 3.4 apresenta um arquivo CSV com os mesmos dados do arquivo ARFF apresentado na Tabela 3.3.

Tabela 3.4 - Exemplo do arquivo no formato CSV.

```
Tempo,Temperatura,Umidade,Vento,Decisão
Sol, 23.9,70,sim,sair
Sol,26.7,90,sim,não sair
Sol,29.4,85,não,não sair
Sol,22.2,95,não,não sair
Sol,20.6,70,não,sair
Nublado,22.2,87,sim,sair
```

Fonte: Vanconcellos (2011).

3.2.3.6.2. Validação cruzada pelo Método Holdout

O WEKA permite realizar o treinamento dos dados e a validação da classificação através de quatro métodos - *use training set, supplied test set,*

cross-validation e percentage split. Neste estudo, foi utilizado o método *percentage split*, que é um método simples e computacionalmente rápido também chamado de método *Holdout* (WITTEN et al., 2011).

No método *Holdout*, que na literatura em português é conhecido como Treino-Teste, $^{2}/_{3}$ dos dados são destinados para o treinamento ou estimação dos parâmetros da árvore de decisão e $^{1}/_{3}$ dos dados são usados para validação do modelo. A Figura 3.18 ilustra esse processo, onde os dados são divididos em duas partes mutuamente exclusivas, ou seja, a realização de um evento elimina a realização do outro. O conjunto de dados de treinamento é utilizado para a construção do modelo e os dados de validação são usados para verificar a acurácia do modelo gerado (OLSON; DELEN; 2008).

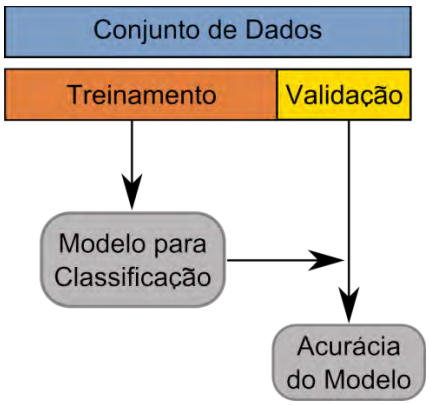

Figura 3.18 - Método Holdout.
Fonte: Adaptado de OLSON; DELEN (2008).

Este método é recomendado para casos que possuem grande quantidade de dados e que permitem reservar pelo menos 1.000 amostras para o teste do modelo. Para casos que possuem uma pequena quantidade de amostras para o teste, esse método pode ser pessimista e os resultados podem ser insatisfatórios. Desta forma, é recomendada a utilização de outro método para validação da classificação desses casos (LEITE, 2007; BATISTA; MONARD, 1998).

3.2.3.6.3. Geração de árvore de decisão para o ENVI usando o WEKA

As árvores de decisão geradas pelo aplicativo WEKA são apresentadas em formato de texto. Nesse tipo de saída, a hierarquia dos nós é representada através de tabulações. O processamento dessas árvores de decisão em aplicativos, como o *Environment for Visualizing Images* (ENVI), é realizada, habitualmente, através da inclusão manual de cada nó da árvore. Entretanto, a inserção de árvores de decisão com dezenas ou centenas de nós, como as geradas neste trabalho, demandariam muitas horas, atrasando o processo de classificação. Além disso, o processo de inclusão manual é passível de erros de digitação, o que causaria grande impacto nos resultados. Por conta disso, foi realizado alterações no código-fonte do módulo de impressão das árvores de decisão do sistema WEKA, para que, além de gerar a saída padrão, fosse gerado um arquivo de árvore de decisão compatível com o ENVI 4.7.

As mudanças realizadas só foram possíveis devido ao código-fonte aberto e de licença livre, que permite a alteração e implementação de novas funcionalidades. Essas mudanças foram feitas por um especialista da área de computação com o acompanhamento da autora desse trabalho. Com estas alterações implementadas, o sistema WEKA adaptado gera automaticamente um arquivo com a árvore de decisão no formato texto puro (extensão .txt) do ENVI, que pode ser facilmente importada. Após a importação da árvore de decisão para o ENVI a imagem é classificada, onde os pixels são separados seguindo a sequência de regras da árvore de decisão e no final são rotulados de acordo com cada classe determinada.

3.2.4. Análise e validação dos resultados

Após a etapa de processamento dos dados, os resultados obtidos foram analisados e avaliados. Para a avaliação da classificação das imagens de 2008 e 2009 através da árvore de decisão, foi utilizado o índice Kappa, que indica a acurácia da classificação através da matriz de confusão. Para isso, é calculada a exatidão global e o índice Kappa de cada classificação, conforme proposto por Congalton et al. (1991).

A matriz de confusão apresenta a distribuição dos dados das classificações. Nesta matriz, encontram-se os resultados dos pixels da imagem corretamente classificados (na diagonal principal) e os pixels que foram erroneamente classificados (no restante da matriz). A estrutura de uma matriz de confusão é apresentada na Tabela 3.5.

Tabela 3.5 - Matriz de confusão adotada para análise da classificação.

		Imagem Referência				
Imagem Classificada	*Classe*	*1*	*2*	*...*	*c*	*Total*
	1	$*X_{11}$	.	.	.	X_{+1}
	2	.	$*X_{22}$	.	.	X_{+2}
	...	.	.	...	.	.
	c	.	.	.	$*X_{cc}$	X_{+c}
	Total	X_{+1}	X_{+2}	...	X_{+c}	*N*

* Pontos da diagonal principal corretamente classificados.

O índice Kappa ($\widehat{K}$) é dado por

$$\widehat{K} = \frac{\theta_1 - \theta_2}{1 - \theta_2} \tag{3.10}$$

em que θ_1 e θ_2 são obtidos das equações

$$\theta_1 = \frac{\sum_{i=1}^{c} xii}{N}\ ; e \qquad\qquad \theta_2 = \frac{\sum_{i=1}^{c} x_{i+}x_{+i}}{N^2} \tag{3.11}$$

onde N é o número total de unidades amostrais contempladas pela matriz; xii é o número de elementos da diagonal principal corretamente classificados; e c é o número de categorias analisadas na matriz.

3.2.5. Dados de campo

Para realização da classificação das imagens dos anos de 2008 e 2009 foram utilizados dados de campo coletados nestes dois períodos, permitindo a validação das amostras de treinamento e da classificação realizada através da árvore de decisão. Além disso, foram utilizados dados do projeto TerraClass disponibilizados pelo Centro Regional da Amazônia – CRA/INPE (www.inpe.br/cra/projetos_pesquisas/terraclass.php).

Em 2008 foram amostrados 76 pontos, que ficaram concentrados na região da "Comunidade São Jorge", localizada no interior da FLONA do Tapajós. As

amostras foram coletadas somente nesta região, pois ela compreende a área das imagens RADARSAT-2. Na Figura 3.19 são destacadas, através de quadrados vermelhos, as localizações geográficas dos pontos coletados sobre a área de estudo.

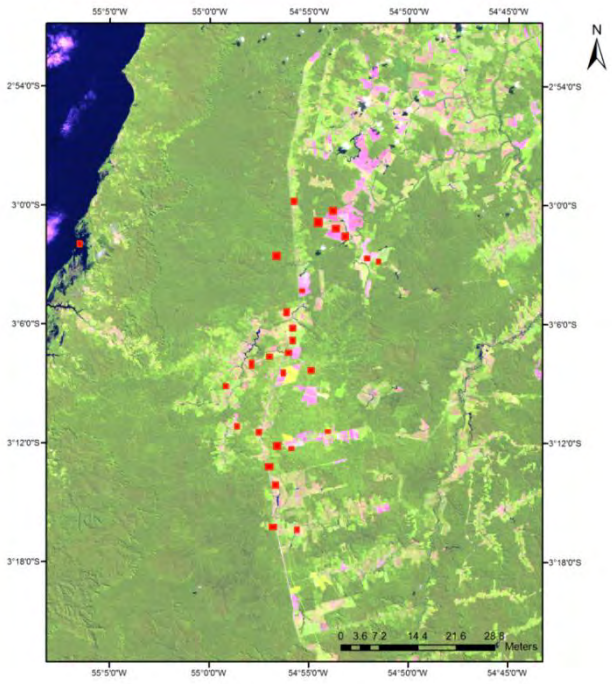

Figura 3.19 - Distribuição dos pontos amostrados em 2008.

Foram amostrados 3.239 pontos nas coletas de campo realizadas em 2009. Esses dados foram obtidos sobre toda a área da FLONA do Tapajós. Na Figura 3.20 são destacadas, com pontos vermelhos, as localizações geográficas das amostras coletadas.

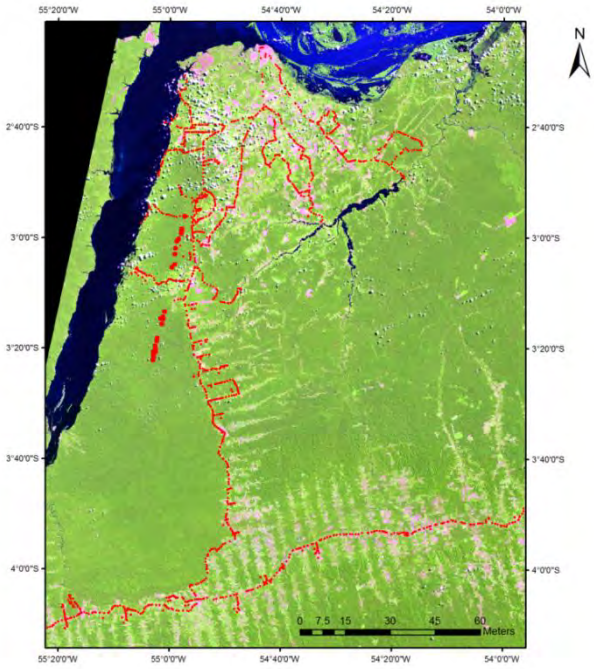

Figura 3.20 - Distribuição dos pontos amostrados em 2009.

A Figura 3.21 ilustra as amostras coletadas de classes de uso e cobertura da terra que foram determinadas para classificação através da árvore de decisão, de acordo com os dados de campo disponíveis para cada ano.

A classe Água refere-se a todos os corpos d'água presentes na área de estudo. A classe Floresta são as áreas de floresta primária, que sofreram pouca ou nenhuma interferência humana. A classe Regeneração são as áreas de floresta secundária em processo de regeneração. A classe Pasto são as áreas com cobertura de pastagem, usualmente destinadas para a criação de gado. A classe Área Urbana são as áreas compostas por casas, ruas e demais construções. A classe Floresta Degradada são as áreas que estão sobre o processo de exploração madeireira e a classe Floresta Secundária são as de floresta que sofreram ação antrópica e que após a sua regeneração já estão consolidadas.

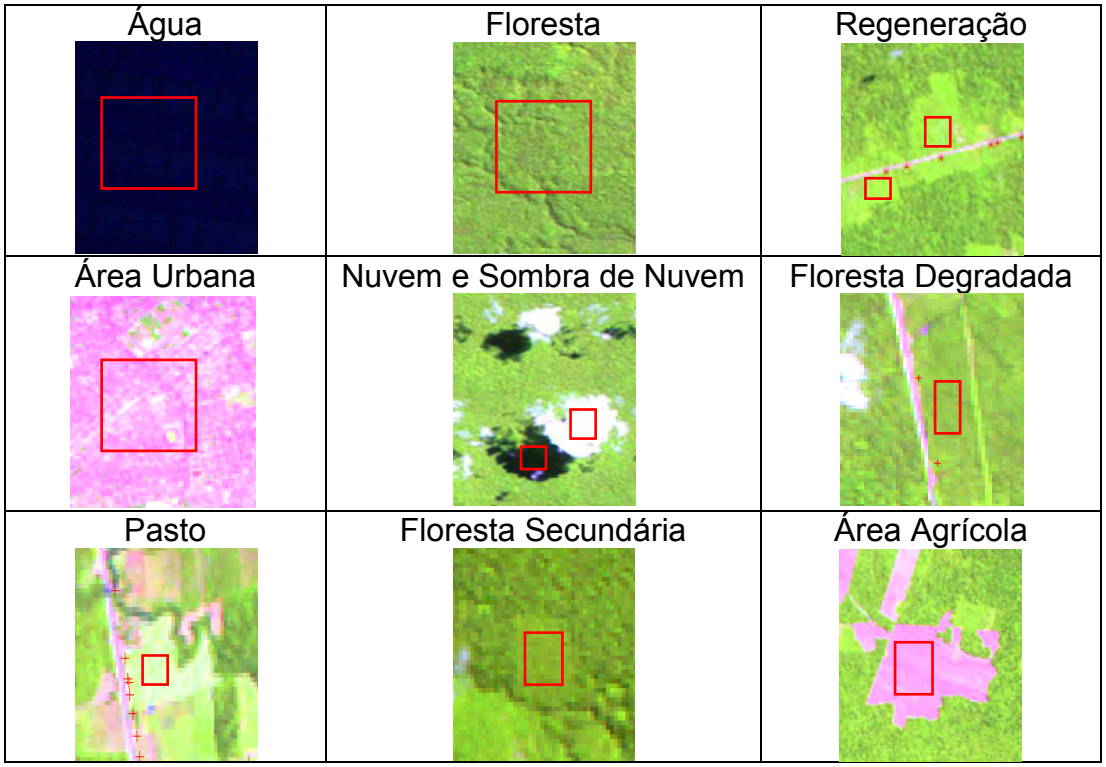

Figura 3.21: Classes e amostras de uso e cobertura da terra, utilizados nas classificações das imagens. Os retângulos vermelhos destacam as áreas onde foram coletados os pixels para formar o conjunto de amostra para treinamento e validação das classificações.

Após a coleta dos pontos das classes de uso e cobertura da terra que foram determinadas para este estudo, baseados nos dados de campos, os pixels das amostras foram embaralhados e posteriormente a escolha dos conjuntos de dados final foi feita através de um sorteio entre esses pixels. Este sorteio foi necessário, pois o aplicativo WEKA apresenta uma limitação quanto ao tamanho do arquivo de entrada para o treino e validação da árvore de decisão.

4 RESULTADOS E DISCUSSÕES

4.1. Classificações

Com intuito de explorar diversas abordagens para a classificação da cobertura vegetal da FLONA do Tapajós, foram utilizadas diferentes combinações de dados de entrada para a construção das árvores de decisão. Com isso, foi possível compreender a contribuição do uso de dados multisensores na classificação da cobertura vegetal da área de estudo. A combinação dos atributos de dados de entrada para o treinamento da árvore de decisão permitiu realizar um total de 7 classificações entre os anos de 2008 e 2009.

4.1.1. Classificações 2009

Devido à maior quantidade de dados coletados em campo no ano de 2009, as análises foram iniciadas por este ano, permitindo uma maior segurança nos testes. Além disso, os atributos foram sendo adicionados um a um, permitindo identificar a contribuição de cada variável na qualidade da árvore de decisão resultante.

A Tabela 4.6 lista as classificações geradas para o ano de 2009, onde a primeira coluna apresenta um nome único dado a classificação para permitir uma maior clareza no texto. A segunda, terceira e quarta colunas apresentam os atributos utilizados como dados de entrada para o treinamento da árvore de decisão.

Tabela 4.6 - Classificações geradas para o ano de 2009.

Identificação	Atributos da Classificação			
2009a	Landsat-5	NDVI, NDWI, SAVI, MLME		
2009b	Landsat-5	NDVI, NDWI, SAVI, MLME	Textura	
2009c	Landsat-5	NDVI, NDWI, SAVI, MLME		RADARSAT-2/Ascendente

4.1.1.1. Classificação da imagem Landsat-5 de 2009

A classificação 2009a foi obtida através da árvore de decisão SimpleCart do sistema WEKA gerada a partir das seis bandas do sensor TM Landsat-5, das

três frações do MLME e dos índices de vegetação NDVI, NDWI e SAVI. A Tabela 4.7 lista os 12 atributos utilizados como dados de entrada no treinamento da árvore de decisão para classificação da imagem de 2009.

Tabela 4.7 - Atributos de entrada na árvore de decisão da classificação 2009a.

Atributos
B1 - Landsat-5
B2 - Landsat-5
B3 - Landsat-5
B4 - Landsat-5
B5 - Landsat-5
B7 - Landsat-5
Fração solo do MLME
Fração vegetação do MLME
Fração água do MLME
NDVI
NDWI
SAVI

Nesta primeira classificação, foram determinadas, de acordo com os dados de campo as seguintes classes: (a) Área Agrícola, (b) Água, (c) Sombra de Nuvem, (d) Regeneração, (e) Pasto, (f) Nuvem, (g) Floresta Degradada, (h) Floresta e (i) Área Urbana. Para cada classe foram destinados 3.500 pixels, totalizando 31.500 pixels, sendo que 2/3 foram utilizados para treinamento da árvore de decisão e 1/3 para validação da classificação, conforme definição do método *Holdout*.

A árvore de decisão gerada pelo software WEKA apresentou grande número de nós, totalizando 335 elementos. Desse montante, 168 eram folhas que representam uma determinada classe. Destaca-se que as classes se repetem nas folhas, pois podem existir diversas combinações de atributos que levam a uma mesma classe. Apesar de resultar em uma árvore de decisão extensa, o software mostrou-se rápido na criação dos limiares que determinam as regras para a separação das classes, com um tempo gasto para a geração desta árvore de decisão na ordem de 19,55s.

Para validação da classificação, foram utilizadas 10.710 amostras. Deste valor, 8.695 pixels, aproximadamente 81,2%, foram classificados corretamente e 2.015 pixels, aproximadamente 18,8%, foram classificados incorretamente pela

árvore de decisão. Os valores corretamente classificados estão posicionados na diagonal principal da Tabela 4.8 e estão destacados pela cor cinza. Os valores que estão fora da diagonal são os pixels que foram classificados incorretamente, estando destacados os valores mais significativos.

Tabela 4.8 - Matriz de confusão da classificação 2009a.

		Imagem Referência								
		Área Agrícola	**Água**	**Sombra de Nuvem**	**Regene-ração**	**Pasto**	**Nuvem**	**Floresta Degra-dada**	**Floresta**	**Área Urbana**
Imagem Classificada	**Área Agrícola**	1.143	0	0	26	12	4	1	1	22
	Água	0	1.201	0	0	0	0	0	0	0
	Sombra de Nuvem	0	0	1.130	0	0	0	1	0	2
	Regene-ração	3	0	0	768	152	0	209	78	2
	Pasto	13	0	0	175	925	0	36	12	9
	Nuvem	7	0	0	0	1	1.181	0	0	3
	Floresta Degradada	1	0	0	235	33	0	469	443	3
	Floresta	0	0	2	81	0	0	345	785	0
	Área Urbana	71	0	0	2	19	4	0	0	1.093

Os maiores erros de classificação ocorreram entre as classes de Floresta Degradada e Floresta, Floresta Degradada e Regeneração e entre Pasto e Regeneração. Estas falhas podem estar associadas ao comportamento espectral muito próximo entre cada um desses pares de classes, tornando difícil a separação através da árvore de decisão. As classes Sombra de Nuvem e Nuvem foram bem classificadas, pois poucos pixels foram incorretamente classificados. Para a classe Água, não ocorreu nenhum erro de classificação, sendo que a facilidade na classificação da classe água por ser associada ao comportamento espectral bastante distinto em relação aos demais alvos da superfície terrestre em análise.

O índice SAVI é o atributo que aparece na raiz da árvore de decisão. A partir deste atributo foi possível separar a classe água das demais classes, como ilustrado na Figura 4.22. Nesta situação, todos os pixels que tiverem o valor do índice SAVI menor que 3, foram classificados como Água.

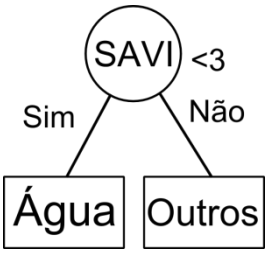

Figura 4.22 - Primeiro atributo usado na árvore de decisão da classificação 2009a.

A Figura 4.23 apresenta o resultado final da classificação através da árvore de decisão gerada pelo sistema WEKA, e a Figura 4.24 apresenta com maiores detalhes a classificação da área da Comunidade São Jorge localizada na FLONA do Tapajós.

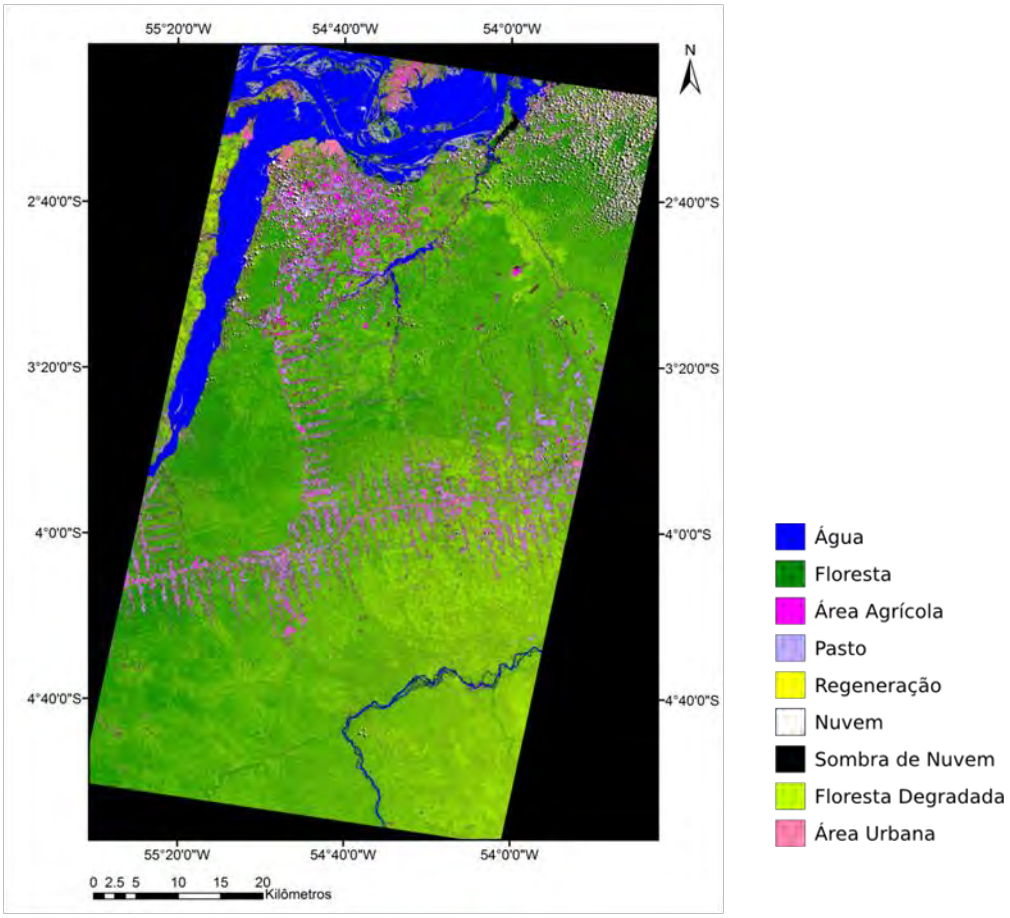

Figura 4.23 - Resultado da classificação 2009a.

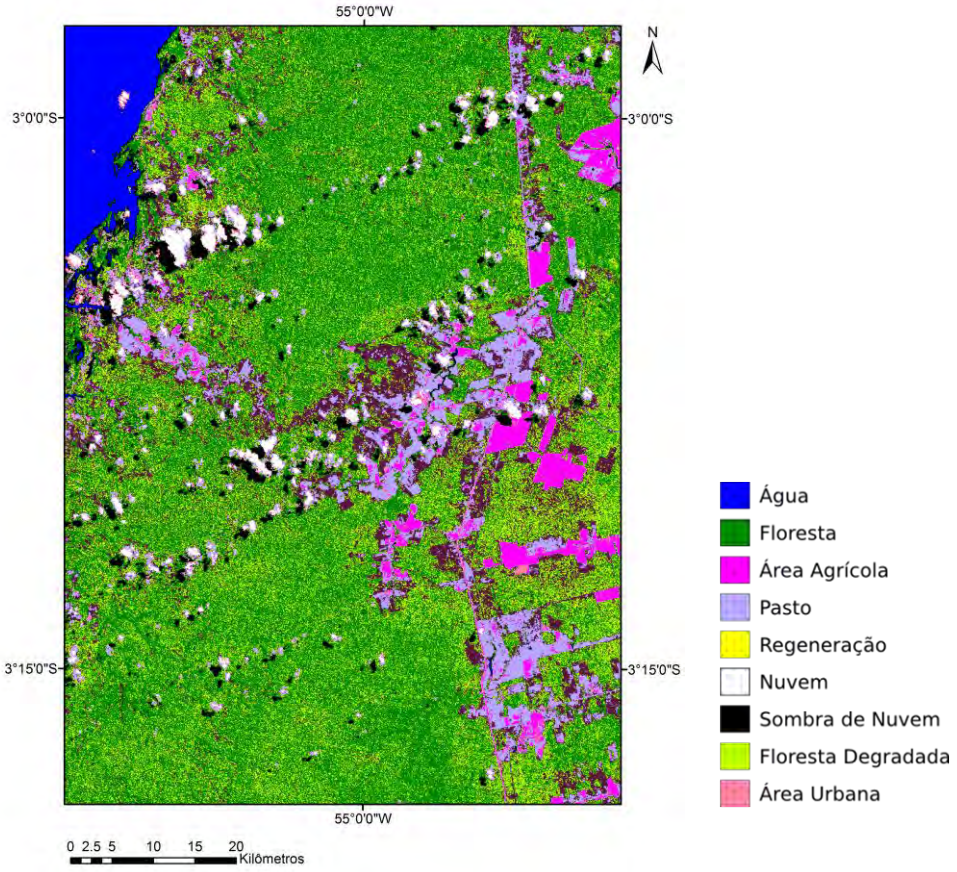

Figura 4.24 - Detalhes da classificação 2009a da Comunidade São Jorge.

4.1.1.2. Classificação da imagem Landsat-5 e dos atributos de textura de 2009

A classificação 2009b foi obtida através da adição dos atributos de textura, obtidos da imagem Landsat-5 de 2009 aos dados da classificação 2009a descrita na Tabela 4.7. Os dados de textura foram extraídos através do software ENVI 4.7 e após testes com diferentes tamanhos de janela de filtros, o tamanho do filtro escolhido foi de 5x5. Este filtro permitiu extrair as informações de textura da imagem sem que ocorressem muitas degradações nas bordas dos alvos em estudo.

Foi utilizado o filtro de ocorrência pelo sistema ENVI 4.7. Esse tipo de filtro é baseado nas medidas de ocorrência, utilizando no seu cálculo o número dos níveis de cinza que estão dentro da janela de processamento. Foram extraídas

5 informações de textura de cada atributo, sendo elas: (1) alcance, (2) média, (3) variância, (4) entropia e (5) assimetria.

Para essa classificação, foram utilizadas as mesmas amostras de pixels e as mesmas classes determinadas na classificação 2009a. A Tabela 4.9 lista os 17 atributos utilizados como dados de entrada na árvore de decisão para classificação da imagem.

Tabela 4.9 - Atributos de entrada na árvore de decisão da classificação 2009b.

Atributos
B1 - Landsat-5
B2 - Landsat-5
B3 - Landsat-5
B4 - Landsat-5
B5 - Landsat-5
B7 - Landsat-5
Fração solo do MLME
Fração vegetação do MLME
Fração água do MLME
NDVI
NDWI
SAVI
Alcance de cada banda
Média de cada banda
Variância de cada banda
Entropia de cada banda
Assimetria de cada banda

A árvore de decisão gerada pelo algoritmo SimpleCart do aplicativo WEKA manteve-se relativamente grande, totalizando 371 nós. Deste valor, 186 são nós folhas, que estão associadas a uma determinada classe.

A adição da informação de textura resultou no aumento do tempo gasto para a criação da árvore de decisão, passando de 19,55s da classificação 2009a para 111,93s nessa nova classificação 2009b. O aumento no tempo de construção da árvore de decisão pode ser explicado pelo maior número de atributos utilizados na classificação.

Para a validação da classificação, foram utilizadas novamente 10.710 amostras, deste valor, 9.917 amostras de pixels, aproximadamente 92,6%, foram corretamente classificadas e 793 pixels, aproximadamente 7,4% foram classificados incorretamente pela árvore de decisão. Os valores dos pixels corretamente classificados estão posicionados na diagonal principal da Tabela

4.10 e estão destacados pela cor cinza. Os valores que estão fora da diagonal principal, são os pixels que foram classificados incorretamente, estando destacados os valores mais significativos na cor azul.

Tabela 4.10 - Matriz de confusão da segunda classificação.

		Imagem Referência								
		Regenera-ção	Pasto	Floresta Degradada	Floresta	Área Agrícola	Nuvem	Sombra de Nuvem	Água	Área Urbana
Imagem Classificada	Regeneração	949	104	153	1	2	0	0	0	0
	Pasto	134	1.031	24	2	9	0	0	0	1
	Floresta Degradada	126	21	914	77	1	0	0	0	1
	Floresta	17	0	53	1142	0	0	0	0	0
	Área Agrícola	14	15	1	0	1.119	0	0	0	21
	Nuvem	0	0	0	0	0	1.192	0	0	0
	Sombra de Nuvem	0	0	0	1	0	0	1.183	0	0
	Água	0	0	2	0	0	0	2	1.211	0
	Área Urbana	0	0	0	0	13	0	0	0	1.176

Os maiores erros de classificação ocorreram entre as classes de Regeneração e Floresta Degradada, Regeneração e Pasto e entre Floresta e Floresta Degradada. A classe Água manteve-se com uma boa classificação, pois novamente não ocorreu nenhum erro de classificação. Além disso, a adição das informações de textura ocasionou uma melhora na classificação dos pixels da classe Nuvem, não apresentando erro de classificação.

A média extraída através do filtro de textura do índice SAVI, é o atributo que aparece na raiz da árvore de decisão. A partir deste atributo, foi possível separar a classe Nuvem das demais classes, como ilustrado na Figura 4.25. Nesta primeira regra, todos os pixels que tiverem a média do índice SAVI maior ou igual a 48,5, são classificados como Nuvem.

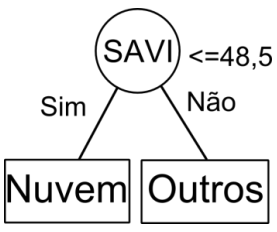

Figura 4.25 - Primeiro atributo usado na árvore de decisão da classificação 2009b.

O índice Kappa apresentou um aumento, pois passou de 0,7883 da classificação 2009a para 0,9167,com a adição das informações de textura. Porém, a classificação final não ficou visualmente nítida, pois foram perdidas as informações de borda dos objetos e seus limites foram degradados. Essas perdas ocorreram principalmente nas áreas das classes de Áreas Agrícolas e de Áreas Urbanas, acarretando em perdas dos limites e a informação exata do tamanho dessas áreas. Nas áreas da classe Floresta, a perda de informação das bordas não acarreta em problemas visuais, devido a sua grande extensão.

A Figura 4.26 apresenta o resultado final da classificação 2009b e a Figura 4.27 apresenta com maiores detalhes a classificação da área da FLONA do Tapajós onde localiza-se a Comunidade São Jorge.

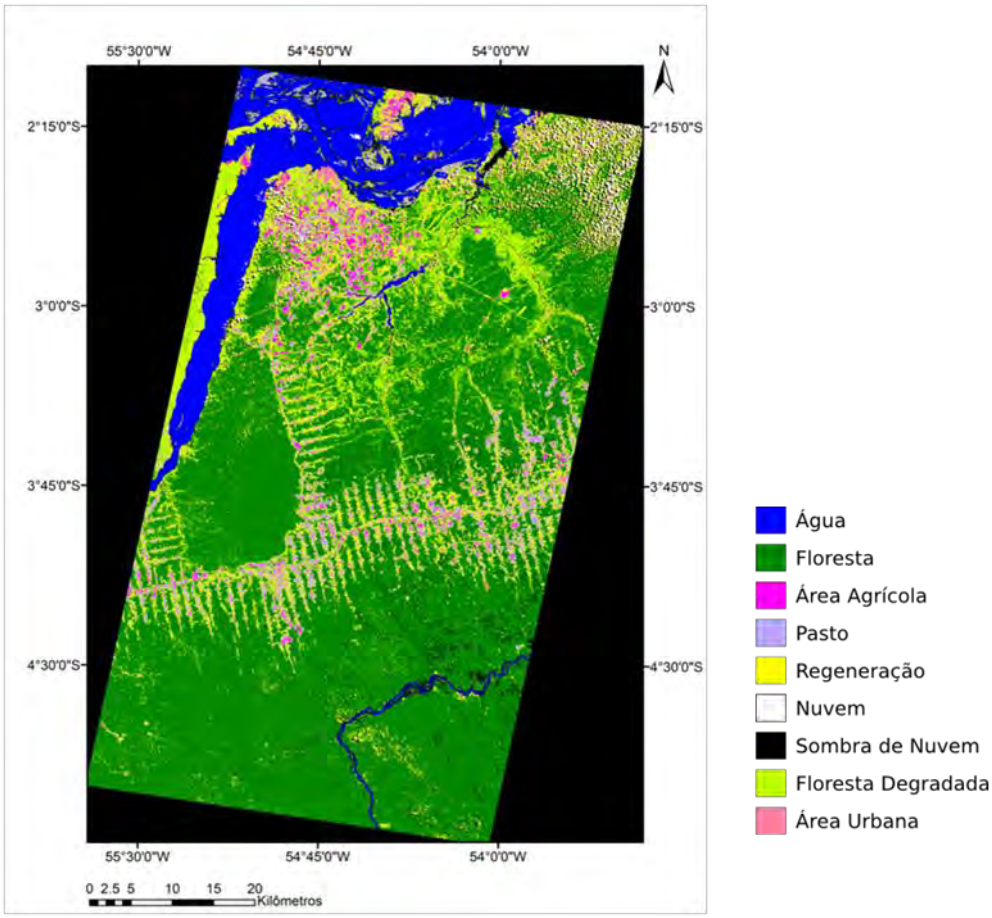

Figura 4.26 - Resultados da classificação 2009b.

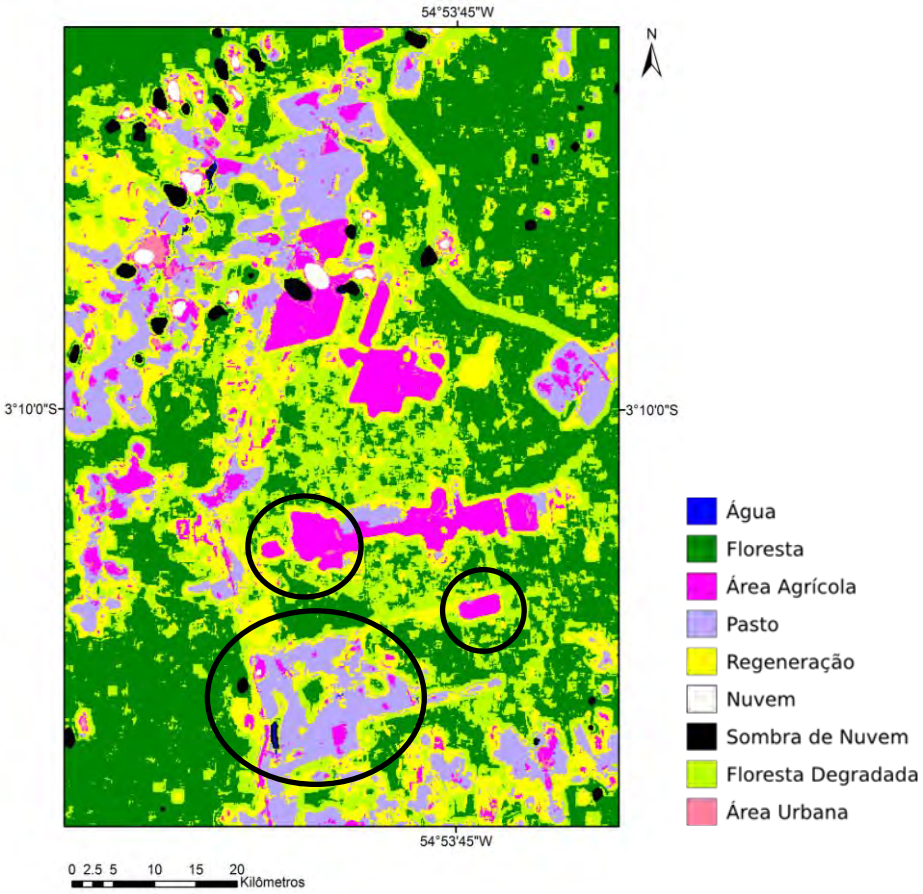

Figura 4.27 - Detalhes da classificação 2009b da Comunidade São Jorge. Os círculos na cor preta mostram algumas áreas que perderam informações de borda.

4.1.1.3. Classificação das imagens Landsat-5 e RADARSAT-2 de 2009

A terceira classificação, 2009c, foi obtida através da adição das imagens intensidade, as diferenças de fase e das 3 componentes de espalhamento P_v, P_d e P_s, extraídas a partir da imagem RADARSAT-2 aos dados da classificação 2009a. A classificação 2009c ficou reduzida a região onde está localizada a Comunidade São Jorge, pois as imagens RADARSAT-2 cobrem apenas essa porção da FLONA do Tapajós. Além da diminuição da área de estudo, também foram reduzidas as classes de estudo, pois as amostras para treinamento e validação da árvore de decisão ficaram restritas às classes que estão na região da Comunidade São Jorge.

Foram testados alguns filtros que estão disponíveis no aplicativo ENVI 4.7 com diferentes tamanhos de janela para reduzir o ruído speckle dos produtos da imagem RADARSAT-2. Os filtros testados foram: Lee, Enhanced Lee, Frost, Enhanced Frost e Gamma. O filtro escolhido para ser utilizado nesse trabalho

foi o filtro Gamma com uma janela de tamanho 7x7, pois apresentou melhor resultado visual das imagens intensidade, das diferenças de fase e das 3 componentes de espalhamento P_v, P_d e P_s.

Nessa terceira classificação, foram determinadas, de acordo com os dados de campo adquiridos na Comunidade São Jorge, as seguintes classes: (a) Área Agrícola, (b) Água, (c) Pasto, (d) Regeneração, (e) Floresta e (f) Floresta Degradada. Para as classes Área Agrícola, Floresta, Regeneração e Pasto, estavam disponíveis 5.000 pixels, enquanto que para a classe Floresta Degradada foram usados 2.695 pixels e para a classe Água foram utilizados 345 pixels, totalizando 23.040 pixels.

A Tabela 4.11 lista os 31 atributos utilizados como dados de entrada no treinamento da árvore de decisão na classificação 2009c.

Tabela 4.11 - Atributos de entrada na árvore de decisão da classificação 2009c.

Atributos
B1 - Landsat-5
B2 - Landsat-5
B3 - Landsat-5
B4 - Landsat-5
B5 - Landsat-5
B7 - Landsat-5
Fração solo do MLME
Fração vegetação do MLME
Fração água do MLME
NDVI
NDWI
SAVI
Imagem C11
Imagem C12
Imagem C13
Imagem C14
Imagem C21
Imagem C22
Imagem C23
Imagem C24
Imagem C31
Imagem C32
Imagem C33
Imagem C34
Imagem C41
Imagem C42
Imagem C43

(Continua)

Tabela 4.11 - Conclusão

Atributos
Imagem C44
Volumétrico
Double-bounce
Superficial

A árvore de decisão obtida pelo sistema WEKA apresentou um tamanho menor quando comparado com a classificação 2009a, passando de 335 para 219 nós. Deste montante, 110 eram nós folhas.

O software mostrou-se rápido na criação dos limiares que determinaram as regras de separação das classes, pois o tempo gasto para a geração da árvore de decisão foi de 34,42s. Apesar do tempo gasto para a geração da árvore de decisão dessa classificação ser maior do que o da classificação 2009a, a adição dos dados RADARSAT-2 possibilitou uma redução no tamanho da árvore de decisão para a classificação dos pixels.

Para validação da classificação, foram utilizadas 7.834 amostras de pixels, deste total, 6.446 pixels, aproximadamente 82,3%, foram classificadas corretamente e 1.388 pixels, aproximadamente 17,71% foram classificados incorretamente pela árvore de decisão. Os valores corretamente classificados estão posicionados na diagonal principal da Tabela 4.12 e estão destacados pela cor cinza. Os valores que estão fora da diagonal são os pixels que foram classificados incorretamente, estando destacados os valores mais significativos na cor azul.

Tabela 4.12 - Matriz de confusão da classificação 2009c.

		Imagem Referência					
		Área Agrícola	Floresta	Regeneração	Pasto	Água	Floresta Degradada
Imagem Classificada	Área Agrícola	1.672	0	0	58	0	0
	Floresta	0	1.500	52	0	0	154
	Regeneração	2	48	1.296	213	0	80
	Pasto	21	0	108	1.591	0	2
	Água	0	7	1	0	108	1
	Floresta Degradada	0	497	140	2	0	279

Os maiores erros de classificação ocorreram entre as classes de Floresta e Floresta Degradada e entre as classes de Regeneração e Pasto. Outro importante resultado encontrado nessa classificação foi que, apesar da adição de dados da imagem RADARSAT-2, a árvore de decisão não foi capaz de classificar corretamente as áreas encobertas por nuvens. Além desses resultados, pode-se analisar que os dados obtidos da imagem RADARSAT-2 foram pouco utilizados nos nós da árvore de decisão para a separação e classificação dos pixels. Caso opte-se por aumentar a uso dos atributos SAR, seria necessária uma interferência do usuário na árvore de decisão.

Os atributos que representam, respectivamente, a fração-solo do MLME e o índice NDVI foram os dois primeiros atributos utilizados para separar a classe Área Agrícola das demais classes, como ilustra a Figura 4.28. Neste caso, os pixels que tiveram valores da imagem-fração solo maior ou igual a 155,5 e o índice NDVI menor que 54,5, são classificados como Área Agrícola.

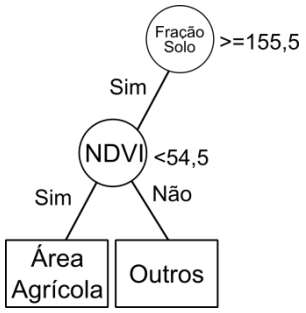

Figura 4.28 - Os dois primeiros atributos usados na árvore de decisão da classificação 2009c.

O valor obtido para o Kappa foi de 0,7764 para essa classificação. Com a adição dos produtos obtidos pela imagem RADARSAT-2, o valor do índice Kappa apresentou uma redução pouco significativa quando comprado ao índice Kappa da classificação 2009a (0,7883).

A Figura 4.29 apresenta o resultado final da classificação 2009c obtido através da árvore de decisão gerada pelo sistema WEKA, e a Figura 4.30 apresenta com maiores detalhes os erros de classificação das áreas encobertas por nuvens.

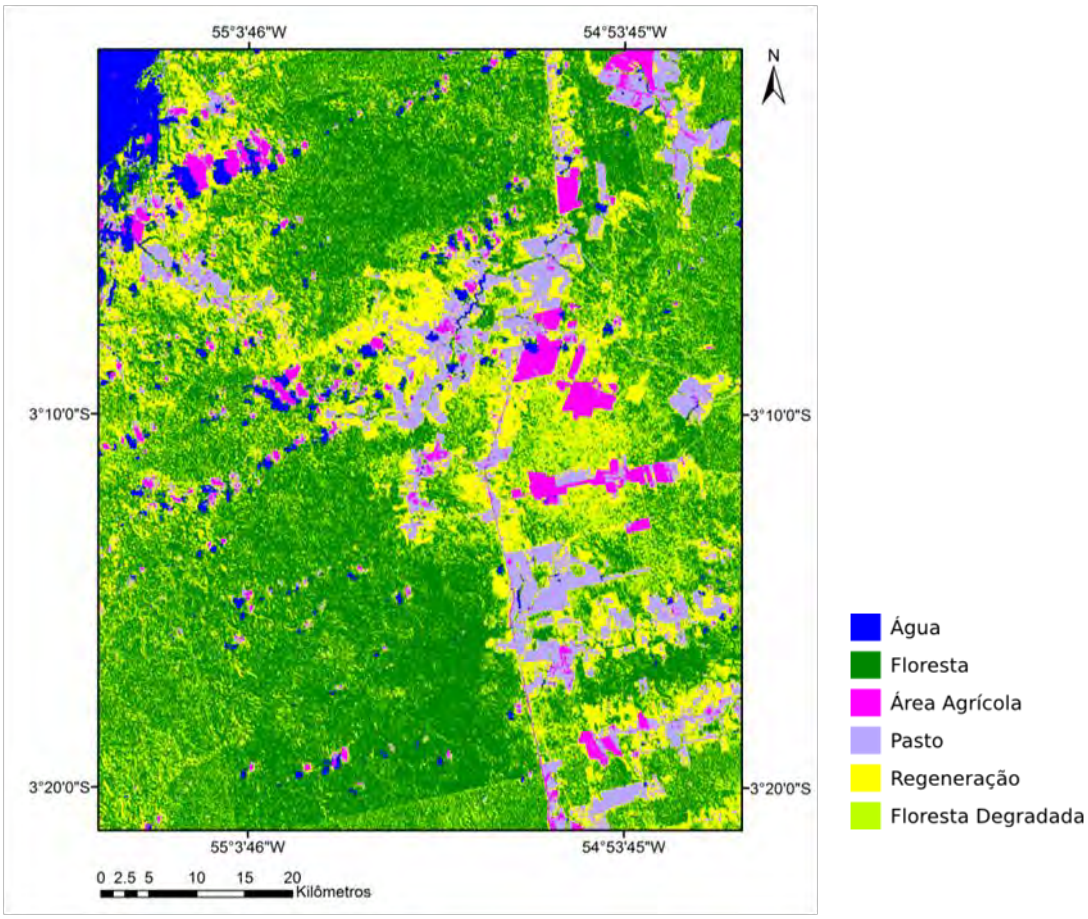

Figura 4.29 - Resultado da classificação 2009c.

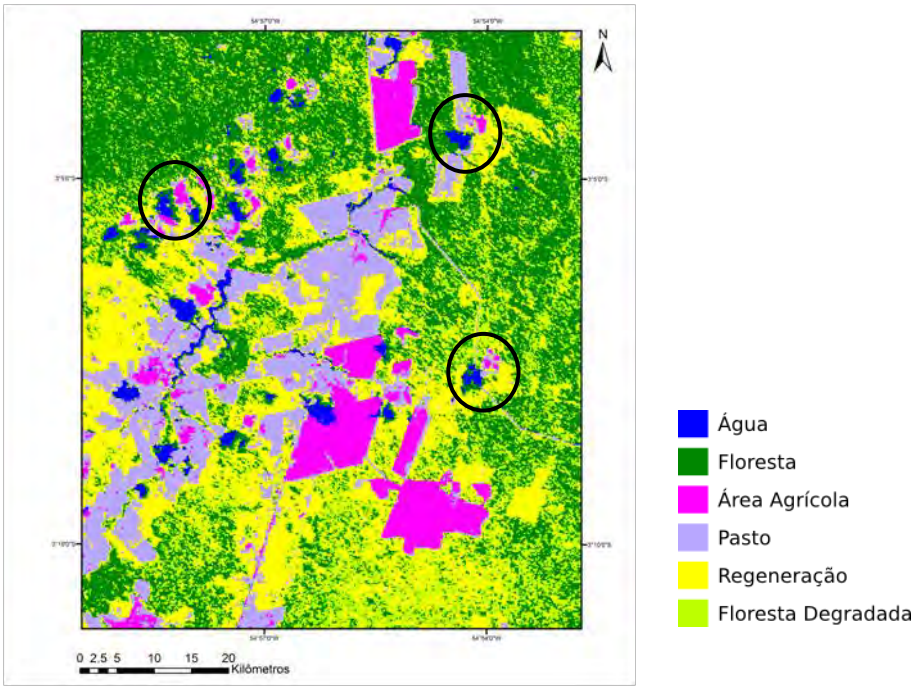

Figura 4.30 - Detalhes dos erros de classificação (círculos pretos) das áreas encobertas por nuvens da classificação 2009c.

4.2. Classificação com dados de 2008

A partir das experiências obtidas com os experimentos realizados para os dados do ano de 2009, foram definidos, a partir de combinações dos parâmetros de entrada, dois testes com os dados do ano de 2008.

A Tabela 4.13 lista as classificações geradas para o ano de 2008, onde primeira a coluna apresenta um nome único dado a classificação para permitir uma maior clareza no texto. A segunda, terceira e quarta colunas apresentam os atributos utilizados como dados de entrada para o treinamento da árvore de decisão.

Tabela 4.13 - Classificações geradas para o ano de 2008.

Identificação	Atributos da Classificação		
2008a	Landsat-5	NDVI, NDWI, SAVI, MLME	RADARSAT-2/Descendente
2008b	Landsat-5	NDVI, NDWI, SAVI, MLME	RADARSAT-2/Ascendente

4.2.1. Classificação da imagem Landsat-5 e RADARSAT-2/Descendente de 2008

A quarta classificação 2008a foi obtida através do algoritmo SimpleCart da árvore de decisão gerada pelo aplicativo WEKA, partir das seis bandas do sensor Landsat-5, das três frações do MLME, dos índices de vegetação, dos dados das imagens intensidade, das diferenças de fase e das 3 componentes de espalhamento P_v, P_d e P_s, extraídas a partir da imagem RADARSAT-2 do ano de 2008. Nessa classificação a área de estudo e o número de amostras também ficaram reduzidas a região onde está localizada a Comunidade São Jorge, pois as imagens RADARSAT-2 cobrem apenas essa porção da FLONA do Tapajós.

Assim como na classificação 2009c, foi utilizado o filtro Gamma com uma janela de tamanho 7x7 para reduzir os efeitos do ruído speckle das imagens intensidade, das diferenças de fase e das 3 componentes de espalhamento P_v, P_d e P_s.

Nessa quarta classificação, foram determinadas de acordo com os dados de campo realizado no ano de 2008, as seguintes classes: (a) Área Agrícola, (b) Água, (c) Pasto, (d) Regeneração, (e) Floresta e (f) Floresta Secundária. Para as classes Área Agrícola, Floresta, Floresta Secundária e Pasto estavam disponíveis 5.000 pixels, enquanto que para a classe Regeneração foram utilizados 1.329 pixels e para a classe Água foram usados 359 pixels. Nessa classificação não foi possível determinar a classe Floresta Degradada, como nas classificações 2009a, 2009b e 2009c, pois o número de amostras dessa classe não era suficiente para o uso do método Holdout. Foi adicionada a classe Floresta Secundária à 2008a, devido ao grande número de amostras dessa classe coletadas no ano de 2008.

A Tabela 4.14 lista os 31 atributos utilizados como dados de entrada para o treinamento da árvore de decisão na classificação 2008a.

Tabela 4.14 - Atributos de entrada no treinamento da árvore de decisão da classificação 2008a.

Atributos
B1 - Landsat-5
B2 - Landsat-5
B3 - Landsat-5
B4 - Landsat-5
B5 - Landsat-5
B7 - Landsat-5
Fração solo do MLME
Fração vegetação do MLME
Fração água do MLME
NDVI
NDWI
SAVI
Imagem C11
Imagem C12
Imagem C13
Imagem C14
Imagem C21
Imagem C22
Imagem C23
Imagem C24
Imagem C31
Imagem C32
Imagem C33
Imagem C34
Imagem C41

(Continua)

Tabela 4.14 – Conclusão

Atributos
Imagem C42
Imagem C43
Imagem C44
Volumétrico
Double-bounce
Superficial

A árvore de decisão gerada pelo aplicativo WEKA apresentou um tamanho menor quando comparado com a árvore de decisão gerada na classificação 2009c, que utilizou os mesmos tipos de atributos como dados de entrada para o treinamento da árvore de decisão. A quantidade total de nós na classificação 2009c foi de 219 e para essa classificação 2008a foi 193, sendo que 97 desses elementos eram nós folha. O software mostrou-se rápido na criação dos limiares que determinam as regras para a separação das classes, pois o tempo gasto para a geração da árvore de decisão foi de 28,93s.

Para a validação da árvore de decisão gerada pelo sistema WEKA, foram utilizadas 7.374 amostras, deste valor, 6.294 pixels, aproximadamente 85,4%, foram classificadas corretamente e 1.080 pixels, aproximadamente 14,6%, foram classificados incorretamente. Os valores corretamente classificados estão posicionados na diagonal principal da Tabela 4.15 e estão destacados pela cor cinza. Os valores que estão fora da diagonal são os pixels que foram classificados incorretamente, estando destacados os valores mais significativos na cor azul.

Tabela 4.15 - Matriz de confusão da classificação 2008a.

		Imagem Referência					
		Floresta	**Pasto**	**Floresta Secundária**	**Área Agrícola**	**Regeneração**	**Água**
Imagem Classificada	**Floresta**	1.367	1	275	0	32	1
	Pasto	17	1.597	19	20	58	1
	Floresta Secundária	299	29	1.351	0	52	0
	Área Agrícola	0	26	1	1.655	2	0
	Regeneração	28	88	129	2	193	0
	Água	0	0	0	0	0	131

Os maiores erros de classificação ocorreram entre as classes Floresta e Floresta Secundária, Floresta Secundária e Regeneração e entre as classes de Pasto e Regeneração. Outro importante erro de classificação encontrado foi a falta de capacidade de classificar corretamente as áreas encobertas por nuvens.

Os atributos que representam, respectivamente, a banda 2 do sensor TM Landsat-5, o índice NDVI, a banda 5 do sensor TM Landsat-5 e a fração-solo do MLME, foram os quatro primeiros atributos utilizados para separar a classe Água das demais classes, como ilustrado na Figura 4.31. Neste caso, os pixels que tiveram valores da banda 2 menor ou igual a 17,5, o índice NDVI menor ou igual 63,5, a banda 5 menor ou igual a 48,5 e valores da imagem-fração menor ou igual a 71,5, são classificados como Água.

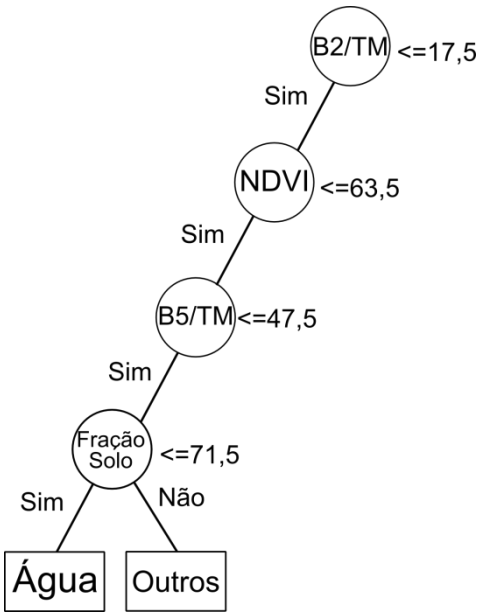

Figura 4.31 - Atributos usados para a separação da classe Água da classificação 2008a.

O valor obtido do índice Kappa foi de 0,8125 para essa classificação. O valor do índice Kappa da classificação 2009c (0,7764), que utilizou os mesmos tipos de atributos como dados de entrada na árvore de decisão, ficou próximo do índice dessa classificação.

A Figura 4.32 apresenta o resultado final da classificação 2008a obtido através da árvore de decisão gerada pelo sistema WEKA, e a Figura 4.33 apresenta

com maiores detalhes os erros de classificação das áreas encobertas por nuvens.

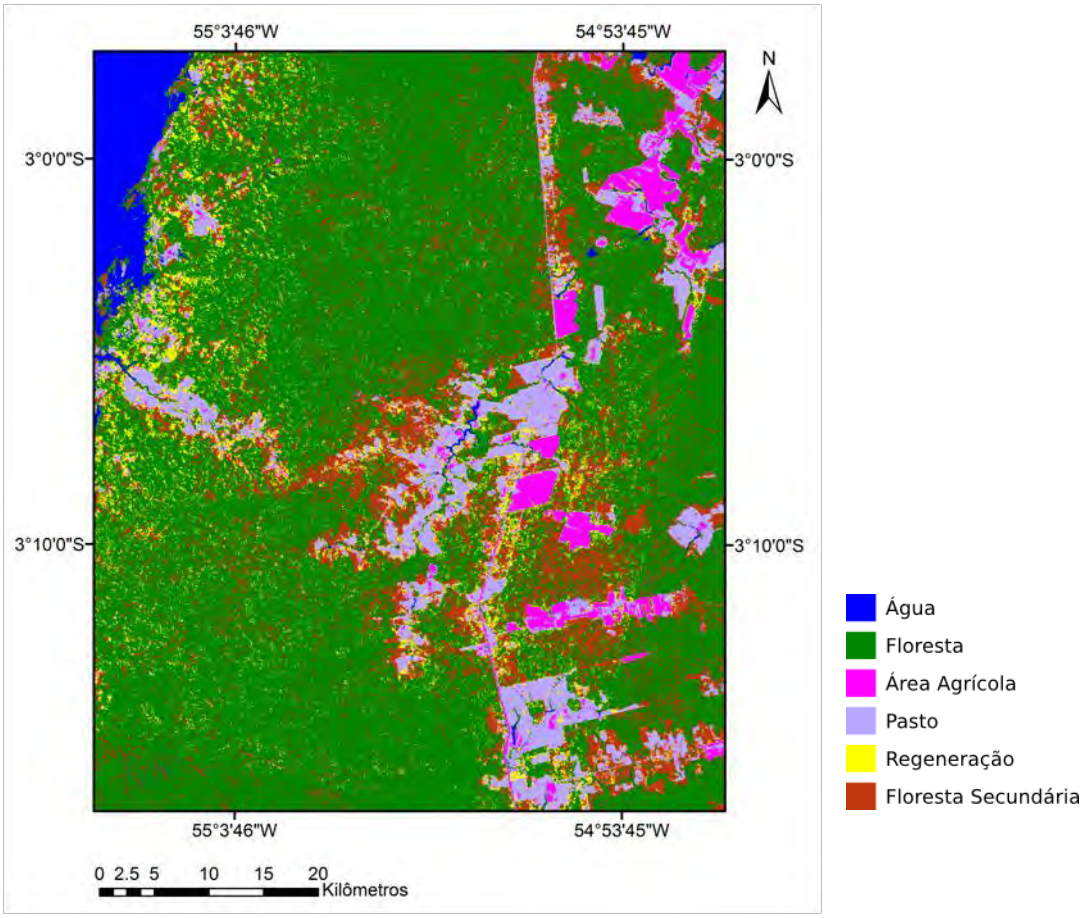

Figura 4.32 - Resultado da classificação 2008a.

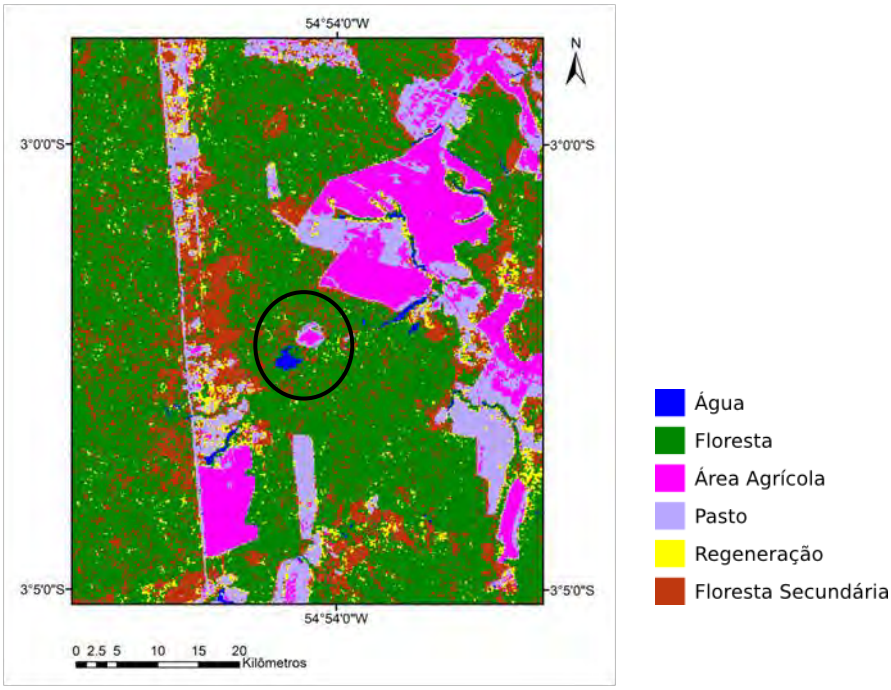

Figura 4.33 - Detalhes dos erros (círculo preto) de classificação das áreas encobertas por nuvens na classificação 2008a.

4.2.2. Classificação da imagem Landsat-5 e RADARSAT-2/Ascendente do ano de 2008

Foi realizada uma classificação baseada na classificação 2008a, porém usando os dados das imagens intensidade, das diferenças de fase e das 3 componentes de espalhamento P_v, P_d e P_s, extraídas a partir de uma imagem RADARSAT-2 de órbita ascendente.Essa classificação é denominada 2008b.

O pixel da imagem RADARSAT-2 foi degradado para ficar do mesmo tamanho do pixel da imagem Landsat-5, com isso, o pixel final ficou com resolução espacial de 30x30m.

A classificação 2008b foi obtida através da integração das seis bandas do sensor TM Landsat-5, das três frações do MLME, dos índices de vegetação, dos dados das imagens intensidade, das diferenças de fase e das 3 componentes de espalhamento P_v, P_d e P_s, extraídas a partir da imagem RADARSAT-2 do ano de 2008 ascendente.

Nessa classificação, foram determinadas, de acordo com os dados de campo, as seguintes classes: (a) Área Agrícola, (b) Água, (c) Pasto, (d) Regeneração, (e) Floresta e (f) Floresta Secundária. Para as classes Área Agrícola, Floresta, Floresta Secundária e Pasto estavam disponíveis 5.000 pixels, enquanto que para a classe regeneração foram utilizados 1.155 pixels e para a classe Água foram usados 286 pixels.

A Tabela 4.16 lista os atributos utilizados como dados de entrada para o treinamento para a árvore de decisão na classificação 2008b.

Tabela 4.16 - Atributos de entrada no treinamento da arvore de decisão da classificação 2008b

Atributos
B1 - Landsat-5
B2 - Landsat-5
B3 - Landsat-5
B4 - Landsat-5
B5 - Landsat-5
B7 - Landsat-5
Fração solo do MLME
Fração vegetação do MLME
Fração água do MLME
NDVI
NDWI
SAVI
Imagem C11
Imagem C12
Imagem C13
Imagem C14
Imagem C21
Imagem C22
Imagem C23
Imagem C24
Imagem C31
Imagem C32
Imagem C33
Imagem C34
Imagem C41
Imagem C42
Imagem C43
Imagem C44
Volumétrico
Double-bounce
Superficial

A árvore de decisão gerada pelo sistema WEKA teve um aumento da quantidade de nós em relação à árvore da classificação 2008a, totalizando 219 nós, sendo que 110 foram de nós folhas. O processamento dos dados para a criação dos limiares que determinam as regras para a separação das classes foi de 31,29s.

Para validação da árvore de decisão foram utilizadas 7.290 amostras de pixels, deste valor, 6.085 pixels, aproximadamente 83,5%, foram classificadas corretamente e 1.205 pixels, aproximadamente 16,5% foram classificados incorretamente pela árvore de decisão. Os valores corretamente classificados estão posicionados na diagonal principal da Tabela 4.17 e estão destacados pela cor cinza. Os valores que estão fora da diagonal são os pixels que foram classificados incorretamente, estando destacados os valores mais significativos na cor azul.

Tabela 4.17 - Matriz de confusão da classificação 2008b.

		Imagem Referência					
		Floresta	**Pasto**	**Floresta Secundária**	**Área Agrícola**	**Regeneração**	**Água**
Imagem Classificada	**Floresta**	1.338	2	304	0	8	0
	Pasto	9	1.666	27	21	46	0
	Floresta Secundária	378	22	1.263	2	34	0
	Área Agrícola	0	65	1	1.617	2	0
	Regeneração	27	101	150	1	109	0
	Água	2	2	1	0	0	92

Os maiores de erros de classificação dos pixels ocorreram entre as áreas de Floresta e Floresta Secundária, Floresta Secundária e Regeneração e entre a classe Pasto e Regeneração.

Os atributos que representam, respectivamente, a banda 2 do sensor TM Landsat-5, o índice NDVI e a fração-solo do MLME. Esses atributos foram utilizados para separar a classe água das demais classes, conforme ilustrado na Figura 4.34. Neste caso, os pixels que tiverem valores da banda 2 maior ou igual a 17,5, o índice NDVI menor ou igual 61,5 e a imagem-fração solo menor do que 82.5, são classificados como Água.

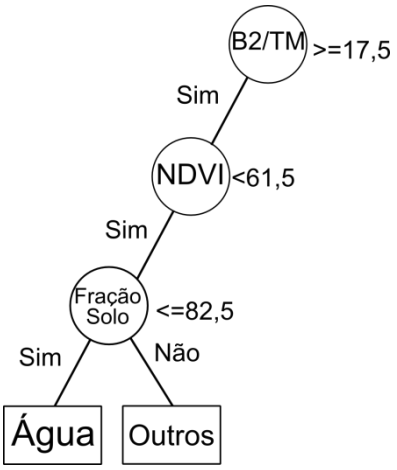

Figura 4.34 - Atributos usados para a separação da classe Água da classificação 2008b.

O valor obtido do índice Kappa foi de 0,7865. Comparando o valor do índice Kappa da classificação 2008a (0,8125) com o valor obtido nessa classificação, nota-se que não ocorreu uma diferença significativa entre as classificações. A Figura 4.35 apresenta o resultado final da classificação 2008b obtido através da árvore de decisão gerada pelo sistema WEKA, e a Figura 4.36 apresenta com maiores detalhes a classificação das áreas de floresta encobertas por nuvens.

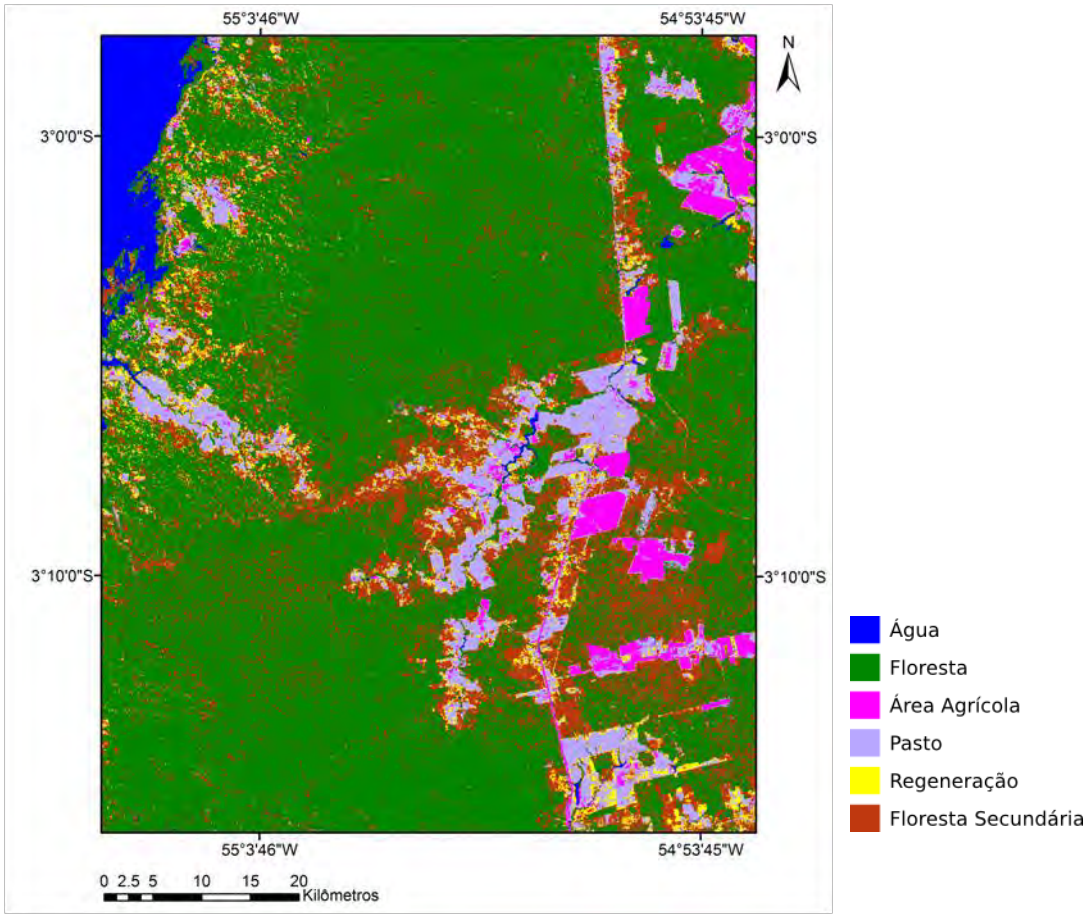

Figura 4.35 - Resultado da classificação 2008b.

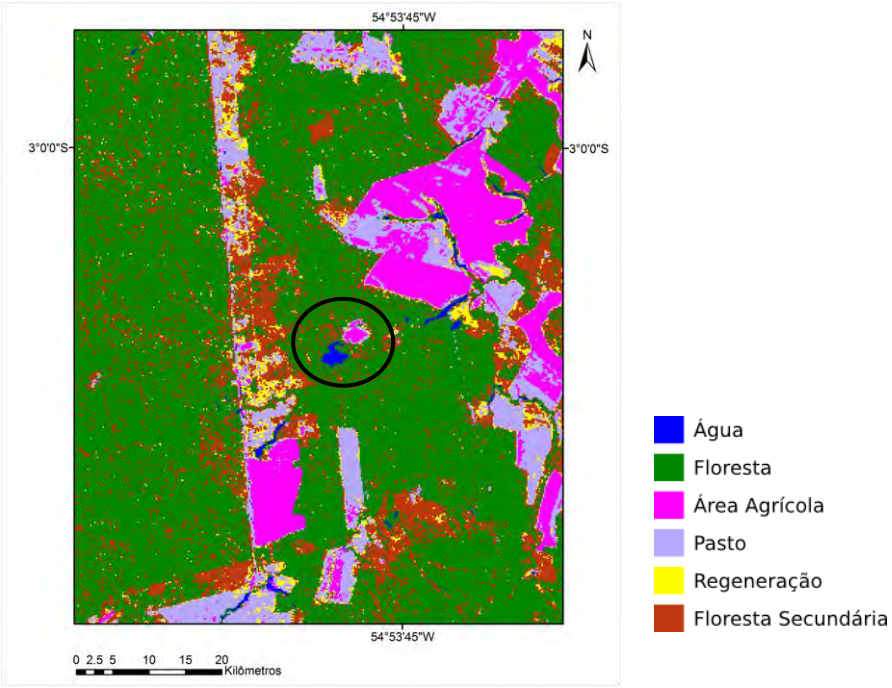

Figura 4.36 - Detalhes dos erros (círculo preto) da classificação de uma área encoberta por nuvens.

4.3. Classificação dos anos de 2008 e 2009 com todos os atributos

A partir das experiências obtidas com as diversas combinações para as classificações da cobertura vegetal da FLONA do Tapajós nos de 2008 e 2009, foram definidas, duas classificações utilizando seis bandas do Landsat-5, imagens frações do MLME, índices de vegetação, dados obtidos a partir das imagens RADARSAT-2 e atributos de textura.

A Tabela 4.18 lista as classificações geradas para o ano de 2008 e 2009, onde primeira coluna apresenta um nome único dado a classificação. A segunda, terceira e quarta colunas apresentam os atributos utilizados como dados de entrada para o treinamento da árvore de decisão.

Tabela 4.18 - Classificações geradas para o ano de 2008 e 2009.

Identificação	**Atributos da Classificação**			
2008c	Landsat-5	NDVI, NDWI, SAVI, MLME	Textura	RADARSAT-2/Descendente
2009d	Landsat-5	NDVI, NDWI, SAVI, MLME	Textura	RADARSAT-2/Ascendente

Para essas duas classificações geradas, não foi possível obter o mapa com as classes de uso e cobertura da terra, pois as árvores de decisão geradas possuem números de nós superior ao limite que o sistema ENVI 4.7 suporta para execução. Por isso, serão apresentados apenas os resultados obtidos pelo aplicativo WEKA.

4.3.1. Classificação da imagem do ano de 2008 com todos os atributos

A classificação 2008c foi obtida através da integração das seis bandas do sensor TM Landsat-5, das três frações do MLME, dos índices de vegetação, dos atributos de textura, dos dados das imagens intensidade, das diferenças de fase e das 3 componentes de espalhamento P_v, P_d e P_s, extraídas a partir da imagem RADARSAT-2 ascendente.

Nessa classificação, foram determinadas, de acordo com os dados de campo, as seguintes classes: (a) Área Agrícola, (b) Água, (c) Pasto, (d) Regeneração, (e) Floresta e (f) Floresta Secundária. Para as classes Floresta, Floresta

Secundária e Pasto estavam disponíveis 5.000 pixels, enquanto que para a classe Área Agrícola foram usados 4.073 pixels, para a classe Regeneração foram utilizados 1.238 pixels e para a classe Água foram usados 355 pixels.

A Tabela 4.19 lista os atributos utilizados como dados de entrada para o treinamento para a árvore de decisão na classificação 2008c.

A árvore de decisão gerada pelo sistema WEKA apresentou um grande número de nós, totalizando 377 nós, sendo que 189 foram de nós folhas. O tempo de processamento dos dados para a criação dos limiares que determinam as regras para a separação das classes foi de 120,24s. O aumento do tempo gasto para a geração da árvore de decisão e no seu tamanho podem ser explicados pela grande quantidade de atributos utilizados como dados de entrada na árvore de decisão.

Tabela 4.19 - Atributos de entrada no treinamento da arvore de decisão da classificação 2008c

Atributos
B1 - Landsat-5
B2 - Landsat-5
B3 - Landsat-5
B4 - Landsat-5
B5 - Landsat-5
B7 - Landsat-5
Fração solo do MLME
Fração vegetação do MLME
Fração água do MLME
NDVI
NDWI
SAVI
Imagem C11
Imagem C12
Imagem C13
Imagem C14
Imagem C21
Imagem C22
Imagem C23
Imagem C24
Imagem C31
Imagem C32
Imagem C33
Imagem C34
Imagem C41

(Continua)

Tabela 4.19 - Conclusão

Atributos
Imagem C42
Imagem C43
Imagem C44
Volumétrico
Double-bounce
Superficial
Alcance de cada banda
Média de cada banda
Variância de cada banda
Entropia de cada banda
Assimetria de cada banda

Para a validação da árvore de decisão foram utilizadas 7.026 amostras de pixels, deste valor, 6.622 pixels, aproximadamente 94,2%, foram classificadas corretamente e 404 pixels, aproximadamente 5,8% foram classificados incorretamente pela árvore de decisão. Os valores corretamente classificados estão posicionados na diagonal principal da Tabela 4.20 e estão destacados pela cor cinza. Os valores que estão fora da diagonal são os pixels que foram classificados incorretamente, estando destacados os valores mais significativos na cor azul.

Tabela 4.20 - Matriz de confusão da classificação 2008c

		Imagem Referência					
		Floresta	**Pasto**	**Floresta Secundária**	**Área Agrícola**	**Regeneração**	**Água**
Imagem Classificada	**Floresta**	1.576	0	97	0	5	0
	Pasto	0	1.647	15	10	21	0
	Floresta Secundária	71	15	1.551	0	58	2
	Área Agrícola	0	10	0	1.417	0	1
	Regeneração	8	30	54	2	324	0
	Água	0	2	3	0	0	107

Os maiores erros de classificação dos pixels ocorreram entre as classes de Floresta e Floresta Secundária e entre Pasto e Regeneração.

Os atributos que representam, respectivamente, a média da banda 3 do sensor TM Landsat-5, a média do índice NDVI e a média da fração-solo do MLME.

Esses atributos foram utilizados para separar a classe Água das demais classes, conforme ilustrado na Figura 4.37. Neste caso, os pixels que tiverem valores de média da banda 3 maior ou igual a 16,5, a média do índice NDVI menor que 62,86 e a média da fração-solo do MLME menor que 90,5, são classificados como Água.

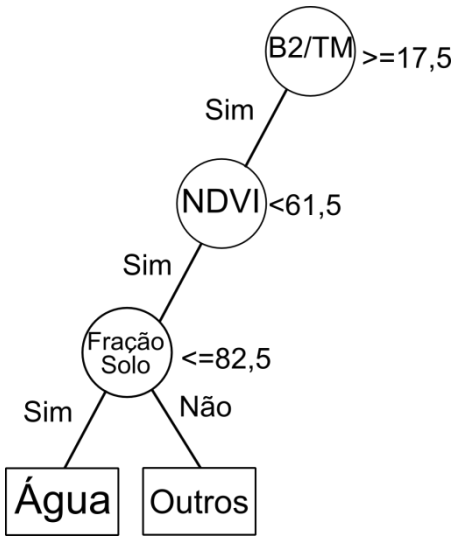

Figura 4.37 - Atributos usados para a separação da classe Água da classificação 2008c.

O valor obtido do índice Kappa foi de 0,9264. Este valor do índice Kappa mostra que a classificação da imagem foi excelente, pois os erros de classificação dos pixels foram muito baixos.

4.3.2. Classificação da imagem do ano de 2009 com todos os atributos

A classificação 2009d foi obtida através da integração das seis bandas do sensor TM Landsat-5, das três frações do MLME, dos índices de vegetação, dos atributos de textura, dos dados das imagens intensidades, das diferenças de fase e das 3 componentes de espalhamento P_v, P_d e P_s, extraídas a partir da imagem RADARSAT-2 ascendente.

Nessa classificação, foram determinadas, de acordo com os dados de campo, as seguintes classes: (a) Nuvem, (b) Área Agrícola, (c) Floresta, (d) Regeneração, (e) Pasto, (f) Água, (g) Floresta Degradada e (h) Sombra de Nuvem. Para as classes Floresta, regeneração e pasto estavam disponíveis

5.000 pixels, enquanto que para a classe floresta degradada foram usados 2.665 pixels, para a classe sombra de nuvem foram utilizados 1.250 pixels, para a classe nuvem foram usados 512 pixels e para a classe Água utilizados 340 pixels.

A Tabela 4.21 lista os atributos utilizados como dados de entrada para o treinamento para a árvore de decisão na classificação 2009b.

A árvore de decisão gerada pelo sistema WEKA apresentou um número grande de nós, totalizando 599 nós, sendo que 300 foram de nós folhas. O tempo de processamento dos dados para a criação dos limiares que determinam as regras para a separação das classes foi de 212,72s. O aumento do tempo gasto para a geração da árvore de decisão e no seu tamanho podem ser explicados pela grande quantidade de atributos utilizados como dados de entrada na árvore de decisão.

Tabela 4.21 - Atributos de entrada no treinamento da arvore de decisão da classificação 2009d.

Atributos
B1 - Landsat-5
B2 - Landsat-5
B3 - Landsat-5
B4 - Landsat-5
B5 - Landsat-5
B7 - Landsat-5
Fração solo do MLME
Fração vegetação do MLME
Fração água do MLME
NDVI
NDWI
SAVI
Imagem C11
Imagem C12
Imagem C13
Imagem C14
Imagem C21
Imagem C22
Imagem C23
Imagem C24
Imagem C31
Imagem C32
Imagem C33

(Continua)

Tabela 4.21 - Conclusão

Atributos
Imagem C34
Imagem C41
Imagem C42
Imagem C43
Imagem C44
Volumétrico
Double-bounce
Superficial
Alcance de cada banda
Média de cada banda
Variância de cada banda
Entropia de cada banda
Assimetria de cada banda

Para a validação da árvore de decisão foram utilizadas 7.877 amostras de pixels, deste valor, 7.272 pixels, aproximadamente 92,3%, foram classificadas corretamente e 605 pixels, aproximadamente 7,7% foram classificados incorretamente pela árvore de decisão. Os valores corretamente classificados estão posicionados na diagonal principal da Tabela 4.22 e estão destacados pela cor cinza. Os valores que estão fora da diagonal são os pixels que foram classificados incorretamente, estando destacados os valores mais significativos na cor azul.

Tabela 4.22 - Matriz de confusão da classificação 2008c.

		Imagem Referência							
		Nuvem	**Área Agrícola**	**Floresta**	**Regene-ração**	**Pasto**	**Água**	**Floresta Degradada**	**Sombra de Nuvem**
Imagem Classificada	**Nuvem**	156	1	0	0	1	0	0	0
	Área Agrícola	0	1.197	0	0	7	0	0	0
	Floresta	0	0	1.548	6	0	0	162	0
	Regeneração	0	0	6	1.559	83	0	25	0
	Pasto	0	7	0	88	1.582	2	2	3
	Água	0	0	1	2	0	98	2	9
	Floresta Degradada	0	0	153	36	1	0	726	0
	Sombra de Nuvem	0	0	0	0	0	7	1	406

Os maiores erros de classificação dos pixels ocorreram entre as áreas de Floresta e Floresta degradada, Regeneração e Pasto e entre Floresta degradada e Regeneração.

Os atributos que representam, respectivamente, a média banda 5 do sensor TM Landsat-5, a média do índice NDVI, a banda 1 do sensor TM Landsat-5 e a média da imagem-fração água do MLME. Esses atributos foram utilizados para separar a classe água das demais classes, conforme ilustrado na Figura 4.38. Neste caso, os pixels que tiverem valores da média banda 5 maior ou igual a 28,18, a média do índice NDVI menor que 79,94, a banda 1 menor do que 41,0 e a média da imagem-fração água menor que 144,24, são classificados como Água.

O valor obtido do índice Kappa foi de 0,9065. Este valor do índice Kappa também mostra que a classificação da imagem foi excelente, pois os erros de classificação dos pixels foram muito baixos.

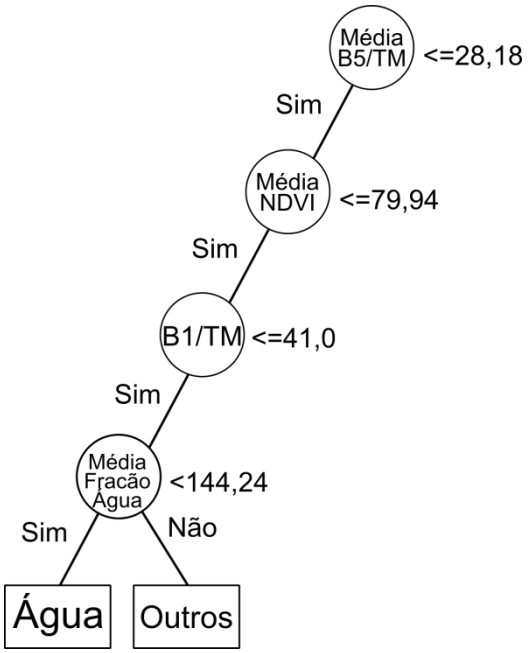

Figura 4.38 - Atributos usados para a classificação 2009d.

A Tabela 4.23 lista uma comparação entre o índice Kappa, os valores dos pixels que foram classificados corretamente, os pixels que foram incorretamente

classificados e o tempo gasto para a geração da árvore de decisão de cada classificação testa nesse trabalho.

Tabela 4.23 - Resumo dos resultados das classificações.

Classificação	Índice Kappa	Porcentagem de Pixels corretamente Classificados	Porcentagem de Pixels incorretamente Classificados	Tempo para criação da Árvore de Decisão
2009a	0,7883	81,2%	18,8%	19,55s
2009b	0,9167	92,6%	7,4%	111,93 s
2009c	0,7764	82,3%	17,71	34,42 s
2008a	0,8125	85,4%	14,6%	28,93 s
2008b	0,7865	83,5%	16,5%	31,92 s
2008c	0,9264	94,2%	5,8%	120,24 s
2009d	0,9065	92,3%	7,7%	212,72 s

5 CONCLUSÕES E RECOMENDAÇÕES

Os resultados obtidos nesse trabalho permitiram avaliar a eficiência do uso de árvore de decisão, uma técnica de mineração de dados, e a integração de dados multisensores, no mapeamento da cobertura florestal tropical da região da FLONA do Tapajós.

Dentro dos objetivos propostos, a árvore de decisão alcançou resultados satisfatórios na classificação da cobertura florestal da FLONA do Tapajós. As diversas classificações testadas e avaliadas geraram classificações com valores do índice Kappa variando entre Muito boa a Excelente de acordo com cada tipo de atributo utilizado como dado de entrada para o treinamento da árvore de decisão.

As classificações apresentaram confusões na separação das classes da cobertura florestal, principalmente entre as classes de floresta, floresta degradada, floresta secundária e regeneração. Essas confusões ocorreram devido ao comportamento espectral desses alvos serem muito próximos, dificultando a criação de limiares de separação dessas classes. Para diminuir esses erros, sugere-se o uso de dados com melhores resoluções espaciais, pois poderiam facilitar a separação das áreas de cobertura floresta da FLONA do Tapajós. Além disso, a utilização de outros atributos a partir dos dados SAR, como por exemplo, índices de vegetação, outros tipos de decomposições poderiam contribuir com uma melhora na classificação.

Nas classificações que integraram os dados ópticos provenientes do satélite Landsat-5 e do RADARDAT-2 como dados de entrada no treinamento da árvore de decisão, poucas vezes foram utilizadas as informações SAR nos seus nós. Sendo que, a maioria dos nós eram determinados por informações do satélite Landsat-5 e de produtos obtidos a partir desse tipo de dado, como as imagens-fração adquiridas do MLME e dos índices de vegetação NDVI, NDWI e SAVI. Para uma maior utilização dos dados SAR nos nós da árvore de decisão, sugere-se a utilização de árvore de decisão que permitam a interferência do usuário (operador) na determinação das regras de separação das classes da árvore de decisão.

As áreas que estavam encobertas por nuvens não foram corretamente classificadas pela árvore de decisão. As nuvens foram, na maioria das vezes, classificadas como áreas agrícolas e as sombras das nuvens foram classificadas como água. Devido à falta de capacidade da árvore de decisão classificar essas áreas, sem a utilização de amostras de treinamento das classes encobertas por nuvens, recomenda-se coletar amostras de pixels dessas regiões e utilizar como dados de entrada no treinamento da árvore de decisão.

O uso da técnica de árvore de decisão apresentou-se computacionalmente eficiente e rápido. Apesar das árvores de decisão geradas apresentarem um grande número de nós, o tempo de processamento foi pequeno.

Por fim, destaca-se que a utilização do aplicativo WEKA, um sistema livre e de código fonte aberto, permitiu alterar suas configurações e adaptá-lo as necessidades exigidas pelo tipo de dados utilizados nesse trabalho.

REFERÊNCIAS BIBLIOGRÁFICAS

AITKENHEAD, M. J. A co-evolving decision tree classification method. **Expert Systems with Applications,** v. 34, p. 18-25, 2008.

ALVES, D. S. Space-time dynamics of deforestation in Brazilian Amazônia. **International Journal of Remote Sensing**, v. 23, n. 14, p. 2903-2908, 2002.

ANDERSON, L. O.; SHIMABUKURO, Y. E.; LIMA, A.; MEDEIROS, J. S. Mapeamento da cobertura da terra do estado do Mato Grosso através da utilização de dados multitemporais do sensor MODIS. **Geografia**, v. 30, n. 2, p. 365-380, mai./ago., 2005.

ANGELIS, C. F. **Análise do comportamento temporal do retroespalhamento na banda L da cobertura da terra na Floresta do Tapajós, Pará, utilizando imagens JERS - 1**. 2000. 154 p. (INPE-9678-TDI/853). Dissertação (Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos. 2000. Disponível em: <http://urlib.net/sid.inpe.br/jeferson/2003/05.15.15.54>. Acesso em: 22 mar. 2010.

ARAGÃO, L. E. O. C. **Modelagem dos padrões temporal e espacial da produtividade primária bruta na região do Tapajós: uma análise multi-escala**. 2004. 283 p. (INPE-11423-TDI/951). Tese (Doutorado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos. 2004. Disponível em:<http://urlib.net/sid.inpe.br/jeferson/2004/07.14.13.33>. Acesso em: 10 jan. 2011.

BATISTA, G. E.; MONARD, M. C. **Utilizando Métodos Estatísticos de Resampling para Estimar a Performance de Sistemas de Aprendizado**. III Workshop de Dissertações Defendidas – LABIC Laboratory of Computational Intelligence, Universidade de São Paulo, 1998. Disponível em: <

http://www.icmc.usp.br/~mcmonard/public/icmcwshG1998.pdf>. Acesso em: 13 mai. 2011.

BITTENCOURT, H. R.; CLARK, R. T. Feature selection by using CLASSIFICATION AND REGRESSION TREES (CART). **Sensors Peterborough**, v. 52, i. 5, p. 1-19, 2003.

BOUCKAERT; R. R.; FRANK, E.; HALL, M.; KIRKBY, R.; REUTEMANN, P.; SEEWALD, A.; SCUSE, D. **WEKA Manual for Version 3-6-2**. 2010. 303 p.

BREIMAN, L.; FRIEDMAN, J. H.; OLSHEN, R. A.; STONE, C.J. **Classification and regression trees.** Belmont, CA: Wadsworth International, 1984. 358p., ISBN 0534980538.

CHANDER, G.; MARKHAM, B. L.; HELDER, D. L. Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. **Remote Sensing of Environment**, v.113, p.893-903, 2009.

CLARO, L. O. T. **Textura de imagens utilizando conceitos de morfologia matemática**. 2010. 68 p. Dissertação (Mestrado em Sensoriamento Remoto) – Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, 2010.

CLEMENT, C. R.; HIGUCHI, N. A. floresta Amazônica e o futuro do Brasil. **Jornal da Ciência**, p. 44-49, 2006.

CONGALTON, R. G. A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data. **Remote Sensing of Environment,** v.37, p. 35-46, 1991.

DAVRANCHE, A.; LEFEBVRE, G.; POULIN, B. Wetland monitoring using classification trees and SPOT-5 seasonal time series. **Remote Sensing of Environment**, v. 114, p. 552-562, 2010.

ELHADI, E. M.; ZOMRAWI, N. Object-based land use/cover extraction from QuickBird image using Decision tree. Journal of American Science, v. 6, i. 2, p. 176 – 180, 2010.

ESPÍRITO-SANTO, F. D. B. **Caracterização e mapeamento da vegetação da região da Floresta Nacional de Tapajós através de dados óticos, de radar e inventários florestais**. 2003. 277 p. (INPE-10133-TDI/898). Dissertação (Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos. 2003. Disponível em: <http://urlib.net/sid.inpe.br/jeferson/2003/12.09.13.07>. Acesso em: 22 mar. 2010.

FONSECA, J. M. M. R. Indução de Árvores de Decisão -HistClass - Proposta de um algoritmo não paramétrico. 1994. 151 p. Dissertação de Mestrado - **Universidade Nova de Lisboa . 1994**

FREEMAN, A.; DURDEN, S. L. A. Three-component Scattering Model for Polarimetric SAR Data. **IEEE Transactions on Geoscience and Remote Sensing**, v. 36, n. 3, p. 963-973, mai.,1998.

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

GAO, B. NDWI - A Normalized Difference Water Index for Remote Sensing of Vegetation Liquid Water from Space. **Remote Sensing of Environment**, v. 58, p. 257-266, 1996.

GUERRA, J. B. Análise metodológica do tratamento de dados sar r99b para discriminar incremento de desflorestamento no sudoeste da amazônia brasileira. 2008. 121 p. (INPE-15677-TDI/1451). Dissertação (Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José

dos Campos, 2008. Disponível em:<http://urlib.net/sid.inpe.br/mtc-m18@80/2009/02.18.14.18>. Acesso em: 20 nov. 2010.

GHOSE, M. K.; PRADHAN, R.; GHOSE, S. S. Decision Tree Classification of Remotely Sensed Satellite Data using Spectral Separability Matri. International Journal of Advanced Computer Science and Applications, v. 1, n. 5, p. 93-101, 2010.

GONG, B.; IM, J.; MOUNTRAKIS, G. An artificial immune network approach to multi-sensor land use/land cover classification. **Remote Sensing of Environment**, v. 115, p. 600-614, 2011.

GUERRA, J. B.; MURA, J. C.; FREITAS, C. C. Discriminação de incrementos de desflorestamento na Amazônia com dados SAR R99B em banda L. **Acta Amazonica**, v. 40, n. 3, p. 557-566, set. 2010. doi: <10.1590/S0044-59672010000300015>.

HALL, F. G.; STREBEL, D. E.; NICKESON, J. E.; GOETZ, S. J. Radiometric rectification: toward a common radiometric response among multidate, multisensor images**. Remote Sensing of Environment**, v. 35, p. 11-27, 1991.

HANSEN, M.; DUBAYAH, R.; DEFRIES, R. Classification trees: an alternative to traditional land cover classifiers. **International Journal of Remote Sensing**, v. 17, n. 5, p. 1075-1081, 1996.

HANSEN, M. C.; ROY, D. P.; LINDQUIST, E.; ADUSEI, B.; JUSTICE, C. O.; ALTSTATT, A. A method for integrating MODIS and Landsat data for systematic monitoring of forest cover and change in the Congo Basin. **Remote Sensing of Environment**, v. 112, p. 2495-2513, 2008.

HARALICK, R. M.; SHANMUGAM, K.; DINSTEIN, I. Textural features for image classification. **IEEE Systems, Man, and Cybernetics Society,** v. 3, n. 6, p. 610-62, 1973

HOUGHTON, R. A. Why are estimates of the terrestrial carbon balance so diferent? **Global change biology**, v. 9, p. 500-509, 2003.

HUANG, Y.; LAN, Y.; LACEY, R. E. Multisensor data fusion for high quality data analysis and processing in measurement and instrumentation. **Journal of Bionic Engineering**, v. 4, p. 53-62, 2007.

HUETE, R. A. A soil-adjusted vegetation index (SAVI). **Remote sensing of Environment**, v. 25, n. 3, p. 295-309, 1988.

IBAMA – Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis. **Floresta Nacional do Tapajós - plano de manejo.** v. 1, 2004. Disponível em: <http://www4.icmbio.gov.br/flona_tapajos/index.php?id_menu=144>. Acesso em: 8 out. 2010.

IM, J; JENSEN, J. R. A change detection model based on neighborhood correlation image analysis and decision tree classification. **Remote Sensing of Environment**, v. 99, p. 326-340, 2005.

JENHANI, I.; AMOR, N. B.; ELOUEDI, Z. Decision trees as possibilistic classifiers. **International Journal of Approximate Reasoning,** v. 48, p. 784-807, 2008.

JENSEN, J. R. **Sensoriamento Remoto do Ambiente:** uma perspectiva em recursos terrestres (Tradução da Segunda Edição). São José dos Campos: Editora Perênteses, 2009. 672 p. ISBN: 97885605070601.

KUPLICH, T. M.; FREITAS, C.C.; SOARES, J. V. The study of ERS-1 SAR and Landsat TM synergism for land use classification. **International Journal of Remote Sensing**, v. 21, n. 10, p. 2101-2111, 2000.

KUPLICH, T. M. Estudos florestais com imagens de radar. **Espaço & Geografia**, v. 6, n. 1, p. 71-96, 2003.

KUPLICH, T. M. Classifying regenerating forest stages in Amazônia using remotely sensed images and a neural network. Forest Ecology and Management, v. 234, p. 1-9, 2006. Disponível em: <http://dx.doi.org/10.1016/j.foreco.2006.05.066>. Acesso em: 22 mar. 2010.

LATORRE, M. L. **Utilização de um método de correção atmosférica para o processamento de dados hiperespectrais do sensor AVIRIS em regiões tropicais.** 1998. 202 p. (INPE-7295-TDI/710). Dissertação (Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São Jose dos Campos, 1998. Disponível em: <http://urlib.net/sid.inpe.br/deise/2000/03.13.07.45>. Acesso em: Acesso em: 10 nov. 2010.

Experimento de Grande Escala da Biosfera-Atmosfera na Amazônia. **The ecological component of an integrated Amazon study: the effect of forest conversion.** Workshop Report, Manaus, mai., 55p., 1995. Disponível em <http://lba.cptec.inpe.br/lba/eng/reports/ManausReport1995.html>. Acesso em: 8 out. 2010.

LEITE, R. M. S. R. **Seleção de algoritmos de classificação – Caracterização interativa baseada no desempenho de amostras.** 2007. 149 p. Tese (Doutorado em Ciência da Computação) – Faculdade de Ciências da Universidade do Porto. 2007

LOCK, R. E. M.; KIRCHNER, F. Classificação de imagens multiespectrais de Landsat TM e feições de textura: mapeamento da cobertura da terra. **Floresta**, v. 27, n. 12, p. 41–58, 1997.

LU, D.; MAUSEL , P.; BRONDÍZIO, E.; MORAN, E. Change detection techniques. **International Journal of Remote Sensing**, v. 25, n. 12, p. 2365-2401, 2004.

Marceau, D.; Howarth, P.J.; Dubois, J. M. Automated texture extraction from high spatial resolution satellite imagery for land-cover classification: concepts and application. In: IGARSS'89 - Canadian Symposium on Remote Sensing, 12., Vancouver, July 10-14, 1989. **Proceedings**. IEEE, 1989, v. 5, p. 2765-2768.

MURRAY, H.; LUCIEER, A.; WILLIANS, R. Texture-based classification of sub-Antarctic vegetation communities on Heard Island. **International Journal of Applied Earth Observation and Geoinformation**, v. 12, p. 138–149, 2010.

NEPSTAD, D.; CARVALHO, G.; BARROS, A. C.; ALENCAR, A.; CAPOBIANCO, J. P.; BISHOP, J.; MOUTINHO, P.; LEFEBVRE, P.; SILVA JUNIOR, U. L.; PRINS, E. Road paving, fire regime feedback, and the future of Amazon forests. **Forest Ecology and Management**, v. 154, p. 395-407, 2001.

NOBRE, C. A.; SAMPAIO, G.; SALAZAR, L. Mudanças climáticas e Amazônia. **Ciência e Cultura**, v. 59, n. 3, p. 22-27, 2007.

OLSON, D.; DELEN; O. D. Advanced Data Mining Techiniques. Editora Springer. 2008. 180 p. ISBN: 3540769161

PACHECO, P.; AGUILAR-STOEN, M.; BÖRNER, J.; ETTER, A.; PUTZEL, L.; DIAZ, M. D. C. V. Landscape Transformation in Tropical Latin America: Assessing Trends and Policy Implications for REDD+. **Forests***,* v. 2, p. 1-29, 2011.

PONZONI, F. J.; SHIMABUKURO, Y. E. **Sensoriamento remoto no estudo da vegetação.** São José dos Campos, SP: Silva e Vieira, 2007. 135p. ISBN 9788560507023.

POULIOT, D.; LATIFOVIC, R.; FERNANDES, R.; OLTHOF, I. Evaluation of annual forest disturbance monitoring using a static decision tree approach and 250 m MODIS data. **Remote Sensing of Environment**, v. 113, p. 1749-1759, 2009.

POWELL, R. L.; MATZKE, N.; SOUZA JR, C.; CLARK, M.; NUMATA, I.; HESS, L. L.; ROBERTS, D. A. Sources of error in accuracy assessment of thematic land-cover maps in the Brazilian Amazon. **Remote Sensing of Environment,** v. 90, p. 221-234, mar., 2004.

POZZER, C. T. **Aprendizado por árvores de decisão**. Disciplina de Programação de Jogos 3D. Universidade Federal de Santa Maria, 5p., 2006. Disponível em: <http://www-usr.inf.ufsm.br/~pozzer/disciplinas/pj3d_decision Trees.pdf>. Acesso em: 14 dez. 2010.

RENNÓ, C. D. Avaliação de medidas texturais na discriminação de classes de uso utilizando imagens SIR-C/X-SAR do perímetro irrigado de Bebedouro, Petrolina, PE. 1995. 111 p. (INPE-10441-TDI/926). Dissertação (Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 1995. Disponível em: <http://urlib.net/sid.inpe.br/jeferson/ 2004/ 03.05.11.22>. Acesso em: 22 mar. 2010.

RENNÓ, C. D.; SOARES, J. V.; FORMAGGIO, A. R.; SÁ, I. B.; CORDEIRO, G. Avaliação de medidas texturais na discriminação de classes de uso utilizando imagens SAR. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 8. (SBSR), 1996, Salvador. Anais... São José dos Campos: INPE, 1996. p. 863-870. CD-ROM. ISBN 85-17-00014-5. (INPE-6197-PRE/2286). Disponível em:<http://urlib.net/dpi.inpe.br/lise/2004/07.20.12.15>. Acesso em: 22 mar. 2010.

ROBERTS, D. A.; NUMATA, I.; HOLMES, K.; BATISTA, G.; KRUG, T.; MONTEIRO, A.; POWELL, B.; CHADWICK, O. A. Large area mapping of land-cover change in Rondônia using multitemporal spectral mixture analysis and decision tree classifiers. **Journal of Geophysical Research**, v. 107, n. D20, p. 40/1-40/18, 2002.

RENNÓ, C. D.; FREITAS, C. C.; SANT'ANNA, S. J. S. A system for region image classification based on textural measures. In: SIMPÓSIO BRASILEIRO

DE SENSORIAMENTO REMOTO, 9. (SBSR), 1998, Santos. Anais... São José dos Campos: INPE, 1998. p. 1195-1202. CD-ROM. Artigo apresentado em Sessões Orais, na II Jornada de Radar, ocorrida paralelamente ao IX SBSR, Santos, 1998. ISBN 85-17-00015-3. (INPE-6952-PRE/2914). Disponível em:<http://urlib.net/sid.inpe.br/deise/1999/02.11.14.26>. Acesso em: 22 mar. 2010.

ROUSE, J. W. JR.; HASS, R. H.; SCHELL, J. A.; DEERING, D. W. Monitoring vegetation systems in the great plains with Erts. **In Proceedings** of the Third ERTS Symposium, v. 1, p. 309-317, 1973.

RUSSELL, S.; NORVING, P. **Artificial intelligence a modern approch**. 2009, p. 1152. ISBN 0136042597.

SAATCHI, S. S.; SOARES, J. S.; ALVES, D. S.Mapping Amazon Deforestation and Land Use in rainforest by Using SIR-C Imagery. **Remote Sensing of Environment**, v. 59, p. 191-202, 1997.

SANTOS, J. R.; GONÇALVES, F. G. Polarimetric responses patterns and scattering mechanisms of forest targets from L-band radar. **Revista Brasileira de Cartografia,** n. 61/04, 2009, ISSN 0560-4613.

SERVELLO, E. L. **Dados polarimétricos do RADARSAT-2 (banda C) na discriminação de uso e cobertura da terra na região da Floresta Nacional do Tapajós**. 2010. 93 p. (sid.inpe.br/mtc-m19/2010/10.26.01.47-TDI). Dissertação (Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos. 2010. Disponível em:<http://urlib.net/8JMKD3MGP7W/38FNABP>. Acesso em: 10 jan. 2011.

SESNIE, S. E.; GESSLER, P. E.; FINEGAN, B.; THESSLER, S. Integrating Landsat TM and SRTM-DEM derived variables with decision trees for habitat classification and change detection in complex neotropical environments. **Remote Sensing of Environment**, v. 112, p. 2145-2159, 2008.

SHIMABUKURO, Y. E.; SANTOS, J. R.; HERNANDEZ FILHO, P.; LEE, D. C. L. Avaliação conjuntural da técnica de abordagem multisensor para o monitoramento da vegetação do Brasil. In: SIMPOSIO LATINOAMERICANO DE PERCEPCION REMOTA, 5., 1991, Cusco, PE. **Anais...** 1991. p. 300-309. (INPE-5326-PRE/1724).

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

SHIMABUKURO, Y. E.; MIURA, T.; HUETE, A.; ARAI, E.; ESPÍRITO-SANTO, F. D. B.; LATORRE, M. L**.** Análise dos dados hyperespectrais do EO-1 obtidos sobre a Floresta Nacional de Tapajós no estado do Pará. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 11., 2003, Belo Horizonte. **Anais...** São José dos Campos: INPE, 2003. Artigos, p. 1099-1106. Disponível em:<http://marte.dpi.inpe.br/col/ltid.inpe.br/sbsr/2002/11.17.13.39/doc/10_309.pdf>. Acesso em: 08 out. 2010.

SHIMABUKURO, Y. E.; MAEDA, E. E.; FORMAGGIO, A. R. Sensoriamento Remoto e Sistemas de Informações Geográficas aplicados ao estudo dos recursos agronômicos e florestais. **Revista Ceres**, v. 56, n. 04, p. 399-409, 2009.

SHIMABUKURO, Y. E.; DUARTE, V.; ARAI, E.; FREITAS, R. M.; MARTINI, P. R.; LIMA, A. Monitoring land cover in Acre State, western Brazilian Amazonia, using multitemporal remote sensing data. **International Journal of Image and Data Fusion**, v. 1, n. 4, p. 325-335, set., 2010.

SNUC - SISTEMA NACIONAL DE UNIDADES DE CONSERVAÇÃO DA NATUREZA Lei nº 9.985, de 18 de julho de 2000; decreto nº 4.340, de 22 de agosto de 2002. 5.ed. aum. Brasília: MMA/SBF, p. 56, 2004.

SOBRAL, A. P. B. Previsão de carga horária – uma nova abordagem por árvore de decisão. 2005. 56 p. Tese de Doutorado – Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2003. Disponível em: < http://www2.dbd.puc-rio.br/pergamum/tesesabertas/9916940_03_pretexto.pdf>. Acesso em: 15 jan. 2011.

TIMOFEEV, R. Classification and Regression Trees (CART) Theory and Applications. 2004. 40 p. Dissertação de Mestrado - CASE - Center of Applied Statistics and Economics, Berlin, 2004.

ULABY, F.T.; KOUYATE, F.; BRISCO, B.; LEE WILLIAMS, T.H. **Textural Information in SAR Images.** IEEE Transactions on Geoscience and Remote Sensing, v. 24, n. 2, p. 235-245, mar., 1986.

ULABY, F. T.; ELACHI, C. **Radar polarimetry for geoscience applications.** Norwood: Artech House, Inc., 1990. 388 p.

VASCONCELLOS, E. C. Árvores de decisão aplicadas ao problema da separação estrela/galáxia. 2011. 98 p. (sid.inpe.br/mtc-m19/2011/06.10.18.36 - TDI). Dissertação (Mestrado em Computação Aplicada) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2011. Disponível em:<http://urlib.net/8JMKD3MGP7W/39RNQ52>. Acesso em: 14 fev. 2012.

VERMOTE, E. F.; TANRÉ, D.; DEUZÉ, J. L.; HERMAN, M.; MORCRETTE, J. J. Second Simulation of the Satellite Signal in the Solar Spectrum, 6S: An overview**. IEEE Transactions on Geoscience and Remote Sensing**, v.35, n.3, p.675-686, May 1997.

WANG, C. **Detection of coal fires in Xinjiang (China) using remote sensing techniques**. 2002. 103 p. Dissertação (Mestrado em Estudos de Desastres Naturais) - International Institute for Geo-information Science and Earth Observation, Países Baixos. 2002. Disponível em:

http://www.itc.nl/library/papers/msc_2002/ereg/chunqing_wang.pdf. Acesso em: 10 dez. 2011.

WITTEN, I. H.; FRANK, E.; HALL, M. A. **Data mining : practical machine learning tools and techniques.** Terceira edição. Editora The Morgan Kaufmann series in data management systems, 2011. 665 p. ISBN 978-0-12-374856-0

WOLTER, P. T.; TOWNSEND, P. A. Multi-sensor data fusion for estimating forest species composition and abundance in northern Minnesota. **Remote Sensing of Environment**, v. 115, p. 671-691, 2011.

ZURITA-MILLA, R.; KAISER, G.; CLEVERS, J. G. P. W.; SCHNEIDER, W.; SCHAEPMAN, M. E. Downscaling time series of MERIS full resolution data to monitor vegetation seasonal dynamics. **Remote Sensing of Environment,** v. 113, p. 1874-1885, 2009.